

**First Language Attrition and Second Language Acquisition in  
a Second Language Environment**

**Cornelia Opitz**

Ph.D. in Applied Linguistics

University of Dublin, Trinity College  
Centre for Language and Communication Studies  
June 2011



## **Declaration**

I hereby declare that this thesis, submitted in candidature for the degree of Doctor of Philosophy at Trinity College Dublin, has not previously been submitted for a degree at this or any other university.

This thesis is entirely my own work, and any assistance is acknowledged.

I am willing to grant permission to Trinity College Library to lend or copy the thesis upon request.

Conny Opitz



## Summary

This thesis is concerned with the outcomes of the parallel processes of first language maintenance and second language acquisition in adult bilinguals resident in a second language environment. Current perspectives on first language attrition and bilingualism makes a strong case for considering L1 attrition as a feature of multi-competence in bilinguals, and for taking into account changes across the range of languages known by a bilingual in assessing proficiency. They suggest that the simultaneous maintenance of several languages by a bilingual may result in trade-offs between those languages, but also that dynamic interactions between languages and a host of other factors will result in very different outcomes for individuals.

In a mixed between-group/within-group design, 27 native speakers of German who emigrated to Ireland as adults (mean age at arrival = 26.8 years; mean LOR = 19.5 years), and two matching control groups of 18 Irish and 20 German L2 users were tested on an extensive test battery of parallel German and English language tests. Participants additionally attempted a linguistic aptitude test and responded to several questionnaires, allowing the comprehensive probing of a wide range of predictor variables for L1 attrition and L2 acquisition.

The thesis provides a comprehensive analysis of the between-group data, assessing participants' L1 and L2 performance across three tasks. Participants' proficiency is thoroughly investigated on quantitative and qualitative measures of complexity, accuracy and fluency at the group and individual levels.

The results of the attrition study show that the bilingual group's performance does not differ significantly from that of the German control group on most individual measures, the exceptions being one fluency measure (percentage of repetitions) and certain error types. However, when all measures are combined in *z*-scores, differences do become significant. There are three individuals who consistently show low performance, who can be considered L1 attriters, while others perform on a par with the native-speaker controls. There is, however, considerable variability within each group, in line with the assumptions of dynamic approaches to language attrition.

The L2 attainment study, on the other hand, shows that on measures emphasising breadth of linguistic knowledge and accuracy the bilingual group performs differently compared to the Irish control group, while on measures focusing on fluency, lexical diversity and idiomatic language use it was comparable. Over half of the bilingual participants have *z*-scores within the control group's range, indicating native-like performance across the three tasks. The significant group difference on the total scores is due to some bilingual participants who have not overall achieved a similar level of proficiency. However, even those participants perform in a native-like manner on some of the measures, and the bilingual group overall performs significantly better than the other group of L2 learners, pointing to successful L2 acquisition on the part of the bilingual group.

A brief consideration of some of the results of the within-group study serves to point out future directions of research in relation to this study, and the chosen line of enquiry.

## Acknowledgements

On the long road to completing this thesis, I was fortunate to be able to rely on the specialist advice, practical assistance and moral support of a great number of people. First and foremost, thanks are due to my supervisor, Prof David Singleton, who guided the process with great expertise, commitment and empathy, who never ceased to believe in the project's success and encouraged and put his trust in me all the way.

Second, to Monika Schmid, for her very hands-on approach to attrition- and data-related problems on more than one occasion, and more generally for setting up the Graduate Network of First Language Attrition, thus creating a platform for learning, sharing ideas and materials, for finding collegial support and forming friendships, all of which has been invaluable over the years. To Barbara Köpke and Kees de Bot, for willingly sharing their vast expertise with kindness, and for providing much-needed encouragement.

To my family, friends and colleagues, not all of whom I have space to mention by name, for your never-failing support and belief in me, particularly during those times of upheaval, for feeding me emotionally or with Green and Black's chocolate and very self-assured blueberry jam, and for occasionally bending over backwards or taking a back seat to allow me to get on with the work. Many of you have become involved in different stages of the process in various practical ways – putting me up, enrolling participants, participating yourselves, transcribing and rating material, being expert consultants on the German or English language and on questions of statistics, reminding me of "my central question" and the progress implicit in putting commas in the right place, being sound-boards for ideas, taking an interest, producing film clips and much, much more.

I am eternally indebted to Brigitte Opitz, Susan Dostert, Gaby Schön and George Courtney for caring and actively supporting me in ways too numerous to list. Denise O'Leary, Charlotte Lary and Jeannette Hemmecke; Sarah Smyth, John Harris, Seán Devitt, Breffni O'Rourke and Lorna Carson; Jenny Bruen, Deirdre Murphy and Feargus Denman; Alexander Koch, Klaus Schwienhorst, Ninel Dadon, Jana Löwe, Christiane Maaß, Hanneke Loerts and Gulsen Yilmaz; Solvejg Opitz and Steffi Gieseler; Angela Leahy, Siobhán McNamara and Annette Simon; Dorota Lubińska and Esther de Leeuw; Gay Conroy, Katrin Eberbach, Tamara O'Connor and Damian McManus; Annette Schiller, Linda Ledwidge, Berni Gunning, Ann Hennessy, Eva O'Cathaoir and Margaret Brady; Gottfried-Wilhelm Opitz, Mira Sonntag, Andrea Hacker, Paula Kelly, Robin Hilliard, Tom Weymes, John Kearns, Ralf Hemmecke, Kai Karnowka, Daniela Kobe and Christof Bußejahn – you have all made a big difference to the project!

I owe a very special debt of gratitude to my participants, who put themselves through a gruelling number of tasks, with great patience and unflinching good humour. The work would not have been possible without you, and I hope sincerely that this thesis – only a first step on the way – does justice to you. You have provided me with special insights, generously opening up your worlds to me, and providing me with enough material to last a lifetime!





I dedicate this thesis to my mum, with love.



## Table of contents

List of tables.....	xv
List of figures.....	xvii
1 Introduction.....	1
2 Literature review.....	9
2.1 Language attrition and bilingualism.....	9
2.1.1 Background.....	9
2.1.2 Points of reference.....	13
2.1.3 Native speakers vs. multi-competence.....	16
2.2 Processes of language attrition and acquisition.....	19
2.2.1 Language attrition – findings.....	19
2.2.2 Learning and forgetting.....	25
2.2.3 Crosslinguistic effects.....	29
2.2.4 Languages in competition.....	34
2.3 Models of language attrition and acquisition.....	36
2.3.1 Activation Threshold Hypothesis (Paradis 1993; Paradis 2001).....	37
2.3.2 The Dynamic Model of Multilingualism (DMM) (Herdina and Jessner 2002; Jessner 2003).....	42
2.4 Predictor variables in language attrition and acquisition.....	49
2.4.1 Age at onset (AOA).....	50
2.4.2 Length of residence (LOR).....	56
2.4.3 Contact with L1/L2.....	58
2.4.4 Social background and education.....	61
2.4.5 Attitudes and motivation.....	64
2.4.6 Personality and aptitude.....	70
2.4.7 L2 proficiency.....	73
2.5 Summary.....	82
3 Experimental design.....	83
3.1 Participants.....	83
3.1.1 Participant search.....	87
3.1.2 Selection criteria and participant characteristics.....	88
3.1.2.1 Age, gender, residence abroad, education level.....	88
3.1.2.2 Region of origin.....	91
3.1.2.3 Social and family setup, L1 and L2 use.....	93
3.1.2.4 L2 proficiency.....	94
3.2 Design of the study.....	96
3.2.1 Longitudinal vs. cross-sectional design – design considerations.....	97
3.2.2 Between-group and within-group design – design decisions.....	101
3.3 Test and elicitation materials.....	105
3.3.1 Questionnaires.....	105
3.3.2 Language tests.....	109

3.3.2.1 C-test .....	109
3.3.2.2 Verbal fluency task .....	113
3.3.2.3 Film retelling .....	122
3.3.3 Linguistic aptitude test .....	125
3.4 Administration of the test battery .....	128
3.4.1 The testing session .....	129
3.4.2 Administration of questionnaires and language tasks .....	136
3.4.2.1 Questionnaires .....	136
3.4.2.2 C-test .....	136
3.4.2.3 Verbal fluency task .....	138
3.4.2.4 Film retelling .....	139
3.4.2.5 Linguistic aptitude test .....	140
4 Results: Language attrition/maintenance – between-group study German .....	143
4.1 C-test German .....	144
4.1.1 Test results .....	144
4.1.2 Item analysis .....	149
4.2 Verbal fluency tasks German .....	155
4.2.1 Test results .....	155
4.2.2 Item analysis .....	160
4.3 Film retelling German .....	165
4.3.1 Narratives .....	168
4.3.1.1 Sample length .....	168
4.3.1.2 Lexical diversity .....	170
4.3.1.3 Fluency .....	172
4.3.1.4 Measures combined .....	180
4.3.2 Scenes .....	182
4.3.2.1 Code-switching .....	182
4.3.2.2 Errors and idiomatic language use .....	187
4.3.2.3 Gem analysis .....	200
4.4 Summary .....	208
4.5 Discussion .....	215
4.5.1 Self-reported L1 proficiency at time of testing (can-do scales) .....	215
4.5.2 Self-reported L1 proficiency over time .....	220
5 L2 attainment – between-group study English .....	225
5.1 C-test English .....	226
5.1.1 Test results .....	226
5.1.2 Item analysis .....	229
5.2 Verbal fluency tasks English .....	240
5.2.1 Test results .....	240
5.2.2 Item analysis .....	243
5.3 Film retelling English .....	246
5.3.1 Narratives .....	247

5.3.1.1	Sample length.....	247
5.3.1.2	Lexical diversity.....	249
5.3.1.3	Fluency.....	252
5.3.1.4	Measures combined.....	260
5.3.2	Scenes.....	261
5.3.2.1	Code-switching.....	261
5.3.2.2	Errors and idiomatic language use.....	264
5.3.2.3	Gem analysis.....	276
5.4	Summary.....	284
5.5	Discussion.....	288
5.5.1	Self-reported L1 proficiency at time of testing (can-do scales).....	288
5.5.2	External proficiency ratings.....	293
5.5.3	Self-reported L1 proficiency over time.....	295
6	Bilingual profiles – within-group study.....	297
6.1	Predictor variables.....	297
6.1.1	Biographical factors.....	297
6.1.2	Language use.....	299
6.1.3	Language attitudes.....	300
6.1.4	Linguistic aptitude.....	302
6.2	Profiles.....	304
6.3	Summary and discussion.....	308
7	Conclusions.....	309
7.1	Limitations of the present study and directions for future research.....	311
	References.....	313
	Appendices.....	325
	Appendix I. Participant data.....	325
	Appendix II. Questionnaires.....	327
	i) Personal background questionnaire (English version).....	327
	ii) Sociolinguistic questionnaire, Part 1.....	330
	iii) Sociolinguistic questionnaire, Part 2.....	333
	iv) Can-do statements (English version).....	335
	v) Attitudes towards foreign language learning/bilingualism.....	338
	Appendix III. Test materials.....	339
	i) Session protocol bilinguals.....	339
	ii) Session protocol controls.....	343
	iii) C-test.....	346
	iv) Film retelling.....	351
	Appendix IV. Tasks not included in the thesis.....	355
	i) Picture description.....	355
	ii) Sentence generation task.....	362
	References.....	369



## List of tables

Table 1: General language effort at different proficiency levels by linguistic environment .....	44
Table 2: Participant background data – overview bilingual participants.....	90
Table 3: Participant background data – primary matching criteria .....	90
Table 4: Language varieties and regions of origin of bilingual participants .....	91
Table 5: English-language variety spoken by bilingual participants.....	92
Table 6: Participant background data – additional criteria (1) .....	93
Table 7: Participant background data – additional criteria (2) .....	95
Table 8: Pros and cons in longitudinal vs. cross-sectional designs .....	97
Table 9: Test Battery .....	103
Table 10: Productivity of different semantic stimuli.....	117
Table 11: Mean scores for German and Irish control groups for all FiCA conditions .....	118
Table 12: Test battery: administration.....	129
Table 13: Overview of testing session.....	130
Table 14: Scoring scheme for C-tests.....	145
Table 15: German C-test mean scores and time taken .....	146
Table 16: Individual German C-test scores and test times* .....	147
Table 17: Individual German C-Test Scores and Test Times continued* .....	148
Table 18: Breakdown of number of correct item solutions by percentage of participants .....	149
Table 19: Problematic items per text across both groups .....	149
Table 20: Selected items from German C-test for detailed analysis.....	151
Table 21: Mean scores for bilingual group and German control for all German FiCA conditions .....	156
Table 22: Comparison of participant scores relative to control group's mean across three tasks .....	159
Table 23: Item analysis of FiCA1 <i>Professions</i> German .....	161
Table 24: Number of German film retellings .....	167
Table 25: Length of German film retellings .....	168
Table 26: VOCD-D values for German film retellings .....	171
Table 27: Fluency measures for German film retellings .....	174
Table 28: Individual results on three fluency measures, German film retelling.....	178
Table 29: Correlations between sample length, lexical diversity and fluency measures.....	181
Table 30: Code-switches film retelling German.....	183
Table 31: Errors film retelling German .....	190
Table 32: Gems resulting in lexical/semantic transfer from English to German.....	194
Table 33: Total errors German film retelling by individual .....	198
Table 34: Gems analysed for lexical diversity and retrieval problems (German).....	200
Table 35: Results of can-do statements German (percentage scores) .....	215
Table 36: Correlations between German can-do statements and C-test for bilingual group .....	218

Table 37: Self-ratings L1 proficiency .....	221
Table 38: Reported changes in L1 proficiency .....	223
Table 39: Self-ratings L2 proficiency .....	225
Table 40: English C-test mean scores and time taken.....	226
Table 41: Individual English C-test scores and test times* .....	228
Table 42: Breakdown of number of correct item solutions by percentage of participants .....	230
Table 43: Problematic items per text across both groups .....	231
Table 44: Selected items from English C-test for detailed analysis.....	232
Table 45: Mean scores for bilingual group and Irish control for all English FiCA conditions.	240
Table 46: Comparison of participant scores relative to control group's mean across three tasks .....	242
Table 47: Item analysis of FiCA1 <i>Professions</i> English .....	244
Table 48: Number of English film retellings .....	247
Table 49: Length of English film retellings .....	247
Table 50: VOCD-D values for English film retellings .....	250
Table 51: Distribution of individual VOCD-D (standard D) scores .....	251
Table 52: Fluency measures for English film retellings .....	253
Table 53: Individual results on three fluency measures, English film retelling.....	257
Table 54: Individual results on three fluency measures, English film retelling, continued .....	259
Table 55: Correlations between sample length, lexical diversity and fluency measures .....	260
Table 56: Code-switches film retelling English.....	261
Table 57: Errors film retelling English .....	266
Table 58: Total errors English film retelling by individual .....	267
Table 59: Gems resulting in lexico-semantic errors .....	273
Table 60: Gems analysed for lexical diversity, idiomatic use and retrieval difficulties (English) .....	276
Table 61: Results of can-do statements English (percentage scores).....	289
Table 62: Can-do statements English, resulting in significant group differences.....	290
Table 63: Correlations between English can-do statements and C-test for bilingual group .....	292
Table 64: Proficiency ratings by external raters by participant .....	294
Table 65: Correlations between bilinguals' z-scores and biographical predictor variables.....	298
Table 66: Correlations between bilinguals' z-scores and language use .....	300
Table 67: Correlations between bilinguals' z-scores and attitude .....	301
Table 68: Mean aptitude scores for all three groups .....	303
Table 69: Correlations between bilinguals' z-scores and linguistic aptitude .....	303
Table 70: Clusters of bilingual profiles based on German/English z-scores.....	306



## List of figures

Figure 1: General language effort in the L1 environment .....	44
Figure 2: General language effort in a monolingual L2 environment .....	45
Figure 3 Schematic representation of proficiency levels in casual condition (after Major 1992) .....	79
Figure 4: Box plots of German C-test scores and times for all three groups.....	146
Figure 5: Scores across three FiCA tasks for bilinguals (cases 1-27, N = 26) and German controls (cases 28+, N = 29).....	158
Figure 6: Scores across three FiCA tasks for bilingual group (descending order).....	160
Figure 7: Features tagged in transcripts of film retellings.....	166
Figure 8: Lexical diversity scores (standard D) for German film retelling .....	172
Figure 9: Fluency measures bilingual group vis-à-vis native-speaker baseline .....	178
Figure 10: Fluency measures control group vis-à-vis native-speaker baseline .....	179
Figure 11: Total errors German film retelling .....	197
Figure 12: Errors bilingual group film retelling German by type.....	199
Figure 13: German groups' performance across all three German tasks .....	213
Figure 14: Distribution of skills scores by participant.....	217
Figure 15: Box plots of English C-test scores and times for all three groups .....	227
Figure 16: Scores across three FiCA tasks for bilinguals (cases 1-27, N = 25) and German controls (cases 51+, N = 32).....	241
Figure 17: Scores across three FiCA tasks for bilingual group (descending order).....	243
Figure 18: Lexical diversity scores (standard D) for English film retelling .....	251
Figure 19: Fluency measures English film retelling vis-à-vis native-speaker baseline.....	256
Figure 20: Total errors English film retelling.....	267
Figure 21: Errors bilingual group film retelling English by type .....	268
Figure 22: Bilingual and Irish control groups' performance across all three English tasks.....	287
Figure 23: Distribution of skills scores by participant.....	291
Figure 24: Scatter plot of bilingual profiles.....	306



# 1 Introduction

Anyone who has learned a second or foreign language (L2) at some stage in their life will, following periods of not using that language, have experienced certain changes – increased difficulties in remembering particular words to use at the right time in conversation, some previously mastered structures of grammar becoming “wobbly” again, an apparent loss of fluency and of the ability to use language idiomatically. In other words, one’s language skills become “rusty”. Few L2 users will deny the reality of this process of deteriorating *foreign* or *second language skills*. Keeping up L2 skills requires effort when the environment does not provide, and the L2 learner/user does not seek regular reinforcement.

An issue for debate, however – both amongst lay people and in the research community – is the notion that under certain circumstances *native language skills* may also undergo such changes, and that first-language/mother tongue (L1) maintenance is not a given under adverse conditions. Such conditions often arise in connection with a person’s prolonged relocation to a different country – a common reality in our time – usually resulting in less frequent contact with other L1 speakers, and consequently in “potential communication problems between individuals and the community of which they consider themselves a member” (Jaspaert and Kroon 1989: 80).

Migration to a different country is also typically associated with pressures to acquire a new L2, further decreasing exposure to and opportunities for using the L1. While in the recent past it has become easier for many migrants to stay in touch with their original communities if they so wish – through more affordable travel, telephone calls and internet/e-mail – the role for active integration into and participation in the new speech community has also grown (e.g. European Commission 2005 Framework for the Integration of Third-Country Nationals in the European Union), raising the question of how bilinguals cope with the cognitive and environmental demands of their languages.

It was my own personal experience that prompted me to start thinking about changes in language proficiency, first and second, and the factors that might bring these changes about. I moved to Ireland in 1995 to take up what became a three-year lectureship in German. During those first three years, my English-language skills expanded rapidly and very satisfactorily, while my German was

kept active through work, my sharing accommodation with another German lector and the occasional trip and phone call to Germany.

Following on from that first job, I started teaching Russian, a language I had studied intensively in school and at college. This, combined with a change in my living situation, meant that I was now speaking either Russian or English in daily life. The first year was difficult in relation to Russian, since I had not practised it during the first three years, and also now needed to learn how to relate Russian and English for teaching purposes, rather than Russian and German. In fact, it has taken me several years to get close again to the level of fluency in Russian I had reached previously, despite the fact that I use the language regularly in work.

Three years on again, I took up additional work as a German-speaking tour guide during the summers. Despite having duly prepared my materials, learned how to manage groups etc., the first tour was a shock, since I had severe problems expressing myself in German. Since then, German has come back into my life in some form or another, but to this day, tourists frequently ask me where I was born. To my sheepish response "in Germany", the usual reply is that they had thought so on account of my native (-like?) accent. However, the fact that the question is put indicates that either some features of my speech are not native-like, or perhaps that people pick up on hesitations and my problems searching for particular words. At this point in time, the language I use most and find it easiest to express myself in, however imperfectly, is English.

My situation is just one example. Many friends and acquaintances have similarly experienced a waxing and waning of their language skills – in all their languages, including the L1. Indeed, there now exists a body of research on this "condition" labelled "first language attrition", the reduction of a person's first-language skills or their mother tongue proficiency. However, the evidence diverges regarding the degree and manifestations of L1 attrition, the causes and "cures", and the underlying processes and mechanisms; and some important questions have only begun to be addressed.

In particular, the role of "the other" language(s) with which L1 is in contact, remains largely unexplored, beyond direct influences on the attriting L1 in the shape of interferences and borrowings. Many researchers now point to the fact that the L1 usually does not deteriorate in isolation but that often another language comes to replace it. Increasingly, it is recognised that a full account of L1 attrition requires investigating changes in the other languages of bilinguals, at

least the host L2, since assessing only the L1 skills of bilinguals ignores the full complexity of their linguistic development, and especially their linguistic achievements.

The same conditions that may result in a reduction in a person's L1 proficiency are arguably conducive to improvements in the L2 proficiency. Living abroad, according to popular belief, almost inevitably leads to a person becoming fluent in the language of the host community, once a certain threshold level of proficiency in the language has been passed. While a great deal of research confirms the facilitatory effect of immersion in an L2 environment for L2 development (e.g. Linck, Kroll et al. 2009; Pérez-Vidal and Juan-Garau 2009), the question of ultimate attainment in adult L2 learners remains an issue for debate (e.g. Birdsong 1992; Bongaerts 1999; Hyltenstam and Abrahamsson 2000; Abrahamsson and Hyltenstam 2009; DeKeyser, Alfi-Shabtay et al. 2010).

Unfortunately, issues relating to adult L2 acquisition and use, and L1 maintenance/attrition in L2 settings have been studied largely independently of one another, while the possibility that there may be interaction between the two processes and languages in question has only recently begun to be explored.

These considerations form the motivation for this work. When it began, there was very limited empirical research attempting to relate the two processes (Hyltenstam and Viberg 1993). Some of this work was undertaken to test the Regression Hypothesis (Jakobson 1941) – the idea that language attrition might proceed as the mirror image of language acquisition processes. The exception to the rule is one article (Major 1993, in Hyltenstam/Viberg 1993) which reported the simultaneous loss of L1-specific VOT values and acquisition of the corresponding L2 values in Brazilian Portuguese and American English.

Eight years on, I know of one completed thesis dealing with null and overt pronouns (Gürel 2002), one thesis in progress which replicates the methodology used in this project (Cherciov to appear), and the above-mentioned immersion study (Linck, Kroll et al. 2009). However, in light of recent theoretical work on multi-competence in second-language users (Cook 2002; 2003) and dynamic interactions between languages in the multilingual mind (Herdina and Jessner 2002; de Bot 2007; Larsen-Freeman and Cameron 2008), as well as newer neuro- and psycholinguistic models of language processing (Paradis 1993; Green 1998; Paradis 2007), the need for refocusing the L1

attrition agenda on the parallel processes of language acquisition and language attrition/maintenance has gained even more pertinence.

The research presented in this thesis goes some way to addressing these issues. It ultimately sets out to establish whether the level of L2 proficiency achieved by adult bilinguals in an L2 setting impacts on the success with which the L1, the mother tongue, is maintained. In other words, it asks whether L2 proficiency could be an independent predictor of L1 attrition as a result of interaction between the language systems, in addition to the environmental pressures and opportunities to learn the language of the L2 environment.

Consequently, the research attempts to identify the separate contributions made by biographical, social, affective and cognitive factors which have been proposed as predictors of language maintenance, while being specifically concerned with a social factor (amount and quality of exposure to either language), an affective one (self-identification and attitudes towards foreign language learning and bilingualism), and a cognitive one (language aptitude). Some of these factors, such as the role of aptitude for L1 attrition, began to be studied only after the start of this work (Köpke 2007; Bylund, Abrahamsson et al. 2010); the study of others, such as language use and attitudes, have yielded inconclusive results.

With the above aims in mind, 27 late bilinguals who are native speakers of German and have settled in Ireland (henceforth the "bilingual group"; mean length of residence (LOR) = 19.5 years) were tested on an extensive test battery of parallel German and English language tests. In addition, the participants responded to several questionnaires, including an in-depth sociolinguistic questionnaire, and attempted a linguistic aptitude test. The test battery yielded proficiency data in respect of the two languages under investigation and probed a comprehensive range of factors linked to attrition/maintenance, such as length of residence and age of first starting to learn the L2, as well as the factors of specific concern to this investigation.

Two matching control groups of 18 Irish native speakers of English and 20 German native speakers ("Irish control group"/"German control group") provided baseline proficiency data for German and English. The control groups met minimum requirements for proficiency in L2 German and English respectively and were distinguished from the bilingual group by being resident in their country of origin and not having lived abroad for extensive periods of time. Since all bilingual participants

emigrated well after the age at which L1 acquisition is considered to be largely complete, any systematic differences between them and the German control group are interpretable as L1 attrition on the part of the bilinguals and/or changes in the German language as spoken in the country of origin. On the other hand, proximity to the Irish control group's performance and self-reports of changes in their L2 proficiency shed light on the L2 acquisition situation.

In the course of collecting and analysing the data, it became apparent that the "big" question of whether highly proficient L2 users are more likely to show signs of L1 attrition than less proficient bilinguals contains several sub-questions:

1. What is the extent of L1 attrition/maintenance amongst the group of adult bilinguals investigated?
2. What is the degree of their L2 attainment?
3. What is the relationship between the level of L1 and L2 proficiency for these speakers?
4. What factors impact on that relationship?

In fact, these four questions represent three separate studies, each worthy of investigation in its own right. The first two studies – the L1 attrition study and the L2 attainment study – are between-group studies entailing the comparison of the bilingual group's performance to the two control groups. They thus relate to sub-questions 1 and 2 and are a prerequisite for answering sub-questions 3 and 4. These sub-questions, on the other hand, are addressed by the bilingual profiles study, a within-group study entailing the investigation of individual bilinguals' language profiles.

In analysing the data, I felt that to do justice to the vast amount of data collected, and by extension to my participants, I ought to provide a detailed investigation of sub-questions 1 and 2. Consequently, this thesis primarily focuses on the results of the between-group studies, while the within-group study is treated briefly in the context of providing directions for future research.

Although the thesis itself thus does not fully answer the original "big" question, the research would not have been conducted in the way it has without having asked that question first. Researching the relevant fields has informed my decisions all the way through, in deciding on the participants' characteristics, the experimental design and materials to be used, as well as shaping the framework for analysis.

As it stands, the work presented here adds to the body of research of several linguistic sub-fields. First, it contributes to the field of L1 attrition research in adult bilinguals by providing a comprehensive analysis of participants' L1 performance across three tasks representing different proficiency dimensions and task demands. The analysis presents a thorough investigation of participants' proficiency along the dimensions of complexity, accuracy and fluency (Larsen-Freeman 2009), combining both quantitative and qualitative approaches. Since some of the materials and participant criteria were the result of collaborative efforts within the Graduate Network of First Language Attrition (2001 – to date, cf. Chapter 2), the results can be directly compared to other studies conducted within that framework.

Second, it contributes to research into the ultimate L2 attainment of adult bilinguals (cf. above) and the "age factor" (Singleton 1989; García Mayo and García Lecumberri 2003), against the view that ultimate attainment is exclusively constrained by age of acquisition, and that nativelikeness is unachievable for adult learners. Since all bilingual participants in this study emigrated well after puberty, their L2 attainment provides reliable evidence for the possibility of successful L2 acquisition in a naturalistic setting and nativelikeness in some domains. The age factor is also an important consideration in L1 attrition studies, thus the results of this study add to knowledge about differences between adult and child attriters.

Third, the study frames L1 attrition as a feature of the simultaneous development of bi- and multilingual competence. A multi-competence perspective proposes that all language proficiency, including first language proficiency, needs to be seen as relative and amenable to change, in contrast with traditional conceptualisations of bilingualism as a different form of monolingualism. The "monolingualist" view has drawn much criticism (e.g. Grosjean 1989; Cook 1997; de Groot and Kroll 1997; Cook 2002), in view of arguments that the process of learning another language fundamentally changes the way our first language is represented in our minds.

A multi-competence view of bilingualism also entails the notion that the two or more language systems of bilinguals interact multi-directionally (Jessner 2003). This interaction can potentially take two forms – positive interaction in the shape of mutual reinforcement of the language systems, or negative interaction in the shape of interferences between the two systems due to competition. The work presented in this thesis contributes empirical evidence for such interaction between



language systems in the bilingual mind, by integrating the two perspectives of L1 attrition and L2 acquisition.

However, and finally, the project ultimately hopes to contribute to our knowledge of how to avoid first-language attrition, or put differently, how to foster active bilingualism, which in turn entails important implications for the area of language policy, in a world where bilingualism is "not the exception anymore" (Harris and McGhee Nelson 1992).

The remainder of the thesis is structured as follows. Chapter 2 summarises findings from the field of language attrition and discusses the theoretical perspective in more detail, Chapter 3 sets out the methodology in relation to the participants and tests. Chapters 4-6 deal with the results of the study: Chapter 4 presents the findings from the attrition study, Chapter 5 those from the attainment study, and Chapter 6 gives a synopsis of data from the profiles study. It also provides an outlook on the next stages of research in relation to this study. Finally, Chapter 7 provides a short conclusion to the thesis, pointing out limitations to the research presented and outlining future directions of research.



## **2 Literature review**

Language attrition research, a comparatively young field of linguistic research, has gained popularity over the past few years. Maybe linked to personal experience, researchers have become curious about the 'intriguing' question of 'whether an individual can really forget a first or second language once learned, and how and why this 'loss' might proceed' (Köpke and Schmid 2004: 2). Similarly, the process of people becoming bilingual has held fascination for many researchers over the years.

As explained in the Introduction, this thesis endeavours to bring together the perspectives of first language attrition and second language acquisition within the framework of some current models of bilingualism. The sections below present relevant findings from these three areas of linguistic research. We first chart the development of the field of language attrition research and discuss some essential concepts to explain the context for the current research. We then go on to present an overview of the main findings in the field of first language attrition to date. The second section further focuses on the development of proficiency in L1 and L2, while the third presents two models which aim to capture those processes. The final section discusses variables that have been proposed as predictors of L1 attrition and L2 acquisition.

### ***2.1 Language attrition and bilingualism***

#### ***2.1.1. Background***

The field of language attrition, or individual language loss, was formally established in the early 1980s with a conference at University of Pennsylvania in 1980 (Lambert and Freed 1982). Some of the various strands of modern attrition research, such as language loss in children and adults; L1 and L2 loss; loss as a pathological condition (in aphasia or dementia) and in healthy individuals; linguistic, sociolinguistic, psycholinguistic and neurolinguistic perspectives, were brought together for the first time. The conference succeeded in establishing a research agenda, de facto launching the discipline, which has since grown notably in terms of the number of researchers involved as well as output.

The conference sparked off a major research project in the US, the Language Skills Attrition Project (Ginsberg 1986), which studied language attrition on the part of selected U.S. populations in Arabic, Chinese, and Japanese. The project tested a large sample of participants (200 per language) on a variety of general and diagnostic tests using a pre-test/post-test design and focusing on attitudes and motivation, language use and exposure between the two measuring points.

In Europe, on the other hand, a two-day workshop at University of Nijmegen in 1982 prompted numerous attrition research projects (van Els 1986: 3), primarily individual case studies of first and second language loss including extensive studies of pathological language loss, notably in the Netherlands, Israel, Sweden and, to a lesser extent, Germany.

For some time prior to those developments, the field of language loss had taken an interest in the factors instrumental in bringing about linguistic change, language shift and language death within communities in language contact situations, such as immigrant communities or minorities. The focus of this branch of research continues to be a concern with intergenerational language loss at the group level, the properties of the language or dialect in question, and the socio-economic factors bringing about these changes.

With the establishment of the discipline of attrition research, the (psycholinguistic) process of intra-generational, individual language loss, changes in individual language use and proficiency, and the impact of personal factors came to the fore, although the fact that individuals are subject to social factors is not ignored.

This change in research focus required new terminology. We follow de Bot's (1995; 1998) use of 'language shift' for the former, 'language attrition' for the latter, and 'language loss' as a cover term for both. Language attrition may thus be defined as "the decline of any language (L1 or L2), skill or portion thereof in a healthy speaker" (Ecke 2004: 322), or, in Yukawa's words, as

"a permanent or temporary regression from a participant's previous linguistic performance or competence at any linguistic level (phonology, morphology, syntax, and pragmatics) in exerting any linguistic skill (speaking, listening, reading, writing and metalinguistic judgement)." (Yukawa 1997: 2)

Today, there are two quite distinct branches of attrition research: first language attrition, which focuses on the individual loss of the mother tongue, usually in migrants, and is the concern of this

work, and second language loss, which investigates L2 attrition as subfield of foreign language teaching.

Another dimension, originally charted in the so-called 'van Els taxonomy' (van Els 1986) (Köpke and Schmid (2004: 8) point out that these distinctions were actually first made by de Bot and Weltens in 1985), is the environment in which the loss happens. Thus, in addition to L1 attrition in an L2 environment, and L2 attrition in an L1 environment, which correspond to the two major directions of research just mentioned, the opposite combinations also obtain, for example in dialect loss (L1 attrition in the L1 environment) and second language loss in elderly bilinguals in L2 environments.

These types are usually applied to "healthy" language attrition, with pathological first and second language attrition, such as in aphasia and dementia, representing a separate field of research (Hyltenstam and Stroud 1989; Paradis 1993).

The bulk of attrition research to date have been descriptive data-based studies rather than theoretical-exploratory works, but there have been attempts to apply theoretical frameworks from linguistics and neighbouring fields to attrition data (cf. Hansen 2001: 60; Köpke 2007 for extensive overviews). Some frameworks tested on attrition data are the regression hypothesis (e.g. Jordens, de Bot et al. 1989; Olshtain and Barzilay 1991; Hansen 1999; Keijzer 2010), markedness theory (Altenberg 1991), learnability theory (e.g. Sharwood Smith 1983; Sharwood Smith and van Buren 1991; Major 1992; Sorace 2000), the critical period hypothesis (Köpke 2004), the matrix-language frame model (e.g. Bolonyai 1999; Schmitt 2000), social network theory (Hulsen, de Bot et al. 2001), sociocultural theory (Jiménez Jiménez 2004) and ethnolinguistic vitality theory (Yağmur, De Bot et al. 1999; Hulsen 2000).

The psycholinguistic branch of attrition research in particular has benefited from increased interdisciplinary activity, borrowing paradigms from related disciplines, such as the savings paradigm (de Bot and Stössel 1999; Hansen, Umeda et al. 2000), the models of the Activation Threshold Hypothesis (ATH, Paradis 2001), the Declarative/Procedural Model (Ullman 2003) and the Inhibitory Control Model (Green 2003).

Recent developments espouse taking a more holistic view of language attrition as an 'integral component of language acquisition' (Köpke and Schmid 2004: 3; cf. also Ng and Wigglesworth

2007: 300: "Attrition is, of course, closely related in many ways to acquisition"). As de Bot and Hulsen (2002: 262) point out, "[l]anguages are never lost in isolation, and L1 attrition typically comes as a by-product of language contact, particularly in migrant settings" (cf. also Ecke 2004: 341).

These two observations have led to the integration of the perspectives also adopted in this publication, namely multi-competence (Cook 1995; 2002; 2003), and dynamic views of multilingual development (Hyltenstam and Obler 1989; Jessner 2003; de Bot 2007; de Bot and Schrauf 2009). A side-effect of this re-orientation is that some researchers now prefer using more neutral terms ("language maintenance", language change) or focus on relearning. Others have argued that attrition could be re-conceptualized as enrichment (Sharwood Smith 1983; 1983; Kaufman 1991; Ammerlaan, Hulsen et al. 2001), recognising the fact that the L1 does not just disintegrate and disappear but that elements of it are replaced with new elements.

There are several "milestone publications" charting the development of the field: Weltens, de Bot *et al* (1986); Seliger/Vago (1991); Hyltenstam/Viberg (1993); Extra/Verhoeven (1999), Hansen (2001), Ammerlaan, Hulsen et al.(2001), Schmid, Köpke et al. (2004) and Köpke, Schmid et al. (2007). This year has seen the publication of a special issue on first language attrition in the journal Bilingualism: Language and Cognition (de Leeuw, Schmid et al. 2010; Keijzer 2010; Ribbert and Kuiken 2010; Schmid 2010), and at least two more are in preparation.

However, even after almost 30 years of research, many questions concerning attrition have remained unanswered or have yielded conflicting evidence. To these belong, amongst others, the issues of how real a phenomenon language attrition (particularly adult first language attrition) is, and whether the changes observed constitute temporary difficulties in language processing or changed linguistic knowledge.

A major obstacle to the development of the field was identified in the early years of this decade (Köpke and Schmid 2004; Schmid 2004): the proliferation of different approaches to language attrition, and the fact that most empirical studies were rather small-scale meant that it was very difficult, if not impossible, to compare findings across different studies. In response to this problem, the Graduate Network of First Language (henceforth "Attrition network"/"Network") was launched in Amsterdam in 2002, with the aim of developing "a common and crosslinguistically

applicable research design" (Schmid 2004: 352), which would then lead to a "common 'pool' of language attrition" data to be made available across the Network (ibid.: 354).

The present author was a member of the Network from the outset, and the work presented here has benefited from the collegial exchange it afforded in many ways. The Network did devise a test battery, the Language Attrition Test Battery (Schmid 2004, henceforth LATB) that is available for future generations of researchers and was in part deployed in this research.

### **2.1.2. Points of reference**

In L1 attrition studies, there are problems in designating the people under investigation, and in finding appropriate controls, arising out of the fact that L1 attriters are at once similar and different in more than one way from other native speakers of their language. Unlike child attriters, whose process of L1 acquisition is halted or reversed, adult L1 attriters were at one point 'fully proficient' speakers of their L1. However, they have lost some of the competence by comparison with other adult native speakers.

On the other hand, L1 attriters, for the most part, become active users of an L2. This knowledge of other languages, coupled with the fact that they live or have lived abroad, potentially impacts on a whole range of dimensions such as cognition, attitudes and language use, along which attriters may not only differ from one another, but as a group differ from 'the' native speaker, making valid comparisons difficult.

Thus, once the potential attrition context has been established for a particular study, referring to the participants as attriters seems a rather negative way of looking at the processes of change to their language(s), not least because the term implies that L1 attrition has in fact taken place while ignoring the (potentially positive) changes in the L2. It would therefore be preferable to take into account the fact that our participants speak a language or languages other than their L1.

An obvious denomination to use would be "bilingual". However, the term is used in different senses and has attracted a number of different sub-divisions, leading to Baker (2001: 15) concluding that "[d]efining exactly who is or is not bilingual is essentially elusive and ultimately impossible".

For example, "bilingualism" has recently been contrasted with "trilingualism" and "multilingualism" (cf. for example Cenoz and Genesee 1998; Cenoz and Jessner 2000; Cenoz, Hufeisen et al. 2001) to denote the knowledge of specifically two languages. These researchers make the point that, in parallel to the differences between monolinguals and bilinguals (speakers of two languages), there are changes to the human mind and the language system as a whole arising out of the acquisition of further languages which are "not additive" (Jessner 2003: 236).

Ng and Wigglesworth (2007: 3ff., with reference to Valdés and Figueroa 1994) summarise the issue by presenting a classification of different descriptive dimensions. First, bilinguals may be distinguished by their degree of bilingualism. Different people hold different opinions of who "qualifies" to be called a bilingual. According to the "maximalist" definition of bilingualism (Bloomfield 1933; cited in Baker 2001; and Ng and Wigglesworth 2007), bilinguals are only those who possess native-like proficiency in two languages. This understanding is usually too restrictive and unrealistic, while the opposite view, the "minimalist definition" (Haugen 1953; Weinreich 1953; Mackey 1962) which includes anybody with just a smattering of phrases in another language (i.e. incipient bilinguals) is for most purposes too broad.

A related understanding is that of "balanced bilingualism", i.e. equal proficiency, vs. "dominant bilingualism". The former is often idealised as equal proficiency at the native level (Baker 2001: 7), although one could of course have a similarly low level of proficiency in both languages. Furthermore, in reality even in people who have highly developed skills in two languages these are likely to differ across domains of language use (Fishman 1972). Fishman also makes the point that societies are only likely to survive as bilingual or multilingual societies, if there is functional differentiation between the languages.

A second relevant criterion for our purposes is the distinction according to the age at which bilinguals acquired their languages, i.e. "early"/"child bilinguals" who grow up with two languages, vs. "late bilinguals" who start learning another language later in life, usually in an instructed context.

The distinction between different contexts of becoming bilingual is further of relevance here (cited in Valdés and Figueroa 1994: 11; Ng and Wigglesworth 2007). "Elective bilinguals" are those who choose to become bilingual, whereas "circumstantial bilinguals" happen to find themselves in a



bilingual context and have no choice in the matter. A similar distinction is that between "naturalistic" and "classroom bilingualism". One may assume that the reasons why one becomes bilingual impact on a number of factors, such as the effort one is prepared to make, one's motivation to acquire, the uses one has for the language etc., and thus ultimately on one's success in becoming bilingual.

Finally, another relevant distinction is between bilinguals who have reached a (more or less) stable level of proficiency in their languages, and those who are still developing with respect to some languages and the overall system.

Given the multitude of meanings and uses of a single term, and the futility of attempting to define bilingualism, Ng and Wigglesworth (*ibid.*) recommend describing the dimensions relevant in each case. Thus, the bilinguals in this study are bilinguals in the sense that they know more than one language. They are late bilinguals, which was a criterion for eligibility. They are also circumstantial bilinguals since they live in the L2 context where they needed to become bilinguals, although there is presumably an element of volition as well, since all bilingual participants emigrated voluntarily. As a result of their prolonged residence in the country, they can all be classed as stable bilinguals.

We did not a priori define the degree of bilingualism (i.e. whether participants are balanced in both, or dominant in one language) participants were supposed to have; they were merely required to meet a minimum level of L2 proficiency. In fact, identifying their degree of bilingualism is one of the aims of this investigation.

We would like to mention two more sets of terms that are current in referring to individuals who know more than one language. The first of these, the terms "foreign/second language/L2 learner", is specific about the fact that the person involved has not started to learn the language in question from birth, but at some later stage. It stresses the dimension of ongoing language learning.

"Foreign language" refers to such learning primarily taking place outside a country/area where that language is naturally spoken, and thus usually implies an instructed context, whereas "second language"/"L2 learner" usually refers to a language contact situation. In this context "L2" may, but does not necessarily, make reference to the place of a particular L2 in the chronological order of

acquired languages by an individual (it could actually be the L4), the order of proficiency levels or the frequency with which it is used.

The second set, the terms "L2 speaker"/"L2 user", focuses on the fact that people make use of languages for communicative purposes, rather than the learning dimension. L2 speakers/users are distinct from L2 learners who "are acquiring a system for later use" in that they exploit "whatever linguistic resources they have for real life purposes" (Cook 2002: 2). Again, in contrast with some uses of the term "bilingual", the question of the actual proficiency level is unspecified. Thus, an L2 user is "any person who uses another language than his or her first language (L1), that is to say, the one learnt first as a child" (Cook 2002: 1).

Thus, while we will refer to our main group of participants as "bilinguals", we may additionally use "L2 user" and "L2 learner" in contexts where those relevant aspects of their persona are focused on. We may also refer to our bilingual participants as 'immigrants'/'migrants' in appropriate contexts. In contrast, with regard to the controls in this study, it is appropriate to use "L2 users" or "L2 learners", although they are also "bilinguals" in the sense of knowing more than one language with a specified degree of proficiency (cf. Methodology).

### **2.1.3. Native speakers vs. multi-competence**

A logical extension of how to denote people with knowledge of more than one language is the question of how to measure proficiency in the L2, and by extension in the L1 of potential attriters. The traditional measure has been the standard of the monolingual native speaker, idealised as "the native-speaker norm". The reason for using such a norm lies in first language acquisition being viewed as being qualitatively different and superior to second language learning. However, there are several problems with this approach, partially in relation to the notion of "norm", and partially with regard to the appropriacy of using monolinguals for the purpose of comparison.

First, it is increasingly being recognised that there is variability in the proficiency level native speakers achieve in their L1, just as there is variability in L2 achievement. As Jaspaert, Kroon et al. (1986: 43f) put it,

‘the L[inguistically] C[ompetent native speaker], just like his home base, the homogenous speech community, is a hypothetical entity that does not reflect everyday linguistic reality. Full [native language] competence simply cannot be defined in an absolute sense, without taking into account the sociological, geographical and historical position each language user is in’.

This may be less obvious with regard to grammar and syntax (any native speaker could be expected to form correct questions, for example) than in relation to the lexicon where greater variability may be found. It is true to say that the variation between (non-attributed) native speakers is less pronounced than between L2 learners, whose achievement is characterised by large inter-individual differences and the common perception of failure. However, it is sufficiently big for the possibility that an individual L2 user may outperform a particular L1 speaker in the L1 of the latter. Moreover, the implicit assumption that L1 proficiency is immutable once acquired is seriously challenged by work on first language variability, L1 attrition and the impact other languages might have on L1. This implies that the benchmark for comparison of either L2 proficiency or attriting L1 proficiency should be typical, rather than assumed normative behaviour.

The second problem with such a normative view is that it implies that second-language learners are deficient by definition and can never achieve native-like proficiency in their L2. Indeed, even extremely proficient L2 users are at best referred to as ‘near-native speakers’. While research has indeed uncovered important differences in L2 users vis-à-vis native speakers of these L2s in terms of acquisition patterns, processing, ultimate attainment and durability of L2 knowledge (cf., for example, Hyltenstam 1992; Hyltenstam and Abrahamsson 2000; Bialystok 2001; Cook 2003; Han and Odlin 2006), there is growing evidence for the capacity of some, even late, L2 learners to reach native-like proficiency (e.g. Birdsong 1992; White and Genesee 1996; Bongaerts 1999; Bongaerts, Mennen et al. 2000; Flege and MacKay 2004; Birdsong 2006).

Cook (2003) argues that there is no intrinsic reason why L1 and L2 ultimate attainment should be the same. He also makes the point that the ability to ‘pass as a native speaker’, which implies performing at the upper level of native proficiency in respect of every last detail, should not be the reason for acquiring an L2, but rather the intention to ‘become an L2 user’, and thus it would be far more appropriate to acknowledge L2 users’ achievements in so doing:

‘One group of human beings should not judge other people as failures for not belonging to their group, whether in terms of race, class, sex or language. People should be measured by their success at being L2 users, not by their failure to speak like native speakers.’ (Cook 2003: 3f)

By the same token, monolinguals could be seen as people who have failed to learn other languages. Finally, the monolingual native-speaker standard is also flawed because evidence is growing that the fact of acquiring a second language leads to changes in the L1 as compared to the L1 of monolinguals: "the first language of people who know other languages differs from that of their monolingual peers in diverse ways" (Cook 2003: 1). Thus, Cook (2002: 4-8; Cook 2003: 5) cites evidence for L2 knowledge impacting on L1 in the following areas: phonology, vocabulary, syntax and reading. He also points out that knowledge of L2 has cognitive effects which go beyond simply knowing another language.

Some of these may be positive, others neutral or perhaps negative. Positive effects cited by Cook in the above works include greater flexibility in thinking and more efficient processing as a result of brain-training, higher language awareness and metalinguistic skills, faster learning to read in L1 and better L1 communication skills. In other words, the linguistic competence of L2 users ought to be seen as a competence qualitatively different from that of monolingual speakers.

While individual competencies of L2 users thus differ – positively or negatively – from those of monolinguals, the combined language proficiency of an L2 user in any case exceeds that of a monolingual and is more than the sum of its parts.

These considerations prompted Cook to propose the notion of *multi-competence*, ‘the knowledge of more than one language in the same mind’, which is ‘not just the imperfect cloning of mono-competence, but a different state’ (Cook 2002: 7f). Consequently, L2 users constitute a group of language users distinct from monolingual users who should be studied in their own right, and their performance – both in L1 and L2 – assessed against appropriate "typical", and possibly bilingual, baselines.

## **2.2 Processes of language attrition and acquisition**

### **2.2.1. Language attrition - findings**

As pointed out above, attrition has been studied from a variety of perspectives; yet, evidence about many facets of the process of L1 attrition remains contradictory. In this section, we consider some of the research findings relating to the whether, what, how, and why of L1 attrition in more detail and in keeping with the general psycholinguistic interest of this research.

To start with, the very possibility of attrition has only been demonstrated beyond any doubt for two populations. The first of these are child bilinguals, who often completely abandon their original L1 in favour of exclusive L2 use when placed in an L2 environment (cf., for example, Mägiste 1979; Leyen 1984; Kaufman 1991; Schmitt 2000). Child adoptees, a special case amongst child bilinguals due to potential emotional trauma and the absence of continued L1 support, have been shown to lose their L1 to the point where no trace was detected even with very finely tuned brain imaging devices (Pallier, Dehaene et al. 2003; Ventureyra, Pallier et al. 2004). The other population are L2 users losing their L2 (Hansen 1999).

Research into the L1 attrition of adults, on the other hand, has so far provided a rather ambiguous picture. Research reported in the literature ranges from finding ‘surprisingly little loss’ (de Bot and Clyne 1989: 87) despite self-perceptions of significant first language decline (Jaspaert and Kroon 1989; Jordens, de Bot et al. 1989; de Bot, Gommans et al. 1991; Jaspaert and Kroon 1992; de Bot and Clyne 1994; Schoenmakers-Klein Gunnewiek 1998) to reports of definite signs of attrition on some or even all linguistic levels (Leyen 1984; Hiller-Foti 1985; van Ness 1988; Altenberg 1991; Olshtain and Barzilay 1991; Major 1992; Ammerlaan 1996; Waas 1996; Köpke and Nespoulous 2001; Schmid 2001; Pavlenko 2003). Thus,

‘[a]fter two decades of first language attrition study’ researchers still disagree ‘*whether* a first language in which a certain level of proficiency has been reached can ever undergo significant attrition, let alone *how* or *why* it might’ (Köpke and Schmid 2004: 1).

Opinions range from viewing L1 attrition as "a ubiquitous phenomenon found wherever there is bilingualism" (Seliger 1991: 227) and as "an all-pervasive problem not restricted to a very limited number of sociocultural scenarios" (Sharwood Smith 1983: 229), to the expectation of stable,

unchangeable L1 proficiency: "In cases other than language pathology, we do not expect an established L1 to deteriorate or diverge from the grammar that has been fully acquired" (Seliger 1996: 616, quoted in Herdina and Jessner 2002: 95).

There are a number of possible explanations for these diverging results. First, differences in methodology rather than the absence of the proposed phenomenon per se may have caused this confusing picture (Köpke and Schmid 2004). For example, some of the studies used tests specifically designed to tap linguistic knowledge, taking care to eliminate factors (to the degree that that is possible) that would impact on processing load (e.g. Jaspaert and Kroon 1989). However, it is now thought that attrition starts as or constitutes primarily a processing or control phenomenon (Sharwood Smith and van Buren 1991; Hulsen 2000), so a verdict of 'no loss' may be true only at the knowledge level.

Further, given that L1 attrition appears to be a rather slow, inconspicuous process, some studies may have failed to uncover evidence for its existence if the instruments used were not sufficiently focused. Moreover, a lot of work has been done on typologically close languages, which may have some part to play in either minimizing or disguising language loss (Köpke and Nespoulous 2001). Conducting the same investigation with a different set of languages (all other things being equal) might yield very different results.

Also, some work still remains to be done to provide a 'workable and testable definition of language attrition - or of what would constitute evidence for its presence' (Köpke and Schmid 2004: 8) which obviously is a precondition for identifying attrition in the first place.

Second, apart from methodological reasons, the above results point to the fact that a disparity exists between perceptions of attrition, as, for example, established by self-assessment instruments, and attrition as established by formal testing. If measured attrition is significantly lower than the reported level, 'contradicting all [...] intuitions on the 'undoubted' and unquestioned existence of the phenomenon we label attrition' (Köpke and Schmid 2004: 7), this may be due to participants overestimating the extent of an unexpected, and therefore quite traumatic, process.

Just how this process is experienced, presumably depends on how one construes first language proficiency – the same for all, fixed and immutable or varying from individual to individual and generally changeable.

There is some scientific support for the 'immutable proficiency' view: Some psycholinguists have proposed a 'threshold of frequency of use' and/or proficiency level beyond which knowledge becomes immune to loss (de Bot 1998: 351), a 'critical mass of language that, once acquired, makes loss unlikely' (Pan and Berko-Gleason 1986: 204, quoted in de Bot and Hulsen 2002: 260). In second language attrition research, high proficiency coupled with an age factor has been pinpointed as a good predictor for language retention (van Els 1986: 9; de Bot and Hulsen 2002: 259; Herdina and Jessner 2002: 97).

The implication for first language attrition is that it should not happen under normal circumstances since most people attain "full" competence in their L1. However, evidence is mounting that

'neither first languages nor second languages are immune to loss. With non-use they fade, and though they keep their place in our memory system, they become less accessible up to the point where the knowledge has sunk beyond reach and is for practical purposes lost' (de Bot and Hulsen 2002: 253)

This then would mean that the issue of L1 attrition is not so much a categorical question (does or can it happen?) as one of degree (how much of it happens). The evidence from the literature would indicate that the actual incidence and rate of adult first language attrition is relatively low in comparison with child L1 attrition, L2 attrition and an affected person's perceptions.

Neither the finding that adult L1 can attrite, nor that the attrition rate is relatively low, should come as a great surprise. From a psycholinguistic perspective, forgetting is a natural process which potentially affects all types of knowledge, including linguistic knowledge: the more time has passed since something was learnt (and practiced), and the more other knowledge is acquired, the harder it is to recall the desired information (Herdina and Jessner 2002: 94). Therefore, it is plausible that L1 should attrite in circumstances when it is not practised.

On the other hand, by the time one has grown up, the L1 is an extremely well-established system and extremely well-practised skill that, our intuition tells us, would require severe neglect probably over many years to become totally unusable. These days, even in an immigration situation, most people have and use the opportunity to keep their L1 active. It is quite possible that most simply do not live long enough for the extreme case (the collapse of the native-language system) to occur.

Herdina and Jessner's (2002) Model of Multilingualism provides specific explanations for these effects, which we will turn to when reviewing the model below.

In addition to the question of who may be affected by L1 or L2 attrition, a further important issue is of course whether language attrition proceeds as a principled, "rule-governed process" (Seliger 1991: 28), and whether there are parts of the language system that are more vulnerable than others. Depending on the theoretical framework employed, many of which, such as generalization, simplification and reduction of markedness (Seliger 1991; Clyne 1992; Jaspaert and Kroon 1992), have been drawn from language acquisition research, different general processes and changes to particular linguistic features have been identified that appear to constitute attrition.

Simplification is assumed to arise out of attriters' need to ease the 'burden on the language-processing mechanism', which results from the 'the new business of coping with two languages rather than just one', and thus strive for 'overall processing simplicity' by adopting into either L1 or L2 structures from the other language which are relatively easier to process (Sharwood Smith 1983: 226). In addition, Clyne has found evidence for changes in attrition to follow the path of or accelerate 'drift' (movement of a language in a particular typological direction) (Clyne 1992; cf. also Kadt 2001), thus linking language attrition in individuals or synchronic language change with change in the system of the language over time (diachronic language change).

However, Köpke and Schmid (2004: 16) argue that a discussion of attrition in terms of simplification can only have explanatory power if related to theories of complexity which are often lacking in favour of a 'random enumeration of what complexity might refer to' (ibid.: 16) (changes from synthetic to analytic form, structures requiring more processing time, late acquired structures...), and refer to the recent employment of models such as Myers-Scotton/Jake's (2000) 4-M-Model (Bolonyai 1999; Schmitt 2000) as welcome departures.

The reference to markedness in particular means that attrition is also a 'selective phenomenon' (Köpke and Schmid 2004: 15), in that it does not affect all structures to the same degree. Although it has been shown that all linguistic levels are vulnerable in principle, there is evidence for the lexicon being more vulnerable than morphosyntax (Schmitt 2000; Köpke and Nespoulous 2001). Within those levels, it is low-frequency and unpredictable ('marked') items and, within grammar,



borderline cases which appear to be affected by attrition soonest (Leyen 1984; Altenberg 1991; Köpke and Nespoulous 2001).

Employing these frameworks, attrition was evidenced in case mergers (dative and genitive); noun-pronoun and noun-article agreement, a reduced number of allomorphs, leading to greater paradigmatic regularity; inflectional morphology of verbs; replacement of synthetic forms by analytic ones or by periphrastic constructions, leading to a less flexible word order and preference for coordinated constructions over embedded ones; syntax in some types of relative clauses as well as loss of plural and gender (Altenberg 1991; Obler and Mahecha 1991; Kadt 2001).

Relating to the linguistic levels of the lexicon and morpho-syntax, robust findings suggest an implicational hierarchy for the relative degree of vulnerability of each level (e.g. Köpke and Nespoulous 2001). Thus, the lexicon is affected earlier and to a greater extent than morpho-syntax. In relation to language skills, empirical proof points to productive skills in L1 being sooner and more severely affected by attrition than receptive skills (Hulsen 2000). This mirrors the development of language skills in L1 and, to a degree, L2 where productive language use sets in only after a lag time (silent period) in which language is acquired receptively. Both findings may be explained with recourse to models of language processing, such as the Activation Threshold Hypothesis reviewed below.

Traditionally, the perspectives just discussed focus on ‘deviant linguistic behaviour’ found in people affected by attrition vis-à-vis ‘normal’ (norm-oriented) language use. However, it is increasingly becoming clear that it is not sufficient to simply identify errors in language use but that a thorough analysis needs to include what is there (positive and negative evidence) as well as what is not, since speakers may, particularly in production, avoid structures they do not feel confident about (Köpke and Schmid 2004: 29). This also means taking into account paralinguistic features such as hesitations and disfluencies (Ammerlaan, Hulsen et al. 2001; Hansen 2001; de Bot and Hulsen 2002).

In considering the types of changes and linguistic items most likely to be affected by attrition, we must make reference to a debate (Sharwood Smith 1983; 1983; 1989; Sharwood Smith and van Buren 1991; Hulsen 2000; Jessner 2003), which centres on the question of whether attrition affects

the knowledge of a language itself ("competence") or merely access to or control of this knowledge ("performance").

Attrition at the level of linguistic knowledge entails restructuring or atrophy of stored presentations (cf. Seliger and Vago 1991, Grosjean and Py 1991). At the level of performance (or processes), attrition results in difficulties in control of that knowledge (which may remain intact) (Ammerlaan 1996: 18)' (Köpke and Schmid 2004: 21). Sharwood Smith describes attrition at the performance level as "deviant or unusual use of psycholinguistic 'traffic rules' working on *established* knowledge" as opposed to "actual changes in underlying competence, i.e., transformed knowledge" (Sharwood Smith 1983b: 224f).

Sharwood Smith and van Buren (1991: 19) point out the possibility that all that will ever be lost in L1 attrition is the access to knowledge while competence itself remains unchanged. In other words, "pure" knowledge would be filtered through increasingly worsening performance. However, as they acknowledge, this is an untestable assumption, since all testing involves performance aspects to a greater or lesser degree which can be consciously minimized but never totally eliminated. In fact, some researchers claim that 'attrition is only traceable through performance' (Jessner 2003: 238).

In their discussion of the methodological implications of trying to disentangle knowledge from processing, Sharwood Smith and van Buren (ibid.) suggest that in order to test linguistic knowledge, a combination of tests which do not test fluency at the same time should be used, minimizing performance effects.

Meanwhile, 'it seems sensible to assume that attrition may occur at both levels, competence and performance' (Sharwood Smith and van Buren ibid.). Elsewhere Sharwood Smith (1983: 225) in a similar vein calls for 'adopt[ing] the working assumption that competences are indeed permeable whether they are complete language systems or interlanguages. Both manifest language transfer'.

He consequently proposes three (two) stages in language attrition – a first stage at which speakers exhibit systematic performance deviations while competence is stable; through a transitional one at which a new 'new externally conditioned variety' of L1 starts to emerge while the speaker is still able to switch back to the old one; to a third stage at which speakers employ qualitatively different competence which may or may not entail a decrease in the speaker's repertoire of structures. These

changes are proposed to be following UG laws (Sharwood Smith 1983: 51; cf. also Sharwood Smith and van Buren 1991).

Current psycholinguistic research, on the other hand, has started to build up evidence and arguments in favour of the position that attrition constitutes reduced accessibility of knowledge and that nothing is ever actually forgotten (Sharwood Smith 1983: 224, citing Penfield). An early study which seemed to support the idea that knowledge is immutable was Fromm (1970, cited in Köpke and Schmid 2004: 21) in which adults under hypnosis remembered a long forgotten childhood language.

This would imply that attrition is a performance rather than a competence phenomenon. Hulsen, de Bot et al. (2001: 162) thus state that

‘[t]he central thought for the psycholinguistic approach to research into language loss is that language loss is in fact nothing more than a decrease in the ability to have immediate access to a word in production and perception’.

While the findings of Fromm's hypnosis study have not been replicated or confirmed by similar studies, the development of more sophisticated methods, such as those within the ‘savings’ paradigm, have made it possible to show up residual or sub-threshold knowledge. This could be quite a promising approach which could go some way to advancing our understanding of whether in attrition we are really dealing with changed knowledge or with impaired access but as yet its systematic application to L1 attrition research is still outstanding (Hulsen 2000: 188). However, given the methodological difficulties in addressing this question, the jury will probably continue to be out for some time to come.

### ***2.2.2. Learning and forgetting***

As Köpke and Schmid (2004: 16) note, "the idea that attrition might be governed by a process that has its roots in the overall cognitive nature of memory, and only indirectly in the linguistic system" has kept its appeal for quite some time in L1 attrition research. This section takes a psycholinguistic perspective on language change processes, which ties language learning, loss and maintenance in with general cognitive processes.

Such a perspective implies that both language acquisition and attrition need to be studied in conjunction since both are subject to particular processing and storage capacities. Furthermore, it moves away from asking whether and which linguistic items of one system may infiltrate and ‘contaminate’ another one by trying to explain processes such as borrowing and code-switching with reference to ‘third-party’ or outside causes common to all facets of language and other learning, for example, processing limitations or psychological salience. Consequently, from a psycholinguistic perspective L1 attrition is hypothesised to result from a need to lessen processing load in accessing linguistic information and the ensuing retrieval difficulties (Ammerlaan, Hulsen et al. 2001).

The main proponents of a psycholinguistic approach to L1 attrition are de Bot (1998; de Bot and Hulsen 2002), Hulsen (2000; Hulsen, de Bot et al. 2001), Ammerlaan (Ammerlaan, Hulsen et al. 2001) and Köpke (Köpke and Nespoulous 2001). We also refer the reader to Ecke (2004) who provides an extensive overview of theories of forgetting. The following section reviews some of the main concepts and assumptions underlying this approach.

Since both language learning and attrition take place in the same mind, it seems prudent to consider the types of linguistic knowledge and the way they are accessed in language use. Due to space constraints, we will not, however, review models of bilingual architecture, such as the Revised Hierarchical Model (Kroll and De Groot 1997), since these focus on representation with less relevance for learning and forgetting processing.

A common distinction is made between declarative and procedural knowledge/representation (memory), and it is the latter which is particularly important for multilingual proficiency as we shall see. Other distinctions include those between controlled and automatic processing, explicit vs implicit knowledge as well as conscious awareness vs. unavailability to introspection (O'Rourke 2002: 71f). Pairs of these are often taken to be synonymous or associated with one another. For example, (Paradis 2001: 10f) associates

- implicit competence – implicit acquisition – unconscious storage – procedural memory – automatic processing (without conscious control) on the hand, and
- metalinguistic knowledge – conscious acquisition – explicit storage – declarative memory – controlled processing on the other.

O'Rourke (2002: *ibid.*) argues that this oversimplifies matters: 'there are logical connections amongst such concepts but [...] they belong to distinct levels of analysis'. Declarative knowledge, for instance, may be accessed with different degrees of control or automatization, and, following Karmiloff-Smith (1986; 1992), explicit knowledge may be conscious or unconscious, but a greater degree of explicitness does not necessarily imply greater consciousness (O'Rourke 2002: 88f).

Ullman's (2003) Declarative/Procedural Model posits that the mental lexicon is stored in the declarative memory system while the mental grammar 'which underlies the computation of complex words, phrases and sentences' is dependent on the procedural system. Moreover, irregular ('unpredictable') grammatical forms are also stored in declarative memory. In other words, declarative knowledge is comprised of a great number of essentially arbitrary items. These are used more or less frequently, and depending on their frequency they may be accessed by more or less controlled processes.

Proceduralised knowledge, on the other hand, is comprised of 'rules' or 'mini-programmes' for the computation of regular grammatical form, of which there may be fewer than items in declarative knowledge but which are invariably accessed by automatic processes (since they are constantly drawn on in language production). Frequency is thus not really an issue for proceduralised knowledge. Ullman's data further show that in late L2 acquisition, learners 'initially rely on lexical/declarative memory for the processing of complex forms that can be compositionally computed by the grammatical/procedural system in L1', while higher-proficiency L2 learners are able to learn those forms additionally in the procedural system (Ullman *ibid.*).

A widely accepted view on language acquisition is based on Anderson's (1983/1995) model which proposes that language learners

- a) acquire declarative (propositional) knowledge which is processed in a controlled and initially conscious manner,
- b) convert declarative to procedural knowledge ('proceduralization') and
- c) increasingly automate the processing of procedures, which involves the ever faster processing of increasingly larger units with less and less conscious control.

This process of moving from controlled information processing corresponding to a declarative stage to automatic information processing (a procedural stage) presumably happens at different

times for different parts of knowledge depending on the recency of acquisition, so that the three stages may actually co-exist in the learner's mind at all times. Karmiloff-Smith's model of Representational Redescription (Karmiloff-Smith 1986; O'Rourke 2002: 65f) describes an alternative or additional process whereby representations that are initially implicit become increasingly more explicit (abstracted) and eventually conscious until they are reportable as metalinguistic knowledge. Similarly, competence or skill development has been described in terms of increased automatization (Shiffrin/Schneider 1977, Shiffrin/Dumais 1981, Schneider/Fish 1982 cited in Segalowitz 1991)

Drawing on Anderson's model, de Bot (1998) argues that if attrition is seen as the reverse of acquisition it should involve knowledge changing state in the opposite direction, i.e. becoming 'deproceduralized' by moving from the procedural to the declarative state, and from automatic to more controlled processing, and predicts the greater vulnerability of declarative knowledge over procedural knowledge (de Bot 1998: 350f.). Empirical evidence for controlled processes being subject to attrition, whereas automatic ones are not comes, for example, from Segalowitz (1991: 74).

It seems plausible that arbitrary items, particularly if they are used infrequently, should be more vulnerable than highly automatized procedures since there is no predictable way to locate these items in memory. Of course, deproceduralisation should theoretically happen to rules as well. However, because of their very nature these are used and 'practised' frequently even in contexts of reduced maintenance.

Turning now to the model of the Activation Threshold Hypothesis (Paradis 1994; 2001; 2007), this provides for differential access to arbitrary, but interconnected, items depending on their activation level. In order to be retrieved, an item needs to surpass a certain activation threshold. Items may be phonological, lexical or any other type, and since retrieval mechanisms are ignorant of linguistic distinctions this would explain why language attrition may potentially happen across all linguistic levels.

If an item is used frequently, its threshold is constantly lowered, making retrieval ever faster and less effortful. Items may have 'competitors' of a similar or higher activation level which may interfere in the retrieval of the desired item. However, if an item is low-frequency to start with,

then lack of maintenance will result in fewer opportunities for them to be activated, particularly in the context of 'stronger' competitors, and in their falling below the minimum activation level required for retrieval.

There is one feature associated with a network of interconnected declarative knowledge items which may save them from becoming truly irretrievable: they may be accessed by a number of different routes (associations), although these 'access routes' may be more or less available. This would not appear to be the case when accessing proceduralised knowledge which can only be accessed by direct route: if they can be located, they will run by themselves. So attrition of proceduralised knowledge may not so much constitute its progressive deproceduralisation but simply whether the routines run or not (Ullman, p.c.). In summary, low-frequency lexical items and low-frequency irregular, unpredictable grammatical features should be particularly prone to attrition which has indeed been borne out by some of the research presented above.

### **2.2.3. Crosslinguistic effects**

As we pointed out before, L1 attrition does not usually happen in isolation, but in the context of another language being acquired: "L1 is generally replaced by another language, and this language is often assumed to influence the process of L1 attrition" (Köpke and Schmid 2004: 17). In keeping with the multi-competence view outlined above, we conceive of the languages of an L2 user not as independent, stand-alone systems, but as being contained in and interacting within the same mind, and amounting to something more than purely the knowledge of L1 + L2 + Ln.

Such a holistic view of the L2 user's mind implies that each of the languages may impact on the other(s) (not just L1 on L2) and that changes to parts of the combined system may be reflected in other parts. This section examines the evidence available for establishing different kinds of such interactions, amongst which, in line with the interest of this investigation, we are particularly concerned with the impact that L2 may have on L1.

As a precondition for such interactions the languages of an L2 user must not just all be contained within his/her mind but also interconnected at some level. In other words, L2 users combine

‘knowledge of two or more languages in one mind’ which ‘must form a language super-system at some level rather than be completely isolated systems’ (Cook 2003: 2).

Cook (2002: 10ff; and 2003: 6ff) discusses the theoretical possibilities for the organisational relationship between the languages in an L2 user’s mind which range from separation through interconnection to integration along a so-called ‘integration continuum’. He concludes, citing relevant evidence, that different areas of language (vocabulary, grammar, phonology) may be integrated to different degrees - while in some areas total separation is possible, some degree of integration is likely in others. For the overall system we have to assume interaction, rather than total separation or integration:

‘Total separation is belied by the use of the same mouth and ears for both languages; total integration is denied by the L2 user’s ability to keep to one language at a time’ (Cook 2002: 12).

While Cook is careful to distinguish between ‘abstract properties’ (Cook 2002: 19) of the languages in question, and the actual physical storage of the languages in the brain, he is hopeful that new departures in neurological research may uncover parallels between these two.

Indeed, apart from the different language levels, dimensions such as the typological relationship between the languages, the psycholinguistic level (conceptual vs. linguistic), interpersonal differences, and finally stage of language development/attrition of each language (i.e. proficiency level) may impact on the degree of interconnectness or integration. The interconnection assumption is consistent with and related to Paradis’s (2001) model of bilingual organisation (Subset Hypothesis), Kecskes and Papp’s (2003) Common Underlying Conceptual Base, and Herdina and Jessner’s (2002) Dynamic Model of Bilingualism, and is supported by research in neurolinguistics.

As was pointed out before with reference to Cook (2003), some effects of knowing more than one language have been shown to be beneficial for the L2 user, whereas others may be negative or simply neutral since some aspects of multilingual performance escape description in terms of being better or worse (for example, differences in aspects of pronunciation). We will now turn to the effects of the interaction between language systems, which we broadly class into the categories of transfer and mutual enhancement as sub-categories of crosslinguistic influence (Sharwood Smith and van Buren 1991), and competition.



Considering what brings about language attrition in a language contact situation, Sharwood Smith and van Buren (ibid.) ponder, reiterating some of the points we made above:

"The changes that occur in language attrition would seem rather to be accompanied by the lack of relevant data. This is interesting because it is usually assumed that mature native languages are typically stable as opposed to interlanguages or developing L1 systems, which are characterized as typically unstable. [...] It would appear from this assumption that once attained, the mature L1 is "fixed" and needs no further input either to disconfirm faulty learner hypotheses or to maintain its final state. [...] Why then should attrition occur? There are two obvious candidates: (a) L1 deprivation; (b) crosslinguistic influence from another language being acquired." (Sharwood Smith and van Buren 1991: 22)

In other words, L1 attrition, and by extension L2 acquisition, may be influenced by two distinct processes - an intra- and an interlinguistic one (Jaspaert and Kroon 1992: 139), or an "internally induced" and an "externally induced" one respectively (Seliger and Vago 1991: 7ff). The first kind is seen to be "motivated by universal principles or related to some fact in the particular grammar of L1", or, in different conceptions of learning and forgetting, general cognitive principles of language restructuring.

In the second, which has variably been called "transfer", "interference", "convergence", "crosslinguistic influence" or "interlingual effects" (cf. Seliger and Vago 1991: 7), an L1 element/form/construction is patterned in analogy with L2 (Seliger and Vago 1991: 7ff). The concept of interference or transfer was defined by Weinreich (1953: 1) as "those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language".

The study of transfer from L1 into L2 has a very long tradition and came to inform a whole branch of second language acquisition research (contrastive analysis). For a long time, transfer (often called "interference") has been seen as the main culprit for L2 users' "failure" to become native-like in L2. However, transfer is a natural and "powerful process that can already take the learner deep into the new system" (Littlewood 2004: 505).

Even though learners sometimes transfer inappropriate knowledge, relying on transfer speeds up and eases the second-language acquisition process. The problem is that there is no way of knowing

when one can and should use pre-existing knowledge, and when it is time to switch to an all-new 'page'.

The opposite influence of L2 on L1, on the other hand, which might be termed "reverse" or "backward transfer" (Cook 2003: 1), or in Schmid and Köpke's (2007) words, *bi-directional traffic*, has only recently begun to be studied, perhaps because this influence tends to be less obvious.

"It becomes blatant only when the first language starts to disappear, for instance when a speaker brings more and more L2 words into his or her first language" (Cook 2003: 1).

Sharwood Smith (1983; 1989) proposed the term "crosslinguistic influence" (CLI) to "cover more than the word "transfer" and includes borrowings, influence on L1 from L2, avoidance of transfer, etc. In other words, it covers a fairly wide range of phenomena" (Sharwood Smith 1989: 185), which we adopt. Thus, one form of CLI which partially allows the L2 user to make positive use of their languages is code-switches, the "alternate use of two languages within the same discourse" (Poplack 1980; cited in Hyltenstam and Stroud 1993: 240).

Code-switching covers a range of phenomena which may be distinguished as deliberate, even purposeful language mixing on the one hand, and involuntary intrusions (interferences) on the other ("functional switching" vs. "derailment" in Hasselmo (1961; cited in Hyltenstam and Stroud 1993: 240). Some researchers prefer to include only such intentional code-switching under code-switching proper (Baker 1996), while others consider purposeful and involuntary code-switches to be varieties under the same cover term (cf. Poplack's 1980 definition above). Nakamizu (2003: 81), for example, defines the term "code-switching" broadly as switches into the other language which are "followed by further communication" in that language.

In relation to the former, Baker (1996: 87f) lists 13 different purposes or aims of code-switching, such as filling lexical gaps where no cultural equivalent exists or a bilingual does not know the corresponding word in a particular domain, communicating changes in social relationship, introducing humour, indicating authority etc. (cf. also Gumperz 1982 for an oft-cited classification of code-switching types).

Involuntary code-switches, on the other hand, have been ascribed to "momentary inattentiveness" (Hyltenstam and Stroud 1993: 240) and are of particular interest for attrition studies. Hyltenstam

and Stroud, with reference to Clyne (1987), describe this type as "psycholinguistically conditioned" as opposed to "sociolinguistically conditioned switching" (Hyltenstam and Stroud 1993: 240).

There are many factors influencing whether code-switching will occur, and to what extent. In some communities, it is a normal and accepted form of communication (the "unmarked choice of language", Baker 1996: 86). In others, code-switching is not encouraged. Further, individuals differ in the extent to which they are amenable to deploying code-switching.

The language mode operative during a conversation (Grosjean 1998), itself subject to a multitude of factors such as the relationship of the interlocutors, their relative proficiency in each language, the social norms, the topic being discussed etc. also plays a crucial role (cf. Section 2.3.1 for a fuller discussion of the concept). Bilingual mode lends itself to code-switching if this is socially accepted, making it harder to distinguish between inadvertent and functional code-switching; monolingual mode, which for various reasons does not sanction the use of the other language (because the other interlocutor does not speak it sufficiently, or because it is inappropriate as in our case), does not. Language mode partially determines whether codeswitches are fluent or flagged as such (Treffers-Daller 1997).

A situation which forces habitual code-switchers to avoid that practice may cause disfluencies and possibly inadvertent code-switches. Drawing on the psycholinguistic literature of the bilingual mind, we may assume that this effect will be more pronounced in people who have achieved a high level of proficiency in the language that may not be used, since it involves rejecting readily available competitors from the other language (e.g., Paradis 1993: 142; Finkbeiner, Almeida et al. 2006; Blumenfeld and Marian 2007).

Regarding the question of whether intra- or interlinguistic processes predominate in L1 attrition, and indeed whether observed changes constitute "full-blown attrition" or are simply "the consequence of [the] subjects' bilingualism" (Köpke and Nespoulous 2001: 233), crosslinguistic influence appears to be most relevant. In a study by Japaert and Kroon, CLI clearly accounted for up ~ 80% and possibly more of the detected deviance, whereas only 10% were not attributable to the "bilinguality of the situation" (Jaspaert and Kroon 1992: 142) and thus had to be due to intralinguistic restructuring.

#### **2.2.4. Languages in competition**

We already referred to the possibility that crosslinguistic processes may play both positive and negative roles in L1 attrition and L2 acquisition. On the positive side, language systems may become mutually reinforced, or acquisition and use may be facilitated. Other effects deriving from the presence of two or more language systems in the same mind and the impact of language learning on more general thought processes fall broadly into the areas of enhanced/more divergent or creative thinking, higher metalinguistic awareness and better sociocultural (including metapragmatic) skills (Bialystok 2001; Cook 2002: 7f; Herdina and Jessner 2002: 60 f.; Jessner 2003; Jessner 2008).

It is in particular the second and third groups where positive interactions between the languages can be most directly established, since L2 users have been shown to possess better literacy-related abilities as well as conversational/pragmatic ones. However, it would appear that the advantages in the above areas do not apply across the board concerning linguistic competence or linguistic achievement, but become apparent only in connection with specific skills and tasks.

So far the discussion has concentrated on interactions between languages which might be seen as by-products of two (or more) sets of knowledge coming into contact with one another in the L2 user's mind. However, one of our interests is to see whether such interaction can lead to L1 attrition, i.e. whether there is a causal relationship between L2 acquisition and L1 attrition. Ecke (2004: 329f.) argues that "achieving bilingual or multilingual competence is by no means an all win situation. Frequently acquisition is achieved only at the cost of (partial) loss and divergence from monolingual norms".

This subsection thus centres on a third type of interaction which proposes a more direct link between the proficiency levels in L1 and L2 and which, for clarity's sake, will be discussed in terms of just two languages, although the points made are valid for any number of languages known by the L2 user. The central notion which we will employ is the one of competition between the languages, as proposed by Seliger and Vago (1991), which is also central to much work on language processing (e.g. MacWhinney 1997; Green 2003; Marian and Spivey 2003; Marian and Spivey 2003; Costa, Colomé et al. 2004):

"The languages spoken by the bilingual may be said, metaphorically, to coexist in a state of competition for a finite amount of memory and processing space in the mind of the speaker". (Seliger and Vago 1991: 4)

This makes sense if one follows some of the implications of the "1 mind – 2 (or more) languages" formula outlined above:

- a) Knowing more language(s) means the need to store and process **more information**.
- b) The mind has **finite resources** for processing and memory space.
- c) As a result, languages **compete** with one another.

Thus, as one language improves and takes up more memory and processing space, the other one(s) come(s) down, constituting attrition. This is in essence the so-called "balance theory" which likens changes in proficiencies to the movement of a set of scales or balloons and which assumes that one language "pushes" the other one(s) "out" on the basis that the overall balance in the mind has to be maintained (Baker 1996).

There is a certain amount of anecdotal evidence in support of this theory concerning the L2s of L2 users, who may realise that they can not keep all their L2s at a high level simultaneously: as soon as they turn their attention to one of them, the other(s) suffer(s). In living-abroad or immigration contexts this situation then may extend to involve L1 as well.

However, balance theory has been rejected on the grounds that it is too simplistic and even "naïve" (Baker *ibid.*: 134ff), and because the issue of limited processing and memory space is likely to be more complex. As Sharwood Smith (1983: 222) points out, "mother tongues do not automatically disappear as L2 structures (and words) are learned". It is unlikely that language size grows in a linear manner, adding more and more items. Instead, increases in proficiency are as much down to a growth in connections between items as they are to increases in automaticity. Moreover, if languages are interconnected at some level and transfer plays as big a role as it is claimed to, then some economies will surely result from tying in new with pre-existing knowledge.

Rejecting balance theory does not imply, however, that the notion of competition between languages for limited resources has to be abandoned. If one language does not "push out" the other, then the other logical possibility for coping with competing demands is to reach a compromise or "trade-off" between proficiencies at the macro level, and convergence on particular linguistic

features at the micro level. Moreover, it has been proposed that competition between the languages is at its highest at high levels of proficiency. Thus, Seliger and Vago (1991) specifically relate the interaction between L1 and L2 proficiency levels to very high levels of L2 competence: "The beginnings of first language attrition can be traced to advanced stages of bilingualism or second language acquisition." (Seliger and Vago *ibid.*: 4).

We will review empirical evidence in relation to this proposition in the section that deals with L2 proficiency as a predictor variable.

### **2.3 Models of language attrition and acquisition**

In the following section, we review two models of multilingual processing and functioning (the Dynamic Model of Multilingualism (Herdina and Jessner 2002) and the Activation Threshold Hypothesis (Paradis 1993). Both models very strongly stress the interconnection between the different language systems and claim each in their own way to be relevant to language attrition research.

ATH: "The hypothesis may account for recency and frequency effects, and priming phenomena, as well as language attrition." (Paradis 2001: 12)

"A differential activation threshold can also account for dissociations between comprehension and production in first and second language acquisition, language attrition, and amnesic aphasia." (Paradis 1993: 139)

DMM: "[T]he process of language attrition will be discussed as one of the key features in the development of a multilingual system." (Jessner 2003: 236)

"In DMM it is [...] assumed that neither language acquisition nor language attrition can adequately be understood if they are discussed as processes in isolation [...]. The two processes have to be seen as an integrated part of an evolving dynamic system, in other words language attrition is a function of language acquisition." (Jessner 2003: 242)

However, in the context of this study, we need to show that the models can also usefully be employed to explain interactions between language proficiencies. Key constructs in this discussion derive from psycholinguistic conceptualisations of language acquisition and change and have already been touched upon above, but may be summarised as follows (Andersen 1982; Paradis 1993; de Bot 1998; Köpke and Nespoulous 2001; Paradis 2001): input; memory constraints; resource limitations; competition; activation/inhibition in neural networks; retrievability; controlled vs. automatic processes; and the declarative/procedural knowledge continuum.

Within this framework L1 attrition has been defined as "both a decline of retrievability of declarative linguistic knowledge and deproceduralization of linguistic knowledge in L1, and an increase of competition by L2 knowledge" (de Bot 2002). This definition gives L2 a clear role in furthering L1 attrition, and the following discussion will centre on just how this might happen in neuro- and psycholinguistic terms.

### ***2.3.1. Activation Threshold Hypothesis (Paradis 1993; Paradis 2001)***

The Activation Threshold Hypothesis (Paradis 1993) forms part of Paradis's Integrated Neurolinguistic Theory of Bilingualism along with the Three-Store and Subset Hypotheses (Paradis 2001). It relates the retrievability of items stored in memory to the frequency of their use and reinforcement in an activation/inhibition framework: "Activation and inhibition mechanisms allow to account for the control of multiple languages in the brain [...] as well as for changing dominance patterns" (Köpke and Schmid 2004: 23).

The Subset Hypothesis (Paradis 1993 and 2001; for a detailed discussion cf. Herwig 2001) proposes that different languages are stored as separate neural networks that form part of a single linguistic system. The networks are made up of the items of the language in question which are connected by memory traces of varying strength and quantity forming nodes. In L2 learning, items of the L2 are normally initially connected with their corresponding L1 items, but as the language grows, it starts separating out and the intralinguistic connections become stronger and more numerous than the interlinguistic ones, implying that the difference between intra- and interlinguistic connections is a quantitative rather than a qualitative one (e.g. Kroll and Stewart

1994 and other interconnection models of bilingual representation; Herwig 2001: 118). Growth of proficiency in a language is thus reflected in an increase in the number and the strength of intralinguistic links, as well as more automated processing (for a discussion of controlled vs. automatic processes cf. below).

The Activation Threshold Hypothesis, on the other hand, concerns the processes by which items thus interlinked are retrieved from memory for use. It proposes that

"an item is activated when a sufficient amount of positive impulses have reached it. The amount necessary to activate the item constitutes its activation threshold. Every time the item is activated, its threshold is lowered and fewer impulses are required to reactivate it."

(Paradis 2001: 11)

Simultaneously with the positive impulses that afford activation of an item, negative (or inhibiting) impulses raise the activation thresholds of other items in the network so they become less available for selection.

A differential level of activation is required for different types of processing (Paradis 1993: 140). Thus, comprehension would require a lesser amount of positive impulses than reconstruction, which requires fewer impulses than production (so-called self-activation). Hence, a higher activation threshold is tolerable in comprehension and reconstruction, and an item still available for comprehension may have too high an activation threshold for production.

Over time, the activation threshold for an item that was used recently gradually rises again. This means that items which are used more frequently have lower activation thresholds and are more available for processing than less frequently and less recently activated items which become more difficult to activate, i.e. to retrieve for processing.

The model as discussed so far applies to both monolingual and multilingual situations. The same principles operate for individual items within and at the level of language-specific networks. It also implies two alternative sources for L1 attrition – one which may arise independently of the increased use of another language, and one which may be linked more directly to interactions with L2 through increased use and proficiency in L2.

As concerns the former, L1 attrition may be predicted as "the result of long-term lack of stimulation" (Paradis 2001: 11f) or "as a natural consequence of lack of use" (Köpke and Schmid



2004: 23). In other words, L1 may decline purely because it is not used, rather than because something else is used more, since the activation threshold of L1 and elements within it rises automatically after last use making it eventually impossible to retrieve elements from L1. Of course, this depends on the original activation level of the item in the first place, and may be expected to occur far sooner with low-frequency items whose activation thresholds would already have been higher than for more frequently used items.

While this cause of attrition may be indirectly linked to increases in L2 use, which is likely to happen in a second language setting, this research is concerned with whether the model can be used to postulate a more direct relationship between the L1 attrition and L2 use/proficiency and whether L2 can "induce" attrition in L1. This will be discussed with reference to a more short-term and a longer-term dimension which correspond to the distinction between a one-off, performance-based aspect, and more permanent changes in knowledge.

To start with the former, we need to consider how L1 activation may be affected by L2 during processing thus creating the conditions for L1 attrition in the longer term. In language processing, an item may be selected if the activation threshold of an item is lower than that of other items and sufficiently low in the first place. This implies the need to inhibit competitors, which is done by raising their activation thresholds sufficiently to prevent selection for production although normally not enough to hinder comprehension.

Crucially, this applies in particular to competitors from other languages.

"When a bilingual speaker elects to speak one language rather than another, the activation threshold of the nonselected language is raised sufficiently to prevent interference during production, but not sufficiently to preclude borrowing and mixing, or comprehension in the other language" (Paradis 1993: 141, with reference to bilingual speakers used to code-switching).

As regards activation and inhibition within the same language, an effect called "activation-spreading" has been proposed, which means that it is not just the item in question the activation threshold of which is lowered but also that of other items connected with it. In other words, while other languages are deactivated or inhibited, a larger set of items of the language to be used than

ultimately selected become activated. This implies a "practice effect" even for non-selected items of that language.

Thus, frequent L2 use impacts on the activation level in L1 by, on the one hand, raising the L1 activation threshold during L2 use while on the other simultaneously keeping larger portions of its own network activated. This creates the conditions which, if consistently present, may eventually lead to L1 attrition:

1. The overall activation threshold for L2 as a result of its frequent use is so low that it cannot be sufficiently inhibited in L1 use, causing interference in production (inappropriate selection).
2. If L2 is very highly developed, this will lead to increased competition by a larger number of competitors with potentially lower activation thresholds, again causing interference and non-retrieval.

The discussion of L1 attrition as a result of competition during processing becomes more complex when one considers the notion of "language mode" developed and subsequently refined by Grosjean in numerous publications (e.g. 1989; 1998; 2001; 2008): "Language mode is the state of activation of the bilingual's languages and language processing mechanisms at a given point in time" (Grosjean 2001: 3).

Language mode is conceived of as a continuum of language activation states within an individual bilingual during language processing (perceptive and productive), ranging from "monolingual mode" in which the base (processing) language is highly active, whereas the other language(s) are (almost entirely) deactivated, through "intermediate mode" to "bilingual mode", during which the base language is highly active and the other language(s) show increasing to very high activation which, however, always falls below the level of activation of the base language. The position on the continuum depends on the situation, the speaker(s) and the conventions between them which might or might not tolerate code-switching and -mixing. Grosjean (1998: 140) makes the point that bilinguals will rarely ever find themselves in pure monolingual mode since the other language(s) will always remain active to a degree.

With reference to L1 attrition this means that a bilingual currently in (largely) monolingual L1 mode who finds items from L2 to be more readily available than items from L1 (competition) will continue to search for an item in L1, which can cause disfluency (Paradis 1993: 142 ). This

situation implies a very high level of L2 activation (since L2 is inhibited in monolingual mode), and failure to retrieve an item and corresponding disfluency might indeed be a sign of language attrition. However, a failure to produce this item in bilingual mode may not, since the bilingual might have opted to select the L2 rather than the L1 item.

Having discussed the ways in which L2 impacts on the L1 activation level during processing, we will now consider the effect of these conditions persisting over a longer period of time, and the effect of L1 consistently being used less than L2, in respect of inducing L1 attrition:

3. L2 use involves raising the activation threshold of L1 (inhibition). With frequent use, the activation threshold for L1 as such and L1 items may become more permanently raised.
4. Simultaneously, the activation threshold for L2 (items) falls and may eventually be below that of L1.
5. The activation threshold of L1 (items) may thus persistently be too high in relation to L2, leading to "reduced accessibility" (Köpke and Schmid 2004: 23) or non-retrieval of items.
6. Due to interference and non-retrieval, the memory traces for L2 items are strengthened, even when using L1.

A high proficiency level is both a function of knowledge and of the ability to access or use it. This is obviously a question of degrees – only if the activation threshold of L2 is so low in comparison with L1 that it consistently wins out in the competition for selection, i.e. so that L1 consistently cannot be accessed, might one speak about a causal relationship between L2 proficiency and L1 attrition.

To summarize, both the short- and longer-term dimensions suggest a relationship between L1 attrition and L2 acquisition, provided L2 is used frequently or more frequently than L1, as is likely to be the case in a second-language environment. The immediate role for competition-induced language attrition, which may be a function of either a highly activated L2 and/or high L2 proficiency, is more noticeable in the first, processing dimension, whereas in the long term the impact of L2 use exceeding L1 use is more visible. As discussed, if competition persists, it will continue creating the conditions which will contribute to L1 becoming less accessible in the long run. It thus follows that L1 attrition 1) may partially be related to L2 use and 2) may be induced by

a sufficiently high level of proficiency in L2. The attribution of causality of L1 attrition to high L2 proficiency is, however, not entirely straightforward.

### **2.3.2. The Dynamic Model of Multilingualism (DMM) (Herdina and Jessner 2002; Jessner 2003)**

The Dynamic Model of Multilingualism (DMM) is based on complex systems theory, adding a dynamic dimension to the notion of "multi-competence" developed by Cook (2003). The model makes the specific assumption that "there are natural cognitive and psychological limits to every multilingual system" although disagreeing with the notion of a "specific limitedness of the language capacity" implied in balance theory (cf. above) (Herdina and Jessner 2002: 28).

In other words, languages are not just "co-existing peacefully" in a multilingual speaker's mind but are in fact in "competition" – for finite resources as regards the amount of language effort required (an alternative, more active view on input). General language effort is needed in both learning and keeping a language at a particular proficiency level (language acquisition and maintenance effort respectively):

$$\begin{aligned} &\text{General language effort (GLE) =} \\ &\text{Language acquisition effort (LAE)} \\ &+ \text{Language maintenance effort (LME)} \end{aligned}$$

Language proficiency in this conception is defined as a learner's ability to use internalised knowledge (competence) for different tasks (Herdina and Jessner 2002: 56).

In relation to our research question of how this effort changes when a person becomes multilingual, the model has the following to say:

"The maintenance of two or more language systems at a similar proficiency level can be seen as more than twice as strenuous as the maintenance of a monolingual system because the multilingual brain is constantly involved in processes of matching and differentiation of two or more language systems. Maintenance work in multilinguals also involves metalinguistic and monitoring processes in order to reduce interference as a processing phenomenon and to ensure a certain speed of recall of information among other aspects.

Psycholinguistic systems containing two or more language systems can therefore be seen as less stable than monolingual ones, and repair or reactivation procedures are constantly required to maintain the system in a steady state." (Jessner 2003: 241, my emphasis)

In other words, multilinguals need to expend a relatively greater amount of language maintenance effort as compared with monolinguals because their minds are more complex. In addition, the model predicts that maintenance effort increases in a non-linear fashion (second order exponential growth) as a function of language proficiency, and that it has to exceed the amount of linguistic growth.

For a monolingual living in his or her home country, this is not a problem – although L1 proficiency is normally close to 100% and thus should require the maximum level of maintenance, the input naturally present and the time spent exposed to this input is normally perfectly adequate, so that people are not even aware they are engaging in language maintenance effort. If a person does know another language to a moderate level and is not keen to improve it, the maintenance effort required for this language should normally also be manageable, although most people experience a "rusting" of foreign language skills once they are no longer trying to learn the language or use the skills (lack of maintenance).

Language acquisition effort comes into play when the person decides to improve their proficiency level in an L2, and this effort will be greater than the maintenance effort required to keep this language at the level achieved. The higher this level, the greater the required maintenance effort which, as we remember, is postulated to increase at the rate of second-order exponential growth.

If an L2 user wishes to simultaneously maintain two or more languages at very high proficiency levels, then the necessary maintenance effort will soon result in competition for the available resources, i.e. the amount of time and energy learners are able and willing to spend on maintaining/learning their languages, which the required maintenance effort is predicted to "soon exceed" (Herdina and Jessner 2002: 113).

The demands on general language effort in the L1 environment as a function of different proficiency levels in two or more languages may schematically be presented as follows. We will look at two scenarios:

1) high L1 proficiency with moderate L2 proficiency and

2) high L1 proficiency with high L2 proficiency.

The graph below expresses magnitude of the components of general language effort for different proficiency situations. The scores were derived at by using an intuitive procedure combined with the predictions of DMM whereby different - and purely symbolic - values were given to moderate vs. high proficiency depending on whether the linguistic environment is congruent with and thus supporting the language, or not, and these values correspond to the height of the bars.

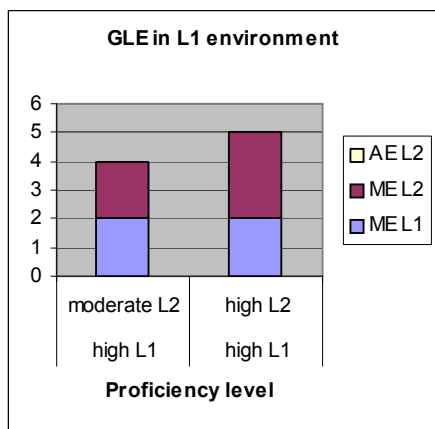
#### Maintenance effort (ME)

- 1 minimum moderate proficiency in the corresponding linguistic environment
- 2 normal moderate proficiency in the opposite linguistic environment
- 2 normal high proficiency in the corresponding linguistic environment
- 3 maximum high proficiency in the opposite linguistic environment

#### Acquisition effort (AE)

- 1 minimum high proficiency in the corresponding linguistic environment
- 2 normal high proficiency in the opposite linguistic environment
- 2 normal moderate proficiency in the corresponding linguistic environment
- 3 maximum moderate proficiency in the opposite linguistic environment

**Table 1: General language effort at different proficiency levels by linguistic environment**



**Figure 1: General language effort in the L1 environment**

As we can see, an L2 user living in his or her first-language environment (which we presume to be monolingual - cf. Herdina and Jessner (2002: 102 f.) for different consequences for multilingual

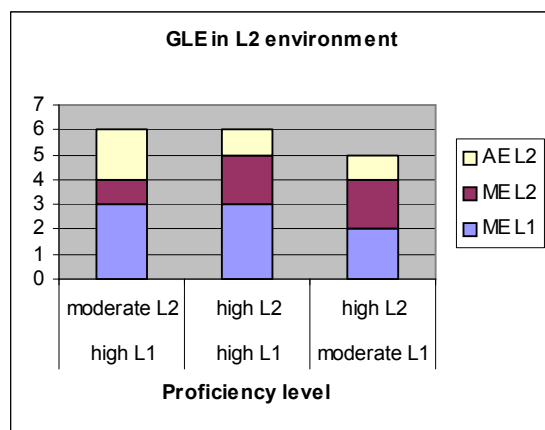
language use of bilingualism in a monolingual community vs. multilingualism in a multilingual community) and wishing purely to maintain his or her languages must expend two types of language effort - maintenance effort for L1 and L2 respectively.

In line with the claims of DMM, in the first situation maintenance effort for L2 is smaller than in the second, since L2 is smaller. Overall, these two situations result in total GLE values of 4 and 5 respectively, which intuitively does not normally seem to be too excessive, since there are many people who manage to maintain an L2 at a high level. However, they do start bordering on values which for some people are too high to support, i.e. for which they either cannot or will not make the necessary resources available (see below for what these might constitute), which will then result in the "rusting" of the L2 skill already referred to (not L1 since this is naturally supported). If a given L2 user started to learn a third language requiring acquisition effort (not graphically represented), the pressures relating to the earlier L2 would be exacerbated.

Let us look now at the corresponding situation in the L2 environment (again a monolingual one).

Here we look at three scenarios:

- 1) high L1 proficiency with moderate L2 proficiency,
- 2) high L1 proficiency with high L2 proficiency and
- 3) moderate L1 proficiency with high L2 proficiency.



**Figure 2: General language effort in a monolingual L2 environment**

As we can see, the problem of lack of resources (motivationally or time-related) for stepping up general language effort is generally more pronounced in a language contact situation (i.e. when a person lives abroad) since this brings with them changed communicative demands and/or needs. The resulting scores of 6, 6 and 5 respectively outweigh the ones from the L1 environment overall, since the L2 user is now faced with a need to become proficient in the language of the host country for which acquisition effort and, as the new system grows, an exponentially increasing maintenance effort are required, while the diminishing opportunities to use the mother tongue call for a step-up in first language maintenance effort.

Of course, in this situation it is theoretically possible to increase L1 maintenance effort sufficiently to cope with the increased demands posed by the acquisition of the new language. Practically, however, these demands may exceed the resources that the speaker is able or motivated to provide. Indeed, the speaker may not even be aware of having a need for their provision. Crucially, in the L2 environment, L1 is not "naturally" supported anymore to the same degree, while the acquisition and/or maintenance of an L2 may be.

From this should follow also, that the most resource-depleting time is when learners are putting time and effort into learning more of the language in which they have already achieved a high level of proficiency. DMM provides that acquisition effort is the inverse of maintenance effort if and only if general language effort is constant, but of course it is also directly dependent on achievement and communicative need. Therefore, although bars 1 and 2 seem to indicate that GLE is the same overall (and higher than in the L1 environment) it is not at all a given that AE drops back at the higher proficiency level, as assumed in the graph. If people continue to put AE into the L2 in which they are already highly proficient, then bar 2 would end up being higher than bar 1.

Again, the situation becomes more complicated, and the demands on resources more pronounced as more languages enter the equation, either to be acquired or maintained at lower or higher levels of proficiency. The same principles apply, and the pressures on maintaining the multilingual system, and L1 in particular, become very clear.

As a result of inadequate language effort, the language(s) in question will either a) stop growing (fossilization), b) deteriorate (attrite) or c) adapt to changed communicative needs by becoming domain-specific which is computationally less taxing (compensation). The proficiency level will



change until the maintenance level provided is sufficient to keep it. This is the situation represented in the third bar in Figure 2.

However, since negative language growth is seen as the inverse of language acquisition, the effort required to keep languages at a particular level drops in an exponential fashion, so a small reduction means greater savings in terms of effort. Language attrition, defined as "a function of language acquisition, with language maintenance providing the necessary link between the two processes" (Herdina and Jessner 2002: 106), is thus a 'key feature' in multilingual systems" (Jessner 2003: 236). However, from the discussion it is clear that the issue of whether the resources provided are sufficient is one of proportions – the relative lack of "natural maintenance" in an L2 environment, coupled with a greater or lesser proficiency in the L2.

Of course, in a dynamic model the processes of positive and negative growth are reversible at any point. However, if adequate maintenance effort is not provided, the model proposes that the language system will gradually return to a monolingual state.

Re-exposure and re-learning constitute a type of language maintenance effort and should contribute to a quick recovery of parts of the language systems being forgotten. Moreover, the L2 user can employ specific counteractive strategies for language maintenance work. These vary greatly from one individual to another, and presumably also in their effectiveness. They are second nature to a lot of people, can be subconscious or explicit meta-cognitive strategies and include anything from consulting a dictionary to comparing notes with a fellow native speaker on some expression or other. However, "most important is the simple fact that use of language counteracts language loss or decay" (Herdina and Jessner 2002: 98).

One issue which we already touched on slightly in the foregoing discussion is the fact that some of the effects of the "natural cognitive and psychological limits" that Herdina and Jessner (2002: 28) envisage may be outside the L2 user's control, such as actual cognitive limitations in connection with the increased demands posed by being an L2 user as compared to a monolingual, whereas others may be more directly caused by the L2 user's psychological limits of not being motivated to step up maintenance effort sufficiently. Continued language use was highlighted as being of utmost importance in counter-acting attrition, but is quite ambiguous in this regard, since although L2

users have agency (Swain p.c.) and can to some extent control the quantity and quality of L1 input, they may be excluded from certain uses because of their social standing.

Let us finally turn to how DMM may be employed to account for findings within the area of L1 attrition, such as why first language attrition might proceed less speedily and noticeably than previously expected on the basis of subjective assessments of its extent.

- a) Herdina and Jessner (2002: 91ff) propose a growth curve for language learning modelled on biological growth. This is a sine curve with a lag time at the beginning (retardation), followed by a time of a steep increase in language development (acceleration), followed by a flattening out of the curve as it gets close to asymptote. They also propose that language attrition (negative language growth) is the inverse of positive growth, which in L1 attrition would start from a level very close to asymptote. From this follows that any decrements in input or maintenance effort will be relatively less noticeable in reduced proficiency than for a system which is rather less developed. It also follows that learners who had a lower proficiency level in L1 to start with will be more affected.
- b) Herdina and Jessner (2002: 94f) further propose that actual loss ("negative growth of the language system" – starting to use an incorrect structure, not being able to retrieve a lexical item) – is preceded by a weakening in the structure of the language system, an increased "scatter" (increased optionality/variability of performance). This is a result of the in-built redundancy of the language system which allows the system to absorb the effect through internal restructuring before settling on a new, lower level of stability, at which point "scatter is reduced" (Jessner 2003: 240).

There also exists some empirical evidence for this claim: in an experiment simulating vocabulary attrition on computer, Meara (2003) found that over a long time the inbuilt attrition events seemed to have no effect on overall proficiency, that "attrited" items could become reactivated virtue to their neighbours in the network ("latching system", constant repair work), but that at some point one more attrition event caused the collapse of the system, the network proving "surprisingly vulnerable to minor damage effects" (Meara 2003).

The same effect of a weakening of the structure before any noticeable decline had been predicted by Lightbown (1979) in his discussion of potential parallels between synchronic language loss and

diachronic language change with regard to grammar: "[S]low piecemeal changes in different parts of the system can eventually lead up to a crisis point where some fundamental principle of grammar is violated" which is then followed by "swift" and "therapeutic restructuring" (Lightbown 1979: 122, quoted in Sharwood Smith 1983: 228).

Unfortunately, it is hard to extrapolate from a computer simulation to how long it might take in human terms to reach this crisis point, so again a given attriter may never in their life time get to the point where the system has become vulnerable enough to disintegrate completely.

c) Finally, Herdina and Jessner (2002: 96) predict that participants are likely to step up their maintenance effort once attrition has become noticeable (cf also Jaspaert and Kroon 1989: 81). Although resources for this are limited, they are not so in an absolute way, since participants may find ways to make more efficient use of them. There is bound to be a great deal of individual variation.

To summarize, DMM is a model which relates the processes of language acquisition and maintenance via a required amount of language effort which is predicted to increase in line with the proficiency level achieved. Expendable or available language effort is limited, albeit not in an absolute way but rather as a function of communicative needs and personal (motivational and cognitive) factors. In situations that require a greater than usual amount of language effort, such as living abroad and acquiring the L2 of the host country to a high level of proficiency, first language attrition will be a necessary outcome, if the speaker for practical reasons is unable to increase language maintenance effort to an adequate level. It thus follows that L1 attrition 1) may be correlated to the process of L2 acquisition and 2) may be induced by a high level of proficiency in L2. DMM thus lends support to the hypothesis that the processes of L1 attrition and L2 acquisition may be interrelated.

#### **2.4 Predictor variables in language attrition and acquisition**

In their overview of attrition research to date, Köpke and Schmid (2004: 15) list the following extra-, socio- and psycholinguistic variables which have been proposed or found to impact on the degree of L1 attrition/maintenance:

- "- age at onset of L2 acquisition
- age at onset of L1 attrition
- time since onset of attrition
- level of education
- attitudes
- frequency, amount and settings of use of the attriting language"

Most of the factors represent extralinguistic, biographical variables. Some of these factors have been framed differently in different context. Thus, "time since onset of attrition" is mostly referred to as "length of residence" in L2 contexts. Further, "L1 use" is often framed as "contact with L1", while "level of education" is often paired with "social status".

Not listed is language proficiency in L2, a variable we are concerned with here, which will be added to the predictor variables to be reviewed in detail. The list also has just one "internal", or psycholinguistic, factor, "attitudes", which is the only factor of this type to have been investigated in any systematic way so far. We shall add aptitude, motivation and differences in cognitive maturation (Sharwood Smith 1983: 50), as well as awareness of attrition to our list of factors for review. A further variable not specifically discussed, but nevertheless to be controlled, is the age of informants at the time of testing because of the possibility of age-induced rather than context-induced L1 attrition.

Although the above-mentioned and a few related factors will be discussed individually in the following two sections, it is clear from the literature that any attempt at a clear-cut categorization is doomed to failure since in practice it is very difficult to isolate some of them and to exclude their interaction with other factors. In particular, age at onset of L2 acquisition, L1 proficiency (L1 literacy) and education, age at onset of L1 attrition and attitudes, as well as education and social class often confound one another. Thus we shall attempt to distinguish each factor's individual contribution to the extent that this is possible, as well as pointing out clusters of factors.

#### **2.4.1. Age at onset (AOA)**

We already mentioned age of onset of attrition in connection with the question about the reality of the process of L1 attrition. To reiterate the conclusion sketched earlier, results for adults are

contradictory but the emerging picture seems to be that "the amount of attrition present in the data is usually surprisingly low" (Köpke and Schmid 2004: 10). Generally, adults' L1 seems to be less vulnerable and far more slowly affected than either children in L1 attrition, or L2 learners in respect of losing their L2, even years after relocating to an L2 environment.

The possibility of L1 attrition in children, on the other hand, is unquestioned, and the amount of L1 attrition displayed by children who are placed in a new linguistic environment pre-puberty, i.e. up to the age of 12, tends to be severe in comparison with first language loss in adults (Kaufman 1991; cf. Köpke and Schmid 2004: 9 for further references). The participants in Schmitt's (2000) study (five Russian boys up to the age of 11 who had been living abroad for six years with their families), for example, displayed a great amount of "convergence" in their utterances (almost 47% of bilingual utterances, the rest being code-switches, and 17% of total utterances), convergence being, according to the author, a sign of significant L1 attrition:

"[C]onvergence is a mechanism of rather advanced language loss where the speaker can no longer distinguish between the levels of lexical structure and the types of morphemes as identified by the 4-M model in the two languages." (Schmitt 2000: 647).

In pre-puberty children an increase in L2 proficiency is often correlated with a decrease in L1 proficiency (e.g. Jia and Aaronson 1999), and some children abandon their original L1 completely and in its place adopt the L2 (Ecke 2004: 327). This highlights the difficulties in disentangling age of onset of L1 attrition and age of bilingualism, a point to which we shall return in the next subsection. The evidence so far would seem to suggest that "[a]ttrition of L1 among older children and adults differs from the L1 attrition process among pre-puberty children" (Kaufman 1991: 185), and that the age at which people relocate to an L2 environment does matter for the extent to which a person's L1 is likely to undergo attrition.

Studies investigating the relevance of age at onset of attrition (Ammerlaan 1996, Pelc 2001) have only been able to show an age effect if including both participants whose exposure to the new language environment began pre-puberty and post-puberty respectively. Leyen (1984), for example, found significant attrition of vocabulary in children, but not in adults. In fact, in these cases age was established as the single most important factor in predicting attrition (Köpke and Schmid 2004: 10), whereas no age effect was found in studies employing only older speakers.

In other words, once people reach adulthood and with it a level of L1 proficiency approximating something of a "full native-speaker competence", their chances of maintaining their mother tongue are a fair deal better than those of younger speakers whose mother tongue was not yet as well established, but do not seem to continue to improve vastly as age at immigration increases.

This is in line with proposals about the Power Law of Language Learning (for a discussion cf. Ellis 2002), which states that practice effects become relatively less important and eventually reach asymptote with increasing proficiency levels, i.e. the better established L1 is, the less difference further practice or lack of practice is going to make.

The age factor is also a central concern in second language acquisition. In a second-language environment, age at onset of L1 attrition and age at onset of L2 acquisition or bilingualism, often overlap, particularly in younger learners. Consequently, the main theory in second language acquisition dealing with the age factor, the Critical Period Hypothesis, or rather the factors "which have led to [its] postulation" (Köpke and Schmid 2004: 20), have been turned to for an explanation of attrition data.

While acknowledging the contentiousness of the existence of a pre-designated "critical" or "sensitive" period for language learning (cf. Singleton 2003; Singleton and Ryan 2004 for a critical review of the evidence pertaining to the Critical Period Hypothesis), Köpke and Schmid suggest that the relative ease with which children acquire L2s might facilitate the forgetting of L1. However, to our knowledge there have been no systematic attempts to investigate the impact of age on retention of L1 across different adult populations.

One problem with postulating age-related differences for L1 attrition between adults and children is that one needs to be careful to distinguish between attrition and non-acquisition. While this is necessary for all age groups, the scope for classifying something as lost which in fact has never been acquired is greater in children, as by definition their L1 is only developing. Although adult L1 systems do not stop expanding either, at least in the lexical area (Singleton 2003: 8), the pace at which this happens is rather slower, and it is generally acknowledged that grammar acquisition is more or less complete around age 15. These differences between lexicon and grammar arise because grammar is a closed system, while lexicon is open.

Concerning grammar, relatively clear notions exist of what constitutes full language competence, i.e. of what every adult ought to know (Jaspaert, Kroon et al. 1986). The fact that L1 takes time to become properly established in the mind is also the likely reason for those differences observed between child and adult attriters since a less complex and less entrenched system is more easily replaced than a fully established one. In psycholinguistics, the notion of a "critical mass" of linguistic knowledge has been advanced to account for these findings: "This raises the question of whether there is some 'critical mass' or threshold level which, if attained substantially reduces the degree of attrition experienced" (Pan and Gleason 1986: 198).

Older learners, on the other hand, often have already acquired L2 to a certain degree before settling in the L2 environment so that age at start of L2 acquisition pre-dates L1 attrition. However, due to the great differences between language learning in one's native country and in the L2 environment, both in terms of input and outcomes, age at time of immigration may actually outweigh the impact of age at start of acquisition, conflating the two variables.

Despite some commonalities between younger and older L2 learners, there is one fundamental difference between pre- and post-puberty, or child and adult, attriters making it essential to distinguish between these, as well as between simultaneous, early and late bilinguals (Köpke and Schmid 2004: 9): the former's L1 is only developing, as is their L2, which makes interactions between the languages, to the extent that L2 takes over L1 and takes over as the dominant language, far more likely than in the latter case, where L1 is a fully entrenched, well-practised system which it is not as easy to displace.

This point was made by Singleton (2003) with reference to Liu, Bates and Li's (1992) acquisition study which found that learners under the age of 13 process sentences in L2 like native speakers, whereas those of 20 plus years of age process them like they process sentences in their own mother tongue. Refuting the contentions of the CPH, Singleton argues that this result can be explained "in terms of the increasing extent to which the L1 influences L2 processing as a function of years of experience of the L1 and the degree to which it is entrenched" (Singleton *ibid.*: 12) rather than being seen as evidence for a qualitative shift in the way language is acquired up to a particular maturational point.

In this context it is worth considering the potentiality for L2 development of living in an L2 environment vis-à-vis language learning in formal instructional contexts in the L1 environment. A recent special edition of *SSLA* on the effects of context-of-learning for second language acquisition, contrasting L2 learning in the L1 environment (formal classroom setting) vs. the L2 environment (through study-abroad/immersion programmes providing formal and informal learning opportunities) brought together evidence for the far greater efficacy of the L2 environment on various measures of not only oral proficiency, but also grammatical and lexical acquisition (Collentine and Freed 2004).

Despite the fact that gains measures displayed a high degree of interpersonal variation, due to complex correlational patterns between learning context, amount of language use inside/outside the classroom, cognitive abilities, initial proficiency level and motivation/attitudes of the learner and his/her environment, the power of the L2 environment for L2 language acquisition would appear formidable. This would be particularly the case for people placed permanently in the L2 environment, who usually have a need and a desire to acquire the L2, leading to the commonly expressed idea that L2 learning in an L2 environment will look after itself. Thus, depending on the mastery of the L1 already achieved prior to relocation to the L2 environment, it will be easier (for adults with a fully formed system) or harder (for pre-puberty children with developing systems) to protect their L1 from the "onslaught" of L2 acquisition.

Where people who started to acquire an L2 in their L1 environment are concerned, a slightly different situation presents itself since L2 acquisition should not have the same disruptive power (for children in particular) on L1 as in the L2 environment as the L1 environment continues to support L1 development. However, in this context, it has been argued, often in relation to the CPH referred to above, that age of starting to acquire an L2 impacts on ultimate attainment in that language (e.g. Birdsong 1992; Hyltenstam 1992; White and Genesee 1996; Abrahamsson and Hyltenstam 2009), as expressed in the oft-invoked credo that "younger is better" in L2 acquisition (Singleton 2003; Singleton and Ryan 2004).

This assertion has an immediate bearing on this study, which has taken L2 proficiency as an independent variable under investigation, thus making age at onset of L2 acquisition a predictor variable for L1 attrition, if the contentions of this study are borne out.



As Singleton (2003) points out

"[t]here is [...] no real quarrel among SLA researchers over the proposition that individuals who begin to acquire an L2 early in life generally do better in the long run than those who begin as adults" (Singleton 2003: 14)

However, things are more complicated than simply younger = better under all circumstances. First, this view does require relativisation, as younger starting age is by no means a guarantee for perfect command of the L2, and there are numerous cases of older learners outperforming younger ones (for an extensive review of arguments and evidence cf., for example, Singleton 2003; Singleton and Ryan 2004).

Second, there are many alternative explanations of differences in L2 achievement between older and younger learners outside of the Critical Theory Hypothesis, as they can for a large part be put down to general maturational changes in cognition (cf. Cochran, McDonald et al. 1999 for evidence of Newport's 'less is more' hypothesis) on the one hand and input differences on the other, since younger learners as a rule have access to a greater amount of better quality input over a longer time because of schooling requirements for children (Singleton 2003).

The latter would also appear to be particularly important in an L2 environment where compulsory schooling in the L2 undoubtedly bestows a vast advantage to children over their elders for acquiring the L2, although the amount and types of L2 input becoming readily available when moving abroad should go some way to obliterating previous age and input differences from earlier L2 learning.

By the same token, for adults acquiring L2 in the L2 environment, higher age, along with family status (being married) and the presence of children, has been shown to be predictive of a lower degree of one facet of L2 proficiency, namely lexical richness (van Hout and Strömquist 1993). These three factors are clearly interrelated - the older the person (the cut-off point in this study was 25 years of age), the more likely s/he was going to be married and have children (and therefore possibly interacting less with the outside world). This underlines the fact that different age groups do tend to have different types of input of varying intensity, but the picture may be very different in individual cases. In general, however, there are a number of factors, apart from age per se, that seem to militate against adults' acquisition of L2 to the same level as children.

This discussion has a number of implications for L1 attrition: Given that a younger age of starting to acquire the L2 tends to be correlated with a higher level of ultimate attainment in L2 for the reasons above, L1 attrition may occur as a function of age, if the hypotheses of this study about possible interconnections between the acquisition and attrition processes are confirmed.

However, age as a variable tends to be profoundly confounded with a number of other variables, already hinted at above, including education, social status, family status, amount of contact, attitudes and probably constructions of personal and ethnic identity; thus, studying its impact in isolation is rather difficult.

#### **2.4.2. Length of residence (LOR)**

Length of residence in the host country, or length of time since onset of attrition, typically coincides (or is assumed to coincide) with the time elapsed since emigration and thus continued exposure to the native language. Given the theoretical and practical possibility of L1 attrition, it would appear to be common sense that skills or knowledge should deteriorate progressively the longer they are not used; however, the picture emerging from the research is more complex. Changes do happen gradually, but the rate may differ at various times.

One change that appears to occur very soon after a person becoming immersed in a different linguistic environment is a shift in language dominance, which is discernible as "access to L1 gradually becoming slower than access to L2, without L2 competence necessarily being native-like" (Köpke and Schmid 2004: 11). Depending on the area of knowledge tested and the tasks chosen, language dominance has been demonstrated to switch after between 3 and 7 years of length of residence (Mägiste 1979; Frenck-Mestre 1993, reported in Köpke and Schmid 2004).

As Köpke and Schmid further point out, a change in language dominance is not per se to be equated with attrition, since differences in processing speeds are hardly noticeable in actual speech situations, but it is "most likely that attrition is preceded by such a reversal [in language dominance]" (Köpke and Schmid *ibid*: 12).

Other signs of language attrition have also been identified after relatively short periods of time, i.e. within the first decade of residence abroad. After this time, attrition levels off, residual knowledge

appears to remain intact, and it becomes increasingly difficult to identify a time effect (Waas 1996). This effect is even more pronounced in L2 attrition where substantial attrition, followed by relative stability, was found within the first two years of leaving the L2 environment or stopping to learn the L2 (cf. Bahrick 1984, Weltens 1988, both cited in de Bot and Hulsen 2002).

With regard to the timeline and extent of attrition, a proficiency effect has been demonstrated: more proficient learners displayed an initial retention plateau in comparison with their less proficient peers. This has been explained with reference to a "critical threshold during learning" (Neisser 1984, in de Bot and Hulsen 2002), i.e. the idea that knowledge learnt to a certain degree may be less vulnerable to loss (Hansen 2000, in de Bot and Hulsen 2002). This idea is formalised in the tenets of DMM (Herdina and Jessner 2002) presented above.

In addition to the impact of initial proficiency, it is likely that other factors, such as the amount of contact with L1 and typological proximity, play a part in the attrition pattern. As an extreme example one might cite the case of an 83-year-old Dutchman (Jaspaert and Kroon 1992) who had been living in the US for 60+ years and whose L1 showed a lot less erosion than one might have expected after such a length of time. The researchers estimated about 5% of his language to be affected (ibid: 142), which should have been far higher if it were true that reduction of contact with the L1 ultimately leads to its disappearance.

This participant used his mother tongue in letters to his family only and was otherwise a prolific writer in L2; he did, however, show a lot of adaptation (in main verbs, nouns, adjectives and adverbs) facilitated by the close typological relationship between L1 and L2. This relationship might be one of the factors for rendering the total disappearance of L1 unlikely. For similar findings cf. de Bot and Clyne (1994), Schmid (2001) and Hutz (2004).

A longitudinal study over 16 years of Dutch immigrants in Australia (de Bot, Gommans et al. 1991; de Bot and Clyne 1994) found that proficiency levels in L1 were even over time for people who had continuous contact with their L1 (if their partner was speaking the same language, and if they had contact with other native speakers at least once a week), while for people who had little contact, proficiency declined in a linear fashion over time (de Bot, Gommans et al. 1991: 94).

This finding mirrors results from L2 acquisition studies, where the impact of the LOR was expected to be higher and was subsequently found to have been "concealed" by the contact variable (van Hout and Strömquist 1993: 169). In other words, LOR only mattered for people with few contacts, and this, in turn, might mean, as has been proposed by de Bot and Clyne (1994), that it is the first 5-10 years abroad which determine whether attrition will actually take place or not.

### **2.4.3. Contact with L1/L2**

Contact with L1 (i.e. the frequency, amount and modes/settings of L1 language use) has already been referred to several times in the previous discussion as a confounding variable for length of residence and age at onset of L2 acquisition, and may well, of all the sociological and biographical factors investigated, turn out to be the decisive one in language contact situations.

In fact, in psycholinguistic (usage-based) approaches to attrition it is construed as the fundamental factor underlying reduced activation of and access to knowledge, and studies conducted from this perspective provide support for the link between lack of use and attrition: "[O]ur data seem to support the idea that even the L1 of a late bilingual may suffer from attrition when it is *no longer practised*, as predicted by the Activation Threshold Hypothesis" (Köpke and Nespoulous 2001: 233, my emphasis).

Such perspectives tend to take the impact of lack of use for L1 attrition for granted:

"Fifteen years of language attrition research in various countries all over the world have made it clear that the real picture is rather different: neither first languages nor second languages are immune to loss. *With non-use they fade*, and though they keep their place in our memory system, they become less accessible up to the point where the knowledge has sunk beyond reach and is for practical purposes lost" (de Bot and Hulsen 2002: 253, my emphasis).

Similarly, Hulsen (2000: 21) states that the impact of language use and contact on the L1 in a migration context is "self-evident" while conceding that

"[i]t is unclear exactly how much use or contact is necessary for a language to be maintained and other factors, such as time since migration, are likely to be confounded with use and contact" (Hulsen *ibid.*: 22.).

The reason why this factor should assume such importance is that language contact situations naturally involve increased use of L2 as well as (more or less) diminished contact with L1, "ultimately leading to 'restriction in language use' and 'break in linguistic tradition'" (Andersen 1982: 87, cited in Jaspaert and Kroon 1992: 139).

While it is doubtful whether these typical consequences of contact situations (and stipulated main causes for L1 attrition) can realistically be separated as neatly as seemingly proposed by Major (2002: 79): "First language loss, or attrition, is frequent in speakers who have continuous L2 exposure and also in those who use their L1 less and less frequently" (my highlights), these two factors would appear to be causally linked (one of the contentions of this study): as L2 use increases, so decrease the time and contexts available for L1 use. This link is not, of course, entirely direct - speakers do have choices about how much they want to communicate in general, and in which language, but there would appear to be upper limits to compensating for reduced opportunities of L1 use in L2 contexts.

Further complications are that such choices of language use are partially dependent on attitude, and are often not choices at all, but "outside the speaker's control for sociogeographical reasons" (Köpke and Schmid 2004: 13). Finally, measures of contact are obscured by the fact that contact may or may not be used constructively for the purpose of language maintenance, and that due to "insignificant retention of the language" (de Bot 1998: 347).

Nevertheless, almost all immigrants will be confronted with fewer opportunities to speak their L1 than had been the case in their L1 environment, different modes (e.g. only informal) and settings (only speaking and listening, watching TV etc.), leading to attrition even in speakers who have a vested interest in maintaining their first language and who could be characterised as "speakers with many contacts" (Olshtain and Barzilay 1991: 139). The role of lack of contact as a major factor in L1 attrition has been repeatedly corroborated (de Bot, Gommans et al. 1991, Köpke 1999; de Bot 1998; Schoenmakers-Klein Gunnewiek 1998; Hulsen, de Bot et al. 2001), prompting the following statement by Seliger and Vago (1991: 4):

"The diminished role of L1 in use and function, exacerbated by separation from the L1 speaking community in the case of immigrants, is one of the significant sociolinguistic variables in the advent and sustenance of first language attrition."

However, other studies have failed to produce an unambiguous link (Jaspaert and Kroon 1989; Schmid 2007) which, as Köpcke and Schmid (2004: 14) suggest, may be due to methodological reasons, since assessment of language contact more often than not relies on self-report data.

The key for language maintenance would appear to be regular and active contact, and use of L1 in "various domains outside the home" (Hulsen 2000: 189). This comment provides a welcome bridge to discussing the interaction between language shift (in the immigrant community) and language attrition (in the individual) (cf. Fishman 1972 and later writings). These two processes are certainly linked, and it is often thought that on the one hand, individual loss of proficiency is a precursor and indicator for shift in the group, and that shift aggravates loss because of a lack of opportunities for use (Fase, Jaspaert et al. 1992: 9f). Thus, in a study investigating proficiency levels of 1st, 2nd and 3rd generation immigrants (Hulsen, de Bot et al. 2001) in the original L1, accuracy and speed in productive and receptive tasks were best for the 1st generation and worst for the 3rd, mirroring actual contact and use patterns in the different generations.

But there are also dissimilarities, and what is an effective means to halt language shift may not be as positive in terms of L1 proficiency. For example, we already discussed the fact that being married with children may be one of the reasons that limits the types of contact a person may have with both L1 and L2. While this clearly has negative implications for L2 acquisition, it would seem to be a good thing in terms of L1 maintenance.

Indeed, Harres's (1989) study of language shift amongst German families in Australia found that women in particular were the ones using the L1 longer as a result of "occupational and social segregation" (Harres 1989: 383), whereas husbands' occupational demands and the needs of children of school-going age tended to initiate language shift in the family. However, the women's L1 use was now restricted to the domain of home, meaning that they had less opportunity to maintain the language in other domains - which in turn implies attrition in those domains and reinforces Hulsen's observation above.

Ironically, fewer L1 domains in an L2 environment, such as home, church, clubs, which are clearly separated from others in which L2 is used, may be the way forward to avoiding complete language shift and maintaining some, albeit different, and in the case of subsequent generations certainly lower, L1 proficiency:

"By maintaining clear boundaries to areas where English will be spoken - an integrative rather than assimilationist strategy - those committed to the retention of German have succeeded in retaining domains in which German is spoken exclusively - with the result that fluency can be maintained" (Kadt 2001: 63).

#### **2.4.4. Social background and education**

This factor, although typically included in questionnaires eliciting participant information, does not appear to be as relevant as one might assume, at least not a lot has been reported about it. The reason for why it would intuitively appear an important variable is that belonging to social class is potentially linked to education level and therefore to personal and occupational setups which may foster or hinder the interaction with the host community.

Irish society, for example, is investing a fair amount of effort and money in encouraging children from designated "socially disadvantaged areas" to take up third-level education, which would not be part of the culture of lower socioeconomic groups. Children with a more prosperous socioeconomic background, on the other hand, are over-represented in education and higher status jobs.

Yet as far as language attrition is concerned, the indication is that social background, if investigated as a distinct factor, plays a far smaller role than either education, proficiency in other languages or length of stay in the host community (Jaspaert and Kroon 1989). An age factor was also investigated and similarly shown to have very little impact, but this was age at testing as opposed to age at onset of either L1 attrition or L2 acquisition. On none of the tasks reported in this study did social background reach significance; in fact, it consistently ranked amongst the lowest three, except for one task where it occupied the fourth last place with hardly any difference between the lowest ranking factors.

The implications of this finding are intriguing. Perhaps immigration encourages people to break with the educational and occupational traditions of their upbringing, or perhaps education overshadows social background as a factor. For example, the Italian immigrants in the Jaspaert/Kroon study (1989) had emigrated to the Netherlands to work in specific low-skill jobs, i.e. as

"granito"/"terrazo" workers, coal miners, textile workers and "ice-cream Italians" (ibid.: 77f), and had been encouraged to leave education early, thus not achieving a possibly critical minimum level of education (cf. below).

Where social background does come into play is in the area of subordination or "recessivism" of L1 to L2 in the affective domains which, according to Seliger and Vago (1991: 4) is associated with notions of prestige, social status (of the speakers and the language), attitude and degrees of acculturation. In other words, perceptions of the L1 as having low social status would lead to negative attitudes towards it. This, in turn, would encourage speakers to abandon the L1 in favour of the higher-status L2, leading to L1 attrition. Further, social status of the parents' occupation has been shown to be an important factor in language shift studies, where a lower status was indicative of a greater probability of active bilingualism, the primary interlocutor for the L1 being the mother (Boyd 1986).

Education level has already been referred to as a factor which is likely to confound other variables, and which it is difficult to operationalise. Some studies, such as the Jaspaert and Kroon (1989) study referred to above, that have tried to establish its individual contribution to language attrition, isolated a high level of education (as measured in years of school attendance) as the main predictor variable for a lower level of L1 attrition, along with length of stay (cf. above) and proficiency in other languages (cf. below).

"This would mean that people with a higher education can maintain their language proficiency in an immigration context longer than people with a low level of education."  
(Jaspaert and Kroon ibid: 92)

The authors speculate that this capacity might, amongst other things, be explained by more educated people having more tools for language maintenance at their disposal, including more insight into the structure of language, possibly receiving higher salaries which would afford them more frequent travel home or other ways of keeping in contact, or being more familiar with and hence possibly readier to use written language, which again would maximise their opportunities for L1 use abroad (cf. also Köpke and Schmid 2004: 21).

In this regard, however, Jaspaert and Kroon's (1989) findings are surprising, insofar as the education factor appears to be far more important than contact with L1, a factor which in other



studies has been found to be very influential and a crucial variable in psycholinguistic models of language learning and attrition processes that employ frequency of input for explaining changes in proficiency (see above).

As a facet of educational level, L1 literacy skills per se or indeed biliteracy have been established as an important precondition for preventing language attrition (Baker 1996: 52ff; Hansen 2001), language shift (Davies 1986) and thus language death (Jahn, in Ammerlaan, Hulsen et al. 2001), partially because of the impact that learning a written code has for the perceived status of the language (Davies 1986: 124; Pan and Gleason 1986: 197), and partially because of its role in further "fixing" the corresponding language in the brain (Köpke and Schmid 2004). This incidentally provides a further argument as to why L1 attrition in very young children (up until the age of 8 or 9) should be far greater than in older children or adults (Olshtain 1986, cited in Köpke and Schmid 2004).

Köpke and Schmid (2004: 10) make the point that higher education may impact positively on linguistic performance in general and on one's capacity to solve linguistic tasks, which would explain better retention and better scores. A further possibility is that more educated people had a higher level of L1 proficiency from the start, although it is unclear whether such variability in L1 proficiency prior to emigration is large enough to account for variable L1 maintenance.

On the basis of Jaspaert and Kroon's (1989) study, unfortunately no claims can be made about the continuing impact of post-secondary education and the question of whether there is a minimum education level that needs to be achieved in order to maintain L1 better. This last question arises because of the possibility that incomplete education in the mother tongue may well have similar, albeit not quite as serious, consequences, for the full acquisition of L1 as a non-L1-acquisition situation brought about by emigration during childhood or adolescence.

If education level is restricted to (secondary) schooling, as in their study, this may create methodological problems for some populations of migrants, particularly those who have moved more recently, because of compulsory education to age 16 or more in most European countries, implying that there will be less sociolinguistic variation in the factor at this level.

Waas's (1996) study recorded education level from primary through to tertiary level and found that better educated people performed better on the verbal fluency tasks, while in Jaspaert and Kroon

(1989) education was a significant factor in relation to three out of four tasks: to a correction task, an editing task and a lexical task, but not to a comprehension task. The tasks in their study were constructed and administered in order of decreasing appeal to metalinguistic knowledge. This means that the most implicit task seemed to be least dependent on education level, surely an intriguing but not too surprising a finding if one considers that schooling is assumed to further the development of abstract thinking in line with commonly assumed cognitive development.

In contrast, implicit learning is associated with earlier developmental stages, so people with a lower educational level have mastered it as well. Since these issues refer to general cognitive changes rather than specifically language acquisition, the question arises (posed for instance in Köpke/Schmid (2004: 10)) whether education level impacts on language attrition only, or on linguistic performance in general, and whether there may be a correlation between task type and education level.

If education is a non-specific factor, then its impact should be seen in both the experimental and the control group, in that people with a higher level of education might be better able to solve some tasks, particularly those involving a metalinguistic dimension. Converging evidence for this contention comes from Yagmur (1997) and Köpke (1999), who, in addition to a verbal fluency task, found its impact in a relativisation task, as well as a task on sentence generation and a grammatical judgement task, but not on general discourse measures (picture description) (discussed in Köpke and Schmid 2004). From this discussion follows that for our purposes there would appear to be a definite need to control for this factor, particularly on tasks with an overtly metalinguistic dimension.

#### **2.4.5. Attitudes/motivation**

In contrast with the sociolinguistic factors just reviewed, which in attrition studies are mostly charted routinely as part of the participants' background data, attitudes have to be assessed either through inferences from biographical data or with specially designed instruments (the same is true for the other psycholinguistic variables), and are notoriously difficult to operationalise. Most commonly informants are presented with statements for ranking or rating.

To counteract problems potentially associated with such self-assessments, Schmid (2004) recommended the inclusion of a matched-guise task in the testing apparatus. In a matched-guise task, participants are presented with recordings of the same text in the languages known by them. These recordings are read by the same person, a highly proficient bilingual, but participants are not made aware of this fact. Instead, they are asked to rate the "two readers" in categories such as friendliness, approachability etc., thus allowing the researcher to gain indirect access to their emotional evaluation of the respective language and culture.

Regardless of which procedure is applied, it is assumed that more positive attitudes imply better language maintenance. However, as Hulsen (2000: 23) points out, this will only be the case if this attitude translates into concrete measures that are taken to support the language.

In discussing attitudes, it seems appropriate to draw a distinction between two facets: a) immigrants' attitudes towards the host community and their own community/background (as well as those of the host community towards them (Boyd 1986)), and b) their attitudes towards language maintenance and language learning, although these undoubtedly interact with one another, making a clear differentiation between them potentially problematic.

Starting with a), different ways have been suggested to operationalise this factor. One is to use measures of the person's ethnic affiliation which in turn may be deduced from one's membership of clubs and churches, active private and possibly work relationships and citizenship (permanent residents vs. naturalized citizens) (Waas 1996), or the extent and type of their social network (Hulsen 2000), on the basis of the assumption that attitudes interact with constructions of identity, the social networks that are being made use of, and ultimately language maintenance (de Bot and Hulsen 2002: 256).

Waas (1996), who studied Germans living in Australia, did find correlations between ethnic affiliation and self-reported proficiency, but not on tested proficiency. "This suggests that attitude, as measured in her study, influences self-perception but not actual proficiency, i.e. language attrition vs. retention" (Köpke and Schmid 2004: 13). Similarly, Hulsen did not find support for the hypothesis that language processing would be affected by characteristics of the informant's social network, such as the ratio between L1 and L2 contacts in the various domains, the domains of L1 contact, the size of the primary and non-primary, the home and the host networks, and suggests that

perhaps social networks relate more strongly to language use than to language processing (Hulsen 2000: 183).

As an aside, we would argue, on the basis of anecdotal evidence, that Waas's measure of citizenship would not appear to be reliably extendable to other contexts: A German living permanently in Ireland or the UK naturally faces the same question as a German in Australia of whether he/she will change their passport when becoming eligible. Doing so could probably be seen as a measure of positive attachment to the host culture. However, not applying for citizenship is not automatically an indication of the neutral or negative attitudes towards the host culture, particularly since within the European Union the only advantage of being granted citizenship is a person's right to vote in all elections, clearly not a priority for everybody.

Further, one major obstacle to deciding to obtain an Irish passport up until very recently was the impossibility of dual citizenship according to German law. This effectively meant having to trade in one's original citizenship although, one could argue, having both probably represents the self-identification of most long-term emigrants best. One could also view changing one's citizenship as an unnecessary bureaucratic step even for people who do not intend to return to their country of origin, as a result of the constitutional alignment of the rights of European citizens in Europe.

Finally, one also has to bear in mind that a strong desire to adopt a new identity is not sufficient for successfully doing so, since it needs to be afforded by the surrounding community. The community must be willing to allow a person to "shake off" their origins and past, and to become "fully assimilated" to the point of changing one's citizenship. The likelihood and indeed desirability of this to happen would appear to be lower in Europe, where people have traditionally retained a strong sense of regional and national difference and have never really subscribed to the "melting pot" metaphor.

Returning now to possible ways of operationalising attitudes, Ethnolinguistic Vitality Theory (Yağmur 1997; Yağmur, De Bot et al. 1999; Hulsen 2000) has recently been explored for this purpose, working from the idea that speakers from communities with high ethnolinguistic group and language vitality scores would be less likely to show attrition (or, conversely, if speakers in a group have experienced little attrition this may be a reason for high ethnolinguistic vitality scores of that group). Ethnolinguistic vitality is linked to attitudes via the distinction between *objective*

*vitality* as gleaned from "available sociological, demographic, economic and historic information" (Hulsen 2000: 31), and *subjective vitality*, i.e. perceptions of the vitality of one's own group and language.

Here it has been shown, that, particularly in unstable bilingual situations, there may be quite a big gap between the objective chances of the group to survive as an independent entity (Giles et al., 1977: 308, in Hulsen 2000: 29), and the perceived fortunes of the group. In such situations, informants may, "as the result of cognitive and motivational factors", over- or underestimate their group's actual strength, and this effect is most pronounced with first-generation immigrants, who, in contrast with later generations, often experience a "drop in status" (Hulsen *ibid.*: 31). This issue naturally ties in with the age-old question of whether to self-segregate or assimilate in the immigration context, and whether it is possible to maintain a distinct identity (and language as part of that) without resorting to building ethnic enclaves.

Thus, it would seem that measures of ethnolinguistic vitality should allow valid inferences to be made about attitudes. However, results have been inconclusive about the link between attitudes as conceived within this approach, and L1 proficiency. Hulsen (*ibid.*: 186), for instance, found high perceptions of L2 vitality to be linked to higher L1 reaction times, while there was no effect for L1 vitality scores, in contrast with what one might expect.

Another facet of the attitudinal spectrum, which we referred to as b) above, is a person's attitudes towards language maintenance. We would argue that this is indeed a distinct part of the complex of attitudinal factors, since a person may well feel very strongly about their ethnic identity and most things that go with it, but still be violently opposed to any measures supporting the language originally associated with that ethnic group.

For example, this is the case for many people living in Ireland, where the Irish language has become marginalised over the centuries despite great bottom-up activism and top-down support. In other words, a particular language may or may not be an essential part of the make-up of personal or ethnic identity, and thus attitudes towards maintaining it may differ drastically from attitudes towards other elements, such as traditions, sports, musical heritage, "ethnic characteristics" etc.

Apart from not seeing a language as an integral part of one's own identity, another possible reason as to why people may feel negatively about maintaining their L1 themselves or imparting it to their

children is that they may have a negative view of the learning of languages in general, possibly brought on by negative experiences. People often seem to subconsciously distinguish between their learning of "foreign languages" and L1 acquisition because they will have had very different experiences in these, and thus may not perceive a link between their "always hating French at school" and L1 attrition.

This proposed link has not so far been addressed in research. However, a tentative argument in favour of proposing such a connection is that L1 maintenance and re-learning in our experience strongly resemble L2 learning, that is the same measures employed to learn an L2, such as creating vocabulary "lists" – mental or physical – on the basis of semantic links, or rehearsing verb paradigms, have been found to be effective in re-activating L1. This point is interesting because anecdotally a great deal of L2 learning models and methodological advice are trying to emulate L1 acquisition processes in the hope of making L2 learning more effective.

This implies that the differing experiences of L1 "acquisition" and L2 "learning" can probably be attributed to the context in, and the age at which one became exposed to these languages, the latter determining the stage of cognitive maturation and thus to some extent the type of learning that one is capable of and that will be effective, rather than an intrinsic and qualitative difference between L1 and L2 acquisition. As a result, a person who has had positive foreign language learning experiences and thus a more positive outlook on language learning may unwittingly or consciously be employing L1 maintenance measures when necessary.

There is some empirical, albeit indirect evidence for the case of L1 attrition from a study of second-generation bilingual children, which found them to fare far better in L3 acquisition than their monolingual peers and which also revealed a positive effect of L3 acquisition on L1 maintenance via improved attitudes and facility in learning/maintaining languages: "Through L3, the pupils' home language maintenance is strengthened, they develop a more general interest in languages, and they engage in 'crossing'" (Clyne, Rossi Hunt et al. 2004).

Not many studies have looked at attitudes towards maintaining L1 separately. Of those who did, Hulsen (2000) did find more negative attitudes towards language maintenance to result in reduced lexical access/recognition in production, in line with the common-sense prediction that the less a person cares about a language the more likely s/he is to forget it.

There is probably no argument about attitudes determining whether a language will be passed on to the next generation, although this in turn has agency in deciding whether they want to take up that language themselves, when they are encouraged to do so and given opportunity to learn, or indeed despite being discouraged. However, it is at least debatable and requiring investigation whether under normal circumstances negative attitudes or disinterest have the capacity to impact on L1 proficiency in themselves, or whether observed attrition is rather a result of "neglecting" one's L1 due to such attitudes.

The interface between attitudes and L1 attrition would indeed seem to be language use, or the lack thereof, since attitudes and motivation likely impact on language choice in daily language use and on a person's willingness to make a special effort to support or learn an attriting L1. Moreover, attitudes interact with a whole set of other circumstances (such as constitutional and perceived status (prestige) of the language, presence or absence of institutional support, personal experiences), and are therefore not only rather difficult to assess but are also likely to be quite changeable, limiting their potential effectiveness.

This would explain why the only study able to unambiguously isolate attitudes as a predictor for the extent of L1 maintenance/attrition was a study involving participants with exceptional and traumatic backgrounds - German Jews who by the time of their emigration to the US before or during the second world war had experienced a greater or lesser degree of persecution (Schmid 2001).

Consequently, they differed in how negatively they viewed their backgrounds and their mother tongue, and in how they construed their subsequent identities. As suggested by Köpke and Schmid (2004), circumstances surrounding the adoption of children may similarly serve to create a stronger than usual antipathy for one or the other language and culture, resulting in attrition in that language and a more definitive link between attitudes and attrition.

A further factor which has not been studied systematically at all, but which we argue cannot be neglected in attrition studies, particularly ones with a psycholinguistic focus, is the extent to which becoming aware of a language being lost may impact on a person to start counteracting this loss.

While in an L2 environment, the context for this study, it is undoubtedly important to master the language of the host community, nevertheless importance has to be attached to retaining the

linguistic and cultural background, at least for reasons of cultural diversity. Ironically, as has been argued by Perdue (1993), language and cultural loss may "be provoked by the very success of the process of settling into the new country".

This could happen without an external impetus, for instance, when a person realises that he or she is not able linguistically to do the things he or she used to do, or attention may be drawn to this fact through outside factors, such as participation in a formal experiment. Not all such realisations of "gaps" will require specific and active contravention, such as checking the form in a dictionary. For example, confrontation with not being able to come up with a particular word may simply trigger a more extensive retrieval chain, which, although possibly not completed in the time allowed in an experiment or at the time of conversation etc., might eventually lead to the word in question, constituting re-exposure and re-learning.

"The confrontation with a certain amount of language loss may be all the stimulus the informant needs to remember aspects of the tested variables in an analogous test situation some months later. Then the design itself has disturbed the natural course of the process it hoped to track down" (Jaspaert and Kroon 1989: 81).

It has been proposed that people possess "agency" in language learning (Swain, p.c.), and by extension in language loss, which means they will differ in the extent to which their awareness of L1 becoming weaker propels them into action. Agency is implied in the notion of language effort proposed in DMM (Herdina and Jessner 2002). However, from this discussion and the nature of the factor, it is clear that it is difficult to operationalise and to measure, making predictions of its effect largely guesswork. Nevertheless it seems a worthwhile area of investigation, particularly in the context of assessing learner agency.

#### **2.4.6. *Personality and aptitude***

Twenty-five years ago, Ginsberg (1986: 21), in a paper on methodological issues in language attrition research, pointed out that neither linguistic aptitude, general intelligence, learning styles and other cognitive attributes, facility in various aspects of English nor test-taking ability had been interrogated as to their contribution to language attrition in a single study to that date, including his own.



A review of the current literature reveals that this situation has not changed substantially: Köpke (2007: 19f.) confirms that language aptitude "has not so far been investigated in this context" (there has since been one study by Bylund, Abrahamson et al (2010) of L1 Spanish-L2 Swedish bilinguals, cf. below), while the other personality and cognitive attributes do not get mentioned in reviews of the state of the art (e.g. Köpke and Schmid 2004; Schmid and de Bot 2004; Schmid 2009). Consequently, at this point we are again at best able to speculate how or indeed whether these factors might impact on L1 attrition, basing these guesses on findings from L2 acquisition studies and the assumption that similarities exist between L2 learning and L1 maintenance.

One lead followed in L2 acquisition studies is that learners with more outgoing personalities have greater success in L2 acquisition, since they would be more likely to practice their skills (Ellis 2000), but no unambiguous link has so far been established. Similarly, the factor field-independence – the degree to which a person can dispense with context in language learning – has been proposed as a predictor variable for more successful L2 acquisition. This is related to the distinction between analytically oriented learners and memory-orientated ones.

In this regard, Skehan's work (1998) seems to suggest that these simply may be alternative paths towards the same goal with similar chances of achievement. However, he further indicates that at high proficiency levels the ability to analyse language becomes relatively less important than the ability to memorize, so memory-oriented learners may have a late advantage over the analytic ones. On the other hand, this advantage may be offset in analytic learners by virtue of their cognitive orientation possibly developing metacognitive skills to a greater extent than memory-oriented learners, since metacognitive skills are thought to facilitate the (L2) acquisition process and to be indispensable for going beyond a certain level of achievement.

The proposed link between the ability to metacognise and achievement has not been confirmed unquestionably; and in fact the direction of cause and effect may well be reversed, given that early bilinguals have invariably been found to possess a higher level of metacognitive skills (Bialystok 2001). In other words, these skills may develop as a result of L2 acquisition rather than the other way round. Similarly, no clear link has been found between metacognitive skills and cognitive style. On the basis of these findings, or their absence, it would appear premature to hypothesise that more outgoing and linguistically aware people would be the better L1 maintainers at this point.

General intelligence and linguistic aptitude have held great interest for L2 acquisition studies, which have researched them extensively (Skehan 1998; Ellis 2000 for references and discussion), but, as pointed out above, have surprisingly been neglected in attrition research. Köpke points to the potential of investigating the concept of language aptitude since it "might be better suited to capture individual variation in attrition than e.g. education level" (Köpke 2007: 20).

The results of the relevant L2 acquisition studies have shown that both general intelligence and language aptitude, which are seen as relatively immutable and probably innate, play a role in enhancing "the rate and ease of [language] learning" (they are not prerequisite to successful L2 acquisition), provided the learner is "sufficiently motivated to learn (i.e. make the effort to use their intrinsic abilities)" (Ellis 2000: 496).

Language aptitude is seen as a "special propensity for learning an L2" (Ellis 2000: 494) and has been confirmed as the "single best predictor of achievement in a *second language*" (Gardner and MacIntyre 1992: 215, my emphasis), but one of its dimensions established by Skehan (1998), which he terms "underlying language learning capacity" (Ellis 2000: 497), also correlates with L1 achievement. Ellis argues that this dimension can be related to Cummins's BICS (basic interpersonal communicative skills, the skills required for "oral fluency and the sociolinguistically appropriate use of a language" (Ellis *ibid.*)), which are developed during L1 acquisition, so that this particular part of language aptitude would be relevant for acquiring any language, L1 or L2(s).

The other aptitudinal dimension identified by Skehan – "ability to handle decontextualised language" – can be related to CALP (cognitive academic language proficiency), which in turn strongly correlates with general intelligence (cf. Ellis 2000 for references). This dimension, then, is not specific to language acquisition but would enhance all learning. Skehan (1998) has also proposed that as a result of the specific strengths of components of aptitude, learners may have distinct aptitudinal profiles (memory-oriented vs. analytic learners) affording alternative ways to enhanced L2 learning. This last proposal touches on the concerns of the next subsection.

On the basis of these findings, we may hypothesize, following Köpke (2007: 20), that people with higher language aptitude will be better L1 maintainers, since aptitude impacts positively on the ease with which L2 (and L1?) is acquired. Similarly, higher linguistic aptitude and higher (linguistic or verbal) intelligence, one of the eight types of intelligence proposed by Howard

Gardner (Holden 2005), should result in better L1 maintenance because of greater facility in acquiring, storing and retrieving linguistic material, thus reducing the language effort required for these purposes.

To date, there has been a single study investigating the role of language aptitude in L1 attrition on the basis of an L1 grammaticality judgement task completed by 25 pre-pubescent bilinguals with L1 Spanish and L2 Swedish (Bylund, Abrahamsson et al. 2010). Participants with above-average language aptitude were found to be more likely to be performing within the range of a matched control group, whereas for participants with lower aptitude scores the amount of L1 use mattered for the success with which L1 was maintained. This led the authors to suggest that "language aptitude has a compensatory function in language attrition, helping the attriter to retain a high level of L1 proficiency despite reduced L1 contact" (Bylund, Abrahamsson et al. *ibid.*: 443).

#### **2.4.7. L2 proficiency**

Finally, we return to the question of whether L2 proficiency may be a predictor for L1 attrition. Here, we review some studies providing converging and attenuating evidence. Like most other factors discussed earlier in this chapter, this variable is an ambiguous one. This may be a result of different operationalisations of the factor, and consequently two essential distinctions have to be made between proficiency in the L2 (the language of the host community) vs. proficiency in L3s, and between the fact of having learned another language vs. the actual proficiency level in that language.

For example, Jaspaert and Kroon (1989: 90f) find what they term "proficiency in languages other than L2" (as opposed to proficiency in L2) to be consistently amongst the top 5 factors (education, stay, proficiency in other languages, nationality of partner, age) for L1 maintenance, and amongst the top 3 after correcting the results for the control group baseline (as opposed to a 100% reference point). Proficiency in languages other than L2 was significant for the two tasks with higher appeal to metalinguistic knowledge (COR and EDIT).

On further reading, however, it becomes clear that we cannot be sure as to which level of proficiency in the L3s is necessary to produce this effect, since participants merely had to state

whether they had or had not at some point learned another language (answering a yes/no question). In other words, from this study it is not so much proficiency in L3 which is a predictor for L1 maintenance, but rather knowledge of an L3. As such, this result may be a by-product of their differential education level reported earlier rather than a finding in favour of the positive effects of foreign language learning for L1 maintenance.

On the other hand, Clyne, Rossi Hunt et al. (2004) report that Australian bilinguals learning a third community language which may or may not be related to their home language, not only tended to be "more effective and persistent learners of the target language than monolinguals" who were able to "benefit from their metalinguistic awareness", but whose home language maintenance is improved through learning the L3, which the authors ascribe to the students having developed an interest in language learning per se.

Thus, it is possible that the mere fact of having been exposed to a different linguistic system or systems might confer a certain advantage for language maintenance in terms of providing insights into linguistic structures which are not available to monolinguals. This argument may be summarized as "the minds (and language) of bilinguals [and by extension multilinguals] differ from those of monolinguals" argument (Cook 2003; Singleton 2003) which was discussed in greater detail above.

We are not aware of other research that has raised the question of knowledge of or proficiency in L3s. Thus, while knowing more than two languages seems to confer an advantage for L1 maintenance, we do not know what the impact of different proficiency levels in L3 might be.

Knowledge of L2, on the other hand, has often been considered as one of the factors instrumental in L1 attrition, following from the observation that L1 attrition typically happens in a context where an L2 is being acquired, and that most L1 effects in attrition can be related to interference from L2 of one kind or another.

However, it is the contention of this study that, possibly in addition to the effect of "knowing" one or several other languages in the sense of having been exposed to them, the level of proficiency achieved in L2 (and by extension any L3s) matters for L1 maintenance/the extent of L1 attrition, which as we shall see has not been addressed very often. The strongest support for our contention so far comes from child studies and the area of phonology in adults.

So far, empirical evidence for the proposed correlation between proficiency levels in L2 and L1 is rather limited. To start with the negative evidence, in Jaspaert and Kroon's (ibid.) study above, participants, in addition to giving information about their knowledge of L3s, were also asked to provide an indication of their proficiency level in their L2 Dutch (by means of self-rating scales).

The results show that overall and amongst all 8 factors investigated (social background, age, stay in the Netherlands, education, proficiency in Dutch, contact with Italians, nationality of partner, proficiency in other languages), L2 proficiency was relatively unimportant and accounted for only a small percentage of the variance, apart from the lexical task (one of the two tasks with lesser appeal to metalinguistic knowledge) where it is found in 4th place without being significant. Otherwise (particularly after correction for the control group baseline) it comes last or second last. Thus, there is a clear difference between the effect of L2 proficiency (next to none) and L3 knowledge (highly significant). Similarly, Sharwood Smith's "Utrecht project" with English-language child attriters in the Netherlands (reported in Sharwood Smith 1983b) failed to produce a clear link between L1 loss and L2 proficiency:

"[T]here was no significant relationship between L1 loss of adverbial placement rules and L2 mastery of the equivalent rules in Dutch" (Sharwood Smith 1983b: 223).

The participants did, however, exhibit significant differences in their languages in comparison with both the English and Dutch control groups and the extent of this L1 loss did correspond to the amount of L2 use at home.

Another child attrition study (Okimura-Bichard 1985) found that "at least for some speakers, increasing proficiency in the societal language need not be accompanied by regression in the mother tongue" (Pan and Gleason 1986: 196). The quote does seem to suggest, however, that the expected development is a pattern of increasing L2 proficiency with simultaneously reducing L1 proficiency.

In this study, a positive correlation was established between years of schooling in L2 English and L1 Japanese with skill in the respective language, but schooling in English did not discriminate between children with high proficiency in both languages, low proficiency in both languages, or high proficiency in L2 and low in L1. The latter pattern would be expected if L1 attrition was a function of L2 acquisition.

These kind of data prompted Sharwood Smith (1983b) to argue against a possible, but simplistic interpretation of the concerns of the present study, namely the assumption that L1 loss should occur because L2 is taking L1's place. In other words, the lack of a significant relationship would suggest that there is no automatic replacement of L1 by L2, although there is a great deal of interaction and selective transfer between the two languages.

However, other research suggests that there may be a link between proficiency levels. To cite findings from another child attrition study, Marino (1983, reported in Pan and Gleason 1986) found that bilingual Chicano children showed consistent growth in English language proficiency between kindergarten and grade 4, language "arrest" or fossilisation in Spanish comprehension, and attrition in Spanish production since fourth-graders performed worse than children in kindergarten. This result is evidence that this is not a case of non-acquisition. This pattern of development, where children eventually become more proficient in their L2 and attrite or do not continue to develop in their L1, is seen as typical (Pan and Gleason 1986: 196).

Jia and Aaronson (1999) found a steady narrowing of the distance between L1 and L2 scores (L1 scores decreasing, L2 scores increasing) with increasing age on arrival until about age 12 at which point the languages cross over. This effect prompted the authors to propose the Dominant Language Switch and Maintenance Hypothesis according to which younger arrivals are likely to switch, and older arrivals to maintain, their language dominance as a function of age on arrival, and original L1 proficiency. ("Dominance" is to be read here as "proficiency" since it was assessed on the basis of grammaticality judgement and other proficiency tasks.)

A similar cross-over effect had been found in Mägiste (1979) who described a dominance switch (dominance was measured in reaction time) after 4-5 years for decoding, and 6 years for encoding. Following the switch, the two languages developed in the same direction, but at a much decreased rate and in some cases producing a parallel development pattern.

In other words, the changes in coding and encoding proficiency appear to be correlated as they change simultaneously at roughly the same rate, and once the proficiency levels have more or less converged, they improve consistently in a coordinated fashion. It would be interesting to know how the RT values relate to monolingual values of the two languages, not to pass judgement about

the L2 users performing better or worse, but to see whether the fact of knowing two languages brings about a convergence on this variable.

Yeni-Komshian, Flege and Liu (2000) also posited an "inverse relationship" between L1 and L2 pronunciation proficiency: "[t]hat is, as L2 pronunciation proficiency increases, proficiency in L1 may decrease, and vice versa" (ibid.: 134). The two proposed causes are L2 learning on the one hand ("increased phonetic learning in L2 should result in more segmental changes in L1"), and L1 disuse on the other ("pronunciation in the L2 may be enhanced because the attenuated L1 will have a weaker influence on the L2"). Generally, the pattern was confirmed, but the effect only held for ages on arrival up until age 12.

Segalowitz (1991) investigated the interactions between L1 and L2 reading skill from a perspective assuming competition between languages wondering:

"might the time spent reading in L2 be considered, in effect, time taken away from just this sort of consistent practice with L1 [required for skillful performance], and does this produce adverse results such as reduced speed of processing (less automaticity and/or reduced effectiveness of strategic processing) in L1 reading?" (Segalowitz 1991: 61).

His results provide evidence for a correlation between high L2 reading skill (an indirect indicator of proficiency) and decrements in L1. Interestingly, only controlled processes were subject to decreased performance, while automatic processes were unaffected. The results confirm the view that high proficiency in the L2 might result in competition for reduced opportunities to maintain the L1, trade-offs between the languages and eventually L1 attrition.

As far as the relationship between the performance/proficiencies in the two (or more) languages of adult L2 users is concerned, very little research exists. One of the few exceptions is Major (1992), who, looking at a specific phonological aspect (VOT) in monitored and spontaneous speech, found a definite pattern for a correlation between L1 and L2 proficiencies. Major's (ibid.: 191) hypothesis was that

"[t]he greater the degree to which a speaker masters one system, the greater the extent to which one might expect it to affect another system. This pattern should be true regardless of whether the system under consideration is L1 or L2. Just as we see a strong L1 influence in most individuals, so too would we expect L1 loss to be greatest in individuals who are

under a strong influence of a L2, that is, in individuals who have mastered the L2 to a relatively high degree."

His participants fell into the following three categories: 1) Two of his participants had not lost the English feature to a great extent, but had also failed to acquire the L2 Portuguese one to a very high degree (all tasks and conditions). 2) One participant performed in a native-like manner in both L2 and L1 for the formal condition of the test, but was not recognisable as a native speaker of L1 in the casual condition. 3) Two further participants did not perform in a native-like way in either language on the investigated feature since their VOT values were half-way between the native values for either L1 and L2 (Major 1992: 193) (cf. Figure 3).

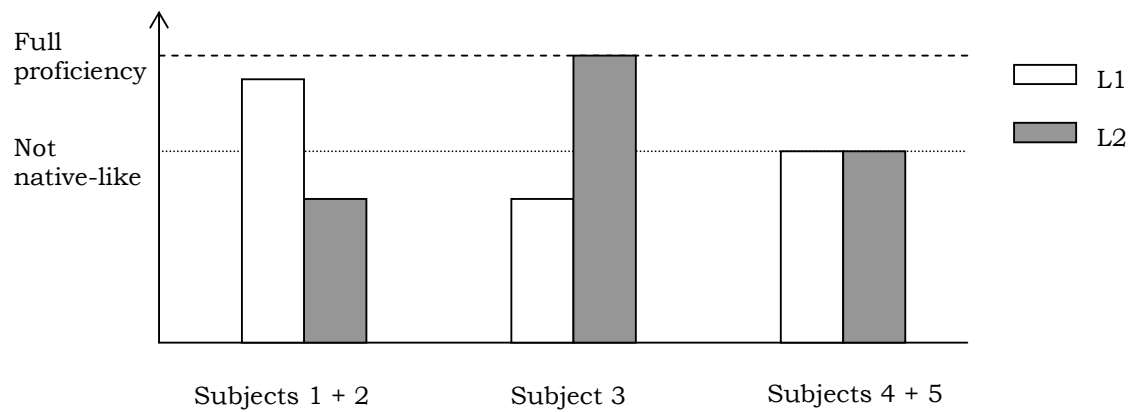
This latter group is reminiscent of "semilinguals" who do not achieve native-like proficiency in any of their languages, although admittedly we are talking here merely about one isolated feature. The effect was more pronounced in the "casual condition" when participants were monitoring their speech less.

Major deduced an implicational hierarchy from the data, whereby loss of the feature in the former condition is predictive of loss in the latter, but not vice versa. Overall it would appear that it is more difficult to retain an L1 casual style, which Major (having ruled out that the formal style is used more often) ascribes to the greater difficulty for monitoring performance.

"Speakers who closely monitor their speech, whether in L1 or L2, can more easily suppress the influence of whatever other linguistic system (L1 or L2) they have acquired which can interfere with the target. Furthermore, the greater the mastery of a linguistic system, the greater its influence will be on another linguistic system, especially in an unmonitored situation." (Major *ibid.*: 202)

Thus, there would appear to be a double-effect: Loss of casual style is both subject to the degree of L2 proficiency AND occurring independently of the degree of mastery of L2.





**Figure 3 Schematic representation of proficiency levels in casual condition** (after Major 1992)

Overall, Major's findings support his hypothesis of greater L1 attrition being implied in a greater degree of L2 acquisition. He deems subjects 3-5 (significant loss at high or very high L2 proficiency levels) to be providing strong support for his hypothesis, whereas subjects 1-2 (little loss, little acquisition) neither support nor refute it.

The results would also seem to indicate that the participants experienced some trade-off in the processing of the two languages in line with the contentions of the previous section – they either kept up their L1 to the detriment of L2, or achieved native-like level in the L2 to the detriment of L1, or kept neither L1 nor achieved native-like level in the L2. This finding could reasonably be explained by the speakers' need to achieve a compromise in a situation when resources are not sufficient to sustain or achieve a native-like level in both languages.

It should be noted that the participants in this study lost their L1 to some degree despite them having a vested interest in maintaining it since all were working as teachers with their L1. That such an interest and the attendant positive attitude towards the language to be maintained alone are not sufficient to prevent L1 attrition was also observed by Sharwood Smith (1983b) and Olshtain and Barzilay (1991):

"There is, in fact, a good deal of evidence that even native speakers with responsibilities as teachers, that is, professional native speakers, manifest interference in their spontaneous speech quite frequently, and also that students of English as a foreign language at the advanced level find their L1 affected" (Sharwood Smith 1983b: 224, my emphasis).

In the case of the first constituency, Sharwood Smith "blames" the frequent exposure to imperfect learner speech for this effect, whereas in the latter case the interference seems to be a side effect of high L2 proficiency, suggesting that the relationship between L2 and L1 proficiency may be different at different proficiency levels. The possibility that correlations between L1 and L2 proficiency levels may only exist at relatively high levels of L2 proficiency makes intuitive sense, since below a certain level there may simply not be enough L2 present to affect any influence on L1, as was already suggested in the previous section.

Empirical evidence for the role of high L2 proficiency comes from Sorace (2000) who in her investigation of the use of overt pronouns still found "residual" optionality in their "L2 endstates", presumably as a result of the persistent influence of L1, but also saw optionality in L1 syntax "emerging" (Sorace 2000: 719).

"Near-native proficiency in English involves the acquisition of a system in which there is no obligatory occurrence of [+Topic Shift] with any pronominal form. As a result, this possibility may also become lost in the null subject L1 grammar: that is, the interpretable feature obligatorily associated with an overt subject pronoun becomes optionally unspecified, and the visible effect is the occurrence of overt pronouns in [-Topic Shift] contexts [...]" (Sorace 2000: 722ff).

Thus, before attrition the overt pronouns in L1 always carried the [+Topic Shift], whereas in attrition the marked option became destabilized and the pronouns optionally unspecified.

To summarize, the evidence accumulated so far seems to suggest that L2 proficiency may, under certain conditions or in certain linguistic areas, be predictive of L1 loss. The few studies that have recorded and analysed the factor, seem to converge on it not being the first and foremost predictor, which, however, does not mean that it can be ignored. It is unclear whether there are differences between adults and children in this regard, and whether in order to produce an effect, L2 proficiency simply has to be higher than a certain minimal level, or whether in fact there is a correlation between the actual proficiency levels after a certain point.

The strongest evidence for an interconnection between high L2 proficiency and L1 attrition comes from studies investigating phonological changes, whereas in the areas of lexical and grammatical competence the link is not as clear. This raises the question whether phonological

proficiency is a competence separate and independent from other sub-components of proficiency. Overall the evidence must be considered as too sporadic and ambiguous. Thus, the jury is still out on the issue of a link between the levels of proficiency in two or more languages, the issue prompting the current investigation.

However, there is one further complication to do with the point at which resources start to become limited, both in general and with regard to L1 in particular. In other words, we need to ask whether there is a level of knowledge above which further increases in that knowledge will put excessive pressure on either memory or processing. (The example of multilinguals who speak several languages fluently, and of highly educated people certainly cast doubt on this proposition, although there is probably a great deal of inter-individual variation.)

So how can we be sure that the interactions described above are in fact due to competition between the systems? Neither transfer, nor indeed code-switching can exclusively be put down to computing pressures. For example, transfer from L1 to L2 is at its highest when L2 is only starting to develop, so surely this is for reasons other than difficulties in storing or processing a new language, such as our knowledge-constructivist tendencies.

One thing does seem to be clear, however: there would appear to be a lower limit below which it would be unreasonable to anticipate strong effects caused by competition. As Carroll (1968: 118, quoted in Sharwood Smith 1983: 222) points out: "[I]n teaching a second language we do not expect our students to forget the first system." It would indeed be counter-intuitive if the teaching happens in the L1 environment and the level of proficiency achieved in L2 is far from advanced.

While this quote reveals our bias towards taking L1 impact on L2 for granted but not the reverse, it also draws attention to the likely possibility that the latter might only make itself felt at a sufficiently high proficiency level in L2. In other words, it only makes sense to talk about competition when the L2 system has developed sufficiently so it can begin to rival the L1 system, particularly when this is aided by an environment where L1 is not naturally supported.

In summary, we would cautiously like to propose that in a situation of highly developed bi-/multilingualism the languages in question "compete" for potentially insufficient resources required for maintaining the languages simultaneously at the desired high level. Thus, because of trade-offs, the proficiency level attained in the L2(s) and the fact of attrition in L1 could be connected.

## **2.5 Summary**

From the predictor variables discussed above, the following picture emerges: The younger a person when placed into an L2 environment and the less contact this person has with L1 subsequent to migration, the more profound and wide-reaching the effects of L1 attrition. Pre-puberty children additionally may stop acquiring L1. Adult migrants suffer less attrition than children, one of the main reasons being their having a far better developed L1 system as well as different choices and opportunities (or lack thereof) for maintaining the language. Frequent opportunities for L1 use per se are, however, not sufficient to prevent attrition. Attrition effects typically show during the first decade of living abroad, after which the pace of attrition slows or stops. Education level is more important than social background, but causation is unclear.

In relation to the psycholinguistic factors, some interesting findings have been made, but overall this is an area requiring a lot more attention. The emerging picture suggests that attitudes and motivation may account for some of the variance, but that as variables they are ill-defined, and results stemming from various conceptualisations are contradictory. Nevertheless, even on the basis of such preliminary findings, Hulsen, de Bot et al. (2001) suggest that any psycholinguistic factors or effects discovered and interpreted in a psycholinguistic manner will most likely be found to offer alternative explanations for or even correspond to sociolinguistic, attitudinal and/or biographical factors already established.

Overall then, this chapter provided an overview of the current state of knowledge in the area of attrition research, particularly L1 attrition, and, where plausible, tied the discussion in with the parallel process of L2 acquisition. As we saw, a lot of the concepts and indeed models can be usefully shared between the two perspectives. Relating the two processes becomes an imperative in the adopted framework of bilingual or L2 users' multi-competence, and by extension within dynamic approaches to bilingualism, including the Dynamic Model of Multilingualism discussed in detail in this chapter.

### **3 Experimental design**

#### **3.1 Participants**

The primary informants are 27 German native speakers resident in Ireland (“bilingual group”; S1-S27) whose data are compared to that of two matched control groups, a German one (20 Germans resident in Germany; S31-S50) and an Irish one (18 English native speakers resident in Ireland, S51-S68). At the early stages of planning this investigation, we envisaged testing a group of Russian native speakers resident in Ireland in addition to the German bilingual group for an additional sociolinguistic dimension and crosslinguistic comparisons (Opitz 2005). However, time constraints and theoretical considerations led to this idea being abandoned in favour of just one group of German-speaking migrants.

As a result, it was possible to more than double the size of the bilingual group, as well as to increase the number of controls (to 27, 20 and 18 as indicated above). Ideally, the three groups would have been equal in size, but this proved impossible within the given timeframe due to difficulties in enlisting controls and logistic problems. The increase in group size impacts positively on the study’s rigour with respect to quantitative parameters. Nevertheless, it is important to recognise that this is still a relatively small cohort of informants with all the attendant dangers that arise when attempting to generalise from such a population to larger ones. This limitation is offset by the wealth of qualitative information yielded by the investigation.

The three groups were matched along a number of primary criteria (age; gender; education level, cf. Appendix I for basic participant data and participant labels) and additionally some secondary criteria to the extent that this was achievable (marital status, partner’s L1 and presence of children as variables modulating the choice of languages in the home, and thus language use; living in rural vs. urban areas; German and English variety spoken; level of L2 proficiency in German or English respectively).

The last criterion points to the fact that although we refer to the group of Germans living in Ireland as the “bilingual group”, and the individuals within the group as “(adult) bilinguals”, in fact all participants in all three groups are late bilinguals. A late bilingual is a native speaker of one language who grew up monolingually but has later in life, through schooling and/or trips to or

residence in a country where another language is spoken, achieved a reasonable level of proficiency in at least one other language.

Thus, in our case, the German participants were all reasonably proficient in English, and the Irish participants in German, operationalised here as a self-rated “good”, or point 3 or better on a 5-point L2 proficiency scale (cf. Section L2 proficiency below). Many of them know further languages, although this was not a selection criterion. While comparable on this criterion, the three groups are systematically different with respect to the amount of time spent abroad in a German or English L2 environment (seven years minimum for the bilingual participants; less than one year for the controls), their current place of residence (L2 environment for the bilingual participants; L1 environment for the controls) and consequently the degree of immersion experienced in their respective L2.

Requiring the controls as well as the migrant group to be bilingual is in deliberate contrast with traditional studies in bilingualism and L1 attrition which compare bilinguals to monolinguals, for three reasons.

First, the study’s design requires at least a basic level of English-language proficiency on the part of the German control group for performing the linguistic aptitude test and the film retelling (cf. “Tests” below).

Second, finding monolingual controls who would match the bilingual participants in all other relevant aspects, such as age combined with education level, would have been impossible due to comprehensive foreign language teaching in both Germany and Ireland. Since 1964, and in some regions even before that, anybody finishing German secondary school has studied a second language, mainly English. Similarly, most Irish people in the Republic of Ireland have at least studied Irish, and usually a foreign language for several years.

Third, and most importantly, as we discussed above, comparing mono- and bilinguals seems to be ill-advised in consideration of evidence for irrevocable changes to a person’s mind as a result of acquiring a second language (e.g. Cook 2002; Michael and Gollan 2005; Schmid and Köpke 2007), on the one hand, and in view of the "ethical" and "fairness" arguments against using the monolingual native-speaker standard (e.g. Cook 2003: 4f), on the other. The latter argument is an implicit consequence of the former: if bilinguals perform systematically differently from

monolinguals, then measures apparently showing a lack of L2 acquisition or reduced L1 performance may simply be a result of differing cognition. Labelling participants as "failed acquirers" or "attriters" on the basis of such measures would then not only be incorrect, but also unfair in so far as the difference would be interpreted as a deficit.

Consequently, we argue that the comparison should be made between groups of participants who have all had the experience of learning another language. In so doing, we can exclude the "bilingualism effect" in our data and rest assured that any differences found are in fact due to attrition or lower levels of acquisition by the participants relative to the control groups.

Having decided to deploy a bilingual comparison measure, an important consideration is the level of L2 proficiency which should be required of the controls. Interestingly, neither Cook nor Grosjean who argue the case of treating bilinguals as a group in their own right, are specific about the level of L2 proficiency that would be necessary to effect a change in cognition.

One possibility would be to try to ensure that controls matched the level of L2 proficiency achieved by the bilinguals. However, given the different contexts the participants in each group find themselves in for acquiring the L2, different factors are likely to play a role in participants' ability to attain a high level of L2 proficiency.

The bilingual participants have a greater chance to experience an input-rich environment, whereas the controls do not have the benefit of "automatic" reinforcement of their second language skills. Instead, they need to have sought out opportunities to practice and develop their second language, by travelling, using the language professionally, attending a language class or developing friendships. In lieu of such opportunities, aptitudinal factors would have to compensate for the poorer L2 input. Otherwise the environment effectively limits the degree of L2 proficiency achievable, rendering any requirement for the controls to be as proficient in their L2 as the bilinguals an unrealistic expectation. This was in fact borne out by our inability to match the L2 proficiency levels of bilinguals and controls (cf. section 3.1.2.4 L2 proficiency).

If, on the other hand, the controls did manage to create a similarly L2-input-rich environment in their home country compared to that experienced by the bilingual participants by virtue of their living in the L2 environment, this might have implications for their status as native speakers of their L1. It could be argued that, like in the case of the bilingual participants, their exposure to and

practice of the L2 is at the expense of maintaining their mother tongue. However, it is rather improbable that a person living in their native environment could experience L1 attrition, since the effort required to support the mother tongue in the congruent environment should be minimal, and reinforcement "automatic" (cf. the discussion of DMM in Chapter 2).

An alternative perspective on the required level of L2 proficiency on the part of the controls involves accepting that the very fact of being exposed to a second language system, irrespective of the level of proficiency achieved, brings about a quantum leap in cognition, not unlike grasping the concept of reading and becoming literate (O'Rourke, p.c.). However, it is likely that the impact on cognitive measures (response latencies, for example) is greater, once an L2 user has improved their proficiency beyond the very basic level.

With these considerations in mind, we opted to choose bilinguals with what we judged was a meaningful level of L2 proficiency, in other words, a minimum rather than an absolute requirement of the level of L2 proficiency. This decision minimises any potential danger of the controls having become L1 attriters in their L1 environment due to lack of L1 input, while excluding participants whose L2 proficiency is too low to be likely to have any impact on their L1 cognitive system.

As a result, the L2 scores of the controls at the time of testing are not dissimilar to those of the bilingual group prior to migration, after discounting the cases of very low initial L2 proficiency. There are two controls who claim to have had excellent L2 skills at the time of testing, a category which is not represented in the bilingual group's initial L2 proficiency. Possibly, this difference reflects the limits of attaining a high level of L2 proficiency in the L1 environment at the time of migration of the bilinguals (on average 19.5 years prior to the time of testing, and in some cases substantially longer) vs. those pertaining now, as a result of easier travel and enhanced digital communication in the recent past.

Specific to this investigation is that participants were required to be late bilinguals of the same language combination (German-English, with German being L1 and English L2 for the bilingual group and the German control group, and vice versa for the Irish control group). This would not strictly have been necessary to address the "changed cognition" argument since this would be satisfied by sufficient proficiency in any other language. However, this requirement allowed for direct and objective comparisons between the three groups with respect to the bilingual group's



dependent variables. Knowledge of other L2s was not controlled for; but the maximum length of residence in countries where that language is the main language of communication was.

### **3.1.1 Participant search**

Potential candidates for both the bilingual and the two control groups were sought through personal contacts, fliers and newsletter items distributed in relevant departments at universities and institutes of technology, the Embassy and the Goethe-Institute, as well as St Finian's Lutheran church, the German Catholic community, St. Killian's German school, the German butcher's shop and the Liebherr factory in Killarney.

The project was framed as a study in "adult bilingualism" in order to avoid mentioning the potentially sensitive issue of "language attrition". Such partial concealment of the purpose of the investigation is recommended in the literature (Wray, Trott *et al.* 1998, citing Saville-Troike 1989: 128 ) in order to counteract biases on the part of the participants in the areas that matter (i.e. "courtesy"/"sucker" bias: participants' potential tendency to either please or deliberately mislead the researcher).

However, a different kind of bias may have accidentally been created as a result of the selection procedure (self-selection) combined with the advertised theme. It is possible that this topic attracted primarily people who are especially interested in and successful at learning languages and maintaining their bi- or multilingualism. We attempted to counter this bias through encouraging a word-of-mouth recruitment campaign. Consequently, many people volunteered as a result of a friend's suggestion rather than their own intrinsic interest in the issues.

After first contact, participants were sent a personal background questionnaire with a view to collecting general biographical data as well as information on L2 acquisition and L1/L2 use. A formal meeting for conducting the tests was then arranged. Sessions mostly took place over one session, but some of the bilingual participants opted for two sessions. The sessions with the bilingual group were conducted between March and July 2006; those with the German and Irish control groups took place between March 2007 and May 2008. Most participants completed a further questionnaire in their own time which contained can-do scales and some statements

describing conceptions of bilingualism and foreign language learning. This questionnaire was handed to participants, with explanations, at the end of their session.

Participants did not receive any kind of material compensation other than being entered into a draw for some small cash prizes, and a small edible “thank-you” token which was given to each informant at the end of their session. The researcher did, however, offer participants the possibility of being kept informed about the progress and the outcomes of the study.

### **3.1.2 Selection criteria and participant characteristics**

#### **3.1.2.1 Age, gender, residence abroad, education level**

The primary informants were 27 native speakers of German (25 Germans, 1 Swiss and 1 Austrian) drawn from migrant communities in Dublin and the west of Ireland. Table 2 gives an overview of basic background data for the bilingual group which are discussed in this and the following sections (with partial replication of the table).

In order to be eligible, the bilingual participants had to meet certain criteria in addition to the late bilingual and L2 proficiency criteria set out above. First, the minimum age at emigration (or age on arrival) was set at 17 years to ensure participants had completed their first language acquisition. Informants who are younger at age of emigration, particularly up to the age of 12, present the problem of a clear distinction between attrited and non-acquired language, particularly with cross-sectional designs. This minimum is in line with other studies of adult migrants (e.g. de Bot, Gommans et al. 1991) and corresponds to the minimum age recommended as part of the common methodology of the Graduate Network on First Language Attrition (Schmid 2004).

Second, the minimum length of residence abroad (LOR), or "time elapsed since emigration" (de Bot and Clyne 1994)) was 7 years. This was to ensure that the most active phase of L2 acquisition was over, so that developing bilinguals had become stable ones (Ng and Wigglesworth 2007: 20). In other words, we aimed for our participants to have reached a "mature" or "steady state" of bilingualism (Hyltenstam and Obler 1989: 4), allowing us to assess L2 proficiency at its maximum. Further, it has been recognised that attrition usually progresses at a very slow pace (Jaspaert and

Kroon 1989, cf. below), so we needed to allow for sufficient time for that process to take place and for any interactions between the two language systems to play out.

Finally, although we did not specify a maximum age at time of testing, we were conscious of the fact that aging in itself may have an impact on language proficiencies and use (e.g. Goral 2004; de Bot and Makoni 2005) which would make very elderly participants hard to compare to younger ones. The oldest participants are 66 (one bilingual participant, one Irish control) which is just outside the range we would have ideally liked. However, since we did not employ psychometric tests, we ultimately decided to accept these participants and to exclude them if it turned out that their performance was markedly different from the rest of the group.

The basic characteristics of the bilingual informants are as follows: all had emigrated as young adults (mean age at emigration 26.8 years) and had lived in Ireland for 19.5 years on average. The majority had emigrated straight to Ireland, but four participants (15%) had spent some time in another English-speaking country (UK, US) before moving to Ireland. People who had lived in countries other than English- or German-speaking ones for periods over six months were excluded, unless their contact with the relevant host language during their residence was judged to be negligible. The mean age at time of testing was 46.3 years (range: 28-66 years).

The majority was female with a ratio of 3.5:1. The males were slightly older on emigration but the difference is not significant ( $t(25) = 1.67, p > .05$ ). Similarly, the differences between males and females in terms of length of residence and mean age at the time of testing do not reach significance ( $t(25) = .30, p > .05$  and  $t(25) = 1.76, p > .05$  respectively).

The bilingual participants are from a variety of backgrounds but are well-educated overall, with one third having secondary education plus vocational training, and the remainder having completed third-level education (mean years in education 16.3). The figures for the men diverge from the average in that five (83%) have learned a trade, while only one (17%) has third-level education. For the females, these figures are almost exactly converse (19% and 81%). The difference in years in education is significant according to the Mann-Whitney test ( $U = 22.00, p < .01$ ) and is taken into account as one of the predictor variables.

	<b>Bilinguals</b>	<b>Females</b>	<b>Males</b>
Number (N (%))	27 (100%)	21 (78%)	6 (22%)
Mean age at time of emigration	26.78	23.76	31.67
Mean length of residence (in years)	19.48	19.23	20.33
Mean age at time of testing	46.26	44.61	52.00
Educational level (mean years in education)	16.26	16.90	14.00
Self-rating of English language proficiency (N (%))			
sufficient	2 (8%)	2 (10%)	-
very good	8 (33%)	6 (30%)	2 (50%)
excellent	14 (58%)	12 (60%)	2 (50%)
Urban residents (N (%))	17 (63%)	15 (71%)	2 (33%)
German partner (N (%))	6 (30%)	2 (10%)	4 (67%)
Children (N (%))	18 (66%)	13 (62%)	5 (83%)

**Table 2: Participant background data – overview bilingual participants**

For the purpose of establishing native-speaker baselines for both German and English, two control groups were recruited, one of 20 German native speakers based in Germany and one of 18 English native speakers based in Ireland. These were matched with the bilingual group in age, gender and level of education as the primary selection criteria (cf. Table 3). In contrast with the bilingual group who had a minimum length of residence in Ireland of seven years while having spent less than six months in other L2 environments, the controls were required to have spent less than 6 months in any L2 environment.

	<b>Bilinguals</b>	<b>German controls</b>	<b>Irish controls</b>
Number (N)	27	20	18
Females (N (%))	21 (78%)	16 (80%)	14 (78%)
Males (N (%))	8 (22%)	4 (20%)	4 (22%)
Mean age at time of testing (years)	46.26	44.15	41.22
Educational level (mean years in education)	16.26	16.70	16.22

**Table 3: Participant background data – primary matching criteria**

The groups are well matched overall, but the Irish control group is slighter younger than the German control group which in turn is slightly younger than the group of bilingual participants, a difference which is not, however, significant (Welch  $F(2, 34) = .75, p > .05$ ). Neither gender split nor educational level yield a significant difference ( $F(2, 62) = .019, p > .05$  and  $F(2, 62) = .39, p > .05$  respectively).

The following sections discuss the secondary matching criteria.

### 3.1.2.2 Region of origin

An attempt was further made to match the bilingual and the German control group in respect of region of origin in order to control for possible language variety differences (in terms of lexical choice and structures) in the L1. The German language varieties and corresponding regions the bilingual participants were drawn from are shown in Table 4.

	Low German		West German	Central	Upper German		
	Lower Saxon	Low Franconian	Ripua-rian	Mosel-Franco-nian	Low Allemanic	High Allemanic	South Franconian; Bavarian-Austrian
	Hamburg Flensburg	Düsseldorf, Essen, Duisburg, Niederrhein	Rhineland: Willich, Cologne	Saarland	Heidelberg, Stuttgart	Baden, Zürich	Franconia; Bavaria, Austria
N	7	8	1	1	4	3	3
%	25.93%	29.63%	3.70%	3.70%	14.81%	11.11%	11.11%

**Table 4: Language varieties and regions of origin of bilingual participants**

In the event, with the exception of two speakers of southern varieties of German (2/27, 7.5%) whose region of origin is clearly identifiable in their speech, the majority of bilingual participants speak educated standard German with few distinct regional markers.

The German control group came mainly from the first three regions: Northern Germany (Hamburg, Osnabrück, Hanover: 9/20, 45%), the Ruhr area (Düsseldorf: 5/20, 25%) and the Rhineland (Cologne, Bonn: 3/20, 15%). Just one participant (5%) came from an Upper German region

(Stuttgart), and two more (10%) from the region of Thuringia, the regional variety of which is East Central German and has no match in the bilingual group.

Thus, only a partial match between the bilingual group and the German control group was achieved. However, given that few participants speak with a distinct local dialect, the importance of exact regional matches was judged to be not as great as originally thought. To compensate for the discrepancies, the help of native-speaker experts from the under-represented regions was enlisted in cases of doubt.

In selecting the English-language controls, we similarly attempted to match the control group to the L2 variety spoken by the bilingual participants. Most of the bilingual participants had learned English for five years or more as part of their schooling, and had since acquired an Irish accent to some degree, as confirmed by native-speaker ratings we will discuss in Chapter 5. A few can definitely be placed in their current area of residence by their accent (cf. Table 5). L1 German influence can be found in a number of participants' speech, but only three have a markedly foreign accent. There is a sub-set of participants who speak with a distinct British English accent, possibly due to prior residence in the UK and/or their schooling, RP being the preferred variety for teaching English in German schools.

	<b>Hiberno-English</b>		<b>British English</b>
	<b>Generic</b>	<b>West of Ireland</b> Kerry, Clare, Limerick	
N	15	7	5
%	55.5%	26%	18.5%

**Table 5: English-language variety spoken by bilingual participants**

The majority of the English-speaking controls come from around Dublin (11/18, 61%) and just one participant from the west (1/18, 6%), but most speak an educated variety of Hiberno-English. Two are British English speakers (2/18, 11%), while the remainder are from Northern Ireland, speaking a northern variety (4/18, 22%). Thus the regional varieties of the west of Ireland are under-represented amongst the controls, while the northern accent is over-represented. We again resorted to native-speaker experts where no direct regional comparisons could be made.

### 3.1.2.3 Social and family setup, L1 and L2 use

In the bilingual group, a small majority of participants (17/27, 63%) live in Dublin and environs, while the remainder live in rural areas in the West of Ireland. These figures differ for men and women: only one third of the male informants (2/6, 33%) are living in an urban centre, as compared to 15 of the 21 females (71%) (cf. Table 6).

In contrast, all English-speaking controls (18/18, 100%) and all except for two of the German controls (18/20, 90%) live in urban areas. The inability to match controls to the bilingual group better on this criterion is connected to the specifics of recruiting the controls and constraints in relation to where testing could take place.

	Bilinguals			German controls	Irish controls
	Females	Males	Total		
Urban residents (N (%))	15 (71%)	2 (33%)	17 (63%)	18 (90%)	18 (100%)
L1 partner (N (%))	2 (10%)	4 (67%)	6 (30%)	19 (95%)	14 (78%)
Children (N (%))	13 (62%)	5 (83%)	18 (66%)	12 (60%)	4 (22%)

**Table 6: Participant background data – additional criteria (1)**

The bilingual and control groups also differ drastically in their choice of life partners. The vast majority of all controls have, or have had, a partner speaking the same L1 as themselves (19/20, 95%, and 14/18, 78% respectively, cf. Table 6), while this case is much rarer amongst the bilingual participants (6/27, 30% overall). The females in particular (14/21, 66%) are predominantly married to/living with an Irish partner. Just two females (10%) have a German partner, while the remaining five (24%), now singles, had Irish partners in the past. The males, on the other hand, are more like the German control group in this regard, since two-thirds of them (4/6, 67%), have (had) a German partner (one has an Irish partner and another, a Spanish girlfriend).

Two-thirds of all bilingual participants (18/27, 66%) have children. This figure is higher for the men (5/6, 83%). The percentage is very similar for the German control group (12/20, 60%), while, somewhat surprisingly, far fewer Irish controls have children (4/18, 22%). This might be due to differences in the Irish control group's age profile, since it contains a greater number of younger

people who have just left college or are still in education; in its higher percentage of single people; and possibly in terms of profession (a higher percentage of teachers/academics).

The choice of life partner and presence of children (particularly those still living in the family) often dictate language choice and use. For the bilingual participants, the relevance of eliciting information on these factors lies in their potential for predicting first-language maintenance in an L2 environment, when L1 use at home and particularly with children creates a linguistic bubble in an otherwise L2-dominated context.

In the control groups, L1 is the language of daily communication for 100% of controls across all areas, even in those cases where the partner has a different L1. However, over half the controls use their L2 either in their private and/or professional lives (11/20, 55% German controls; 11/18, 61% Irish controls). Professional use is higher in the German control group (13.6% private vs. 86.4% professional use), while private L2 use (L2 partner, phone contact with friends abroad) is more dominant in the Irish control group (72.8% private vs. 18.2% professional use).

For the bilingual participants, the L2 has become the main language of communication and is used across all domains. However, they continue to use their L1 to varying degrees, both in Ireland for communication within the family and with friends, and for staying in contact with their country of origin, by phone, letter and travel. None of the bilingual informants has lived or stayed in Germany for durations of more than a few weeks since emigration.

L1 use is generally higher in relationships where both partners speak the L1, and particularly in those with children. Just three participants rarely use their L1 with family or friends, but two of those use it professionally. There is a large sub-group (14/27, 51.8%) of the group overall that uses German professionally on a daily or weekly basis (this figure is similar for both men and women: 50% and 52.4% respectively).

#### **3.1.2.4 L2 proficiency**

Although all bilingual participants had learnt English in school (mostly for five years or more, with some exceptions), their first significant exposure to the language (immersion) usually began with their migration, sometimes preceded by a time of more intensive English-language use through holidays, e-mail, written and phone contact with a future partner.



In the preliminary recruitment questionnaire (Personal background questionnaire), all participants were asked to rate their current L2 proficiency on a scale of 1-5 (“very low” to “excellent”; the minimum level was 3 “sufficient”). In the case of the bilingual participants, this allowed us to exclude anybody with insufficient knowledge of English to test our hypothesis. (For logistic and time reasons, not all participants were able to fill out the questionnaire prior to the session, in which case the proficiency information was elicited with the sociolinguistic questionnaire.)

The bilingual participants rated themselves either as 3 (“sufficient”, 2 ex 24; 8.3%), 4 (“very good”; 8 ex 24; 33.3%) or 5 (“excellent”; 14 ex 24; 58.3%) (cf. Table 7), so we are dealing with proficient or highly proficient bilinguals. There is a small gender bias in that none of the men rated themselves as a 3 (“sufficient”), but a slightly higher percentage of the females awarded themselves the top mark 5 (“excellent”).

	Bilinguals			German controls	Irish controls
	Females	Males	Total		
Self-rating of L2 proficiency (N (%))	N = 20	N = 4	N = 24	N = 20	N = 17
low	-	-	-	1 (5%)	1 (6%)
sufficient	2 (10%)	-	2 (8%)	9 (45%)	4 (24%)
very good	6 (30%)	2 (50%)	8 (33%)	8 (40%)	7 (41%)
excellent	12 (60%)	2 (50%)	14 (58%)	2 (10%)	5 (29%)

**Table 7: Participant background data – additional criteria (2)**

The controls were also required to have at least a good level of proficiency (3 or more) in their respective L2 (English for the German controls, German for the Irish ones), for the reasons outlined above. However, not entirely surprisingly, it proved rather difficult to find controls who rated their proficiency in their L2 (German or English) sufficiently high, while meeting the residence abroad criterion of less than six months. Further, in enrolling the Irish controls, we were partially relying on recent graduates of the Germanic Languages department at Trinity College Dublin, who routinely spend 7-9 months in Germany on their year abroad. Consequently, it was decided to relax the time limit for residence abroad to one year for the controls to increase the pool of eligible candidates.

Overall, the controls in both groups rated their L2 proficiency lower than the bilingual participants (cf. Table 7). Fewer controls (10% and 29% German and Irish controls respectively, vs. 58% of bilinguals) rated their L2 proficiency as “excellent”, and a much higher proportion thought their L2 competence was just “sufficient” (45% and 24% vs. 8%). Two participants, one in each control group, rated their level of L2 proficiency as “low” (2), i.e. less than the required minimum, but for a number of reasons it was decided to proceed.

Thus, the three groups compare less well on this criterion. The differences in the self-ratings of the controls vs. the bilingual group are quite large; in the case of the bilingual vs. the German control group they are highly significant ( $F(2, 58) = 8.59, p < .001$ ). Possibly, the level of L2 proficiency in the controls reflects the typical level achievable without extensive immersion in the L2 environment. However, on the whole the controls satisfy the minimum requirement for degree of bilingualism, thereby addressing the “cognitive” and the “unfair comparison” arguments against comparing mono- and bilinguals.

### **3.2 Design of the study**

The study presented here is a cross-sectional investigation into the linguistic development of adult bilinguals in a foreign-language setting. As such, it is a type of field study - a "natural" or quasi-experiment which is distinguished from a "true" experiment in that we do not control the independent, or predictor variables:

"A *true experiment* attempts to isolate cause and effect and occurs when an independent variable is manipulated and all other salient variables are controlled, including the random allocation of participants to conditions." (Coolican 1999: 50).

On the other hand, in quasi- or natural experiments, potentially relevant variables are selected amongst a host of naturally occurring ones (ibid.: 83). Because of the nature of these variables which will be discussed in greater detail below, random allocation of participants to groups was not an option, and indeed there are no "conditions" of these variables to which participants were subjected. Furthermore, the independent variables are unlike the "treatments" that are administered (or not administered) in experimental studies.

### 3.2.1 Longitudinal vs. cross-sectional design – design considerations

In this section, we consider important issues and choices in designing the present study as well as its characteristics. First, given that concepts such as "development", "attrition" and "acquisition" imply processes rather than states, a cross-sectional design raises the question of its capacity to capture change in a person's language proficiency. Intuitively, a longitudinal design would seem better suited to investigate this question. We therefore begin by discussing how the inherent problem of studying change in cross-sectional designs is addressed.

The following table sets out the points discussed in what follows:

<b>Pros (+) and cons (-)</b>	<b>Longitudinal design</b>	<b>Cross-sectional design</b>
Suitable to investigate language change	+	(-)
Time	-	(-)
Test/re-test effects	-	n/a
Participant loss	-	n/a
Arbitrariness of sampling points	-	(-)
Point of reference	(-)	-
Language change	(-)	-

**Table 8: Pros and cons in longitudinal vs. cross-sectional designs**

The principal distinction between the two types of design boils down to whether sampling is conducted on a repeat basis with the same participant(s) (longitudinal testing), or one-off with (groups of) individuals (cross-sectional testing). In the former, participants' "performance at Time 1 is usually compared to performance of the same individual or group of individual at Time 2", in the latter, "different individuals or groups of individuals are compared to each other at the same point in time" (Marian 2008: 16). In longitudinal designs, between-group differences are minimized, whereas in cross-sectional designs it is harder "to control for extraneous factors" (ibid.: 16).

By definition, therefore, longitudinal designs are better able to capture on-going change than one-shot cross-sectional designs. A cross-sectional design, on the other hand, allows judgements to be made about the outcome of change processes. This distinction is not, however, absolute. Some types of longitudinal study could be seen as a series of cross-sectional studies, whereas cross-

sectional investigations may find tools retrospectively to shed light on the processes that preceded and led to the current state.

The type of investigation opted for also has an impact on resource requirements (time and number of participants required), partially dictating the type of data elicited (qualitative and/or quantitative) and the breadth and depth, or "descriptive detail" (Gass and Selinker 2008: 55) of analysis. Thus, in weighing the pros and cons of one design over another, both methodological and resource issues need to be considered (cf. Jaspaert, Kroon et al. 1986 and; Jaspaert and Kroon 1989 for an overview of issues relevant to attrition research).

Jaspaert and Kroon discuss several potential problems with regard to both types of design. In relation to longitudinal designs, a crucial factor to be considered is time. Jaspert, Kroon et al. (ibid.) argue that adult L1 attrition is a very slow process as compared to L2 acquisition, L2 attrition and child L1 attrition, since a mature mother tongue is a very well-established, very well-practised system. They estimate that the gap between pre- and post-test in needs to be "at least several years" (ibid.: 38) for major changes to show. However, de Bot and Clyne (1994), in one of the few longitudinal studies of adult migrants, found that after a lapse time of 16 years the mother tongue of those participants who had still been fairly fluent at the time of first testing remained largely unchanged.

Another reason for deploying sizeable intervals is the need to minimize test-retest effects. In other words, there is a danger that as a result of (first or repeated) testing, participants may realise that their native language proficiency has diminished, and consequently increase their maintenance effort. The "natural" attrition process would thus be "disturbed" (Jaspaert and Kroon 1989: 81), and the data collected subsequently would be different than if the earlier testing had not taken place. However, the long time intervals required for the reasons just mentioned can cause real practical problems for researchers (Marian 2008: 16). Similarly, Jaspaert and Kroon conclude that "a longitudinal design is not feasible in research on L1 loss" (1986: 39).

Jaspert and Kroon (1986) mention two further problems with longitudinal studies. First, they are associated with a risk of participants not being available for re-testing through various circumstances, particularly where adult participants are involved ("attrition" of participants, Marian

2008: 16). The "non-random" (de Bot and Clyne 1994: 21) loss of participants presents a problem for certain analyses.

Second, it is increasingly becoming clear that neither language acquisition nor language attrition proceed linearly over time (Gass, Madden et al. 1989; Jessner 2003; Victori and Tragant 2003; Larsen-Freeman 2006; 2007). Consequently, it is unclear at what stage important changes happen, and, given that the points at which testing will take place have to be chosen arbitrarily, these may be missed. The arbitrariness of selecting the point at which to test applies to cross-sectional designs as well; however, since these do not make any claims about ongoing change, but rather attempt to establish the state or level of a criterion at a given point in time, this is less problematic.

While cross-sectional designs avoid the problem of the time factor, they raise another very important issue – that of establishing the point of reference against which the data is measured (Jaspaert, Kroon et al. 1986: 38ff.; Jaspaert and Kroon 1989: 81ff.). Longitudinal designs have the advantage of providing their own reference point with the first set of data collected from the participants, though external comparison data may additionally be employed. In cross-sectional designs, on the other hand, an external reference point inevitably has to be established to confirm what changes have taken place in the language of the informants.

Several options exist: comparing the primary data with actual controls, against "the native-speaker standard" (Cook 2003)/the "linguistically competent speaker" (Andersen 1982), or with other sources, such as census data, which serve as a hypothetical control group baseline (Jaspaert, Kroon et al. 1986: 43).

Using the "linguistically competent speaker" of a language as a reference point is a popular approach but one that is problematic given that this speaker is a hypothetical, idealized construct (Jaspaert, Kroon et al. *ibid.*: 43; Davies 2004: 431, 447; Ng and Wigglesworth 2007: 24). Even where empirically established norms exist (such as developmental stages for children), these are usually normed on monolingual, socio-economically different populations making the tests inappropriate for bilinguals and placing them at a serious disadvantage (Ng and Wigglesworth *ibid.*: 38). For our constituency, the "native-speaker standard" may well lead to an over-estimation of "L1 deficit" in the bilingual participants, and an under-estimation of their L2 proficiency.

Thus, it is preferable to employ an actual, matched control group of native-speakers whose data will serve as a baseline. Actual performance data, unlike a normative standard, will show the variation in performance present in "normal" controls and is a more reliable gauge for what constitutes "normal" vs. "changed" behaviour. The implicit assumption in our study is that the L1 proficiency of bilingual participants and German controls would be comparable if the bilingual participants had not emigrated. Any difference could then potentially be ascribed to language attrition. This assumption can only be met if participants are matched closely on educational, socio-economic and other criteria, such as region of origin, impacting on the language variety spoken by a particular individual.

A further problem with regard to the point of reference arises for both types of design. With the passing of time, the language used in the original speech community and that used by people living outside that community are bound to diverge due to language change. While émigrés are often cut off from language developments in their country of origin, their own language develops in line with the communicative requirements in their new speech community.

As a result, the features of their language become harder to interpret. Does difficulty in discussing a particular topic indicate forgetting, or may it instead point to non-acquisition of the requisite linguistic means? Could the use of a loan word constitute adaptation to a new code spoken in the L2 environment? This problem calls for careful interpretation of any effects, and potentially implies the use of complementary points of reference (Jaspaert, Kroon et al. 1986: 38).

In summary, both designs have their pros and cons, in some cases the same. Longitudinal designs do allow a detailed investigation of change processes, but they present problems for sampling, data validity and time-frames. For the purpose of this investigation, the latter was a major factor in deciding against a longitudinal design.

Cross-sectional designs on the other hand by definition investigate states, or the outcomes of processes, rather than the processes themselves, and great care is necessary in choosing the reference point against which the primary data will be compared. This was achieved by careful matching of the controls with the bilingual participants across a range of criteria (cf. Section "Participants").

So how then is it possible to reconcile the task of studying the processes of L1 attrition and L2 acquisition with the current design? In addition to various tasks that measure the current level of proficiency in both languages, we asked participants to report retrospectively about the development of their languages. While this method yields less rich data than a longitudinal investigation would, and while all self-report data are subject to some limitations regarding their reliability (Clark 1982), it seems to us currently our best option to get an insight into both processes.

### ***3.2.2 Between-group and within-group design – design decisions***

We already mentioned above that the current study is a quasi-experiment because of certain features that are missing in comparison with true experiments (control over independent variable, random allocation of participants to groups or conditions, characteristics of independent variables). In actual fact, this investigation combines two studies: a between-group study, and a within-subject, in order to answer the research questions set out in the Introduction.

The between-group, or inter-group study, is similar to an independent samples design in psycholinguistic experiments, in that we have three groups that are being studied – the bilingual group and two control groups, one for each of the languages tested. In such designs, the equivalence of groups is of utmost importance, requiring careful sampling. As discussed in Section 3.1.1 above, the groups are closely matched on a number of personal characteristics but are distinguished by their place of residence (the independent variable): The bilingual group resides in an L2 environment, whereas both control groups reside in their L1 environment.

Residence in one or the other environment thus resembles an experimental "condition", though it is difficult to conceive of it as a "treatment". For the purpose of the between-group study, the controls provide the baseline data for their respective L1. The dependent variables are thus the scores on the German language tests for the comparison between the bilingual and the German control groups, and those on the English tests for the comparison between bilingual and Irish control group.

The within-subject, or within-group/intra-group, study is similar to a repeated measures design in psycholinguistic experimentation, insofar as the bilingual group tackled a test battery of equivalent tasks in two languages ("conditions") – the L1 and the L2. While the measures are not absolutely

identical, they are designed to be equivalent and could thus be seen to be administered repeatedly (twice). Since all the bilingual participants are residing in their L2 environment, the independent variables that potentially play a role in our bilinguals' L2 acquisition and L1 maintenance (after controlling for education and sociolinguistic background) are gender, age of arrival, length of residence, amount/frequency of language use, linguistic aptitude, demographic and attitudinal variables, while the dependent variables again are the scores on the German and the English language tests, now comparing the two scores achieved by the same person.

A typical problem with within-subject designs are possible order effects discussed in more detail below, which call for the counterbalancing of the order in which the two "conditions" – L1 and L2 – are administered. Implementing counterbalancing creates an additional design layer, or an extra independent variable that needs to be taken into account in the analysis.

With these considerations in mind, the materials used in this study were designed to elicit a variety of data:

- a) demographic data, including language learning history;
- b) language use data;
- c) proficiency data in the two main languages spoken by the informants (German and English);
- d) language attitude data;
- e) language aptitude data.

The demographic data (a) were needed for descriptive and matching purposes, while L1 and L2 proficiency data (c) which constitute the dependent variables, allow us to assess the degree of L2 acquisition and L1 maintenance (between-group study), and to establish the proficiency profile for individual participants (within-subject study). Finally, the language use, attitude and aptitude data (b, d and e), along with some demographic factors, constitute the potential predictor variables for the levels of L1 and L2 proficiency (within-subject study).

Materials combine self-report with more or less formal tasks. The "language tasks" were collected with a test battery containing parallel sets of written and oral tasks in L1 and L2 (the principal focus being on spoken data), ranging from a broad, global assessment of language proficiency through lexical and syntactic tests to measures of lexical access and fluency. Tasks also vary along the dimensions of response constraints (from relatively free spoken data in a prompted story telling



task to very rigidly controlled data in a C-test), of processing demands (from untimed, open-ended tasks, such as the retelling tasks, to tasks involving strict time constraints, such as a verbal fluency task), and consequently, of task complexity.

Below is an overview of the instruments used.

---

Questionnaire instruments:

Personal background questionnaire <sup>2)</sup>

Sociolinguistic questionnaire <sup>2)</sup>

Part 1: Language use questionnaire

Part 2: Language proficiency/attitudes questionnaire

Can-do scales <sup>2)</sup>

Attitudes towards foreign language learning/bilingualism <sup>3)</sup>

---

Linguistic aptitude test (LAT) (English version) <sup>1)</sup>

---

Language tests:

C-test <sup>2)</sup>

Verbal fluency tasks (FiCA, Fluency in Controlled Associations) <sup>1)</sup>

Film retelling <sup>1)</sup>

Picture description <sup>3)</sup>

Sentence generation task <sup>3)</sup>

---

**Table 9: Test Battery**

<sup>1)</sup> published material

<sup>2)</sup> Attrition network

<sup>3)</sup> author

Some of the tests and instruments are in the public domain and were employed with minor modifications: the Lognos linguistic aptitude test battery (LAT test) (1994); (Meara, Milton et al. 2001, English version, parts B, C and D), the verbal fluency tasks (FiCA, Fluency in Controlled Associations) based on Fishman and Cooper (1969) and Benton, Eslinger et al. (1985); and one of the free-spoken data tasks, the Charlie Chaplin film retelling which was borrowed from the SFI procedure (Broeder, Extra et al. 1988; Perdue 1993; Perdue 1993).

Other tests were constructed within the Attrition network between 2002 and 2004 (principally at the time: Monika S. Schmid, Merel Keijzer (cf. Keijzer 2007), Susan Dostert (cf. Dostert 2009), and this researcher) with the aim of harmonizing methodologies across several studies investigating various aspects of L1 attrition (Schmid 2004; Schmid 2004), and were adapted to this particular project: the personal background and sociolinguistic questionnaires; the can-do scales, themselves based on the can-do statements developed within the Council of Europe; the C-test and the picture description task.

Finally, some instruments were developed by the researcher specifically for use in this study: the attitudes questionnaire (Conceptions of L2 learning/bilingualism); the sentence generation task which builds on a grammaticality judgement task devised within the Attrition network but also employs stimuli used by Köpke (2001) in a sentence-generation task.

For all instruments, parallel English- and German-language versions were developed, except for the sentence generation task which only existed in German, and the linguistic aptitude test LAT. The questionnaires were administered in L1, while the language of administration for the language tasks mirrored the language version being used. The LAT deploys an artificial language as test material, and an intermediary language (either English or French) to present the test material. Here, the English-language version was used with all participants.

Only the bilingual participants solved all language tasks in both languages. The control groups completed the set of language tasks in their L1 only, with the addition of the C-test in their L2 and the aptitude test (cf. Section 3.4.2 below).

In this thesis, we do not report the results from all the instruments used. The tasks not included in the analysis at this stage are the picture description task and the sentence generation task. We took this decision mainly since in doing the analyses it transpired that we had collected a large amount of data, which to combine into a single profile would have been quite complex. Reducing the number of tasks to be amalgamated reduced the complexity somewhat.

Further, the sentence-generation task was only conducted in German. Designed to tap syntactic processing, it would have given us more information about syntactic attrition without complementary acquisition data for the L2. The picture description task, on the other hand,

partially replicates the film retelling in eliciting spoken production data, so is not expected to change the overall picture drastically.

While it would have been of interest to include the results from the very fine-grained lexical analysis of the picture description task, and the syntactic analysis of the sentence generation task, these constitute a further analytic level beyond the quantitative and qualitative measures elicited with the other tasks. So for the purpose of this thesis, we decided to concentrate on a more thorough analysis of the questionnaires, C-test, FiCAs and film retellings at the expense of adding another layer of complexity to the analysis.

The following sections provide more detail about the development of individual instruments as well as the implementation/administration of the test battery and individual tests (cf. Appendix for details of the construction of the picture description and sentence generation tasks).

### **3.3 Test and elicitation materials**

#### **3.3.1 Questionnaires**

The study employed several questionnaires and self-report instruments (cf. Table 9 above): a personal background questionnaire; a sociolinguistic questionnaire comprising two distinct parts; a questionnaire eliciting attitudes towards foreign language learning/bilingualism, and can-do scales for both languages under investigation (cf. Appendix for the actual questionnaires).

- The personal background questionnaire ("Participant questionnaire"/"Teilnehmer-Fragebogen") comprised 32 questions relating to demographic information, language background and use. This questionnaire was completed by the participants prior to the actual study and returned by either email or post, allowing the researcher to screen participants for eligibility.
- The sociolinguistic questionnaire contained a total of 65 questions in two sections: part 1, the language use questionnaire (36 questions), and part 2, the language proficiency and attitudes questionnaire (29 questions). This questionnaire was completed by the researcher, who used it as a script for an in-depth semi-structured interview lasting up to 1.30 hours.

- The can-do statements (Kann-Aussagen) were optionally completed by participants after the experimental session in their own time. A total of 43 statements in 4 sections relating to participants' perceived proficiency in German and English were to be rated on a 5-point Likert scale.
- Finally, a separate attitudes questionnaire ("Ideas about foreign language learning and bilingualism"/"Meinungen zum Fremdsprachenlernen und Bilingualismus") was distributed in the same document and was administered in the same way. It contained 16 statements, which were also to be rated on a 5-point Likert scale.

Parallel language versions were developed for all questionnaires, except for the sociolinguistic questionnaire, which was not administered to the control groups. The questionnaires were administered in the participants' respective mother tongue.

The questionnaires were designed to serve five main purposes:

- a) to obtain demographic data with a view to describing the participants and to ensuring inter-participant comparability, such as social background and education level;
- b) to elicit concrete, quantifiable information about participants' language learning history and use, such as length of residence in the host country, amount, modes and settings of contact with either language since emigration, age of first starting to learn the L2;
- c) to elicit self-report data on participants' language proficiency;
- d) to elicit information about participants' attitudes towards language learning in general and towards integration with the host community in particular; and
- e) to provide a recruitment/screening instrument.

Furthermore, the personal background questionnaire provided a starting point for the follow-up interview based on the sociolinguistic questionnaire, providing a bridging function. Similarly, two items (Question 34 and 36) in the sociolinguistic questionnaire, asking about whether the participants had enjoyed learning languages and which areas they found difficult, were included to help introduce the linguistic aptitude test in a more empathetic way. In summary, the questionnaires served descriptive, exploratory and explanatory purposes (Brown 2001: 19).

Before constructing and implementing a tool like a questionnaire, alternative ways of obtaining the same information need to be considered. For bio-/demographic data, these alternative means are

principally interviews, which according to Brown (2001) have the advantage of being relatively personal and flexible, providing richer data and achieving good “return rates”. On the other hand, they are time-consuming, geographically restricted, less easily standardized and have a greater potential for subconscious researcher bias as well as for participants’ self-censorship because of the impossibility of complete anonymity. Alternative means for eliciting language use data, such as diaries and third-party reports (Clark 1982), were not really viable in the context of this study.

Considering the advantages as well as disadvantages of each of these methods, and the importance of validating information through various means, the question was not so much which one of these means to use, but of how to maximise the benefits from each method. Thus, it was decided to adopt a comprehensive approach, i.e. to combine questionnaire and semi-structured interview (“spoken questionnaire technique”, Wray, Trott et al. 1998: 183).

Semi-structured interviews are preferable to other types since they allow the researcher to both control process and contents, as well as providing the flexibility to explore unexpected issues further (Nunan 1992: 149ff.; Coolican 1999: 140ff.). Of the possible range of interview styles (e.g., Gubrium and Holstein 2002), we adopted the personal style both by personal preference and the methodological advantages, such as richer answers.

In constructing the questionnaires, we drew on several more sources to inform that process. Brown (Brown 2001) and Baker (1996; 2001) provide sets of questions in relation to language use, language attitudes as well as background scales. In addition, we were able to consult several questionnaires from the area of applied linguistics (e.g. the CLCS Modern Languages Research Project Questionnaire Singleton and Little 1991; Perdue 1993; El Aissati and Schaufeli 1999; Dewaele and Pavlenko 2001-2003). Finally, we were able to conduct an in-depth conversation with a person whose profile fitted the pre-determined participant characteristics as described above very closely, following Brown's (2001) suggestion that interviews are very well suited to exploring what the relevant issues are.

Several decisions were prompted or informed by this process. The first related to the manner and timing of eliciting the relevant information (by participant vs. by researcher, written vs. oral presentation, paper-and-pencil version vs. computer version; pre-, during or post-experiment). As a result, questions relating to the demographic information required to screen participants were

included in the personal background questionnaire, to be completed by participants themselves in advance of the study. All other questions were to be contained in the sociolinguistic questionnaire, except for the can-do statements which do not require the interviewer's presence, and some additional attitude questions which were deemed to be too time-consuming and could provide a reliability check on those included in the questionnaire (cf. below).

The decision to administer the sociolinguistic questionnaire as a semi-structured interview prompted a split of the questionnaire into two parts, one to be administered at the beginning of the experimental session, the other at the end. The first part deals with amount, types and modes of L1 contact; frequency of language use and language choice across various domains; primary interlocutors and social networks; as well as with participants' views of themselves as language learners/bilinguals. In other words, it elicits primarily factual, quantifiable information.

The second part was to be administered at the end of the testing session and focuses on the current level of proficiency and changes therein in relation to all languages known by the participants; on participants' attitudes to language learning and language maintenance; and the experience of acquiring a language in an L2 environment. This part elicits mainly subjective, open-ended data. The split provided a logical, coherent frame to the testing session.

Originally developed independently, the personal background and sociolinguistic questionnaires were aligned with the questionnaires contained in the "Language Attrition Test Battery" (Schmid 2004: 8, 15ff, 41ff), henceforth LATB, in order to ensure that they contain all the questions relevant to the broader enquiry of the Attrition network. The main difference between the questionnaires contained in the LATB and the current sociolinguistic questionnaire lies in the number and the ordering of the questions. Given that the Attrition network version was extensively pre-tested, we felt that a small-scale pilot would be sufficient in the present case.

The questionnaires were continually refined with particular attention to

- the relevance of all questions to the research purpose;
- the clarity of instructions/questions;
- economy in design, so that the questionnaires would be at once user-friendly and would elicit a maximum of relevant information without being too taxing;
- coding and scoring requirements (Kipp, p.c.) and

- the operationalisation of the variables (e.g. re-framing the concept of “initial exposure to the foreign language” as “first significant exposure”, (White and Genesee 1996: 241ff)).

We also adopted the can-do scales contained in the LATB (Schmid 2004: 15, 22, 58ff) for both languages, in addition to the global self-assessment of L1 and L2 proficiency in the sociolinguistic questionnaire. The can-do scales are drawn from the Council of Europe’s descriptors of language functions (ALTE 1990) and allow participants to assess their own language proficiency across the four areas of listening, reading and oral comprehension as well as spoken production.

The attitudes questionnaire was developed as an additional, more easily quantifiable tool than the relevant attitudes questions on the sociolinguistic questionnaire. We felt that emotive responses may result in more response bias than the reporting of hard facts, particularly in an interview situation. Asking participants to complete some of those questions on their own afforded them greater anonymity (Takona 2002: 99f) and allowed us to cross-check the responses given during the interview. On the downside, since this was an optional questionnaire, some participants either did not return it, or returned an anonymous one, which meant that those answers could not be factored into that particular participant's profile.

The statements to be rated in the attitude questionnaire were derived from views encountered by the researcher in the media, the literature and personal conversations. They focus on questions relating to values associated with the two languages under investigation (e.g. Davies 1986: 123: functional, symbolic, sentimental), to different types of motivation for learning foreign languages (and by extension maintaining the mother tongue in an L2 environment), as well as to attitudes, hopes/beliefs and strategies towards bilingualism, language shift and maintenance (personal and with regard to the next generation). The attitude questionnaire was trialled in a small-scale pilot along with the sociolinguistic questionnaire.

### **3.3.2 Language tests**

#### **3.3.2.1 C-test**

C-tests are a type of written integrative task tapping both receptive (reading) and productive (writing) language skills. Hence they are usually employed to assess "general" or "global" language proficiency (for a qualification, cf. Grotjahn and Stemmer 2002; Raatz and Klein-Braley 2002: 75).

The 'classical C-test' was introduced in 1981 as an attempt to overcome shortcomings of the cloze test common in L2 testing (Raatz and Klein-Braley 1981; Raatz and Klein-Braley 2002).

Both cloze test and C-test are gapped texts which exploit the principle of reduced redundancy as a measure of the "efficiency of a global language processing competence" (Grotjahn 2000: 159). This principle was first explored in Spolky's Noise test, testing native speakers' and language learners' ability to comprehend aural messages under increasing levels of white noise (Spolsky, Sigurd et al. 1968). Thus, they emulate natural language processing in situations requiring a language user to restore distortions in incoming text streams caused by, for example, background noise or bad radio signals, by relying on the natural redundancy of text. The repairs happen routinely and often without the user realising this. This skill appears to grow with language proficiency (Klein-Braley 1997; Raatz and Klein-Braley 2002; several contributions in Grotjahn 2006).

Cloze tests contain gaps created by the removal of whole words. The non-objective removal procedure is usually guided by didactic considerations and tends to prefer some word types (typically content words) over others. It also tends to imply the use of just one, typically quite long text, capable of testing language knowledge across a limited range of topics and themes (Klein-Braley 1985).

C-tests, on the other hand, make use of an objective distortion procedure, the so-called "rule of two" (Raatz and Klein-Braley 2002: 75), similar to recurring white noise distortions. For the canonical C-test, this rule prescribes the removal of the second half of every second word (except numerals and proper names) starting with the second sentence (if a word to be distorted has an uneven number of letters, half the letters plus 1 are removed). This results in a high number of gaps covering all parts of speech in relatively short stretches of text, compared to the cloze test.

This method of reducing redundancy has been criticised (e.g. Alderson 2002) on the grounds that, rather than using a distortion procedure that's sympathetic to the structure of the particular language, and the meaning of the text, C-tests "fire shotguns at text, and hope by some pseudo-random sampling to delete items that will require language knowledge to restore". Nevertheless, C-tests do correlate well with other measures of language proficiency (cf. numerous publications by Klein-Braley, Raatz, Grotjahn).



To further objectify the C-test, four to five short (ideally complete) authentic texts covering different topics with 20 or 25 gaps each are used. The texts are arranged in ascending order of difficulty. Within each text the first sentence is left intact, and there is normally a run-out after the last gap to the end of that or the next sentence.

This principled distortion procedure was first developed and tested for German and English. Subsequent research was devoted to assessing the suitability of this procedure for different languages (French, Spanish, Russian, Italian and Turkish, amongst others) and different populations of test-takers. Modifications to the distortion procedure were found necessary to cope with particular properties of other languages, such as compounds in Italian and German and contracted forms in French and Italian, agglutination in Turkish and pictorial writing systems (Grotjahn, Klein-Braley et al. 2002: 96f). In other studies, variations of the deletion procedure were tested with a view to adjusting the level of difficulty of a particular C-test for specific learner populations. For example, Daller, Treffers-Daller et al. (2002) tested the impact of deleting every third half of words vs. every third syllable/morpheme vs. the middle of every second word in a test designed for Turkish returnees.

The C-test has been found to be a highly reliable test. Extensive testing against a whole range of other, commonly used instruments revealed that it consistently loads on general language proficiency as the most significant explanatory factor (Klein-Braley 1997). Results typically also correlate highly with verbal proficiency (verbal intelligence), making it a particularly useful test for adults, since, according to Raatz (2002) maximum proficiency in this area is only achieved after age 16. The test also draws on logical abilities, but elicits primarily micro-level processing because of the shortness of the texts (ibid: 173) while at higher proficiency levels top-down processing plays an increasing role (Grotjahn and Stemmer 2002; Klein-Braley 2002).

Despite its provenance in foreign language testing, the C-test is equally suitable as an L1 test, and particularly as a test for studying adult L1 attrition, since it has been found to differentiate well at higher levels of proficiency in L2 learners. To avoid ceiling effects (Grotjahn, Klein-Braley et al. 2002: 97), modifications to the distortion procedure, such as speeding the texts, deliberately selecting very difficult texts (Raatz 2002: 79), adjusting the length of the texts, the number of gaps to be filled, and formal properties of the texts, need to be considered.

The German and English C-tests used in this study were specially developed by the Attrition Network as part of the Language Attrition Test Battery (Schmid 2004: 11f, 15, 21f, 50ff), in line with the recommendations by the "parents" of the C-test (Grotjahn, Klein-Braley et al. 2002; Raatz and Klein-Braley 2002, part. p. 84). They each consist of 5 short texts which differ in topic and genre as well as in the degree of formality of the language used (Biber 1988). Since one of the purposes of the Attrition Network was to gather L1 attrition data over a wide spectrum of language pairs using the same instruments, all language versions were designed as native-speaker tests but were used for both L1 and L2 populations.

In two parallel pilot studies conducted in December 2003 (by this author and M. Schmid), a large number of 'general' and 'formal' German-language texts of varying lengths were administered to volunteer native speakers of German. The pilot had a dual purpose: to test the impact of some variations on the basic design principle in order to create a test that would be suitable for native speakers of German, i.e. one that is sufficiently difficult, and to identify five texts to be included in the study proper. The standard distortion procedure was applied but the number of gaps was higher than in a standard C-test (54 on average for the formal texts, 65 for the 'general' texts).

The results from the two pilots were compared and discussed at a one-day conference of the Attrition network in Amsterdam in December 2003. As a result, it was agreed to provide the first letter of the second part of German compounds. Further, it was agreed to allow 5 minutes for the completion of each text in the test, which seemed long enough not to create undue pressure for fully proficient speakers. Finally, a lower cut-off point for inclusion of a text in the test of 87% accuracy (as achieved by the native speakers providing the baseline data) as opposed to the usual one of 90% was set, as an indication of a higher level of difficulty of the text. The rationale was to avoid ceiling effects and to increase differentiation between the German native speakers in the bilingual and control groups which, though representing similar constituencies, we would expect to differ in some ways.

The final German version of the C-test used in this study is a selection of 5 of the texts from the pilot phase. The English and other language versions (cf. Appendix for the German and English C-test) were constructed in an analogous manner. The final versions were piloted on another set of native speakers under the same conditions as in the actual testing session, but no further formal

statistical evaluation of reliability, discrimination or dimensionality were conducted since "in numerous studies C-Tests have proved to have extraordinarily good statistical qualities even though they were used in the study for the first time" (Grotjahn, Klein-Braley et al. 2002: 98).

### **3.3.2.2 Verbal fluency tasks**

Verbal fluency tasks (also "word naming tasks" (Fishman and Cooper 1969: 277; Baker 2001: 30f), or tests of "fluency in controlled associations"/FiCA (Köpke and Schmid 2004: 25; Schmid 2004: 14, 21f)) are a test of lexical retrieval. "Verbal fluency can be defined as the ability to retrieve members belonging to a specified category within a limited time period." (Rosen, 1980, cited in Lupien, Roch Lecours et al. 1994: 2896). They thus tap lexical access as a function of the relative activation levels of languages under timed conditions.

Two types – semantic and formal verbal fluency tasks – and several sub-types exist. In each case, participants are instructed to say as many words as they can think of within a given time frame (usually 1 or 2 minutes) in response to a particular stimulus.

In semantic fluency tasks, the verbal stimulus is a semantic category label, such as *animals, fruit, items of clothing* etc., and the required responses are members of the category (category naming task). This task is sometimes altered by instructing participants to associate freely (word association task, e.g. Lambert 1955).

In formal fluency tasks, participants are given a letter and are asked to verbalise either words starting with that letter (letter or orthographic fluency task), or words starting with the sound produced by the letter (phonemic category task). This choice elicits different responses: in both cases the sound of the letter is likely to act as the main cue, but in the first case participants will additionally check their associations for correct spelling where different graphemes produce the same phoneme, such as the letters *f* and *v* in German, which may both be pronounced as /f/.

Regardless of type, the measure obtained by verbal fluency tasks is initially a number of words or phrases produced within the time allowed. The number of producible words is influenced by the absolute set size of a given stimulus on the one hand, and the familiarity of a person with a particular category on the other. "Familiarity" would logically seem to entail both the dimensions of first familiarisation/learning, and recent (re-) exposure. Seeing as both "lexical access" and

"lexical retrieval" imply that something can be retrieved to start with, the stimuli chosen are usually relatively common and ideally culture- and gender-neutral categories.

The total number of words verbalised allows conclusions to be drawn about the representation of the relevant category in memory and/or the ease of accessing those representations. Reduced access, repetitions, interferences (e.g. words produced in a non-target language) and certain types of errors may be indicative of patterns of brain organisation and processing mechanisms. For this purpose, qualitative analyses of semantic clusters and chains and, where appropriate, the overlap of words between different languages or trials (Roberts and Le Dorze 1997) are sometimes conducted.

Verbal fluency tasks have been used extensively in the study of neurological conditions, such as aphasia, Parkinson's and Alzheimer's diseases, as a method of gaining a better understanding of brain organisation and language processing (e.g. Goulet, Joanne et al. 1997; Beausoleil, Fortin et al. 2003). Other studies have focused on healthy monolinguals, for example, children and ageing people (e.g. Lupien, Roch Lecours et al. 1994; Goral 2004). Finally, the tasks have been deployed in the study of bilinguals, both healthy and neurologically impaired (e.g. Roberts and Le Dorze 1997; Gollan, Montoya et al. 2002), to assess bilingual dominance and to study differences in bilingual language organisation and processing vis-à-vis monolinguals, and between different groups of bilinguals.

In relation to the cognitive process tapped by verbal fluency tasks, Köpke and Schmid (2004: 21f) observe:

Note that word finding difficulties are [...] the most common disorder in all types of aphasia, in healthy aging (see Goral, submitted), as well as one of the most frequent performance problems observed in young healthy unilinguals (particularly in case of fatigue, stress, etc.), suggesting that lexical retrieval is one of the most vulnerable aspects of language processing.

This also suggests that verbal fluency tasks might also be a fruitful avenue for the study of L1 attrition. Lower performance on a verbal fluency task indicates reduced lexical retrieval, which has been found in the majority of studies on L1 attrition as one, sometimes the only, manifestation of a reduction in L1 (cf. Chapter 2, e.g. Olshtain and Barzilay 1991; Ammerlaan 1996; Waas 1996;

Hulsen 2000; Köpke and Nespoulous 2001; Schmid 2001; Goral 2004). Consequently, several studies on L1 attrition have used the task (e.g. Waas 1996; Goral 2004; Dostert 2009).

Even if this task lends itself to the study of L1 attrition, the results obtained have to be interpreted cautiously. As pointed out above, verbal fluency scores reflect familiarity with a given category. Consequently, they are not a direct expression of overall proficiency in a language (Lambert 1955). Low scores may be due to a lack of acquisition/representations or due to impaired access if the overall activation of the language is low. This implies a need to carefully choose stimuli for which participants are likely to have built up sufficient representations in the past.

An additional problem in interpreting the results arises out of the fact that L1 attriters are bilinguals. As Gollan, Montoya et al. (2002) show, there seem to be systematic differences between bilinguals and monolinguals in naming speeds (bilinguals producing fewer verbalisations on all trials, particularly semantic ones). Partially in consideration of this problem, the controls in the present study were bilinguals rather than monolinguals in order to reduce the possibility of differences in performance between the bilingual group and the controls being due to the factor of bilingualism rather than attrition.

Using verbal fluency tasks as a measure of individual bilingual dominance/balance is less problematic. Their utility for this purpose was first shown by Lambert (1955) who had established bilingual balance by way of a different measure (response-time differences) but found high correlations ( $r = 0.82$ ) between this and the verbal fluency measure ("index of verbosity index"). Several studies have subsequently deployed the task for this purpose (Lambert, Havelka et al. 1959; Fishman and Cooper 1969; Roberts and Le Dorze 1997; Langdon, Wiig et al. 2005).

### Tasks and Stimuli

In this study, we used two semantic fluency tasks (Fishman and Cooper 1969) – FiCA1, a category naming task with two stimuli or conditions in which participants named either *professions* or *pieces of furniture*, and FiCA3, word-listing by domain (WLD) which required participants to verbalise free associations with one of *street*, *church*, *school* and *kitchen* (four stimuli). Further, we used one letter fluency task (FICA2; Benton, Eslinger et al. 1985), FiCA2 (three stimuli), in which participants named words starting with either *F*, *A* or *S*.

The purpose of using a variety of verbal fluency tasks was to explore whether certain types of naming tasks are experienced as more difficult than others, i.e. phonological/orthographic vs. different types of semantic criteria. Rohrer, Wixted et al. (1995: 1136) for example have argued that semantic fluency tasks are more "natural" than phonemic ones, and Rosselli, Ardila et al. (2000) have found differential effects between semantic and letter naming. Secondly, we sought to establish whether dominance patterns within bilinguals differed depending on the type of task, and whether both access routes were equally available in L1 and L2.

Though not the focus of this study, it was possible to pursue these questions since verbal fluency tasks are very fast to administer while being highly reliable (Schmid 2004: 360). An "accidental" advantage of the decision to use several types and of the speed of their administration was that they could be used as a "breather" or "switching task" between the more demanding language tasks in the test battery.

#### Category Naming Task, FiCA1

As pointed out above, two stimuli were used for the category naming task (FiCA1): *professions/Berufe* and *furniture/Möbel*. "Professions" might have been better given as "occupations" or "jobs" to avoid the narrow interpretation of the stimulus as "the professions" (i.e. an occupation requiring formal specialist training and qualification, and one whose affairs is overseen by regulatory bodies (Bullock and Trombley 1999: 689)). However, except for one bilingual participant and one Irish control, who initially interpreted the stimulus in the strict sense and were additionally prompted with "jobs", all other participants interpreted "professions" in the broader sense of "a paid occupation, esp." (but not limited to) "one that requires advanced education and training" (Hornby 1995: 883), i.e. any occupation. This was evidenced from their responses which included *bakers, electricians, dancers, chimney sweeps* and *cleaners* without further prompting after the general invitation to "name any professions at all".

The choice of stimuli for this task was motivated by the fact that on the German side two very similar groups were being compared – native speakers of German who are not monolinguals. The most commonly used stimuli in studying aphasia and dementia (Roberts and Le Dorze 1997) are *animals* and *food/fruit and vegetables*, which was the reason for some of the researchers in the

attrition network (e.g. Dostert 2009) to employ the same stimuli. However, since these stimuli are among the highly productive categories (Roberts and Le Dorze 1997: 419, and below), we felt that they were going to leave very little room for variation and might produce a ceiling effect. We therefore checked the literature to identify two stimuli that were comparable to each other, but had lower productivity than either "animals" or "fruit and vegetables".

Sabourin (1988), using a 2-minute elicitation task, tested 10 different semantic and 10 letter stimuli in French and classified them into three categories by productivity or number of verbalisations (high, average, low). His set of semantic stimuli included *animals*, *fruit*, *vegetables* (individually) and *furniture*. As expected, *animals* were highly productive (mean = 27.6 words/2 min), and *fruit* and *vegetables* each came in the average category (17.2 and 17.1 respectively/2 min). This pattern was replicated for English by Gollan, Montoya et al. (2002), who, on a 1-minute elicitation task, tested a similar number and similar types of semantic and letter stimuli (*animals* 20.5, *fruit* 16.3, *vegetables* 12.7/1 min). However, if *fruit* and *vegetables* were combined as one category, one would expect the productivity to increase. This is confirmed in Dostert (2009) where *animals* on the one hand, and *fruit and vegetables* combined on the other, achieve similar mean scores (21.7 and 22.2 respectively, 1 min).

Gollan, Montoya et al.'s set also included *occupations*, here restricted to the ones that require an advanced degree, which showed low productivity (7.3/1 min), while *furniture* had average productivity (20.5/2 min) relative to the other stimuli in Sabourin's study. Again, if *occupations* was broadened out, one would expect an increase in the level of productivity.

<b>Stimulus/ Productivity</b>	<b>Animals</b>	<b>Fruit</b>	<b>Vegetables</b>	<b>Fruit/Veg combined</b>	<b>Occupations</b>	<b>Furniture</b>
<b>Sabourin</b>	high	average	average	-	-	average
<b>Gollan et al</b>	high	average	average	-	low*	-
<b>Dostert</b>	high	-	-	high	-	-
<b>Opitz</b>	-	-	-	-	average- high**	average

**Table 10: Productivity of different semantic stimuli.**

\* restricted set, \*\* broad set

We have been unable to confirm this pattern independently for German, but given that these stimuli are culturally and linguistically relatively neutral, we would not expect an influence of the language in which the task was performed in. Our own results from the control groups provide evidence for both stimuli being of comparable productivity in German (*professions* 19.9; *furniture* 18.6 per 1 min interval, cf. Table 11 below, which also provides the results for the other two types of stimuli). However, there was a significant difference between English *professions* (23.2/1 min) and English *furniture* (18.2/1 min) ( $t(33) = 2.38, p = .023$ ). When comparing the stimuli across languages (German *professions*/English *professions* and German *furniture*/English *furniture*), they were not significantly different. This nevertheless implies that certain precautions have to be taken when comparing the bilingual group's scores within tasks (cf. Chapter Analysis).

FiCA	Stimulus	German control group				Irish control group			
		N	Mean	Range	StD	N	Mean	Range	StD
FiCA1 – Category Naming	<i>Professions</i>	14	19.9	13-28	4.1	17	23.2	13-34	6.4
	<i>Furniture</i>	15	18.6	14-22	2.6	18	18.4	12-31	4.8
FiCA2 – Letter Fluency	<i>F</i>	7	15.6	6-22	5.4	15	18.2	11-27	4.7
	<i>A</i>	10	18.0	8-27	6.2	9	19.6	10-32	7.2
	<i>S</i>	12	22.3	19-29	2.5	9	20.9	8-28	7.4
FiCA3 – Word Listing by Domain	<i>School</i>	8	26.1	18-33	5.4	13	27.1	18-37	6.6
	<i>Kitchen</i>	9	25.7	18-35	5.1	3	30.3	21-35	8.1
	<i>Street</i>	8	22.8	17-29	4.0	12	21.9	15-34	5.7
	<i>Church</i>	4	19.3	16-21	2.4	6	26.2	19-32	6.1
<b>Scaled Total Score</b>		<b>20</b>	<b>68.22</b>	<b>46-90</b>	<b>10.6</b>	<b>18</b>	<b>74.2</b>	<b>44-108</b>	<b>16.6</b>

**Table 11: Mean scores for German and Irish control groups for all FiCA conditions**

There is no apparent reason for this between-category difference in one language only. It may possibly have to do with the fact that the basic set of professions is larger than that of furniture where continuous verbalisations are only possible if one includes sub-categories of the basic set (types of chairs, beds etc.). Apart from English *professions*, the stimuli satisfy the criterion of being of lower productivity than *animals* and *fruit and vegetables* in Dostert's (2009) study, whose participants were selected according to the same criteria as those in the present study. However, the



difference is not as large as had been expected on the basis of Sabourin's (1988) and Gollan, Montoya et al.'s (2002) studies.

In hindsight, instead of using different stimuli, one could have avoided a possible ceiling effect by altering the procedure for *animals* and *fruit and vegetable* and using a longer time interval (Roberts and Le Dorze 1997). As Rosen (1980) suggests, different cognitive processes predominate during different phases of semantic verbal fluency tasks. During the first 30 seconds, verbalisations are accessed by way of automatic processes, following which a "less automatic and more organized exploration" (Joanette, Goulet et al. 1988: 57) of the lexicon takes place. It would be reasonable to assume that the automatic process is similar between the bilinguals and controls, whereas difficulties might be expected for the bilingual participants during the more conscious word retrieval phase. By widening the time interval, group differences should become more visible.

#### Letter Fluency Task, FiCA2

The three stimuli for FiCA2, the letter fluency task, *F*, *A* and *S*, were chosen because the FAS test is one of the "most commonly used forms of the letter fluency task" (Lacy, Gore et al. 1996: 304). Usually, all three letters are administered to participants; however, Harrison (2000) found that results from the three-letter version correlate highly with those obtained when just a single letter is administered, questioning the advantage of using the longer version. In light of this, and since we needed different stimuli for the two language conditions for the bilingual group, we decided to administer just one of the letters in each case.

In contrast with the category naming task, we would expect a difference between the stimuli within each language and across the language pair, since the absolute set size/productivity of each letter is likely to vary (Sabourin 1988; Gollan, Montoya et al. 2002). A crude measure of category or set size (suggested by de Bot, p.c.) – space taken up in a dictionary, here the Duden Oxford (Scholze-Stubenrecht and Sykes 1990) – reveals that *A* and *S* are comparable in German (85 and 84 pages respectively), while *F* is less productive (36 pages). This expected pattern is partially borne out in the scores of the German controls who verbalise most to *S*, followed by *A*, followed by *F* (cf. Table 11 above). The difference among the three conditions is significant ( $F(2, 26) = 4.81, p = .017$ ) owing to the difference between letters *F* and *S*, which has implications for the analysis of the data.

In English, on the basis of the same dictionary measure, *S* is the most productive letter (104 pages), followed by *F* (80 pages), followed by *A* (37 pages). This hierarchy of productivity is partially supported by Gollan, Montoya et al. (2002) who in their comparison of different letter tasks obtained the following baselines: *S*: 16.7, *F*: 13.5, *A*: 13.1 per 1 min interval, where *S* is indeed the most productive, but the difference between *F* and *A* would have been expected to be bigger. However, in the present study the Irish control group achieved fairly similar results across the three conditions, which again might point to a ceiling effect. It is unclear why our and Gollan, Montoya et al's results differ in this regard.

Overall, the letter fluency task was experienced as the hardest of the three verbal fluency tasks by the two control groups (lower scores compared to the other two types).

### Word Listing by Domain, FiCA3

The WLD (word listing by domain) task was first used by Fishman and Cooper (1969) in a study assessing the efficacy of different measures of bilingualism. Fishman and Cooper administered a word-naming task (category naming task) and a free associative task covering essentially the same five domains or contexts, to their Spanish-English bilingual respondents. In the first task, these were *kitchen*, *neighbourhood*, *religion*, *education* and *work*. In the second task, "[r]espondents were [...] asked to give continuous associations, within 1-min periods, to each of the following stimulus words: *home*, *street*, *church*, *school*, *factory*, *casa*, *calle*, *iglesia*, *escuela*, and *factoria*." (Fishman and Cooper *ibid.*: 277).

The stimulus words across the two tasks were chosen to represent the same domains and are presumably supposed to be systematically related: the stimuli in the word-naming task are usually more abstract (*religion*, *neighbourhood* etc.), and those in the free associative task are a concrete manifestation of that same concept (*church*, *street*). This relationship is shared by the stimuli *home* and *kitchen*; however, surprisingly *kitchen*, arguably the more concrete term, occurs in the word naming task, while *home* is the stimulus for the free association. It also remains unclear to what extent the related stimuli actually succeeded in eliciting different kinds of responses, given that those for the word-naming task are rather broad categories while the narrower terms were used with the "freer" task.

In this study, we follow Langdon, Wiig et al. (2005) in using four of the domains or contexts (*kitchen, street, church, and school*) in a word-association task. However, we did not administer all four stimuli to each participant, but again chose one for one language condition, and another for the other language condition.

As in the study conducted by Fishman and Cooper, Langdon, Wiig et al.'s purpose was to examine and compare different measures for the assessment of language dominance. Despite the WLD task being "a classic and commonly used measure for assessing language dominance" (Langdon, Wiig et al. *ibid.*: 319?), they found that there was a low correlation (52%) between the WLD and self-ratings of language dominance. The self-ratings, on the other hand, correlated 100% with three of five Alzheimer's quick tests.

It is not quite clear what causes the discrepancy with Fishman and Cooper's results, and the low correspondence with self-rated language dominance. Self-ratings of proficiency and dominance have in the main been shown to be fairly reliable (Fishman and Cooper 1969; de Bot and Lintsen 1986; Jia and Aaronson 1999; Yeni-Komshian, Flege et al. 2000) so they should correlate better with the WLD, provided that test measures what it is thought to measure. It is true that productivity of semantic conditions depends on an individual's familiarity with the relevant domain. However, the WLD categories are quite broad and participants can associate freely which should compensate for lack of familiarity.

On the other hand, this again may serve to produce a ceiling effect allowing all participants to verbalise fluently during the relevant time interval regardless of activation level of the particular language. In our study, the WLD did elicit more verbalisations than the category-naming task. This is not surprising given that in the latter words had to be category members, thus constraining verbalisations, while in the WLD task any association was valid.

Given the breadth of stimuli on this task, we expected the controls (who responded to one or two of the stimuli only) to achieve fairly similar results across the four stimuli. This is in fact borne out in the absence of significant effects between the conditions for the control groups. However, the numbers of controls for each condition was by necessity lower than for the other two tasks (cf. Table 11 above) so that it would have been quite difficult to obtain significant results. The mean scores reveal a trend for more verbalisations on average being produced for *kitchen* and *school* than

either *church* or *street* in both languages. Across the control groups, *church* just reaches significance ( $t(6.914) = -2.505, p = .041$ ) due to unequal variances. In view of this, the results for each condition will be discussed separately.

### **3.3.2.3 Film retelling**

Film retellings, along with picture descriptions and story retellings, belong to the group of written or spoken production tasks (also referred to as "reconstruction"; Larsen-Freeman and Long 1991: 29). They occupy an intermediate position between tasks that elicit spontaneous speech on the one hand, and elicited production tasks or structured exercises on the other with regard to the constraints placed on participants' productions.

While they allow the speaker relative freedom of expression and would be seen as more "natural" than elicited techniques, they simultaneously place constraints on the range of possible responses through controlling the topic of the production, thus ensuring greater comparability of the elicited speech samples compared to spontaneous speech samples. They may also "force", or at least make more probable, the production of particular language features, which speakers otherwise may avoid or simply not produce enough of (cf. also Larsen-Freeman and Long *ibid.*: 26f; Perdue 1993; Schmid 2004).

In this study, we used an oral film retelling task in a bid to complement the more constrained productions of the C-test and verbal fluency tasks. The task was originally developed by members of the European Science Foundation project (ESF) on adult language acquisition (Perdue 1993). Later, it was adopted as one of the instruments of the LATB (Schmid 2004).

In the original version, a 20-minute sequence from Charlie Chaplin's silent film *Modern Times* consisting of two main episodes was used. Researcher and participant watched the first episode together; the researcher then left the room, and the participant watched the second episode by him-/herself. After the clip was finished, the researcher re-entered the room, and the participant retold the story of the second episode. This procedure served to establish common ground as well as to create a genuine communicative gap between researcher and participant. In cases where the activity was repeated with the same researcher at a later time in the study, retelling motivation and

richness of data suffered, since the researcher was known to be familiar with the story the second time round (Perdue *ibid.*: 106).

The SFI procedure was adapted slightly for the Attrition network's methodology, and then further modified for this study (cf. below). Only the second episode is shown to participants. Participants are, however, briefed on the events leading up to the relevant scene. In some studies deploying this version of the task (e.g. Dostert 2009), the researcher, in line with the SFI procedure, left the room while participants were watching the clip, or at least turned the laptop computer so that the clip would only be watched by the participant, with the intention of increasing retelling motivation.

In this study, on the other hand, we did not attempt to keep up the pretence of not knowing what the clip was about, but preferred to establish common ground between the participant and researcher by watching and enjoying the film together. To avoid the problem of reduced retelling motivation, participants were told that people unfamiliar with the clip, an external audience in other words, would be listening to the stories: other researchers in the Network as well as "Brian", an imaginary English native-speaker colleague and primary analyst of the English texts (cf. Section "Administration"). In other words, participants were asked to create a narrative that would function as a text in its own right outside the situation in which it was retold.

The episode shown is just under 10 minutes in length (9 mins 56 secs) and starts about 33 minutes into the film with Charlie Chaplin showing up at a shipyard in search of a job with a letter of recommendation from the sheriff. The following is a summary of the plot of the sequence from Schmid (2004: 13):

[...] C(harlie) C(haplin) is released from prison with a letter recommending him as an honest and trustworthy man. He takes this letter to a shipyard [...], is accepted for work there but messes up rather badly and leaves. He walks through the city where he meets a young girl who's just stolen a loaf of bread. She is apprehended by the police, but he tries to claim it was him. However, a bystander says it was her, so he is released again. He goes into a restaurant, eats a lot of things and then says he can't pay, so he is arrested again. After a bit more to-ing and fro-ing, he's loaded into a police van, into which the girl is then also put. During an accident they manage to escape, and then walk through the suburbs. They sit in front of a house, and CC starts fantasizing how nice it would be for them to live in a house like that. They wake up back to reality and realize they're very hungry, and there is a policeman standing behind them. They get up and walk away, [...]. (Schmid 2004: 13)

Since we needed parallel productions in German and English for the bilingual participants, we decided to split the sequence into two sections, breaking at the point where Charlie is finally put into the Black Maria (at 4 mins 32 secs). The two sections were always shown in the same order, as the happenings in the second section make sense only with knowledge of the first. However, the order of the language sets was alternated between participants (cf. Section 3.4), so that roughly half the participants started with the English set, re-telling the first sequence in English and the second in German, and vice versa for the other half, allowing us to control for training effects.

Instead of splitting the sequence, another possible avenue for developing parallel language conditions for this task would have been to identify a comparable 10-minute sequence from another silent film. However, this was dispreferred for a number of reasons. First, it would have necessitated finding a clip of similar complexity and possibly a similar theme which, given Charlie Chaplin's unique take on universal themes would have been difficult. Second, sticking to the same clip means that comparable data are available, even if our retellings are shorter than those elicited within the Network and the SFI project. We also decided against administering the whole 10-minute clip in both languages to avoid training effects and reducing motivation in repeating the exact same task.

As Perdue (1993) points out, this task has the capacity to yield rich linguistic data:

[This film retelling task] is a complex verbal task, in which the learner retells a series of events whose relationship to each other must be specified. Within each event, he has to say who did what to whom, introducing new characters and maintaining characters who are already on stage. The main characters are male (Charlie Chaplin) and female (the young girl), and they act and are acted upon. Therefore, the informant has to deal with referent introduction and maintenance in a wide range of semantic functions. (Perdue *ibid.*: 6f)

The story line by necessity limits the linguistic features that can be elicited (for example, a focus on 3<sup>rd</sup> person speech rather than personal speech, questions, orders or deictic reference (Perdue 1993: 106)), which is seen as a "drawback" by the authors of the test (*ibid.*). This would seem to be of less relevance in our study since we are not dealing with beginning learners, although a comparison with other types of speech data could be interesting. The main advantage of using an elicited task such as this lies in controlling what can be retold, and partially dictating the linguistic means which may be used to accomplish the task.

As in other studies using this task, participants first retold the story with minimal interruption by the researcher, producing a monologue. Occasionally, participants had to be prompted in order to ensure all major points were covered, or if they got stuck remembering what came next.

Given that the task was used to test bilinguals in their mother tongue, we decided to challenge them a little more and to attempt to elicit some more complex/abstract speech (Robinson 2007; Robinson and Gilabert 2007). Therefore, we followed the monologue with some questions concerning the reasons for Charlie's and the girl's behaviour and their relationship to one another (cf. Appendix). In the resulting dialogues participants were thus required to make inferences about cause-effect relationship, intentions, assessment of various situations and character, which goes beyond the mere description of concrete objects and events.

### **3.3.3 Linguistic aptitude test**

Part of the interest of this study was to establish whether linguistic aptitude, in addition to be promoting the successful acquisition of the second language by our adult participants, has a role to play in their first language maintenance (cf. Chapter 2). To this end, we employed three of the five sub-tests of the Swansea Language Aptitude Test (Meara, Milton et al. 2001; LAT parts B, C and D).

This test battery, like most expansions and adaptations of the earliest linguistic aptitude tests, such as the MLAT (Carroll and Sapon 1959), was developed to predict likely success in second language instruction (Abrahamsson and Hyltenstam 2008: 486), since linguistic aptitude is usually considered to play a minor role in first-language acquisition:

These tests are considered by the authors to be reliable indicators of an academic propensity for foreign language learning in adults and, if used responsibly, can form part of a professional system for the selection of personnel for training or students for language learning. (Meara, Milton et al. 2001: 6f)

However, this preoccupation with L2 learning does not render this test unsuitable for our purposes. The LAT tests the same aptitude components originally identified by Carroll (1981): a) phonetic/phonemic coding ability, b) grammatical sensitivity, c) rote learning ability, d) inductive learning ability. By their nature, these abilities are independent of the particular language(s) a

person knows. On the basis of the literature, we would thus expect an interaction between these abilities and second-language proficiency, while also hypothesising an effect for the level of L1 maintenance.

Following Spolsky (1995), the LAT test belongs in the category of "synthetic" aptitude tests, insofar as the abilities are gauged on the basis of "minilessons in an artificial or foreign language" (ibid.: 324). It contains five subtests, for phonetic memory (LAT A), lexical memory (LAT B), grammatical inferencing skills (LAT C), aural memory for unfamiliar sound sequences (LAT D), and the ability to form sound-symbol associations (LAT E) respectively.

Of these, we employed only LAT D, B and C (in this order) for reasons of economy. First, LAT A and D seem to be testing the same construct (phonetic/phonemic coding ability), and in a similar manner (exposure to unfamiliar sound sequences), but LAT A is a self-assessment task (test takers decide for themselves how well they did), whereas LAT D is an objective test (test takers need to indicate whether or not they recognise a sound sequence; the correctness of the answer depends on whether the sequence had previously been presented).

Second, we excluded LAT E (ability to form sound-symbol associations), partially, because it, like LAT C, "tends to favour people with formal training in languages or linguistics" (Meara, Milton et al. 2001: 21), while our interest lies primarily in the "natural" abilities of our participants. A second reason was that although the ability to handle written language is undoubtedly important for achieving advanced proficiency levels in a language (Meara, Milton et al. ibid.: 21), it may be hypothesized that in an immersion learning context reduced ability or deficits in this area may be partialled out by the (potentially) rich spoken context. Consequently, in our test battery we primarily elicit spoken rather than written data, which further reduced the need to include this subtest since we do not have any directly relatable proficiency data.

Each of the subtests returns a score out of 100%. The scores are not easily combinable since the normal distribution curves for each subtest fall in different areas. This is expressed in the manual by stating for each subtest where the top 10%, next 20%, middle 40%, next 20% and bottom 10% of all scores fall on that particular test. Thus, average scores have to be used with caution, taking into account the particular test taker's profile: While in practice most test takers achieve either consistently good or consistently bad results, "[i]n theory it is possible for test-takers to display



good scores on some tests and poor scores on others", thus displaying areas of "strengths and weaknesses" (Meara, Milton et al. 2001: 23). The LAT results programme plots the profile of a each test taker onto a histogramme-style graph, with a dividing line cutting across the bars at the same point between average and good scores, displaying clearly whether the profile is consistently good, average, bad or "broken" (ibid.: 23).

The LAT test is available in a French and an English version. Some studies (e.g. Abrahamsson and Hyltenstam 2008; Abrahamsson and Hyltenstam 2008) have used versions of the LAT test that were adapted to other languages (in this case Swedish, the L2 of the participants). As Abrahamsson and Hyltenstam (ibid.: 504) point out:

"It is worth stressing, however, that the actual language materials were not based on Swedish but taken from so-called rare languages of which the testees had no knowledge."

In other words, in neither version is the actual test language the participants' mother tongue but a specially constructed artificial language referred to as "Novish" which contained both made-up nonsense words, and material from languages the participants were not familiar with. English/French are used in order to present the test material (in the follow-up version of the LAT, the LLAMA tests (Meara 2005; Meara 2005), this becomes unnecessary by introducing pictorial elements).

Although it would have been preferable to use an aptitude test in German/a German version of the LAT test, or indeed a test that makes no reference to a particular mother tongue, at the time of preparing the study, none were available. However, we realised that the amount of English language knowledge required to access the LAT test materials and the properties/size of the language to be learned are rather limited, and therefore felt justified to use the English version of the test.

Thus, the vocabulary to be learned in LAT B is restricted to ten verbs of varying degrees of abstractness (to drink, to sit, to sleep, to think etc.). The total vocabulary of LAT C consists of three nouns (three animals), four verbs, four adjectives, one case marker and a plural demonstrative/definite article (subject position). There is just one syntactic pattern (SVO in English, VSO in Novish), and the Novish grammar is limited to one form for verbs, nouns and

adjectives each regardless of case and number, while the copula 'to be' and other articles are not expressed. LAT D works without reference to any language other than Novish.

According to our eligibility criteria, the German participants were required to have at least a good level of English. As a result, the actual level of proficiency attained by our participants far outweighed the level necessary for accessing the test materials. We therefore considered that the use of the English version would not militate against the performance of the German participants on LAT B and C, the two parts involving English as the contrastive language.

Nevertheless, the three groups' results for LAT B turned out to be significantly different owing to the better performance of the Irish controls ( $F(2,59) = 9.524, p = .000$ ). While not significant for the other two sub-tests, the Irish controls' mean is one band higher than the German groups' means on LAT C. This limits the potential use of these two measures, since we cannot exclude the possibility that the difference is an artifact of the test, rather than a genuine difference.

### **3.4 Administration of the test battery**

The test battery was administered in three distinct phases (cf. Table 12). After first contact, potential participants were sent and asked to complete the personal background questionnaire, which collected general biographical data and data in respect of L1 and L2 acquisition, and also functioned as a screening questionnaire.

Researcher and participant then met for an individual formal testing session during which the bulk of the data was elicited. At the end of this session, participants were given the attitude questionnaire and can-do scales with a request to complete both in their own time and return them to the researcher.

Only the bilingual group tackled the entire test battery. The two control groups were administered a shortened version that did not include the sociolinguistic questionnaire or any language tasks in the participants' L2, except for the L2 C-test. This test was included to obtain a second measure of L2 proficiency in addition to the self-ratings, allowing us to describe the degree of bilingualism of the control groups with more confidence. We refrained from administering the other L2 tests to the controls in the interest of respecting participants' time, and since it was not our aim to study the bilingualism of the controls. The controls did, however, tackle the aptitude test LAT which we

considered useful to include in order to establish any aptitude differences between the groups in question, rather than comparing the bilinguals' scores against absolute norms.

<b>Time of administration</b>	<b>Instrument</b>	<b>BL*</b>	<b>GC/IC*</b>
Prior to testing	Personal background questionnaire	✓	✓
Testing session	Sociolinguistic questionnaire	✓	—
	C-Test	✓	✓
	Verbal fluency tasks (FiCA, Fluency in controlled associations)	✓	L1 only
	Film retelling	✓	L1 only
	Sentence generation task	✓	GC only
	Picture description	✓	L1 only
	Linguistic aptitude test (LAT)	✓	✓
Post testing	Can-do scales	✓	✓
	Attitudes towards foreign language learning/ bilingualism	✓	✓

**Table 12: Test battery: administration**

\*Completed by: BL – bilingual participants, GC/IC – German/Irish control group

### **3.4.1 The testing session**

The formal testing sessions took place either in a location chosen by the participant, usually their own place of residence, or a room provided by the researcher at university or a friend's house. Sessions were normally conducted in one block, but in the case of some bilingual participants they were divided into two separate meetings for personal or logistic reasons.

The sessions with the bilingual group averaged four and a half hours in length; those with the German and Irish control groups had an average session length of three hours. Breaks were taken at suitable points between sections (cf. below), or occasionally between tasks, if requested. Table 13 below and the session protocols included in the Appendix set out the structure of the sessions for each of the three groups. The timings given for each task are approximations; the actual duration partially depended on participants' personality, their interest in sharing experiences (with regard to the questionnaire) and their attention to detail (with regard to the longer spoken tasks, film retelling and picture description).

Testing session	Time	Completed by*
<b>1. Interview (sociolinguistic questionnaire, part 1)</b>	<b>0:30</b>	<b>BL</b>
<b>"English half"</b>		
2. FiCA – semantic (CAT)	0:02	BL, IC
3. C-Test	0:25	ALL
4. Charlie Chaplin film retelling	0:10	BL, IC
5. FiCA – phonological (FAS)	0:02	BL, IC
6. Picture description	0:15	BL, IC
7. FiCA – semantic (WLD)	0:02	BL, IC
<b>8. LAT-Test (D, B, C)</b>	<b>0:35</b>	<b>ALL</b>
<b>Break</b>	<b>0:15</b>	
<b>"German half"</b>		
9. FICA – semantic (CAT)	0:02	BL, GC
10. C-Test	0:25	ALL
11. Charlie Chaplin film retelling	0:10	BL, GC
12. Sentence generation	0:10	BL, GC
13. FICA – phonological (FAS)	0:02	BL, GC
14. Picture description	0:15	BL, GC
15. FICA – semantic (WLD)	0:02	BL, GC
<b>16. Interview (Sociolinguistic questionnaire, part 2)</b>	<b>0:30</b>	<b>BL</b>
<b>17. Conclusion/debriefing/request to complete attitudes questionnaire and can-do scales</b>	<b>0:10</b>	<b>ALL</b>

**Table 13: Overview of testing session**

**\*BL – bilingual participants, GC/IC – German/Irish control group, ALL – all three groups**

The session (like any other contact with the participant) was conducted in the participant's L1. In contrast, each set of language tasks (“German half”/“English half”) was conducted in the language to be tested. Thus, all participants used both English and German at some stage during the session. Furthermore, they knew that the study was in the area of bilingualism and had signed up in their capacity as bilinguals. Finally, they were aware that the researcher herself was a bilingual herself. All of these factors combine to move participants' towards what has been termed the bilingual language mode during the session.

We already referred to Grosjean's (e.g. 1989; 1998; 2001; 2008) concept of language mode, in Section 2.3.1. The importance of taking language mode into account is explained as follows:

"Because the mode a bilingual is in corresponds to a state of activation of the bilingual's languages and language processing mechanisms, it has an impact both on language production (maintenance or change of the base language, amount and type of language mixing that takes place, etc.) as well as on language perception (speed of processing of a language, access to one or to both lexicons, role of the less activated language, etc.)." (Grosjean 1998: 137; 2008: 252)

Thus, with increasing activation of the non-base language(s), higher amounts of code-switching and borrowing behaviours are expected, although it is acknowledged that individual bilinguals vary in the extent to which "they travel along the continuum" (Grosjean 1998: 251). Empirical evidence for the existence of the modes, and their impact on bilingual performance is building, and consequently it is a variable that needs to be controlled for.

For our purposes, we aimed to elicit language proficiency independently in the two languages of the participants for later comparison and correlation. We were also interested in the (involuntary) interaction between the two languages, which can only be teased apart from conscious and deliberate language switching in monolingual mode (Grosjean 1998: 137; Grosjean 2008: 40). Hence, we needed to ensure that participants were, in so far as possible, approaching monolingual mode during the two sets of language tasks.

The first measure was to elicit the language proficiency data in designated sets, as described above, with all instructions and materials in the language to be tested. (The one exception is the silent film used during the film retelling task which showed two English-language intertitles in the first sequence.) Additionally, before the start of the L2 "half", participants were involved in a short conversation in that language. However, as Grosjean (1998: 140; 2008: 257) point out, while this serves to "set" a particular language as the base language, it does not guarantee a particular position on the language mode continuum.

Therefore, as a further measure, and since the option of "creating" the monolingual mode by employing monolingual interviewers and experimenters, and consequently a truly monolingual context for the study was closed to us, we invoked absent future recipients of the data. All participants were informed at the beginning of their session, and reminded at the start of the English language set that the English-language data would not in the first instance be analysed by the researcher since she is not a native-speaker of English, but rather by a colleague named Brian,

who did not speak any German. Similarly, the German data would be scrutinized by expert German native speakers who did not live in a bilingual context. Furthermore, the researcher from the outset presented herself as a "language purist", who was careful to avoid mixing languages herself and created the expectation that only one language would be spoken at any one time, which tied in plausibly with the study's design as it was explained to participants.

Grosjean (1998: 140; 2008: 257) comments in relation to a study conducted by Soares and Grosjean (1984), which also employed instructions, practice sentences and stimuli in the relevant test language only, that this "probably pushed the participants towards the monolingual endpoint of the continuum", despite the fact that participants were used to mix languages with their interlocutors. Similarly, we would expect the above measures to be an effective "anti-dote" against the factors promoting a (more) bilingual language mode. Judging by participants' behaviour and comments, the implicit instruction to deactivate the language(s) which were not being tested at the time seems to have been taken on board by all participants, since code-switches and borrowings were generally avoided and only used as a last resort, usually following a longer unsuccessful lexical search.

At the beginning of the session, participants were shown an outline of the various sections of the session to be covered, with an indication of overall time and breaks. They were told that tasks varied in how formal, fun etc. they were, but that an attempt had been made to keep the tasks varied and interesting. Participants were reassured that all data would be treated confidentially, and that they could withdraw from the study at any time. Participants were also reminded that there were no financial rewards arising out of participating in the study.

All parts of the test battery were introduced and explained by the researcher in the appropriate language. Some of the language tasks (e.g. the C-tests and the sentence generation task) additionally had written instructions (cf. Appendix), which some participants read after having been talked through them by the researcher.

Tasks and stimuli were either provided in hard-copy or shown on the researcher's laptop, a Sony Vaio PCG-Z600NE running Windows XP. During the session, participant and researcher sat at a table, the participant usually at the head of the table with the researcher at right angles to his/her left. Occasionally this arrangement was not possible, so that researcher and participant sat at

opposite sides of the table facing each other. The laptop stood in the middle of the table at about 80 cm distance from the participant to give space for the tasks involving paper stimuli; it was angled slightly towards the researcher to allow her to check the recording (cf. below). For those tasks that were administered on the computer, it occasionally was moved closer until the participant was comfortable and confirmed that they were able to see correctly.

The whole session (with the exception of the C-test which was done silently) was recorded via a stereo microphone (Skytronic 173.623) clipped unobtrusively to a pen or similar object which was standing in a tumbler with other pens on the opposite side of the laptop relative to the researcher. One side of the microphone was pointing towards the participants to ensure that their speech could be heard clearly since many participants spoke more quietly than the researcher. The microphone plugged into a portable DAT recorder (Sony DAT Walkman AVLS) which in turn was connected to the laptop. The laptop ran a 2006 version of Total Recorder software, and the session was thus recorded simultaneously on the DAT recorder and the computer to guard against data loss. A mini disc recorder, second microphone and spare batteries for all devices were also always carried, but were not needed in the event.

The session proper started with a quick review of the personal background questionnaire which the participants had filled out beforehand and either posted or mailed to the researcher, to clarify certain points or fill in missing information. For the bilingual group the review naturally led into part 1 (Language use) of the sociolinguistic questionnaire. Following this, the bilingual group tackled either the English or German set of language tasks, whereas the control groups moved on to the first set of language tasks immediately after the initial questionnaire review.

For all groups, the order of the language sets was alternated, or counterbalanced, between participants: roughly half the participants did the German set first, the other half the English set, while the other language was tested after a longer break/non-language specific task (the aptitude test). In the case of the controls, the L2 set only involved the C-test, so their "L2 half" was much shorter than the corresponding "half" for the bilingual group who attempted all language tasks twice, in L1 and L2.

The purpose of counterbalancing tasks is to avoid order effects, that is, the impact on the performance on a task resulting from completing an earlier task. As Mitchell and Jolley (2009:

477) point out, there are four types of order effects: 1) practice effects as a result of which participants may show improved performance on a subsequent task; 2) fatigue effects which may lead to decreased performance as a result of waning motivation and enthusiasm or tiredness; 3) treatment-carryover effect (a treatment administered earlier is still effective in later trials); and 4) sensitization, i.e. changed performance (deliberately or unconsciously, better or worse) as a result of participants having guessed what the tests are designed to test.

For our purposes, it was crucial to control practice effects from one language to the other. If all bilingual participants had done the German language set first, and the English second, then their English performance might well have been better than it would have been without "practicing" that type of task in German before. If, on the other hand, English had always been presented first, and German second, we might not have been able to detect the true level of L1 maintenance/attrition, since German performance would be elevated following the English language trial. Consequently, our assessment of the bilingual profiles at group level would have been flawed.

In addition, we needed to control for fatigue effects. Given that participants were exposed to a very extensive test battery, it was likely that they would feel more tired, or less interested in attempting the "same" task the second time round, leading to lowered performance. Thus, by alternating the language set order, we avoid a systematically lowered performance for one language, and again the impact this would have on the bilingual profiles, at group level. The fourth type of order effect (sensitization) is also effectively controlled in this manner, while the third (treatment carry-over) is not relevant to our situation.

Of course, counterbalancing does not avoid either practice, fatigue or sensitization effects in a given participant, but by alternating the order of the language sets we balance those effects between participants and thus rule out any systematic group differences arising out of the repeated administration of tasks in the two languages.

In addition to the effect of the order of the language sets, the individual tasks within each language set are also subject to order effects. However, counterbalancing a test battery as extensive as ours, with up to five tasks and up to four conditions in each task, would have made for an enormously complex and challenging design both for the purpose of administration and analysis. Order effects add a further dimension to statistical analyses and imply the need for at least twice (in the case of



the bilingual group), if not three times (control groups), as many participants, if one wishes to conduct statistical analyses with any degree of confidence (Mitchell and Jolley 2009: 483).

Instead, we opted for a principled task arrangement according to tasks demands (written vs. oral mode, with/without time constraints, free or elicited data, complexity and duration of task) which lead to a general structure of the "big tasks" (C-test, picture description, film retelling and sentence generation tasks) being interleaved with "quick tests" (the three FiCA tasks). This arrangement, we felt, could both help keep participants interested and motivated, and was theoretically justifiable.

First, scores of individual tasks are only compared to those of the equivalent tasks which have the same position in the other-language set. They are not, however, compared to others within the same set of language tasks - in that case, one would have needed to exclude training or sensitization effects. Second, the tasks are all designed to tap one or another aspect of proficiency, and composite scores are produced for the purpose of bilingual profiles. By keeping the task order constant, any effects arising from the particular arrangement of tasks within each set are congruous for all participants, so the same task order within each set would at least appear to have no negative impact. Third, we did use random allocation of stimuli to participants in order to balance out training effects from one particular stimulus on one task to another.

In hindsight, it would have been possible to counterbalance the order of the "big tasks" on the one hand, and of the "quick tests" on the other while still keeping to the structural framework. This would have addressed order effects within each language set, while still adding to design complexity, although to a lesser degree than counterbalancing all tasks within a set. It is unclear whether this extra layer would have been worthwhile given the preceding discussion.

After the first "half" of language tasks, participants either took a break or went on to do the linguistic aptitude test. The controls, because of their reduced schedule, usually took the linguistic aptitude test in the shorter section, i.e. with the C-test in their L2. The bilingual participants usually took the aptitude test with the English-language set of language tasks (i.e. before or after the "English half", depending on whether it came first or second in the session), for two reasons:

First, the "English half" was slightly shorter since the sentence generation task was only administered in German. Second, the English-language version of the aptitude test was used with all participants in the absence of a German-language version for this particular test. Thus, the test

was employed to form a “bridge” between the two language “halves” by keeping participants in or transitioning them into English-language mode. Instructions and explanations for the aptitude test, on the other hand, were usually given in the L1, so for the German groups both languages would have been activated (bilingual mode), the intention being to facilitate access to either language for solving the task.

The second set of language tasks was followed by part 2 of the sociolinguistic questionnaire, which focused on achievements and changes in language proficiency across the spectrum of languages spoken by the individual, as well as attitudes to multilingualism, language policy and the participant's linguistic identity.

The sessions concluded with a general wind-down, including a thank you, a promise to return the favour, and a question about how to make the results of the investigation available to the informants. Also at that point, informants were handed the final questionnaire (can-do scales/attitudes towards foreign language learning and bilingualism) along with a stamped envelope, and asked to fill this out in their own time if they wished.

### ***3.4.2 Administration of questionnaires and language tasks***

#### ***3.4.2.1 Questionnaires***

As discussed above, the questionnaires were either completed by the participants in their own time, or administered as a semi-structured interview. No further instructions in addition to the information provided on the actual questionnaires (cf. Appendix) were given. When handing participants the can-do scales and attitude questionnaire, the provenance of the can-do statements was explained. In relation to the attitude questionnaire, the controls were asked to imagine themselves in a migration situation when rating the statements, since otherwise they would have made little sense.

#### ***3.4.2.2 C-test***

As explained above, participants were talked through the individual tasks, but written instructions for the major language tasks were also provided (cf. Appendix). The test was provided in hard-

copy with a cover sheet containing the instructions, and each text printed on a separate page (single-sided), and was to be filled out by hand by the participant.

The C-test was introduced as the only formal test in the battery, which required participants to complete 5 short texts. The task was described by drawing a comparison with the situation of trying to understand someone in a noisy pub. The distortion principle was made explicit, and participants were reassured that several solutions were often possible. In a bid to reduce anxiety levels and improve motivation, typical native speakers were described as finding about two gaps per text difficult to solve. This did indeed lead to a motivational boost, when participants felt they were performing close to native-speaker standard on the L2 test.

Participants were told that there was a comfortable time limit of 5 minutes per text, with the investigator giving a 1-minute warning and indicating when time was up. The time taken by each participant to complete each text was recorded. Participants were not allowed to return to a text after having moved on, in line with the standard administration procedure of C-tests. In the English-language test, the American text was pointed out when participants got to it.

Participants were instructed to solve the test at a comfortable pace, giving themselves sufficient time, but without delaying moving on unnecessarily. Participants were advised to skip gaps for which they could not immediately think of a solution, and return to these once they had filled in the rest.

A few participants misunderstood elements of the instructions (5 minutes for the whole test as opposed to each text, no option to return to a text) but this was clarified during the test. Consequently, some participants who had raced through a text and wanted to move on before having attempted all gaps were encouraged to use the time allowed. In some cases, participants were not cut off after time was up when they had got fixated on a single gap and insisted on finding the solution. In the interest of giving these participants a sense of achievement, a note (unseen by the participants) was made of the gaps unsolved at full time. If participants wanted to know the solution after time was up, the researcher obliged.

### **3.4.2.3 Verbal fluency task**

As could be seen in the overall outline of the session, the verbal fluency tasks "sat" between the larger ("real") tasks (C-test, film retelling, picture description, sentence generation task). Their purpose was given as acting either as a warm-up or a break from the more demanding task on account of their short duration, as appropriate.

The bilingual participants were presented with one stimulus per task each in L1 and L2. The controls were asked to perform this task in their L1 only and again were presented with one stimulus per task. Some of the controls performed the task twice, responding to two stimuli for each task, to match the bilingual group's sample size on each task.

Within the relevant language set, participants always started with one of the category-naming stimuli, followed by a letter, followed by one of the domains. In the case of the controls who did the task twice, the category-naming and letter-naming tasks were followed by the second category-naming task, followed by a domain task, then another letter and another domain. The purpose of this order was to keep the two types of semantic fluency tasks maximally apart, or at least allow one type to be finished before starting with the other, to avoid confusion.

For FiCA1, the category naming task, participants were asked to say all the words that belong in the category that will be printed on a card, or all "types of" the word that will be shown. As an example, "clothes" was given as a possible stimulus, and "all types of clothes", such as "shirt" and "trousers", as a possible response.

For FiCA2, the letter fluency task, participants were asked to say all words starting with that letter, and for FiCA3, word listing by domain, they were instructed to say anything at all that came to their mind in relation to that stimulus. With regard to the latter, it was pointed out that they could simply list their associations which would be the faster way, but that some people preferred verbally stringing associations together as a tool for forming associative chains.

The time given for the verbalisations was 1 minute, and participants were warned that this may seem like quite a long time. The stimuli were printed on card and kept covered by the researcher, and once full instructions had been given, participants then either drew or were given one of the cards. In order to get a definitive start point, participants were instructed to say the word on the card as soon as they saw it, and then start to verbalise immediately. The verbalisations were timed

from one second after they said the stimulus word. The researcher stopped the time allowing for some spill over so as not to cut participants off mid-word, having taken a note of the last word/phrase uttered within the one-minute interval. The verbalisations were recorded and subsequently transcribed.

#### **3.4.2.4 Film retelling**

The film retelling task was the second major task after the C-test and was described as a "fun" task during which the participants would have a chance to relax. Two extracts of roughly 4.30 mins from Charlie Chaplin's "Modern Times" each were shown on the researcher's laptop; the bilingual participants saw and retold one in each language set in the original order. Thus, if they started with German, they retold the first section in German, and the second in English, and vice versa.

The controls on the other hand retold both extracts in their L1. This was necessary since we needed comparative narratives for both extracts and felt that showing and asking participants to retell the second extract without their having seen the first one would change the narratives too much.

Procedurally, all three groups thus did two film retellings, even if one group (the bilingual group) changed languages between them, meaning that all three groups were subject to the same training effect during the second extract. A further advantage of this procedure was that we were thus able to elicit a sufficient number of narratives from the controls to even out the numerical difference between controls and bilingual participants.

Before the first extract, participants were given some background to the film extract to be watched, including the director, title, time and place of the setting and the action leading up to the point where the extract starts. They were then invited to watch and enjoy and be prepared to retell story afterwards. As explained above, the researcher was present during the presentation of the extracts in order to establish common ground. This also allowed the retelling to take place immediately after viewing. Participants understood without necessarily being prompted that there would be no talking while watching the extracts. Anecdotally, some people used this time to prepare their retellings linguistically, as gleaned from occasional comments, such as "What's that in German?!", which were acknowledged non-verbally by the researcher.

The retellings were narrated to the researcher who adopted the role of an empathetic listener, but with participants being cognisant that other parties would be listening to the retellings. At first, participant created a complete narrative of the extract they had seen with minimal interaction with the researcher (monologue). Occasionally a participant required or called upon the researcher for prompting. This was followed with questions (dialogue), aiming firstly at eliciting missing or clarifying misunderstood essential parts, and secondly at eliciting further information as described above in Section "Language tests" (cf. Appendix for the list of follow-up questions and instructions).

#### **3.4.2.5 Linguistic aptitude test**

The test was introduced as a task to learn a novel (artificial) language called "Novish". The subtests were administered in the order LAT D (aural memory for unfamiliar sound sequences), LAT B (lexical memory), and LAT C (grammatical inferencing skills), and were described as tasks to remember words by sound, learn pairs of words and work out how the language works respectively (cf. Appendix for written LAT instructions). In part 1 of the sociolinguistic questionnaire, participants had been asked whether they had enjoyed learning languages in the past, and which aspects they had found easier or harder to master. These answers were used as a guide for reassuring participants in relation to one or the other of the subtests.

The LAT test is designed for administration by computer; test items are either presented aurally through headphones and/or in written form on screen, requiring test takers to select their responses by mouse clicks. For all subtest, participants were talked through the interface and how the information would be presented. In the case of LAT D, the researcher stayed around until after participants had responded to the first stimulus, having explained beforehand that participants would do the actual tests without her being present.

For LAT B, the principle of the word pairs being presented to be learned and subsequently tested was explained. Potential snags which might influence the test outcome were also pointed out. For example, participants might not notice that a new stimulus had been presented after clicking an answer, since the answer click is not actually acknowledged by the programme but simply serves to switch to the next stimulus. Again, the researcher waited for the first test screen to appear before

leaving the room, to ensure participants were able to do the task without getting confused about the manner of presentation or testing.

Finally, with regard to LAT C, the researcher first demonstrated the vocabulary look-up, then gave participants a little time to familiarise themselves with the vocabulary and to verbalise anything they discovered – these verbalisations were accepted without further comment – and then explained the procedure of deciding between two test sentences, first during a learning phase, and later in an application phase. Again, the researcher observed the first decision, and then left participants to it, who alerted her when the tests were completed.





## **4 Results: Language attrition/maintenance - between-group study German**

In this and the following two chapters, we present and discuss the data elicited with reference to the 4 sub- or research questions identified in the Introduction, which we reproduce here for ease of reference:

1. What is the extent of L1 attrition/maintenance amongst the group of adult bilinguals investigated?
2. What is the degree of their L2 attainment?
3. What is the relationship between the level of L1 and L2 proficiency for these speakers?
4. What factors impact on that relationship?

Chapters 4 and 5 present the results of the bilingual group on three of the language tasks – C-test, verbal fluency task FiCA, and film retelling – and compare them with those of the two control groups (inter-group comparison). Chapter 4 discusses the German results; Chapter 5, the English ones (research questions 1 and 2). The results for each test are presented separately, and the emerging picture is summarised at the end of each section. In the analysis, we use a mixture of quantitative and qualitative analyses to arrive, wherever possible, at a comprehensive assessment of performance.

In Chapter 6, we compare each bilingual participant's performance across the two languages (research question 3). This intra-group analysis results in bilingual profiles which are further explored by analysing the remaining instruments, the questionnaires and the linguistic aptitude tests, in order to identify the role played by a number of candidate factors, particularly language use vs. linguistic aptitude (research question 4).

In this chapter, the comparison between the two groups of German native speakers will allow us to gauge whether and to what extent the bilingual participants have managed to maintain their L1, or whether they can be said to display signs of L1 attrition. L1 attrition was defined above as a decline in an individual's L1 proficiency.

Given the cross-sectional nature of this study, we are unable to state for each individual their actual level of L1 proficiency prior to emigration, although we may assume that all participants had a fully established, mature L1 system at time of emigration since all emigrated well past puberty. To

overcome this problem, we compare the bilingual participants' performance with that of the matched control group, and operationalise L1 attrition as performance by any bilingual participant that falls outside the native-speaker range as established by the German control group (at the relevant end). At the group level, the presence or absence of significant differences in the group means will serve the same purpose.

De Bot and Hulsen (2002) have characterised how an attriter's performance may be affected:

'In attrition, there may be less attention left for monitoring because production is effortful, the internal system that serves as a basis for the comparison on output and intention may have deteriorated, there may be more errors, and the speaker may need more time and may use specific strategies to gain time.' (de Bot and Hulsen 2002: 263)

With this in mind, we specifically put forward the hypotheses that the bilingual group will

- a) achieve lower scores in and/or take longer to complete the C-test,
- b) achieve lower scores in the verbal fluency tasks,
- c) demonstrate lower lexical diversity in the verbal fluency tasks/the film retelling and
- d) display a higher rate of errors, code-switches and hesitation phenomena in the film retelling compared to the German control group.

Since part of the concern of this thesis is to relate the outcomes of the processes of L1 attrition and L2 acquisition to one another, we will later test the same hypotheses against the L2 data.

## **4.1 C-test German**

### **4.1.1 Test results**

The C-test contained 100 gaps across five texts, and the maximum score obtainable was 100 (one mark for each correct answer). The scoring scheme shown in Table 14 was used following the recommendations formulated within the Attrition network (Schmid 2004: 12):

In line with the Network's procedure, only categories 0-5 were counted as errors and awarded no mark, whereas suitable variants of the target word were accepted as correct, and spelling errors

were not taken into account. The results were compiled in an Excel spreadsheet and recomputed into a score ex 100. As with the remainder of the language tests, results were checked by suitably qualified native speakers of each language. The time taken to complete the test had been recorded during the session and was taken account of in the analysis.

---

0	empty
1	incorrect lexical stem and incorrect word class
2	incorrect lexical stem but correct word class
3	correct lexical stem but incorrect word class
4	correct lexical stem, correct word class, agreement error
5	all of above correct, but still slightly wrong

---

6	acceptable variant with spelling error
7	correct word with spelling error
8	acceptable variant
9	correct word

---

**Table 14: Scoring scheme for C-tests**

The C-test was taken by all three groups. Table 15 and Figure 4 summarise the scores and time taken to complete the test for each group. The comparison of the group mean scores yielded a significant difference (Kruskal-Wallis:  $H(2) = 37.86$ , 2df,  $p < .001$ ), which is due to the performance of the Irish control group. The bilingual group had a mean score of 90.5 (ex 100) and the German control of 93.6; this difference is not significant (Mann-Whitney with Bonferroni correction:  $U = 217$ ,  $p > .017$ ). In contrast, the Irish control group scored significantly lower than both the German control group ( $U = 0$ ,  $p < .0003$ ) and the bilingual group ( $U = 9$ ,  $p < .0003$ ) with a mean score of 50.9.

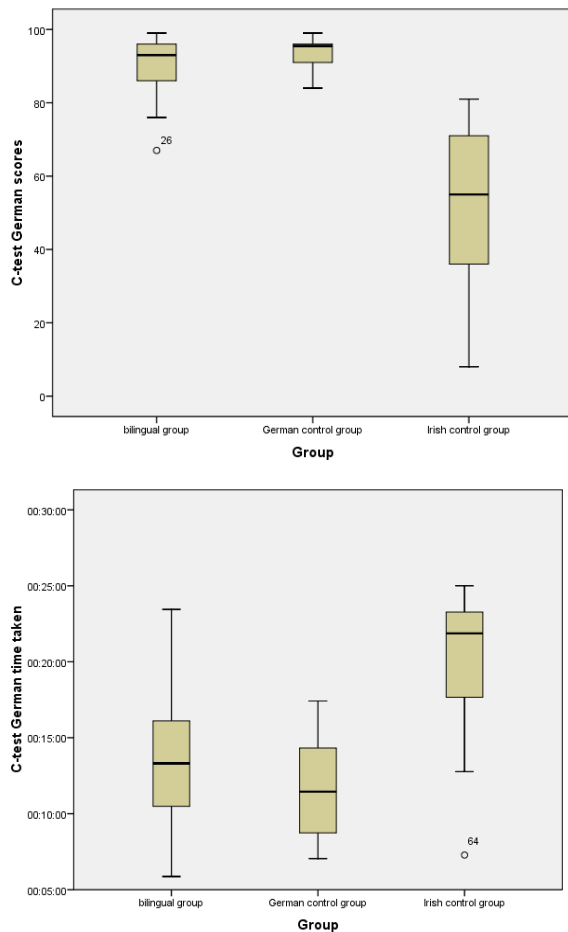
The picture is similar for the time it took participants to complete the test: the difference between the bilingual group and the German control group is not significant, whereas the Irish control group took significantly longer than the other two groups (One-Way ANOVA: Welch  $F(2, 38) = 19.12$ ,  $p < .001$ , Games-Howell).

<b>C-test German</b>	<b>Bilingual group (N = 27)</b>	<b>German controls (N = 20)</b>	<b>Irish controls (N = 18)</b>
Mean scores (ex 100)	90.5	93.6	50.9***B
Median	93	95.5	55
Range	67-99	84-99	8-81
StD	8.0	3.9	21.6
Mean time taken (min:sec)	13:39	11:34	20:06***
Median	13:19	11:27	21:52
Range	5:52-23:27	7:03-17:25	7:17-25:00
StD	4:48	3:12	17:43

**Table 15: German C-test mean scores and time taken**

\*\*\*B Results are significant at  $p .0003$  (Bonferroni corrected  $p$ ).

\*\*\* Results are significant at  $p .001$ .



**Figure 4: Box plots of German C-test scores and times for all three groups**

These results mean that we cannot detect any L1 attrition on the part of the bilingual group at the group level; both German groups achieve similar results in their L1 and complete the test in a

similar amount of time, although there is a non-significant trend for the bilingual group to score lower and to be a little slower. On the other hand, there is a marked difference between the native speakers and the L2 learners (the Irish control group) on both measures.

In the remainder of the discussion, we will mainly focus on the two groups of German native speakers, leaving aside the Irish control or L2 learner group. If we explore the contrast between the bilingual group and the German control further, we note that the bilingual group has a noticeably larger range of scores (67-99 compared to 84-99). The lowest bilingual score is an outlier (S26, Figure 4); the second lowest score is 76 (S1 and S10). In other words, the bilingual group's performance is characterised by heterogeneity, a feature usually associated with L2 learners rather than native speakers, in contrast with the more homogeneous results achieved by the German control group.

While the majority of scores of the bilingual participants is within the native-speaker range (a result unmatched by the Irish controls, none of whom achieve a native-like score in their L2), there are several individuals in the bilingual group (N = 5; 18.5%) whose scores fall outside that range (Table 16). Three of these (S1, S10 and S26) score well outside the range, the other two (S7 and S15) perform just outside.

S1, S10 and S26, the three bilingual informants with the lowest German C-test scores of all, also have test times exceeding the native-speaker range by five minutes or more, while S7 and S15 have test times within the native-speaker range, albeit at the slower end. In line with our operating definition of attrition as reduced performance outside the native-speaker range, we would have to conclude that these five bilingual individuals suffer a degree of L1 attrition, as detected on this test.

<b>C-test German</b>	<b>German controls</b>	<b>S1</b>	<b>S7</b>	<b>S10</b>	<b>S15</b>	<b>S26</b>
Mean score (Range) (ex 100)	93.6 (84-99)	<b>76</b>	<b>82</b>	<b>76</b>	<b>83</b>	<b>67</b>
Mean test time (Range) (min:sec)	11:34 (7:03-17:25)	<b>23:27</b>	15:40	<b>22:27</b>	13:22	<b>22:56</b>

**Table 16: Individual German C-test scores and test times\***

\*Results given in bold fall outside the native-speaker range.

There are a further four bilingual individuals whose test times fall outside the German native-speaker range – three perform more slowly, one faster. However, the three slower participants

(S13, S14, S25, cf. Table 17) are only about one minute slower than the maximum time used by the control group, and their scores fall well within the native-speaker range (92, 95 and 93 respectively).

<b>C-test German</b>	<b>German controls</b>	<b>S11</b>	<b>S13</b>	<b>S14</b>	<b>S25</b>
Mean score (Range) (ex 100)	93.6 (84-99)	98	92	95	93
Mean test time (Range) (min:sec)	11:34 (7:03-17:25)	<b>5:52</b>	<b>18:39</b>	<b>18:25</b>	<b>18:23</b>

**Table 17: Individual German C-Test Scores and Test Times continued\***

\*Results given in bold fall outside the native-speaker range.

Although we find a moderate negative correlation between C-test score and test time for the bilingual group ( $r_s = -.585, p = .001$ ), indicating that people who took longer to complete the test tended to have lower results, test time cannot unequivocally be linked to proficiency since time spent would appear to depend both on objective difficulty (the more problems one has, the longer one takes) and on effort (one person may take time to check his/her answers without needing to change anything; another may rush through the test and end up with a higher error rate).

Moreover, factors such as age and personality are likely to play a role. Indeed, there is also a significant correlation between the time taken on the C-test in German and in English ( $r_s = -.495, p = .009$ ), indicating that people who were slow on the test in one language were also slow in the other. Nevertheless, the three individuals with slower test times may be on the verge of L1 attrition. In contrast, bilingual participant S11 outperforms everybody else in terms of speed (though there was no requirement to complete the test as fast as possible), and does so without any negative impact on quality (at 98, her score is one of the highest). She could therefore be classed as a superior native speaker.

In summary, while the bilingual group overall has comparable scores and test times compared to the German control group, some individuals fall outside the native-speaker range on one or both C-test measures. Three participants score below the native-speaker range and also have test times well outside the native-speaker range. These would be candidates for L1 attrition as gauged by this test. Another two have lower scores with normal test times, while three more have normal scores with slower test times. These may be individuals on the verge of attrition.

### 4.1.2 Item analysis

While the incidence of L1 attrition as shown by this test is very low overall, a qualitative analysis can further pinpoint differences between the two groups, and potentially the beginnings of L1 attrition. The generally high level of scores indicates that few errors were made overall. In fact, 63 items were solved correctly by all controls, and 41 by all bilingual participants (cf. Table 18).

<b>Number of items solved correctly by % of participants</b>	<b>Bilingual group</b>	<b>German control group</b>
all participants	41	63
90-99%	24	21
80-89%	20	7
70-79%	8	3
60-69%	2	1
50-59%	2	2
40-49%	3	2
30-39%	–	1
<b>Total number of items</b>	100	100

**Table 18: Breakdown of number of correct item solutions by percentage of participants**

The successfully solved items partially overlap between the groups – 39 items did not present a problem for any participant across the two German groups, and a further 27 resulted in an error for fewer than 10% of participants in either group. 12 items were problematic for up to 15% of participants, and the remaining 22 items presented problems for 15%-65% in either group. The distribution of these problematic items across the five texts is represented in Table 19.

<b>Text 1</b>	<b>Text 2</b>	<b>Text 3</b>	<b>Text 4</b>	<b>Text 5</b>	<b>Overall</b>
4	7	4	3	4	<b>22</b>

**Table 19: Problematic items per text across both groups**

The groups do not just concur in many of the items which were found problematic, but also in the reasons for which these items were found problematic, and the types of solutions they came up with. However, a Wilcoxon Signed-Rank Test reveals significant differences between the two

groups concerning the success with which items were solved ( $z = -4.084, p = .000, r = -.29$ ). There were 38 ties, i.e. both groups were experiencing similar levels of difficulties on these items; 15 items on which the bilingual participants outperformed the German control group, and 47 items on which the bilingual participants performed worse than the control group.

In other words, despite the fact that the overall scores for the C-test did not reveal any significant differences between the groups, there is a trend for the bilingual participants to have more problems on the individual items than the German controls. Further, several items presented a problem for the bilingual group only, and generally the bilingual participants tended to propose a wider range of solutions than the control group. This trend holds for all error types (lexico-semantic and morpho-syntactic).

For the purpose of discussing the most interesting items, Table 20 below sets out the 22 problematic items mentioned above, plus three additional ones. The texts of the German C-test are reproduced in the Appendix, along with a translation of the intended meaning of the texts, to provide the context for the discussion. The relevant items are put in bold face and are indexed with the number from the table below.

<b>Item</b>	<b>Error type*</b>	<b>Bilingual group</b> % (Number) correct ex 27	<b>**</b>	<b>German controls</b> % (Number) correct ex 20
<b>Text 1</b>				
1) + 2) <i>aus Teilchen aufgebaut</i> ("made up of particles")	LEX	48% (13)	=	50% (10)
	LEX	48% (13)	=	40% (8)
3) <i>tabellarische Aufstellung</i> ("tabular record")	LEX/GR	74% (20)	<	100% (20)
4) <i>Periodensystem</i> ("periodic table")	LEX	59% (16)	<	80% (16)
<b>Text 2</b>				
5) <i>weicht ... ab</i> ("differs")	LEX	85% (23)	>	70% (14)
6) <i>Version</i> ("version")	LEX	81% (22)	=	85% (17)
7) <i>das</i> ("which")	LEX/GR	78% (21)	<	95% (19)
8) <i>also</i> ("i.e.")	LEX	48% (13)	=	55% (11)
9) <i>zusammenkommen</i> ("get together")	LEX	70% (19)	<	85% (17)
10) <i>über</i> ("about")	LEX	81% (22)	=	85% (17)
11) <i>rätselhaftes</i> ("mysterious")	LEX/GR	85% (23)	<	95% (19)



Item	Error type*	Bilingual group % (Number) correct ex 27	**	German controls % (Number) correct ex 20
<b>Text 3</b>				
12) <i>verwendete man</i> ("were used")	GR	70% (19)	<	100% (20)
13) <i>unterschiedlichen</i> ("different")	GR	89% (24)	<	100% (20)
14) <i>beansprucht</i> ("claims")	LEX/GR	89% (24)	>	75% (15)
15) <i>Hilfsmittel</i> ("instrument/means")	LEX/GR	74% (20)	<	95% (19)
16) <i>Wasseradern</i> ("underground watercourses")	LEX/GR	70% (19)	=	75% (15)
<b>Text 4</b>				
17) <i>Fachkräften</i> ("qualified/trained personnel; specialists")	LEX	67% (18)	=	60% (12)
18) <i>unsachgemäÙe</i> ("inappropriate")	LEX	74% (20)	<	100% (20)
19) <i>zweckentfremdet</i> ("used inappropriately")	LEX/GR	81% (22)	<	100% (20)
20) <i>Schäden</i> ("damage")	GR	85% (23)	<	100% (20)
<b>Text 5</b>				
21) <i>Kulturen</i> ("civilisations/cultures")	GR	89% (24)	<	100% (20)
22) <i>aus</i> ("from; derive ~")	LEX	63% (17)	>	35% (7)
23) <i>oft</i> ("often")	LEX	52% (14)	=	45% (9)
24) <i>bestimmende</i> ("determining")	GR	81% (22)	=	90% (18)
25) <i>verloren</i> ("lost")	LEX/GR	70% (19)	<	80% (16)

**Table 20: Selected items from German C-test for detailed analysis**

\* LEX = lexico-semantic error, GR = morpho-syntactic error

\*\* < GC outperforming BG, = tie, > BG outperforming GC

**Items 1, 5, 8, 9, 14, 17 and 22-25** are examples of both groups encountering similar levels and types of problems and coming up with similar, if not identical solutions. The table also illustrates that several of the items that created no or minimal problems for the control group were more problematic for the bilingual group (**Items 3, 4, 7, 11-13, 15, 18-21**). Some of these items will be discussed below.

**Items 1 and 2**, *aus (Teilchen) aufgebaut* "made up from (particles)" presented a problem for both groups, possibly because of the interdependence of the gaps. Participants produced a number of verbs which, however, require a different preposition for their object than one starting with *a-*, besides being factually incorrect. While some of the more plausible solutions are shared between

both groups (*aufgespalten* "split up", *aufgeteilt* "divided" and *aufgeladen* "charged"), four bilingual participants' solutions do not result in a meaningful clause (*aufgeführt\** (*aufgeführt?*) "listed"?, *aufgelistet* "listed", *aufgesetzt* "put on"?). This may reflect retrieval difficulties for contextually appropriate words in the bilingual group.

**Item 4**, *Periodensystem* "periodic table", a compound word missing its second part, also presented a problem for both groups. The target item is strictly speaking a technical term, but one that is known generally, even by people who have never studied chemistry. Treating this word as non-technical is supported by the fact that where participants were told the solution after completing the test, they invariably indicated recognition. Two of the three controls who could not think of the word in question provided a plausible and grammatically correct solution (*periodische System* "periodic system", *Periodenbuch* "periodic book"). Conversely, ten of the eleven bilinguals for whom this item posed a problem left the gap incomplete, indicating retrieval failure. Three of these filled in the bridging *-n* necessary for forming a compound, and several participants commented that they could not remember the word in question.

Retrieval difficulties, and potentially changed underlying L1 knowledge (merged concepts), are also evident in the case of **Item 18**, *unsachgemäße* "inappropriate". This gap was correctly filled in by all controls, but six participants from the bilingual group used an adjective with the same stem but a different meaning: *unsachlich* "subjective". The remaining informant used a different, incorrect derivation of the target stem: *unsachmäßige\**, from *-maß* "measure" which is both semantically and phonologically similar to the target. The original meaning of *-mäßig* "~ in (x) measure" can usually not be isolated from a compound's meaning, e.g. *regelmäßig* "regular", *zahlmäßig* "numerous", *übermäßig* "excessive", *verhältnismäßig* "relative", *anteilmäßig* "proportionate". *-gemäß*, on the other hand, does retain its basic meaning, "corresponding to", in compounds, e.g. *wahrheitsgemäß*, *erfahrungsgemäß*, *standesgemäß*, *erwartungsgemäß*, i.e. "according to the truth", "experience", "status" and "expectations", respectively.

**Item 19**, *zweckentfremdet* "(used) inappropriately/for a different purpose than it was meant for", again caused no difficulties for the control group. Two bilingual participants, however, formed a non-existing, partially ungrammatical lexeme: *zweckenteilig\** "as per purpose"? (which is also semantically wrong) and *zweckenunmäßig\** "inappropriate for the purpose at hand"? In both cases,

participants apparently did not notice the *-en* after *zweck-* or ignored it in their processing of a possible solution. Another participant used the present tense participle instead of the past tense participle of the target word (*zweckentfremdend*), whereas in *zweckentfremd\** the participle ending is missing altogether resulting in an ungrammatical solution. The final solution is grammatical, but logically contradictory: *zweckentsprechend* "according to/fit for the purpose". This solution points to insufficient monitoring of the wider context.

Two of the proposed solutions by bilingual participants for **Item 15**, *Hilfsmittel* "means" (literally "means of help"), similarly violate the semantic constraints of the text: *Hilfswerk* is "a charitable organisation" and *Hilfsleistung* "an act of help". The latter solution also contains the unusual compound bridge *-s* instead of *-e* (*Hilfeleistung*). *Hilfsleistung* is a possible, but depreciated use of the word which sounds more concrete and less idiomatic than *Hilfeleistung*; it is not listed in current editions of either Duden or Wahrig (Wahrig, Hermann et al. 1986; Duden-Redaktion n/a), two authoritative reference works on the German language. *Hilfswerk* similarly seems to be a result of using *-werk* "act" in its direct, rather than idiomatic meaning ("organisation that acts as a means") as part of the compound.

Conversely, two further solutions – *Hilfsinstrument* "additional means" and *Hilfsgerät* "supplemental tool" are concrete nouns in standard German use only – both refer to devices used in lieu of other tools or in order to facilitate a particular, usually complex, process, whereas in this context the text simply requires a "tool" or "means". Incidentally, "means" may be rendered as "instrument" in an abstract sense in English, and it would appear that the solution *Hilfsinstrument* might be the result of semantic transfer. The final solution proposed by the bilingual group was the target word in the wrong case and/or an incorrectly formed plural: *Hilfsmitteln* (should be acc/sg; is dat/pl). The one error in the control group resulted from a semantically inappropriate circumlocution (*Hilfsperson und -materialien* "helper and auxiliary materials").

**Item 6**, *Version* "version", **Item 11**, *rätselhaftes* "mysterious", and **Item 20**, *Schäden* "damage", are interesting from a lexical viewpoint but the latter two are also examples of problems not being restricted to the lexical realm, already seen with Item 15. For **Item 6**, both groups produced several acceptable alternatives; however, the bilingual group additionally proposed several less plausible words. One of these, *Verstehen* "understanding", is semantically related to one of the acceptable

variants (*Verständnis* "understanding"), but denotes the process of understanding rather than its result, such as having understanding, having a concept of something. As a noun, it is rather uncommon. This solution may point to incomplete retrieval or difficulty in mapping concepts to lexemes, and/or to crosslinguistic influence from L2 possessing one word for two concepts, maybe as a result of inter-lingual translation. **Item 11** produced two gaps (compared to one in the control group), one case/gender error (*rätselhaften* = masc acc rather than neuter acc; possibly the gender of *Phänomen* "phenomenon" was unavailable), and one semantically related but divergent solution: *rätselndes* "guessing".

**Item 20** similarly produced a word derived from the same stem in German that does not quite fit the bill and, like Item 6, may have resulted from a crosslinguistically induced concept merger: *Schädigungen*, like the target item *Schäden*, translates as "damage", but the latter is the general term, whereas the former is used for neurological, genetic or tissue damage, or for the process of being damaged or exerting damage. The other three solutions all used the target item in the singular, *Schaden*, which leads to an agreement error with the preceding adjective and is interesting against the fact that the English equivalent "damage" for this sense is used in the singular. (Indeed, eight of the twelve Irish controls who did not leave a gap, produced *Schaden* in the singular.) The same phenomenon holds for **Item 21**, *Kulturen* "civilisations/cultures", with three bilingual participants using the target item in the singular, *Kultur*, although in this case the English equivalent does suggest the use of the plural. (However, again 11 of 18 Irish controls also produced *Kultur* in the singular.)

Crosslinguistic influence further appears to have played a part in deciding the tense for **Item 12**, *verwendete* "used": While pragmatically the German context requires the past tense, which was adhered to by all German controls, eight bilingual participants used the present tense. For comparison, 12 out of 16 Irish controls who filled in a solution, also did.

Syntactic differences and possibly retrieval difficulties are also evident in the solutions by seven bilingual participants for **Item 3** (*tabellarische Aufstellung* "table/tabular record") (all controls solved this gap correctly). The bilingual participants produced six different solutions between them, four of which do not match the "one word criterion", i.e. the requirement to complete the gap with just one word which is distinct from the following word (i.e. cannot form a compound with

the next word): *tabelle (von)* "table (of)", *tabellen* "tables-comp", which causes orthographic and syntactic difficulties. The other three solutions are one incorrect adjectival derivation: *tabellare\** "table-comp", one case/agreement error: *tabellarischen* (gen/dat sg instead of nom/sg), and one gap left incomplete.

Agreement errors are also evident in **Item 13**, *unterschiedlichen* "different-pl/dat", solved correctly by all controls. Three bilingual participants produced *unterschiedliche* (several possible gender/case combinations) or *unterschiedlicher*, possibly f/dat indicating that problems are not confined to the retrieval of concepts and lexical parts of words.

In summary, although only few bilingual participants were shown to display signs of L1 attrition based on quantitative measures, we were able to glean more evidence for differences in the L1 proficiencies of both groups on the basis of the analysis of specific items solutions. As one might expect in an attrition situation, many of the proposed solutions point to either failed, incorrect or partial retrieval of lexical items, possibly combined with a change of conceptual representations. However, others highlight difficulties at the morpho-syntactic and pragmatic levels, such as in the application of case and gender endings, and the choice of the correct word class, the appropriate participle or tense, which contradicts notions that attrition is solely lexical. Crosslinguistic influence from English seems to play some part, but cannot explain all of the divergent solutions, which may simply be due to inaccessibility or insecurity about the correct/appropriate solution following lack of practice/exposure.

## **4.2 Verbal fluency tasks German**

### **4.2.1 Test results**

Before comparing the bilingual group's scores to the German control group, we tested for significant differences between the conditions within each task on the basis of the control group scores. While the conditions in FiCA1 and FiCA3 are comparable, there are significant differences ( $F(2, 26) = 5.36, p < .05$ ) among the three conditions on FiCA2, the letter fluency task, owing to the difference between letters *F* and *S*. As we discussed in Chapter 3, these differences are probably related to absolute category size/productivity.

Consequently, we give results for each stimulus separately (cf. Table 21). Please note that the number of participants/scores differs from that on the C-tests due to missing data in the bilingual group, and owing to administering more than one stimulus on each task to some controls, in order to match numbers on each condition more closely. The scores are the total number of words named within one minute, after exclusion of repetitions, non-category errors, code switches and comments, but without applying any weighting to different kinds of words.

FiCA German	Stimulus	Bilingual group					German control group				
		N	Mean	Me- dian	Range	StD	N	Mean	Me- dian	Range	StD
FiCA1 – category naming	<i>Professions</i>	17	19.7	20	13-27	4.2	14	19.9	20	13-28	4.1
	<i>Furniture</i>	8	18.1	18	14-26	4.0	15	18.6	19	14-22	2.6
FiCA2 – letter fluency	<i>F</i>	6	14.7	14.5	7-21	5.4	7	15.6	17	6-22	5.4
	<i>A</i>	11	15.1	16	8-26	4.9	10	18	17	8-22	6.2
	<i>S</i>	9	20.9	21	13-30	5.2	12	22.3	22	19-29	2.5
	<i>School</i>	7	25.7	25	15-36	6.4	8	26.1	26	18-33	5.4
FiCA3 – word listing by domain	<i>Kitchen</i>	8	23.5	22.5	19-33	4.7	9	25.7	27	18-35	5.1
	<i>Street</i>	7	18.9	17	14-26	4.7	8	22.8	22	17-29	4.0
	<i>Church</i>	4	18.8	20.5	11-23	5.3	4	19.3	20	16-21	2.4
<b>Scaled total score</b>		<b>25</b>	<b>63.8</b>	<b>64.34</b>	<b>46-93</b>	<b>12.1</b>	<b>29</b>	<b>68.22</b>	<b>67.81</b>	<b>46-90</b>	<b>10.6</b>

**Table 21: Mean scores for bilingual group and German control for all German FiCA conditions**

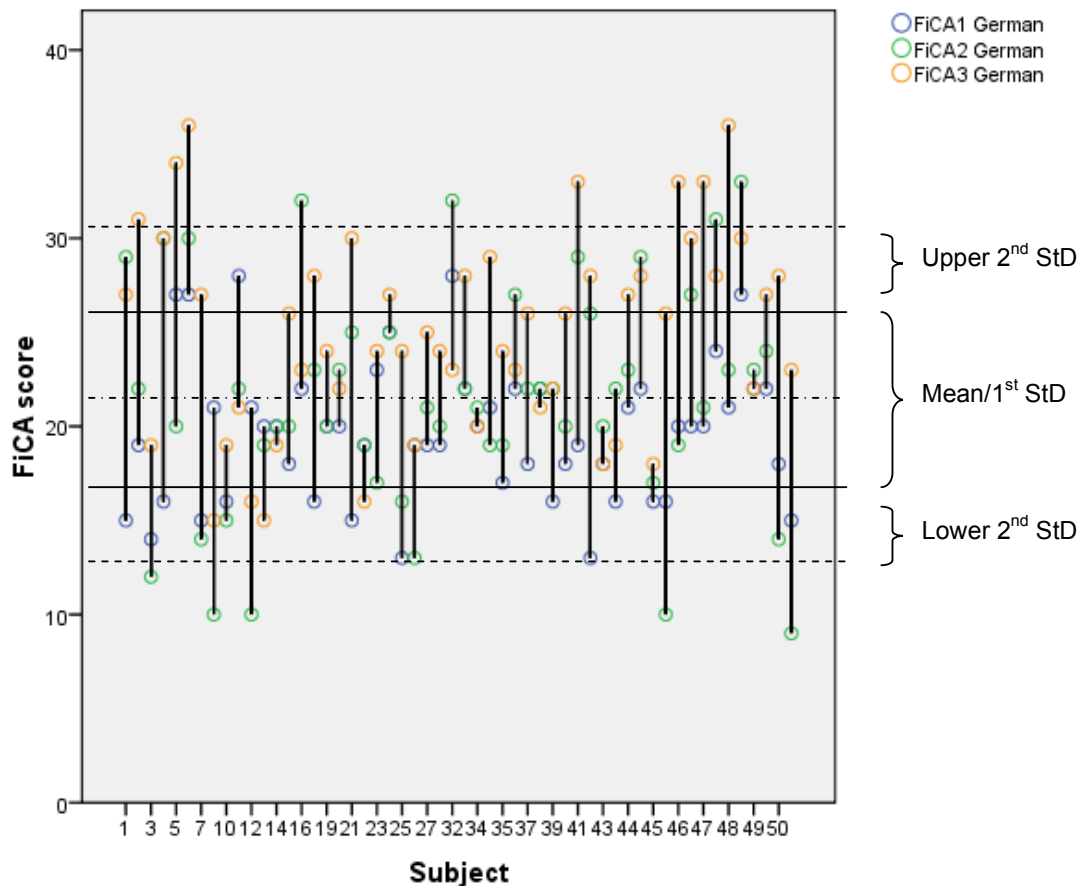
On the basis of the inter-group comparison for each stimulus, we did not find any significant differences between the bilingual and the German control group (one-tailed *t*-test: FiCA1 *Profession*:  $t(29) = -0.14, p = .445$ ; FiCA1 *Furniture*:  $t(21) = -0.35, p = .366$ ; FiCA2 Letter *F*:  $t(11) = -0.30, p = .385$ ; FiCA2 Letter *A*:  $t(19) = -0.97, p = .172$ ; FiCA2 Letter *S*:  $U = 44.5, p = .257$ ; FiCA3 *School*:  $t(13) = -0.14, p = .448$ ; FiCA3 *Kitchen*:  $t(15) = -0.91, p = .190$ ; FiCA3 *Street*:  $t(13) = -1.74, p = .053$ ; FiCA3 *Church*:  $U = 7, p = .457$ ). However, we observe as a trend that all averages for the bilingual group are a little lower than those of the control group. The range of scores is generally quite similar, with the exception of three FiCA3 conditions (*School*, *Street* and

*Church*), as well as Letter *S* (FiCA2), apparently presenting greater difficulty for some bilingual participants.

The absence of significant effects in some conditions, particularly in FiCA3, may be a result of the relatively small numbers participating in these conditions. *Street* almost reaches significance, which may mean that the groups are performing in a less similar fashion than we have been able to show statistically. The lower averages at similar score ranges generally might indicate that relatively more bilinguals than controls are at the lower end of the fluency spectrum, which we investigate below. Of interest also is whether those bilinguals that do fall below the control group's range are the same or different individuals in each case, in order to establish whether some individuals have low verbal scores relative to the rest of the group throughout.

In view of the fact that some of the conditions within each verbal fluency task were not directly comparable to one another, it is necessary to standardise scores across the three tasks in order to arrive at a more global assessment of participants' performance and in order to calculate summary scores. Without this step, a person may erroneously be considered to have reduced verbal fluency when this effect was in fact due to a combination of the conditions which elicited fewer verbalisations on each task. We therefore scaled the scores within each task and language to the level of the most productive condition. The scaling factor was determined on the basis of the relevant control group's means for each condition.

Figure 5 graphically represents the scores achieved by each individual across all three tasks after scaling. The figure confirms the impression that the two groups largely overlap in their performance, and that there is considerable variation amongst individuals, with some individuals in both groups having at least one very low score, and others achieving very high scores.



**Figure 5: Scores across three FiCA tasks for bilinguals (cases 1-27, N = 26) and German controls (cases 28+, N = 29)**

After computing the mean value and associated mean standard deviation for the control group across the three tests (cf. Table 22), we find that there are more individuals in the bilingual group than amongst the controls whose lowest score falls below the 2<sup>nd</sup> or the 1<sup>st</sup> standard deviation (StD) (a total of 14 bilinguals (54%), as opposed to 11 controls (38%)), and fewer bilinguals whose highest score falls outside the higher 2<sup>nd</sup> StD or above the 1<sup>st</sup> StD (a total of 9 bilinguals (36%), compared to 13 controls (45%)). The number of bilinguals with consistently high scores (values around mean and above) is lower compared to the controls (10 bilinguals (38%) vs. 16 controls (55%); this includes one very fluent participant in either group with all scores in the upper 2<sup>nd</sup> StD or above), while that with consistently low scores (below average) exceeds that of the controls (8 bilinguals (31%); 3 controls (10%)). There are two bilinguals, and no controls, who have two very high scores, but quite a low third score.

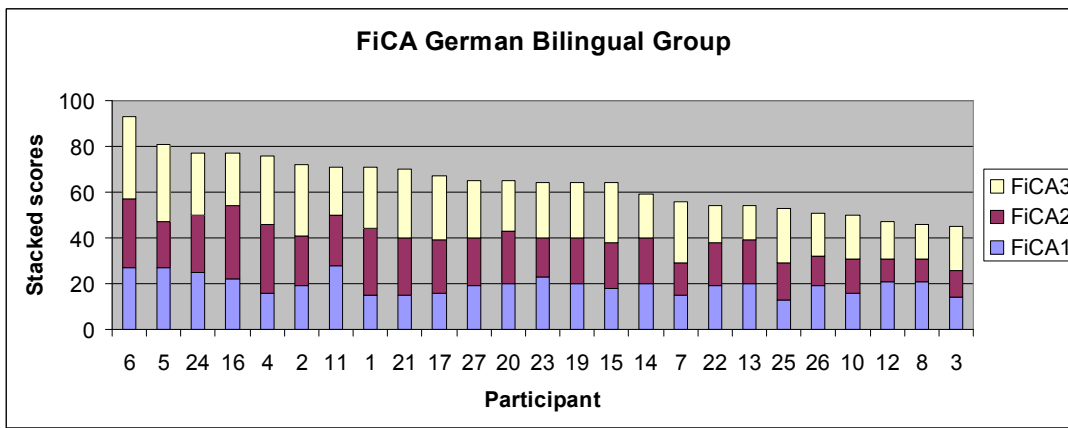


<b>FiCA German</b>	<b>Bilingual group</b>	<b>German control</b>
Number	26	29
Average score across three tests	n/a	22.7
Range of 1 <sup>st</sup> StD	n/a	27.3-18.1
Range of higher 2 <sup>nd</sup> StD	n/a	27.4-31.9
Range of lower 2 <sup>nd</sup> StD	n/a	13.5-18.00
Number with lowest value below lower 2 <sup>nd</sup> StD	5 (19%)	3 (10%)
Number with lowest value below lower 1st StD	9 (35%)	8 (28%)
Consistently low scores	8 (31%)	3 (10%)
Number with highest value above upper 2 <sup>nd</sup> StD	3 (12%)	6 (21%)
Number with highest value above 1st StD	6 (24%)	7 (24%)
Consistently high scores	10 (38%)	16 (55%)

**Table 22: Comparison of participant scores relative to control group's mean across three tasks**

A further interesting difference can be observed amongst those individuals whose lowest score falls below the 2nd StD – the highest score of four of the five bilinguals in that category is low average (19, 21, 19 and 19; the 5th score is above average at 24), whereas the three controls' highest scores are above average (28, 27 and 23), indicating that the controls' lower performance on one FiCA test may have been a "blip" (like for the fifth bilingual participant, whose highest score is 24).

Figure 6 combines the scaled scores for the bilingual group for all three tests, and presents them in descending order by overall score. Participants S3, S8, S12, S25 and S26 are the five individuals whose lowest score is below the 2nd StD. S10, S13 and S22 are individuals whose scores are all relatively low. There is only one control, S45, who matches that profile. The bilingual individual with the highest scores throughout is S6. S24 has high scores across all three tasks, while S5 has one, and S16 has two average scores. Three controls, S32, S47 and S48, who has the highest score overall, equally have very high scores throughout.



**Figure 6: Scores across three FiCA tasks for bilingual group (descending order)**

A *t*-test on the overall scaled scores (three individual scores summed up) of bilingual and German control groups (cf. Table 21) yielded a non-significant result ( $t(52) = -1.43, p = .080$  one-tailed). Further, on the basis of the range of aggregate scores we do not find any individuals who fell below the native-speaker range and could therefore be classed as an attriter. In summary, we can state that performance for both groups is similar, despite a trend for the bilingual group to verbalise somewhat less, and the presence of a greater number of less fluent bilingual participants, some of whom have low scores throughout.

#### 4.2.2 Item analysis

In this section, we concentrate on the qualitative analysis of FiCA1, the category naming task (stimulus *Professions*), only, since the intrinsic differences between the stimuli on FiCA2, the letter fluency task, and the small group sizes on some conditions of FiCA3, the WLD task, make this type of comparison difficult and possibly not very meaningful.

Different approaches have been used to take the analysis of verbal fluency tasks beyond merely recording the number of correct words produced. For example, the time course of the verbalisations, the number of errors and repetitions, the number of comments, the common content between two sets of data, the number and length of clusters of words in semantic or phonetic associations as well as the number of switches between categories have been investigated in studies

of bilingual and monolingual populations (e.g. Joannette, Goulet et al. 1988; Roberts and Le Dorze 1997; Troyer, Moscovitch et al. 1997).

Adapting some of these procedures, we first analysed the number of types and tokens, and of common contents between the verbalisations of the bilingual and the control group (cf. Table 23). This comparison shows that the number of types between the two groups is nearly identical for the category *Professions*, while, presumably owing to the smaller number of participants in the control group, the number of tokens is lower for the German control group. Not included in the type count are morphological derivations of the standard occupations like feminine or plural forms (*Arzt – Ärztin – Ärzte*, "doctor", "woman doctor", "doctors") which were employed infrequently and in a similar manner by both groups. Almost half of the types (about 42%) are shared between the groups. Further, the overlap mostly concerns the more frequently named items, which could thus be seen as the more prototypical occupations.

FiCA1 German	Bilingual group	German control
Category	<i>Professions</i>	
Number	17	14
Types	151	144
Tokens	328	267
Type-token ratio	0.46	0.54
Common contents	41%	43%
Basic occupations (N tokens (%))	297 (90.5%)*	209 (78.3%)*

**Table 23: Item analysis of FiCA1 *Professions* German**

\*\*\* The difference is significant at  $p = .000$ .

For example, all of the items that are named amongst the top 10 items by the German control group, and all but two of the top 25 items are also mentioned by the bilingual group. (The top 10 items were named by at least six bilingual participants or five controls, whereas the top 25 were named by at least four or three participants respectively.) Five of the controls' top 10, namely *teacher*, *doctor*, *cabinet maker/joiner*, *salesman* and *nurse*, are also amongst the top 10 for the bilingual group. Another three items – *painter*, *baker* and *professor* – occur in the bilingual group's

top 25, whereas the remaining two – *nursery school teacher* and *secretary* – are less frequently named by the bilingual participants.

Four of the items of the control group's top 25 (from 11-25) – *engineer, plumber, bricklayer* and *farmer* – are also amongst the bilingual's 11-25<sup>th</sup> most frequent items, whereas the remainder – *architect, butcher, carer for the elderly, construction worker, tiler, mechanic, physiotherapist, solicitor* and *fitter* – occur further down the list in the bilingual group. *Practice nurse* and *office worker* are not mentioned by the bilinguals.

On the other hand, the bilingual group have five items among their top 10 which either are much less frequent in the controls' list – *gardener* and *tradesman* – or do not get named at all – *cook, translator/interpreter* and *bus/coach driver*. The bilingual group's top 25 additionally contain *roofer, electrician, carpenter, hairdresser* and *policeman*, which are mentioned less frequently by the controls, and *tailor, guide* and *glazier*, which do not get mentioned by the controls at all.

This comparison reveals some trends for both groups. Trades and labels for self-employed people such as *businessman* feature more prominently in the bilingual group while recognized white-collar occupations requiring formal training have a greater share of the control group's verbalisations. This is perhaps due to differences in the social context our participants find themselves in – most of the German controls live in towns and work in white-collar jobs, whereas more than a third of the bilingual participants live in rural parts of Ireland.

A further reason may lie in the fact that Germany as a country is a lot more urbanised than Ireland, people tend to be employed rather than self-employed, and great emphasis is placed on appropriate training. Ireland, in contrast, has a less stringent qualifications system for trades than does Germany, so in principle anybody can take up a new trade at any stage. Judging by the evidence from the questionnaires, traditional trades were in fact the occupations that many of the bilingual participants practised when they first came to Ireland (some still do), often as a result of a conscious decision to leave the more hierarchically structured and less flexible German market behind. Regarding more recent job opportunities resulting from the economic boom, both *computer expert* and *software engineer* are mentioned by the bilingual group (one each), and not by the controls.

Of course, one would expect the verbalisations to reflect participants' experience and profile. In this context, it is interesting to note that *translator/interpreter* has rank 7 in the bilingual group (named by seven participants), whereas it is not mentioned by the control group. This would indicate that more of the bilingual participants have been involved in language related work, although roughly equal numbers have some foreign language or Germanic studies-related training in both groups (nine bilinguals and eight controls, or 33%/40% respectively; of these, three bilinguals and one control have specific translator/interpreter training). Similarly, *tour guide* ranks 16.5 (five mentions, plus a specialist *language tour guide*) and again is not named by the controls. These, however, list *travel agent* – a reflection of the reality of Ireland as a tourist destination.

Related to the more rural nature of Ireland, and a higher percentage of bilinguals living in rural areas (nine (33%) vs. two (10%)) is probably the fact that agricultural occupations feature more highly in the bilingual group: these include *farmer* (five mentions by bilinguals, three by controls), *vet* (two vs. one mention), *veterinary nurse* (one mention by both groups), *shepherd* (one mention vs. not mentioned by the controls), several *breeders* of different animals (three vs. none) and finally *gardener* which features as highly as rank 4 (nine mentions) while being only mentioned by one control. Several food and/or tourism-related occupations also occur – *cook* (rank 5, eight mentions), *wine grower*, *pub owner*, *hotel manager* (one mention each), without a match in the control group.

Not as obvious are the reasons for why the bilinguals specifically list various kinds of drivers *bus/coach driver* (rank 9.5, six mentions), *taxi driver* (three mentions), *train driver* (two mentions), *lorry driver*, *tram driver*, *ambulance driver* (one mention each), whereas the controls just have a generic *driver* and a *lorry driver* (one mention each).

The controls, on the other hand, list many more occupations which reflect the hierarchical nature of the German market, and the requirement for specialist training. For example, where the bilingual group simply lists *Kaufmann* "trader/businessman", the controls specify specialisations: *Bürokaufmann* "office clerk" (two mentions), *Industriekaufmann* "industrial clerk/manager"/*Industriekaufmann für Bürokommunikation* "~ for office communications" (two plus one mentions), *Versicherungskaufmann* "insurance salesman" (two mentions) *Reisebürokaufmann* "travel agent" (two mentions), *Bankkaufmann* "banker" (one mention) and

*Handelskaufmann/Einzelhandelskaufmann* "tradesman"/"retailer" (one mention each), as well as the generic term.

Similarly, in addition to the generic *teacher* (the most popular profession in both groups with 13 mentions each) and *doctor* (the second most popular profession – twelve and nine mentions), the controls list several sub-divisions such as the *maths* and *French teacher*, a *surgeon* and *gynaecologist*. Finally, terms that specify further qualifications in a particular trade or profession, such as various master tradesmen, and terms denoting the higher echelons of power (*manager*, *boss*, *CEO/managing director*, *head of Board of Directors*) feature more strongly in the control group's verbalisations.

These comparisons raise another issue – the extent to which the bilingual group differs from the controls on linguistic measures, such as prototypicality and lexical diversity. For the latter, we calculated the type-token ratio (TTR) for both groups. There is a small difference in TTR (0.46 for the bilingual group, 0.53 for the control group), which is unlikely to be due to the small difference in sample size, indicating that the control group does use a more diverse vocabulary.

With regard to prototypicality, it has been argued that language is simplified and marked forms are lost in attrition situations (e.g. Seliger and Vago 1991). We therefore assigned a code of either 1 "generic/basic label for occupation" or 2 "specialist occupation, requiring additional training and/or implying higher than usual levels of responsibility" to all items. A Mann-Whitney test returned a highly significant result ( $U = 38414.5$ ,  $p = .000$ ). In other words, the bilingual group does indeed display signs of a more restricted, or in this case, more prototypical/less diversified repertoire.

Additionally, there were a few instances of retrieval difficulties/errors and crosslinguistic influence in the bilingual group. For example, all the controls refer to *doctor* as *Arzt*, whereas the bilinguals variably use *Arzt* (eight times) and *Doktor* (four times). *Doktor* does exist as a word but is not the official professional label but rather an informal one, and it would not be surprising if this choice resulted from transference from English.

Further, one participant produced *Kindergartenmädchen* "child in kindergarten" instead of *Kindermädchen* "nanny" or *Kindergärtner(in)* "(female) nursery school teacher" after a lengthy retrieval process; another was trying in vain to remember *Rechtsanwalt* "solicitor" and gave up. Another participant produced an incorrect suffix on *Zoowärter* "zookeeper", saying *Zoowärtner*

instead, and one named *Bahnfahrer* "train driver" in analogy with other drivers (*Straßenbahnfahrer* "tram driver", *U-Bahnfahrer* "metro driver"), which, however, without a further designation of what type of train is being driven is usually understood as "train passenger", whereas the unambiguous term for a train driver is *Zugführer* or *Lokomotivführer*.

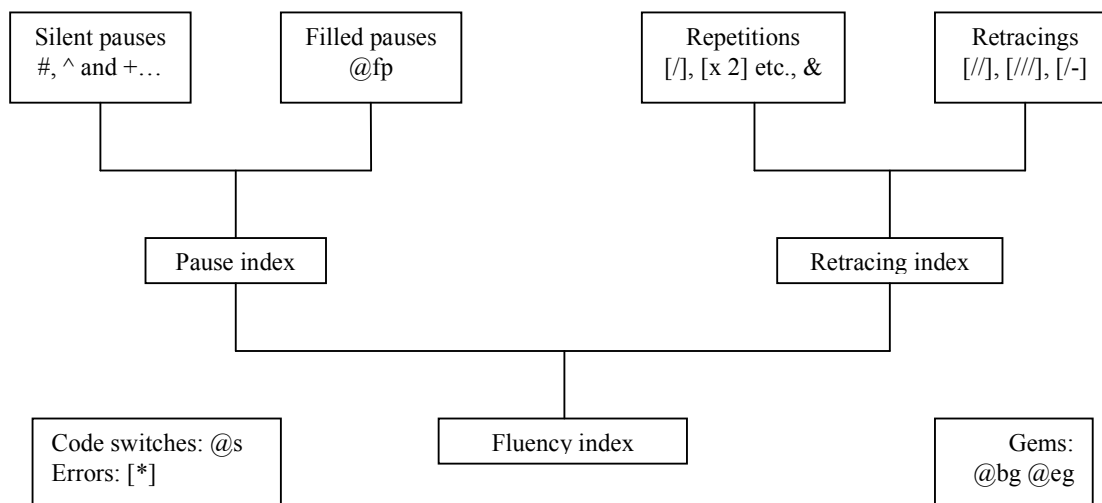
However, overall error rates are low, and the main difference between the groups lies in the diversification of the occupations named, and the slightly lower verbalisation rates, either as a result of lack of input (attrition) and/or of the different environments encountered and biographical paths taken by the two groups.

### **4.3 Film retelling German**

The retellings elicited during the Charlie Chaplin retelling task were transcribed and coded according to the conventions of the CHAT transcription system (MacWhinney 2000; and updates on the associated website <http://childes.psy.cmu.edu/>: MacWhinney 2009; MacWhinney 2009), which was originally developed for the analysis of child language acquisition data gathered in the CHILDES corpus. In contrast with analysing written discourse, this involved identifying utterances, defined as stretches of speech "delineated by pauses and tonal contours" (ibid.: 13), as the basic unit for analysis. The advantage of using the CHAT transcription conventions lies in enabling the researcher to use a range of automated tools (the CLAN suite, available from the CHILDES site) to analyse the speech data both quantitatively and qualitatively.

The transcripts were produced in a three-stage process. First, the actual speech was transcribed, noting code-switches, overlaps and pauses. In a second run-through, these transcripts were converted to CHAT format, checked, and coded for retracings, repetitions etc. Finally, "gem" tags (a "gem" = a section of interest) were added for the qualitative/scene analysis reported below.

Each retelling consisted of a monologic part (the actual retelling, the "narrative") and a dialogic part (follow-up questions). The narratives were fully coded for silent and filled pauses marked #, (word-internal) ^ and @fp respectively; repetitions [/], [x 2], [x 3] etc., retracings [//], reformulations [///], false starts [/-]; fragments &, blocks (several attempts to start a word)/drawls ^ and trail-offs +.../+.?. The complete retellings including the dialogic part were tagged for errors [\*], code-switches@s and "gems" @gb/@eg (Figure 7).



**Figure 7: Features tagged in transcripts of film retellings**

During coding, it was not always clear whether something should be coded as a false start or a retracing, and so retracings proper, reformulations and false starts will collectively be referred to as retracings, and several other features have similarly been combined under broader headings (cf. Figure 7). Pauses were not precisely measured since we lacked both time and equipment to do this in a rigorous manner, but were noted if they were noticeable as pauses (1-2s). For longer pauses, the # mark was used repeatedly to reflect the longer duration of the pause. Silent pauses were only coded if they were "disfluent" or not "noble" (Hilton 2007), i.e. did not occur at utterance boundaries, "at grammatical junctures where commas are general used" or "other sites that are determined by discourse rules" (MacWhinney 2009: 61).

Adding these features allows us to more fully describe the performance of the two groups along a dimension of proficiency not analysable on the basis of the C-test and the FiCAs, namely fluency, which is directly impacted by pausing and various types of retracings. Pausing serves many functions: physiological, discursive, linguistic and social. For our purposes, the linguistic functions, which include discourse planning, syntactic planning and lexical retrieval, are of relevance, in particular lexical retrieval, which is thought to be impaired in attrition. Research has also linked the amount, place and type of pause use to (L2) proficiency (e.g. Hilton 2007). Thus, pauses can indicate linguistic insecurity, which again makes analysing them relevant for this study.



Repetitions and retracings have similarly been tied to planning and retrieval processes (Hilton 2007). On the basis of our data, they would appear to serve slightly different functions: retracings seem to be operate mostly at the discourse/syntactic level, as well as the semantic level (through lexical elaboration/monitoring), while repetitions seem to be more closely connected to the lexical and phonological levels (as evident in difficulties in retrieving or pronouncing individual words), though they may also serve a monitoring or planning function. Repetitions often seem to occur involuntarily, whereas retracings often represent a conscious effort aimed at enhancing the narrative, for example by providing more semantically fitting words. Similarly, false starts often occur because the narrator has remembered some detail which should be told first. Nevertheless, these features all impact on the flow of discourse in a particular way, and so in line with Hilton (2007), they are ultimately combined into a more global retracing factor, the "retracing index" (cf. Figure 7 above).

As was explained in Section Administration above, the bilingual participants retold either the first (Charlie 1) or the second part (Charlie 2) of the film sequence in German, while all of the controls retold both sections of the clip. One recording was lost, and several had to be excluded, either totally or from the analysis of the narratives, due to bad sound quality and various problems during the retelling. A total of 26 retellings from the bilingual group, and 36 from the German control were analysed across the two parts (cf. Table 24).

<b>Film retelling German</b>	<b>Bilingual group</b>	<b>German control</b>
Number of narratives/complete files, Charlie 1	13/15	17/18
Number of narratives/complete files, Charlie 2	10/11	18/18
Total number of narratives/complete files	23/26	35/36

**Table 24: Number of German film retellings**

At the beginning of this chapter, we had formulated some hypotheses for confirming L1 attrition at the group level: our prediction was that the bilingual group would have a lower degree of lexical diversity and a higher rate of hesitation markers (pauses, repetitions and retracings), as well as a higher rate of errors and code-switches. In the sections below, we analyse the following variables:

1. sample length (time and tokens), lexical diversity (VOCD-D), pauses, repetitions and retracings on the basis of the narratives (monologues),
2. errors, code-switches and specific solutions to a range of lexical items/scenes ("gems") on the basis of the complete retellings.

### 4.3.1 Narratives

#### 4.3.1.1 Sample length

As a first and very basic step in describing and comparing the retellings, we checked how long participants took to complete their narrative, excluding the follow-up questions. Charlie 2 is objectively slightly longer than Charlie 1 (5:17 vs. 4:39, or 17 vs. 12 scenes/sub-scenes), which is borne out in the slight difference in retelling length of the control group. We found both parts to be comparable across groups (Charlie 1:  $U = 78.500, p = .127$ ; Charlie 2:  $t(11.013) = 1.065, p = .310$ ), with the bilingual group talking slightly less than the controls on Charlie 1, and somewhat longer on Charlie 2 (cf. Table 25).

Film retelling German	Bilingual group	German controls	Bilingual group	German controls
	Charlie 1		Charlie 2	
Number of narratives	13	17	10	18
Mean length of retelling (min:sec)	03:24	04:09	05:41	04:43
Median	03:13	03:47	05:00	04:37
StD	01:02.60	01:20.44	02:44.34	01:13.03
Range	01:42-4:52	02:34-07:42	02:15-10:25	02:56-07:45
Mean tokens	535.77	634.53	892.30	737.67
Median	551.00	594.00	852.50	695.00
StD	191.10	182.41	406.05	163.32
Range	284-968	371-970	399-1830	515-1029

**Table 25: Length of German film retellings**

Although one might have expected the bilingual group to talk for longer to retell the same contents if our hypothesis about the bilinguals pausing more and using more retracings were true, the

absence of any significant differences is in fact not illogical: the time it takes to complete a retelling depends not just on (lack of) fluency, but conceivably also on effort, memory and personality traits (talkativeness, attention to detail etc.) which may either compound or cancel one another out across groups and individuals. One individual may be less fluent, thus pausing more, reformulating more and therefore taking relatively longer to complete his/her narrative, while another may be fluent, but may have difficulty remembering the details of the sequence, while a third may produce an exhaustive retelling of story, taking care to use precise descriptions and to elaborate the story. On the other hand, it may be possible to compensate for a lack of fluency by using simpler language. Of course, these tendencies may also combine. Consequently, sample length as measured in time is not a very direct or illuminating descriptor of the narratives.

Another way of looking at sample length is to compare the number of tokens produced. Table 25 gives the means, StD and range of the total number of tokens produced, including retracings of all kinds. The controls produced longer narratives for Charlie 2 relative to Charlie 1, in line with the relative objective length of the sequence. The bilinguals' narrations were shorter than the control group's on Charlie 1, but exceeded the latter's productions on Charlie 2 by a wide margin. Although the StDs and ranges on all tasks indicate that neither group performs very homogeneously, this is most noticeable for the latter task.

The very high difference in the mean in favour of the bilingual group is related to an outlier, S11, who produced almost 650 tokens, or one-third of her tokens more than the person with the next highest token count (1830 tokens compared to S24's 1184 tokens), and over 1400 more than the person with the lowest token count (S19, 399 tokens). S11, however, also has the highest proportion of repetitions and retracings, which we will discuss in more detail below. *t*-tests on the token counts are not significant, irrespective of whether S11 is included or excluded (Charlie 1:  $t(28) = -1.433, p = .163$ ; Charlie 2:  $t(10.645) = 1.154, p = .274$ ).

Comparing the two length measures, we find that they correlate highly (Charlie 1:  $r = .874, p = .000$ ; Charlie 2:  $r = .884, p = .000$ ) though not perfectly, which presumably goes back to the impact of the aforementioned additional factors, such as personality traits, lack of fluency or effort. S11, for example, also has the highest value for the time she took to retell the story, which is not surprising giving her exceedingly high token count, and the amount of retracing. There were others

in the bilingual group, however, who came close in terms of the time they took to narrate the story, at much lower token counts, illustrating the multifarious relationship between the different measures.

#### **4.3.1.2 Lexical diversity**

Next, we tested the degree of lexical diversity in each set. The traditional measure for lexical diversity is the type-token ratio TTR; however, as is now widely accepted, it is not a very reliable measure owing to its critical dependence on sample size. In order to achieve a high level of reliability on this measure, our speech samples would have needed to be quite short since TTR values reduce with increasing text length, and of equivalent length, neither of which is the case (cf. Table 25 above).

A lot of effort has of late gone into developing alternative measures, the best of which for our purposes appears to be D, which is computed by a programme included in the CLAN suite called VOCD (henceforth VOCD-D) (McKee, Malvern et al. 2000; Malvern, Richards et al. 2004). A detailed description and explanation of the theory behind, and the algorithm and method used for, calculating VOCD-D may be found in MacWhinney (2000). Although VOCD-D is influenced by text length to some degree also (McCarthy and Jarvis 2007), it has been shown to be a fairly robust measure of lexical diversity (Silverman and Ratner 2002; McCarthy and Jarvis 2010).

We computed D twice using the relevant CLAN command. In the first analysis, we excluded hesitation words such as erm, er, uh, and repetitions/retracings since these will be looked at separately, in order to get an idea of "raw" lexical diversity. The results of this analysis are reported in Table 26. A high D value indicates a higher degree of lexical diversity.

As Table 26 shows, the D values and ranges are quite similar across the two groups and sections. A *t*-test confirms that the two groups are indeed comparable on this measure (Charlie 1:  $t(28) = -.110$ ,  $p = .203$  one-tailed; Charlie 2:  $t(26) = -.124$ ,  $p = .451$  one-tailed). The bilingual group's results for the two portions of Charlie are almost identical, while the control group has slightly lower results on Charlie 2 compared to Charlie 1, and to Charlie 2 in the bilingual group. It is possible that this may be an artefact of the testing procedure, which required the control group to retell both parts in

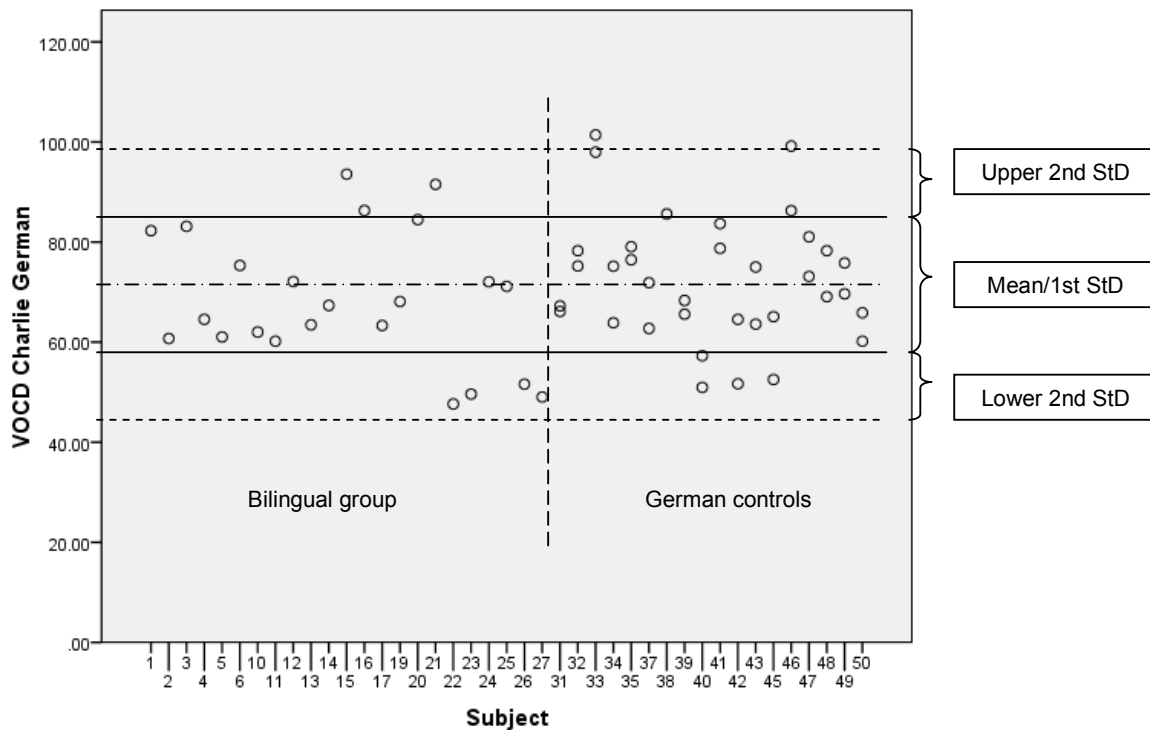
the same language, possibly resulting in reduced motivation and/or a greater scope for tying the second narrative back to the first one.

Film retelling German	Bilingual group	German controls	Bilingual group	German controls
	Charlie 1		Charlie 2	
Number of narratives	13	17	10	18
Mean VOCD-D, excluding hesitation words and retracings	75.33	79.88	77.5	76.24
Median	68.76	81.05	78.04	75.91
StD	16.71	12.85	13.14	13.42
Range	53.76-100.30	57.25-104.71	53.56-100.73	51.82-98.10
Mean standard VOCD-D	68.75	73.27	68.67	70.89
Median	63.46	69.65	67.72	71.86
StD	15.68	12.67	10.61	12.25
Range	47.63-93.56	50.96-101.40	49.61-84.47	51.67-97.97

**Table 26: VOCD-D values for German film retellings**

After reintroducing hesitation words and retracings/repetitions in the second analysis (standard D), the picture does not change much (Table 26). Overall, the values are lower, which is to be expected given that taking account of repetitions will increase the token count and thus reduce the lexical diversity measure. The control group now has a slightly higher average score on both parts of Charlie compared to the bilingual group; however, this difference again fails to reach significance (Charlie 1:  $t(28) = -.874$ ,  $p = .195$  one-tailed; Charlie 2:  $t(26) = -.652$ ,  $p = .260$  one-tailed). The same picture obtains when the data from the two parts of Charlie are pooled.

The bilingual group overall thus cannot be said to show reduced lexical diversity, which would be indicative of L1 attrition, relative to the control group on either measure. As far as individuals are concerned, we can identify two participants, S22 and S27, who fall below the native speaker range, and none who have higher scores on this measure than the controls. However, a graph of all participants' lexical diversity scores (standard D, Figure 8) shows that both S22 and S27 are still, even if just about, within the lower 2<sup>nd</sup> StD of the control group's mean and that two more bilinguals and three controls have only slightly higher scores.



**Figure 8: Lexical diversity scores (standard D) for German film retelling**

The graph also highlights the fact that there are large inter-individual differences in both groups – the top value of 101.40 is roughly twice the lowest values (S22: 47.63, S27: 49.03, S40: 50.96). As we saw before, there is a slight trend for the bilingual group's scores to be lower – this is also noticeable at the top end of the range where two of the controls, S33 and S46 are unmatched. However, this trend is not strong enough to manifest itself in a statistical difference. Nevertheless, there are some important qualitative differences masked by the group figures, which we will discuss as part of the gem analysis below.

#### **4.3.1.3 Fluency**

The next analyses focus on several fluency/hesitation measures (Table 27). First, we provide separate figures for silent pauses (including pauses within or at the start of words) and filled pauses. Then, we combine these figures into a global "pause index" (cf. Figure 7), since it would seem to be largely an individual matter whether a person pauses silently or by using a hesitation word. The use of hesitation words may additionally depend on pause length: Hilton (2007) reported

that pauses of more than 3s length are experienced as uncomfortable and will usually result in the conversation partner's intervention. In the absence of the experimenter's intervention, participants may feel the urge to fill the gap with a hesitation word. Thus, a greater number of hesitation words may indicate a higher incidence of longer/more complex pauses.

Second, we give an overall figure for repetitions and fragments (which were coded as repetitions) on the one hand, and retracings, reformulations and false starts on the other. Repetitions and retracings are then combined into a general "retracing index" (Figure 7). We also provide a percentage for the amount of material (tokens) in retracings relative to the overall number of tokens, which is based on the tokens CLAN excludes as retracings and covers both repetitions and retracings in our sense. Not surprisingly, this figure correlates highly, and partially overlaps, with the retracing score derived from the relevant CLAN tags ( $r = .864$ ,  $p = .000$ ), and is not used in further computations.

Finally, we combine pauses, repetitions and retracings, but not the amount of retraced material into a global "fluency index". As Hilton (ibid.) points out, different types of retracings often co-occur forming "hesitation groups"; however, apart from giving a total figure for these markers, we will not explore this aspect further. Following Dostert (2009), all figures are given as percent of overall tokens to overcome the effect of differing text lengths. Some of the relevant features, such as the silent pauses #, are not included in the normal token count, so the underlying total number of tokens has been adjusted where relevant.

Table 27 illustrates that, again, the two groups behave very similarly in relation to most of the fluency measures. The control group's scores are very similar for both parts, showing consistent linguistic behaviour between the first and the second retelling. The bilinguals who retold Charlie 1 in German seem to be somewhat more fluent than the controls since their averages for both types of pauses, retracings and the overall hesitation index are below the controls'. The bilinguals narrating Charlie 2, on the other hand, have a slightly lower average for those measures. The only measure on which the bilingual group consistently has a higher score than the control group is repetitions.

<b>Film retelling German</b>	<b>Bilingual group</b>	<b>German controls</b>	<b>Bilingual group</b>	<b>German controls</b>
	<b>Charlie 1</b>		<b>Charlie 2</b>	
Number of narratives	13	17	10	18
Mean silent pauses (%)	3.16	3.87	4.33	3.84
Median	2.92	4.45	3.83	3.42
StD	1.06	1.57	1.89	1.47
Range	1.21-4.83	1.14-6.34	2.32-7.61	1.96-6.50
Mean filled pauses (%)	4.69	5.05	5.14	4.61
Median	5.15	4.36	5.38	4.61
StD	1.87	2.56	2.47	2.52
Range	1.60-8.30	0.61-9.71	0.25-7.99	0.29-9.31
Mean pause index (%)	7.85	8.92	9.47	8.45
Median	8.07	8.60	8.92	8.69
StD	2.03	3.10	3.47	2.58
Range	4.71-12.50	3.22-15.86	4.78-15.32	2.50-12.80
Mean repetitions (%)	1.45*	0.81*	1.68*	0.99*
Median	1.43	0.62	1.52	0.87
StD	1.17	0.87	1.23	0.70
Range	0.00-4.22	0.00-3.63	0.00-4.10	0.00-2.68
Mean retracings (%)	2.13	2.46	2.39	2.26
Median	2.38	2.18	2.03	2.38
StD	1.21	1.16	1.43	0.90
Range	0.23-4.01	0.41-4.66	0.86-5.68	0.31-3.69
Mean retraced material (%)	4.94	5.43	5.52	4.68
Median	5.26	4.88	4.82	4.77
StD	3.53	2.71	3.45	1.99
Range	0.45-14.18	1.23-10.84	1.00-13.50	0.61-9.34
Mean retracing index (%)	3.58	3.26	4.07	3.25
Median	4.06	3.07	3.76	3.07
StD	2.05	1.37	2.55	1.21
Range	0.68-7.69	1.02-5.81	0.98-9.78	1.22-6.02
Mean fluency index (%)	11.44	12.18	13.54	11.70
Median	11.93	10.96	13.52	12.27
StD	3.71	4.02	5.06	3.12
Range	5.60-20.19	4.24-21.67	6.27-21.52	4.27-17.46

**Table 27: Fluency measures for German film retellings**

\* Group differences are significant at the .05 level.



This pattern is borne out in *t*-tests/Mann-Whitney tests which return non-significant results for all measures, except for repetitions (repetitions: Charlie 1:  $U = 66.500$ ,  $p = .033$  one-tailed; Charlie 2:  $U = 55.500$ ,  $p = .05$  one-tailed; we decided against quoting individual *t*-test scores for the other measures given that they are not very illuminating). Since existing differences may not be borne out statistically when small group sizes are present, and since the data on all measures seem to be very comparable across tasks as well, we then pooled the data for the two parts of Charlie for all measures. The results are essentially the same; the significance or non-significance of the group difference is even clearer in the pooled data (repetitions: Charlie pooled:  $U = 254.000$ ,  $p = .009$  one-tailed).

This is an intriguing result. As we discussed above, simple repetitions, particularly of function words, tend to be connected to lexical retrieval processes (Hilton 2007), while retracings tend to involve discursive, morpho-syntactic or semantic processes. It would therefore appear that the single area in which the bilingual group is experiencing a greater degree of problems is that of lexical retrieval, while at other linguistic levels it performs as well as the control group. However, we ought to point out that despite the significant difference between the groups on this measure, the number of repetitions in the bilingual group (< 5%) would not cause participants' speech to be considered disfluent (Hegde and Hartman 1979).

Pauses, particularly long and complex pauses, on the other hand, have also been tied to lexical retrieval, as well as speech planning; but on this measure the bilingual group appears to perform at the same level as the control group. There is, however, some evidence to suggest that the number of silent pauses and the amount of repetition present in speech might constitute a mutual trade-off. In a study by Beattie and Bradbury (1979) reported in Howell and Sackin (2001) participants were "punished" for pausing. Participants were able to reduce the amount of silent pauses without reducing overall speech rate, but in so doing increased the number of word repetitions. Howell and Sackin (*ibid.*) continued this line of enquiry and showed that the change in the rate of repetitions occurred only on function words, which serves to "buy time" for "difficult" content word planning. Similarly, our participants may maintain apparent fluency, which is important for keeping the floor in a conversation, by compensatory repeating of words.

However, there may be an alternative explanation for why the control group pauses as often and as extensively as the bilingual group. While currently we cannot provide any more than anecdotal evidence, it appears, on the basis of the transcripts, that the bilingual group pauses more in order to retrieve words (Examples 1 through 3, all from the bilingual group), while the control group takes more time to plan and enhance/elaborate/correct their narratives (Examples 4 through 6, all taken from the control group).

- 1) \*XYZ:       und er # < &pf > [/] ähm@fp pfeift da aus der Hintertür raus [: hinaus] .  
                  (and he # &wh erm whistles out of the back door)
  - 2) \*XYZ:       und in diesen # Wagen mit den anderen Verhafteten # getan  
                  (and (is) # put into this # van with the other prisoners)
  - 3) \*XYZ:       aber jedenfalls ähm@fp &=laughs äh@fp +//.  
                  (but anyway erm &=laughs er)
- \*XYZ:       wie sagt man da ?  
                  (how do you say for that?)
- \*XYZ:       äh@fp mir fehlen die Worte .  
                  (er I'm lost for words.)
- \*XYZ:       ähm@fp war # sein Atem [!] &=laughs < ganz > [//] ein bisschen # stark !  
                  (erm his # breath [!] &=laughs was very [//] a little # strong!)
- \*XYZ:       und # gräuslich !  
                  (and # horrible!)
- \*XYZ:       < würd(e) i(ch) > [//] # wollte ich sagen .  
                  ((is) what I would [//] # wanted to say)

In Example 1, the problem could be either with remembering, or (phonetically) producing the exact shape of the word *pfeifen* ("whistle"). In Example 2, the words chosen after the pause ("the van" and "put into") are quite general compared to the usual solutions: "the prison van", "the Black Maria", "the paddywagon" and similar for "van"; and "pushed on" or "brought in" for "put on the van". It is possible that the participant settled for a generic word after failing to retrieve a more specific one. Example 3 illustrates an explicit retrieval difficulty, of which we will discuss more instances below.

- 4) \*XYZ: < ein > [//] # ich würde mal sagen, eine Hartz\_vier\_Empfängerin # geht um die Ecke .  
(a # I would say a welfare recipient # goes around the corner)
- 5) \*XYZ: und rümpft äh@fp die Nase über < die anderen äh@fp Passagiere, die da > [//] oder die anderen Gefangenen, die da drin(nen) sitzen .  
(and turns up erm his nose at the other er passengers, who, or the other prisoners who are sitting in there)
- 6) \*XYZ: < die bei einem # Bäcker > [//] # die vor einem Bäckergeschäft aus dem < Bäcker^auto > [//] oder aus dem Auto, was die Backwaren anliefern, ein Brot stiehlt.  
(who from a # baker [//] # in front of a bakery out of the baker's # van [//] or out of the car which delivers the baked goods, steals a loaf of bread)

Example 4 shows the opposite tendency to Example 3 – the speaker pauses and then produces a very specific assessment of the background and appearance of the young girl, when most people simply referred to her as "the (young) girl" or "the (young) woman". In Example 5, the speaker either corrects or explicates "the passengers" to "the prisoners"; and in Example 6 the precise location from which the bread was stolen is gradually specified.

- 7) \*XYZ: und jetzt weiß ich nicht mehr weiter .  
(and now I don't remember how it went on)
- \*XYZ: und dann +...  
(and then +...)
- \*XYZ: dann kam also < die Wache > [//] der Wagen an .  
(then the guards [//] the van arrived up)
- 8) \*XYZ: ähm@fp # &j@fp und in +//.  
(erm # y-- and in)
- \*XYZ: halt, nee [: nein], jetzt muss ich g(e)rad(e) überlegen .  
(no, sorry, I must have a quick think)
- \*XYZ: < und in der > [/-] # äh@fp dann sieht man ihn jedenfalls in [x 2] so (ei)n Lokal gehen .  
(and in the # er in any case, then you see him go into [x 2] a kind of eatery)

Examples 7 (taken from the bilingual group) and 8 (taken from the control group) are instances of individuals pausing while trying to remember what happened next, which is equally representative of both groups.

Turning now to the question of whether any bilingual's scores fall outside of the native-speaker range, we first plotted the bilingual group's fluency measures relative to the control groups means and limits of the upper 2<sup>nd</sup> StD (cf. Figure 9).

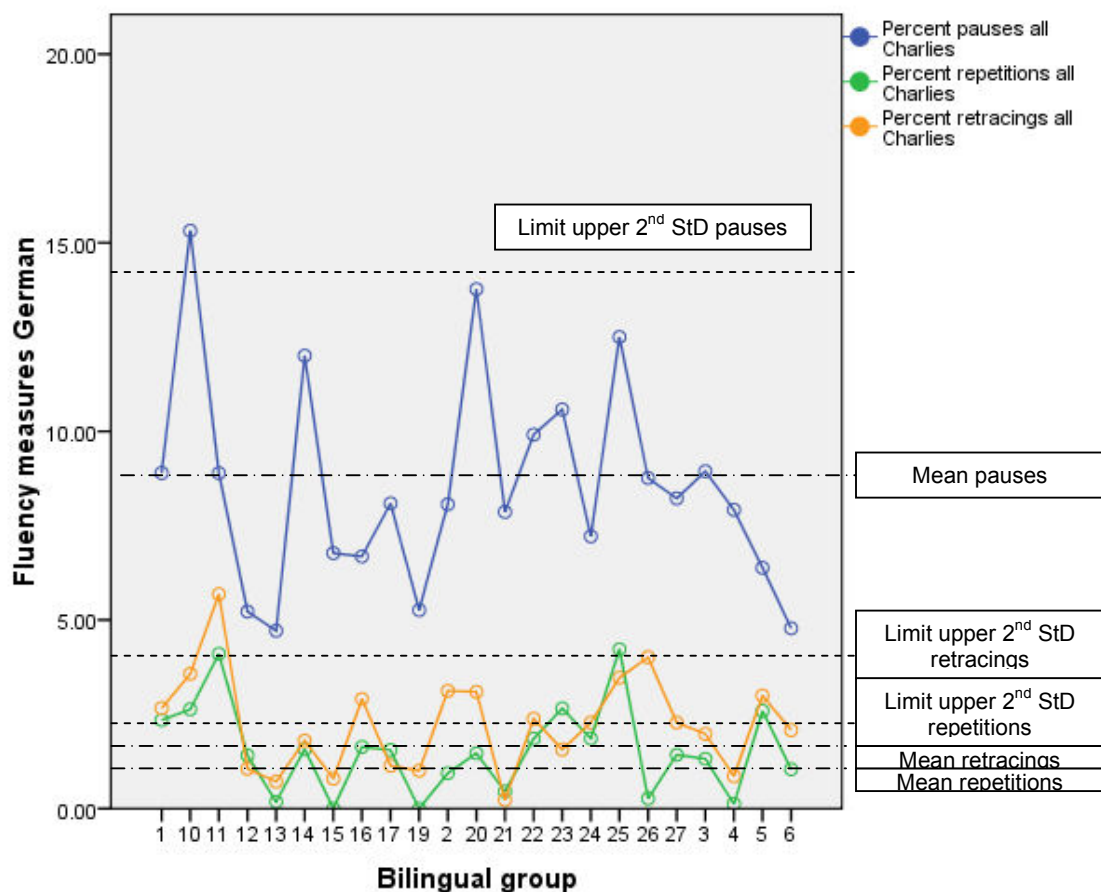


Figure 9: Fluency measures bilingual group vis-à-vis native-speaker baseline

Participant	Pauses		Repetitions		Retracings	
	Mean: 8.7%	Upper limit 2 <sup>nd</sup> StD: 14.3%	Mean: 0.9%	Upper limit 2 <sup>nd</sup> StD: 2.5%	Mean: 2.4%	Upper limit 2 <sup>nd</sup> StD: 4.4%
S5	6.38%		<b>2.59%</b>		2.99%	
S10	<b>15.32%</b>		<b>2.63%</b>		3.57%	
S11	8.89%		<b>4.10%</b>		<b>5.68%</b>	
S20	<b>13.77%</b>		<b>2.47%</b>		3.09%	
S23	10.58%		<b>2.66%</b>		1.56%	
S25	<b>12.50%</b>		<b>4.22%</b>		3.47%	

Table 28: Individual results on three fluency measures, German film retelling

Figure 9 confirms that most of the scores fall within the native-speaker range, as already shown by the similar score means, StDs, ranges and the absence of significant differences in all except the repetitions measure. However, there are some individuals who exceed or are very close to the limits of the 2<sup>nd</sup> StD on one or more of the measures (cf. Table 28; high values are put in bold). S10 exceeds the native-speaker range for pauses, for which she has the highest score, and repetitions. Her retracings are within, but at the higher end of the 2<sup>nd</sup> StD. S20 has the second highest score for pausing at the high end of the 2<sup>nd</sup> StD, her repetition rate is also elevated, while her retracings are closer to the mean. S11 and S25 are those with the highest scores for repetitions; S11 also has the highest score for retracings and an average score for pauses. S25, on the other hand, has the third highest rate of pauses, and an elevated retracing count. Two more individuals, S5 and S23, have repetition scores just outside the 2<sup>nd</sup> StD, but S5 has a low rate of pauses and a raised score for retracings, whereas it is the other way round for S23.

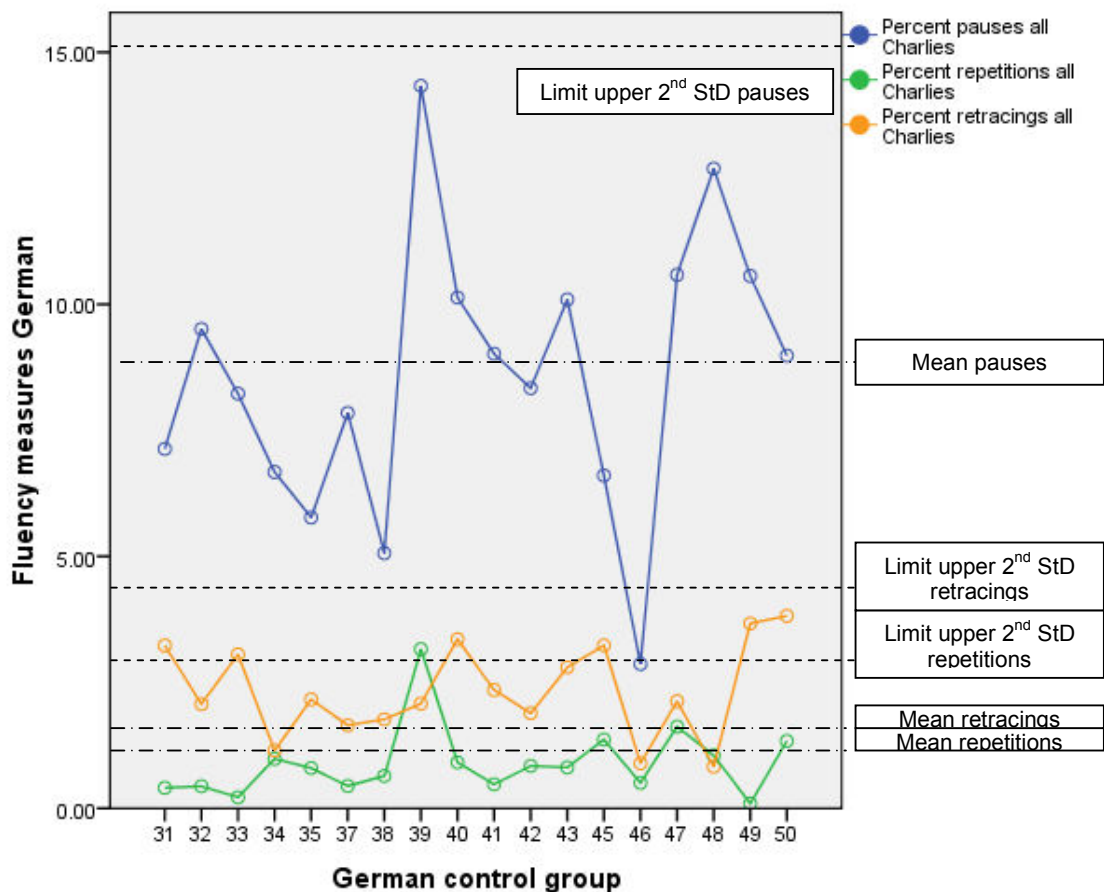


Figure 10: Fluency measures control group vis-à-vis native-speaker baseline

Although some bilingual participants thus have relatively high scores on all measures, and others relatively low ones, this is not necessarily so, as this comparison shows. S10, S11, S20 and S25 would appear to be the most obvious candidates for L1 attrition on the basis of this analysis, and we will return to these when summarizing the results. S11 may be a case where a slightly lower rate of pausing is achieved at the expense of repetitions, as was suggested above. Again, the picture shows that there is a lot of variability between individuals, and indeed for individuals between different measures, no more so than for the pause measure where the highest score is three times the lowest score.

There also is a great deal of variability amongst the control group (cf. Figure 10; the points represent the mean score for each participant). However, all scores bar two, which both belong to S39, are within the range of two StDs for the relevant measure. S39 would appear the most disfluent native-speaker, with a pause score of 14.33% and a repetitions score of 3.16%. S48 also has a high score for pauses (12.69%), but the other measures are rather low; which again may be due to one measure off-setting the other. S49 and S50 have the highest retracing scores (3.67% and 3.82% respectively), which was connected to conscious monitoring and the desire to retell the story precisely.

#### **4.3.1.4 Measures combined**

To conclude the analysis of the narratives themselves, and to check the idea that some of the measures might be reciprocally related, we ran correlations on the various measures under discussion: sample length (time and tokens), lexical diversity (raw VOCD-D) and fluency measures (pauses, repetitions, retracings and material in retracings) (Table 29).

There are no real surprises in this analysis. The total time it took to retell the story is highly significantly correlated, albeit with varying correlation strengths, with all measures except lexical diversity. The highest correlation is, not surprisingly, with the total number of tokens, followed by retraced material, retracings, repetitions and pauses. Lexical diversity has no significant effects (these become significant medium effects when standard VOCD-D is used).

Charlie German		Time taken	Total tokens	Raw VOCD-D	% pauses	% reps	% retrace	% retr mater
	Pearson Correlation	1.000	.897**	-.054	.393**	.412**	.420**	.453**
	Sig. (2-tailed)		.000	.685	.002	.001	.001	.000
Time taken	N	58	58	58	58	58	58	58
	Pearson Correlation	.897**	1.000	.041	.170	.333*	.363**	.385**
	Sig. (2-tailed)	.000		.762	.201	.011	.005	.003
Total tokens	N	58	58	58	58	58	58	58
VOCD-D excluding hesitation words & retracings	Pearson Correlation	-.054	.041	1.000	.129	-.084	-.165	-.137
	Sig. (2-tailed)	.685	.762		.333	.530	.215	.305
	N	58	58	58	58	58	58	58
	Pearson Correlation	.393**	.170	.129	1.000	.445**	.277*	.390**
	Sig. (2-tailed)	.002	.201	.333		.000	.035	.002
Percent pauses	N	58	58	58	58	58	58	58
	Pearson Correlation	.412**	.333*	-.084	.445**	1.000	.315*	.519**
	Sig. (2-tailed)	.001	.011	.530	.000		.016	.000
Percent repetitions	N	58	58	58	58	58	58	58
	Pearson Correlation	.420**	.363**	-.165	.277*	.315*	1.000	.863**
	Sig. (2-tailed)	.001	.005	.215	.035	.016		.000
Percent retracings	N	58	58	58	58	58	58	58
	Pearson Correlation	.453**	.385**	-.137	.390**	.519**	.863**	1.000
	Sig. (2-tailed)	.000	.003	.305	.002	.000	.000	
Percent retraced material	N	58	58	58	58	58	58	58

**Table 29: Correlations between sample length, lexical diversity and fluency measures**

\*\* Correlation is significant at the 0.01 level (2-tailed).

\* Correlation is significant at the 0.05 level (2-tailed).

All results, except for pauses, are negative implying that longer narratives with more retracings and repetitions will have a lower lexical diversity. Pauses correlate moderately with repetitions, and less so with retracings and retraced material. This correlation seems to speak against the suggestion of a trade-off between pauses and repetitions made above. Finally, repetitions correlate moderately with retracings, and the retracing measures correlate highly amongst themselves. The absence of a stronger effect between repetitions and retracings would point to different uses of these measures.

### **4.3.2 Scenes**

In this subsection, we will look at errors, code-switches, and "gems", i.e. scenes and people/objects of potential interest, across the full range of the retellings. These three parameters partially overlap: the gems are interesting precisely because they often gave rise to errors, as well as code-switches, and hesitation behaviour as discussed above. We identified 46 "gems" across Charlie 1 and Charlie 2 (cf. Appendix for the complete list), but in the interest of brevity, we shall restrict the discussion to the most interesting items. The reader is referred back to Chapter 3 for the description of the plot of the film sequence.

#### **4.3.2.1 Code-switching**

As we discussed in Chapter 2, code-switches are a common occurrence amongst bilinguals in many settings. They are interesting because they show that more than one language may be active at any given time in a bilingual's mind, which raises questions about the mechanisms of language selection and, ultimately, the nature of the bilingual mind.

Although some researchers prefer to reserve the term for deliberate, purposeful language mixing, we use the term for both functional and unintentional code-switching regardless of the duration of the switch to highlight the fact of any other-language use, rather than hypothesize about its causes. The intended purpose and organisation of our study strongly implied that using a language other than the one being tested during the language tasks was undesirable and inappropriate, so any such use which did not clearly serve discursive functions would be indicative either of a failure to inhibit the other language sufficiently (interference), or of failed retrieval, prompting the participant to resort to code-switching.

We do distinguish it from the use of borrowings or loan words, however, which are other-language lexical forms which have been integrated into a given language as inflectable elements. In our data, we find a few examples of English-language borrowings into German, which have become accepted in standard language and even occur in dictionaries of German: *Gentleman* and *gentlemanlike*, *Bowler (hat)*, *Outfit*, *overdressed*, *happy (family)* and *Loser*. The borrowings mentioned here are used by both groups (bilingual group: 7 instances across 26 retellings, one per



every 3.7 narratives; control group: 20 instances across 36 retellings, one per every 1.8 narratives) and will not be considered further.

All instances of English-language material in the German narratives were coded with @s and subsequently identified using the `FREQ` (frequency) `CLAN` command. All occurrences were checked and classed into the following five types of code-switching behaviour (Table 30).

"Quotes" refers to participants repeating the text that was shown on the intertitles in the film, or occasionally what they understood an actor to have said ("I am hungry." or "Come on!").

"Reference to actor" refers to applying English-language labels to the main characters: *der Sheriff*, *der Officer*, *der Constable*. (*Sheriff* is ubiquitous since that character is named in the letter that is shown at the beginning. We therefore excluded that item from the count.)

"Lexical gaps" are situations where a participant flagged not knowing or searching for a word, which we will return to below. "Self-talk" refers to instances of participants wondering what came next, turning to the investigator for assistance or commenting on their state of knowledge of certain terminology. Finally, "interferences" refers to the production of English-language words or (subsequently corrected) word-fragments, which serve no discernable purpose. The first three types constitute conscious, deliberate uses of code-switches; the remaining two are inadvertent and/or, in the context of this study, inappropriate uses. Figures correspond to total occurrences across the transcripts.

Type of code-switch	Bilingual group (N = 26)		German controls (N = 36)	
	Instances	Individuals	Instances	Individuals
Quotes	2	S3, S7	9	S31, S35, S41, S42, S43, S46, S47, S48
Reference to actors	0	—	2	S40, S46
Lexical gaps	9	S1, S2, S3, S10, S21, S27	1	S35
Self-talk	4	S10	0	—
Interferences	16	S1, S3, S7, S8, S10, S11, S25, S26	0	—
<b>Total</b>	<b>31</b>	<b>11</b>	<b>9</b>	<b>9</b>

**Table 30: Code-switches film retelling German**

The data paint a really clear picture: the control group uses code-switches sparingly and almost exclusively to quote from the film; the bilingual group on the other hand shows evidence of both lexical retrieval problems and "interlopers" from English. In all, eleven bilinguals, but only one control manifested these kinds of problems.

This distribution constitutes a fairly clear division between deliberate and involuntary/unwelcome (when flagging lexical retrieval problems) use of English between the two groups. Based on the figures above, and also on the rate of English-language borrowings in both groups, it would appear as though the control group is more at ease using English in their speech, which mirrors the very common use of English words for special effect in modern German, particularly in the media. The bilingual group, on the other hand, being subject to more persistent and comprehensive English-language influence, apparently tries to avoid English in their German production, correcting code-switched material and even borrowings wherever they were noticed, and using code-switches as a last resort when unable to think of an expression. The bilingual group would also be less familiar with English-language vocabulary that has become accepted in German. It is possible that the awareness of English-language influence on their mother tongue has prompted the bilingual participants to attempt to keep their languages separate. However, as we will see below, such attempts have not always been successful.

The majority of the control group's quotes referenced in Table 30 above come from the dialogic part (seven of nine). In exploring the why and how further, the controls quoted back the text of the intertitles without being prompted to use English. One of the follow-up questions concerned the phrase *Determined to go back to jail*. (**Gem 11**; accounting for two of nine code-switches), which the majority of the controls (12 ex 15) had misunderstood as *Charlie was destined to go back*, or *It had been decided that he must go back*, rather than that he desperately wanted to go back. This appears to be due to a learned one-to-one correspondence between *determined* and *bestimmt* (he is destined/it is decided; several controls use this lexeme when rendering the original phrase in German).

As mentioned above, two controls used an English-language label for the policeman, and one control felt that Charlie's wiping his hands on the curtain was not "in style", probably a

misremembered phrase. This was annotated as a lexical gap owing to the prosody of the phrase, but it is also possible that this phrase was used for effect, i.e. deliberately.

On the part of the bilingual group, two participants also quoted from the intertitles, one instance of which occurred during the narrative, while the other was prompted by the investigator. The four instances of self-talk all come from the same person, S10, whose English is highly active, causing interferences and retrieval problems. Ten more bilinguals also encounter either one or the other, or both.

The inadvertent production of code-switches in the bilingual group mainly concerns nouns, particularly cognate ones (Example 9), function words (prepositions) (Examples 10 and 11), the conjunction *and* (Example 12) and the interjection *well*. Verbs and adjectives are not immune, though (Examples 13 and 14). Interestingly, the function words tended to be corrected by participants, whereas the other involuntary uses went unnoticed. This possibly goes back to the suggestion discussed under the fluency features above that function words are often used to buy time, and so a switch into the other language may be more noticeable than in other contexts. The code-switch is marked @s.

- 9) \*XYZ: < und dann > [-] ähm@fp < sie > [/] < sie > [-] da kommt auch direkt < ein > [/] ein Strahl Milk@s < &dir > [//] < in den > [/] in den Krug hinein .  
(and then erm erm she she [-] a spurt of milk came directly into into the jug)
- 10) \*XYZ: er wurde also aus (de)m Gefängnis entlassen < mit einem > [/] äh@fp mit einem Begleitschreiben from@s [//] ähm@fp # vom Sheriff .  
(so he was released from prison with an erm with a covering letter from erm from the sheriff)
- 11) \*XYZ: und die Trauben wachsen da gerade < &out@s > [//] # vor dem # Fenster .  
(and the grapes are growing just &out [//] outside the window)
- 12) \*XYZ: and@s pflückt sich (ei)nen Apfel von +//.  
(and picks (himself) an apple from +//.)
- 13) \*XYZ: weil äh@fp ich nehme an, < die > [/] < die &dr@s > [//] die fahren sehr schnell und # < kürvig > [//] < kurvig > [\*] .  
(because erm I assume they they &dr [//] they drive very fast and winding\*)
- 14) \*XYZ: ich mein(e), die < laten@s > [//] späten < &se > [/] sechziger, siebziger Jahre, ja ?  
(I think, the late [//] late sixties, seventies, yes?)

This category also includes expressions used by two participants, which, on the face of it, simply contain a loan word, i.e. a regularized form which should not be classed as a code-switch anymore. On closer inspection, these expressions turn out to be calques, although it is unclear whether they were made up on the spot, or are part of that person's vocabulary, in which the loan word is used in a different sense than that which it has been borrowed for (Examples 15 and 16).

- 15) \*XYZ:        typisch # Charlie Chaplin äh@fp macht er # (eine)n herrlichen # Job@s d(a)raus .  
                   (typically Charlie Chaplin er he makes a lovely job of it)
- 16) \*XYZ:        und < er > [x 2] # er äh@fp äh@fp ## versucht da herauszufinden, ob sie ok@s ist  
                   (and he [x 3] er er tries to find out, whether she's ok)

*Job* has been borrowed into German meaning "an occupation", "an opportunity to earn money", "a (computer) task" (Wahrig, Hermann et al. 1986), but not in this idiomatic context where it is supposed to mean *make a mess of something*. Similarly, *ok* is used widely in many different contexts, but to express the meaning of the phrase *she is ok* (i.e. she is not injured), one would use a phrase like *es geht ihr gut, sie ist unverletzt, es ist alles mit ihr in Ordnung*. Like some other cases, this might constitute a borderline case in terms of acceptability; however, against the background of the overall characteristics of this participant's speech we felt justified in determining this use as transfer.

Another case is the use of *Counter* by one bilingual participant, who uses it to refer to the tobacconist's counter. *Counter* has indeed been borrowed into German in the context of *airline counter* or similar, not a *pub* or *shop counter*. The participant seemingly also considers the word as not quite fitting, since she subsequently corrects it by supplying a more appropriate, though not entirely correct German term, *die Anrichte* ("sideboard"). This use of the English-language word was classed as a lexical gap as a result.

Some of the words which resulted in a lexical gap for some participants (*shipyard*, **Gem 1**; *trustworthy*, **Gem 2**, *wedge*, **Gem 4**; and *loitering*, **Gem 41**) will be discussed as part of the gem analysis below. Other cases are *support*, *impact*, *tray*, a way to describe the girl's attitude towards Charlie (possibly *to flirt*), and *jump*. The last example is very interesting in that the participant was

evidently looking for a very specific (Swiss) German word (*gumpen*), as a result of which she could not think of other ways to express this meaning.

Before considering further examples of lexical retrieval difficulties as well as some issues relating to lexical diversity by discussing the "gems" we identified, we shall briefly present the results of the error analysis.

#### **4.3.2.2 Errors and idiomatic language use**

An error analysis is not as straightforward as it sounds, particularly in relation to spoken data, for three reasons. First, since speaking (and listening) is a real-time online task, the sense units handled tend to be shorter, and certain features produced in response to processing constraints may become acceptable that would be considered errors in written text. Second, in spoken conversation, it is impossible to delete what has been said, and for the listener it may not be obvious whether what has been said is still part of the current utterance, or whether the speaker has changed his/her mind and started a new utterance, particularly since this may happen very fluently. Thus, in the analysis, segments may erroneously be construed as units and found to contain errors. Third, in general, it is impossible to know the speaker's mind and therefore the intention behind certain (fragments of) phrases and consequently the cause for certain errors. This may create problems for their analysis, for example, the assignment of error codes.

The first reason mainly concerns issues of word order. German main clauses containing verb phrases (finite verb and complement), for example, must take the form a so-called "verb bracket": the finite verb is in second place, the complement in last place, and all other elements of the clause are arranged around these two positions (with some arrangements being more marked or suitable for emphasis than others). However, since the finite verb may be an auxiliary, the distance between verb and complement may be large, in fact, too large to handle in spoken conversation. Thus, in speaking it is an oft-seen tendency to shorten the verb bracket while allowing other elements, which should strictly occur within the verb bracket, to follow. Since spoken text is divided into utterances, this arrangement tends to be acceptable – it also serves to lighten the processing load

for the listener. In writing, on the other hand, where it is possible to monitor one's production, it is not.

On the other hand, other phenomena, such as lack of gender/number agreement, use of an incorrect case etc., though possibly caused by the same constraints on monitoring, would not be as readily accepted as correct or "excusable", just because they occur in a spoken context. The elements to be agreed usually occur side by side in the same utterance, so monitoring for congruence is much easier, and a mismatch would usually be followed by a retracing to correct the clash, since these kinds of "violations" stand out, both for speaker and listener. Anecdotally, even though this type of error usually does not interrupt communication too much, its presence would be thought of as lack of proficiency on the part of the relevant speaker, since using case and gender/number correctly is a very basic and highly automated/intuitive skill in native speakers.

Similarly, inappropriate lexical and semantic choices can often be "corrected" by the listener him-/herself without passing comment; but even if the communicative purpose was achieved (by virtue of an accommodating listener), the word or phrase in question still retains its usual (inappropriate) meaning, and a less accommodating listener might choose to explicitly correct the choice of word/phrase. Thus, some occurrences of what would traditionally/in the written mode be classed as errors, are relatively more clear-cut than others.

Given the inherent difficulties in performing an error analysis on spoken data, we took extra care to verify our decisions. Potential errors were checked by at least one, in cases of doubt two, native-speakers of German and trained Germanists. Additionally, reference works on the German language, e.g. the Duden and Wahrig Deutsches Wörterbuch (Wahrig, Hermann et al. 1986; Duden-Redaktion n/a), as well as online sources (linguistic fora) were consulted to verify whether a certain word or expression has become standard use. Each case was checked against the notion of "Would one hear this? Has this become regular use?", and if so, it was not counted as an error.

With regard to word order violations as discussed above, essentially elements outside the verb bracket, these were not counted as errors if prosodically they were distinguishable as separate utterances. With others, like verb second violations, there are arguments speaking for and against a classification as error. On the one hand, verb second is a, if not the most, basic feature of German word order, and violations cannot be explained by constraints arising out of long processing chains.

Thus, command and correct parsing of this feature, even in demanding processing conditions, could be expected. On the other hand, in spoken German the occasional positioning of some adverbs and modal particles such as *natürlich* ("naturally") or *eigentlich* ("in a sense", "for all intents and purposes") before the main clause (effectively putting the verb in third place), as a means to provide focus and emphasis, is attested, though again one would expect some prosodic feature to mark off this element as separate (through intonation and/or pausing). Since this word order arrangement can be expected to be promoted by English-language influence, we felt it would be useful to mark these occurrences as errors, so as to be able to check for differences in distribution.

All errors were coded according to an exhaustive coding scheme (courtesy of Monika Schmid; cf. Appendix), encompassing the whole range from lexical and semantic errors through every kind of morphological error to syntactic and discourse errors. The sub-types are based on previous analyses of the speech of German speakers in an English-language environment (Schmid 2001).

For our purposes, these were summarized into three main categories (Table 31 below). Within the category of lexico-semantic errors, there are three sub-categories: lexical errors (non-standard forms of words, inappropriate words for context etc.), semantic errors (semantic transfers, slips of the tongue) and function words. Morpho-syntactic errors comprise errors of form (case, gender/number in noun phrases; verb tense, voice, mode), syntax (syntactic transfer, negation, passive) and word order (verb second, verb last, verb bracket). Finally, discourse errors refer to anaphoric reference (failing to make reference to the various protagonists explicit for an audience that had not seen the film) and unclear references.

Again, a very clear picture presents itself: The bilingual group has a significantly higher rate of errors overall and for two of the three categories, lexico-semantic and morpho-syntactic errors (total errors:  $U = 311.500$ ,  $p = .012$  one-tailed; lexico-semantic errors:  $U = 295.500$ ,  $p = .006$  one-tailed; morpho-syntactic errors:  $U = 333.500$ ,  $p = .023$  one-tailed). The difference on discourse errors, on the other hand, does not reach significance ( $U = 415.500$ ,  $p = .203$  one-tailed), indicating that there was a similar level of difficulty across both groups (bilingual group: 10 ex 26 participants = 38%; German control: 18/26 = 50%; the problem arose for a higher percentage of individuals in the German control group).

	<b>Bilingual group (N = 26)</b>	<b>German controls (N = 36)</b>
Total number of errors	149	135
Mean total errors	5.73*	3.78*
Median	5.00	3.50
StD	4.03	2.50
Range	0-19	0-12
Mean lexico-semantic errors	2.81**	1.50**
Median	2.00	1.00
StD	2.10	1.36
Range	0-7	0-5
Percent of total errors	49%	40%
Mean morpho-syntactic errors	2.31*	1.47*
Median	2.00	1.00
StD	2.53	2.43
Range	0-13	0-8
Percent of total errors	41%	39%
Mean discourse errors	0.62	0.81
Median	0.00	0.50
StD	0.90	0.98
Range	0-3	0-3
Percent of total errors	11%	21%

**Table 31: Errors film retelling German**

\* Group difference is significant at  $p = .05$ .

\*\* Group difference is significant at  $p = .01$ .

Within the sub-categories, the difference comes down to two error types: lexical interferences/semantic transfers ( $U = 259.000$ ,  $p = .001$  one-tailed), and word order ( $U = 343.000$ ,  $p = .009$  one-tailed). This result is in line with the expectation to see an influence from L2 on L1 of the bilingual group. In other words, there was no difference on the use of function words, morphological/form errors or morpho-syntactic errors. We will discuss lexico-semantic errors in more detail below.

Concerning word order, as discussed, we logged verb second violations to see if there are any distributional differences between the two groups. Two controls do indeed place an adverb or



modal particle before the main clause (Example 17, control group). In both cases, the discursive function is plausible (the other occurrence involves *jedenfalls* ("anyway")).

17) \*XYZ: ja, < irgendwie der Polizist war > [\*] anderweitig beschäftigt, ne [: nicht] ?  
(yes, somehow the policeman was busy elsewhere, right?)

18) \*XYZ: und < dann sie # nimmt > [\*] ihm den Hut ab und alles !  
(and then she # takes the hat from him and everything!)

In the bilingual group, seven individuals violate verb second in one way or another; and one of these, S10, has five occurrences in total, three of which involve the lexeme *natürlich* ("naturally") in different positions. The "pragmatic" use, as attested in the control group, accounts for six out of the eleven occurrences. In two more, *naturally* occurs after the subject, pushing the verb into third position. In another case, subject and verb were not inverted after an initial sub-clause, and the remaining two occur after *und dann* ("and then") (Example 18, bilingual group).

All three structures are licensed by English word order, confirming our hypothesis that transfer might impact on the frequency with which these structures occur in the two groups. The use of *und dann* in this position in particular seems to be a case of transfer, since there is no obvious pragmatic reason for this word order. Note that in Example 18 the subject is produced as one unit with the preceding *and then* before a pause. In the retellings by the bilingual group, *und dann* is by far the most common way to start a new utterance, and there are a few cases where a participant corrected and post-positioned a subject that had been produced immediately after the *and then* phrase.

In contrast with our results, we would have expected L2 influence on the use of function words by the bilingual group, since prepositions in particular, and conjunctions to an extent, have been found to represent a vulnerable word class (e.g. Köpke 2002). However, this is not borne out in the data. Similarly, we would have expected syntactic transfer to play a role; yet again, the difference in morpho-syntactic errors is not significant.

On closer inspection, however, it becomes apparent that, despite similar rates of errors on this error type across the two groups, the problems partially lie in different areas: whereas in the bilingual group, there is indeed evidence for syntactic transfer (Example 19), the controls violate different

syntactic constraints, such as subject continuity with to- (-infinitive) clauses (Example 20). On the other hand, there are some incorrect uses of passives and negation in both groups (Example 21, control group; Example 22, bilingual group), which otherwise might have been claimed to be instances of syntactic transfer.

- 19) \*XYZ: und ähm@fp < dass > [//] wenn (e)s dann < zum [: zu dem] Bezahlen kommt > [\*]  
hat er natürlich kein Geld .  
(and erm that [//] when it comes to paying [\*] he has no money of course)
- 20) \*XYZ: und Charlie bietet ihr < ihr > [//] sein # buntes # Taschentuch an .  
(and Charlie offers her her [//] his colourful handkerchief)  
\*XYZ: < um > [\*] die Tränen zu trocknen und zu schnäuzen .  
(in order to [\*] dry the tears and to blow (her) nose)
- 21) \*XYZ: < das Mädchen # wa(r) > [//] dem er vorher geholfen hatte , < war > [\*]  
(he)reingeführt .  
(the girl was [//] whom he had helped earlier on, was [\*] brought in)
- 22) \*XYZ: +" das is(t) < nich(t) (ei)n > [\*] Keil .  
(that is not a [\*] wedge)

In Example 19, the phrase that was transferred from English (*when it comes to paying*) should have been something like *als/wenn es ans Bezahlen geht* (literally: "when it goes to the paying"). In Example 20, the to-clause is incorrect because structurally it requires that the subject is the same as in the main clause (Charlie), whereas the contents implies a different subject (the girl). A correct solution would have employed a different type of clause such as *so dass sie sich die Tränen trocknen und schnäuzen kann* ("so that she would be able to dry her tears and blow her nose").

Example 21 is not entirely ungrammatical in so far as this type of passive exists in German ("Zustandspassiv" = stative passive, forms the passive with the auxiliary *sein* (to be)), but pragmatically the other type, "Vorgangspassiv" = dynamic passive", which forms the passive with the auxiliary *werden* (become), is the appropriate choice in this context: *das Mädchen, dem er vorher geholfen hatte, wurde hereingeführt*. Finally, in Example 22, the negative article *kein*, instead of the analytic construction *nicht ein* (not a), should have been used. These occurrences reinforce the point made in Köpke and Nespoulous (2001) and Köpke (2002) that the most vulnerable grammatical structures in attrition, and apparently in less proficient non-attributing native

speakers, are "borderline" ones – clearly, German has the grammatical means to express negation analytically; however, accepted use requires the synthetic expression.

In a similar vein, the errors that occur in the lexical and semantic domains belong to different kinds in both groups. The bilingual group, as expected, has a number of instances of lexical interferences, and semantic transfer from English (Examples 23 through 25), while this error type was not found with the control group.

23) \*XYZ:     der < sitzt > [\*] natürlich # unter anderen Sachen d(a)runter .  
                  (that [the wedge] is of course sitting [\*] under other things)

24) \*XYZ:     und natürlich is(t) auch gleich die Polizei < an der Szene > [\*] .  
                  (and of course the police is immediately at the scene [\*])

25) \*XYZ:     und jede Menge Essen < gehabt > [\*] hat .  
                  (and had [\*] plenty to eat)

In Example 23, *sitzen* ("to sit") cannot really be used to denote inanimate objects; instead, *feststecken* ("to be stuck") or some such verb should have been used. In Example 24, instead of the calque of *at the scene*, an idiomatic way to formulate the same idea would have been *vor Ort*. In Example 25, the problem is with the verb (*Essen*) *haben* ("to have (food)"), for which in German one has to use the equivalent of *to eat* (*essen*, *Essen zu sich nehmen*).

Other examples of semantic transfer include the use, instead of *frisch verheiratet* ("newly wed"), of *neu verheiratet* which in German means "re-married", and coining *institutionalisiert* ("institutionalized") in the sense of "become accustomed to living in an institution" (prison), "to have lost self-reliance" (**Gem 22**). (Standard German uses *institutionalisiert* in the sense of "to become/be made an institution" only.) The latter case seems to represent a genuine lexical gap in the German language: the problems associated with leaving prison tend to be discussed in terms of the "reintegration of society", without labelling the cause. Consequently, the controls used circumlocutions such as *Charlie does not cope with life, he does not have to expose himself to real life* or *he is not fit for simple tasks*. Not frequent, but also indicative of transfer is the use of *rennen* ("to run/flee") instead of *wegrennen* (**Gem 31**) and "*das Brot nehmen*" ("to steal/take the bread")

instead of *das Brot stehlen/klauen* or *sich das Brot nehmen* (with the reflexive *sich*) (**Gem 14**, cf. Table 32).

<b>Gem tag</b>	<b>Description</b>	<b>Target</b>
3 foreman	person in charge at shipyard	Vorarbeiter, Kapo, Polier
5 ship	object being built	Schiff
14 steal	stealing the bread	das Brot stehlen/klauen
22 prison	reason for Charlie to go to prison	wird versorgt; ist sicher; das Leben ist einfacher
26 seat	Charlie offering the girl his seat	den Platz anbieten
27 hankie	Charlie offering the girl his handkerchief	das Taschentuch geben
31 flee	the girl's chance to get away	wegrennen, weglaufen, fliehen

**Table 32: Gems resulting in lexical/semantic transfer from English to German**

In relation to lexical transfer, an interesting case is the fact that two of the five bilinguals who mentioned the handkerchief Charlie offered the girl (**Gem 27**) at the first attempt produced *Handtuch* ("towel"). (Both subsequently noticed the mistake and corrected to *Taschentuch*.) Further, in contrast with the overwhelmingly uniform use of the collocation *er bot ihr seinen Platz an* ("he offered her his seat", **Gem 26**) by all controls and most bilinguals, two bilinguals use *Sitz/Sitzplatz* ("seat") instead of *Platz*, probably being influenced by the English cognate.

Lexical interference is probably also the cause for six (of 15) bilinguals referring to the foreman/boss on site as *Vormann* instead of *Vorarbeiter* (**Gem 3**). It is possible that they intended to use *Vormann* in its meaning as a generic boss, for which it is sometimes found in journalism; however, for this meaning it was not used by any of the controls who referred to the *Chef* or *Leiter* (boss) instead (otherwise using *Vorarbeiter*). One bilingual used *Vormann* and *Vorarbeiter* for the same person, indicating either interference or semantic transfer, which probably also applies to the remainder of the group. Those bilinguals who did use *Vorarbeiter*, all paused either immediately before the word or between prefix and root (five bilinguals), which could be related to suppressing the unwanted lexeme.

Not an error, but a divergence from native-speaker norms in the frequency of the use of words, most likely as a result of the bilingual group adapting to the relevant L2 norms, is **Gem 5**, the

reference to the object being built at the shipyard. While the vast majority of controls refer to the ship as *Schiff* ("ship"), the bilinguals are divided between *Schiff* (seven bilinguals) and *Boot* ("boat"; six bilinguals).

The official distinctions between ships and boats (size, number of officers, conventional use of *boat* for submarine despite the fact that it is a ship) are virtually identical in German and English; however, English also uses *boat* affectionately for *ship*, which German does not. A Google search as a rough estimate of word frequency and a check on the WebCelex database (Baayen, Piepenbrock et al. 1995) confirms that in German *ship* is three to four times more frequent than *boat* (12.5 million vs. 4.4 million Google; 575 vs. 155 WebCelex), while in English, *boat* and *ship* are used with similar frequency (165 million vs. 160 million Google).

A difference in frequency of use was also observed in relation to **Gem 13** (girl). The most common labels are *junges Mädchen* ("young girl") and *junge Frau* ("young woman"), alongside the much rarer *junge Dame* ("young lady"). Interestingly, and without any obvious explanation, the bilinguals use both labels with roughly equal frequency (11 and 9 times respectively), whereas the controls have a clear preference for *junge Frau* (4 and 14 times respectively).

The control group, on the other hand, produced several instances of errors which are occasionally made by native speakers, but would be corrected by teachers of German as a native language, and which have found their way into web fora promoting the correct use of German: *zumindestens* (at least) which is a contamination of *zumindest* and *mindestens* (both meaning "at least"); *Essenstisch* ("dinner table", for *Esstisch*). *miesmutig*, another contamination between *mies* ("bad") or *miesepeterig* ("crabby, cranky") and *missmutig* ("querulous"), appears to be novel or idiosyncratic. There are also slips of the tongue, e.g. using *Werft* ("shipyard"), instead of *Schiff* ("ship"); *zum Fenster hinaus* ("out of the window"), instead of *zum Fenster hinein* ("into the window"), and word uses which are simply not correct in the context, e.g. *zwiespältig* ("contradictory"; of the people in the prison van), instead of *zweilichtig* ("shady"), *Rotation* ("rotation") or *Straucheln* ("tripping", usually of horses), for the movement of the prison van, instead of *Schwanken* ("rocking motion").

Errors on form are the second biggest category of errors after lexical and semantic errors in both groups. Here, few systematic differences in use were found. Many of the errors in both groups occur in the context of lexical searches and constitute articles or verbs uncorrected for case or

number after choosing a non-matching word (Example 26, control group; Example 27, bilingual group). As we argued above, native speakers normally attempt to avoid this type of mismatch, since it is very noticeable. Another common error, which is quite certainly connected to processing load, is the use of a pronoun of the wrong gender but in the correct case when referring to one of the main characters (cf. Example 20 above where the error was corrected; more often speakers did not notice it).

26) \*XYZ: ja, könnte ja auch noch < (ei)ne > [\*] # Grund gewesen sein .  
(yes, (that) might have been another-fem # reason-masc [\*])

27) \*XYZ: und äh@fp < ^d [x 2] > [//] äh@fp während < die Polizei > [//] der Polizist wieder  
irgendwie jemanden < anrufen > [\*] .  
(and er while the police [//] the policeman are [\*] somehow calling somebody  
again)

A borderline, and indeed quite tricky case is the agreement of the pronoun with *das Mädchen* ("the girl") which is grammatically neuter, but pragmatically feminine. Few people in either group avoided referring to the girl (if they had chosen that lexeme) as *she*, and this was accepted as long as it did not occur in a dependent clause where the gender clash would have been obvious.

Further, there is evidence of wrong plural formation, *mit Hammern* ("with hammers") instead of *mit Hämmern* (bilingual group), *Zimmern* ("rooms") instead of *Zimmer* (bilingual group), *Vagabunde* ("vagabonds") instead of *Vagabunden* (control group), as well as verb formation, e.g. *schwank* ("waved", with *hat*) instead of *schwenkte*, i.e. using a non-existent "strong" past tense form instead of the normal "weak" one (control group).

Two types of form errors interestingly are only attested in the control group: use of past perfect instead of simple past or perfect, and confusion of dative and accusative objects. Both have been noted in some dialects, e.g. the former is generalized in some Central German dialects, and the latter is systematically confused in the Berlin dialect, and fused in some low German dialects (Clyne 1995), which might explain the source of these errors. However, given that the bilinguals and controls were matched on their region of origin, one might have expected these errors to occur in both groups.

If regional influences are not the source of these errors, an alternative explanation to consider for the use of tense is that it reflects the individual's perspective in relation to the order of events in the narrative. Jumps alert the listener to an interruption in the chronological retelling, which can be put to good expressive use. However, if this was the intention behind the uses of past perfect that occur in our data, this was not successful – rather, they distort the actual chronology (Example 28), or present a contradictory combination of conjunction and tense (e.g. *während* (while) plus past perfect: *während sie sich die Augen trockengetupft hatte* – "while she had dabbed her eyes dry").

28) \*XYZ: < und sie > [//] < er ist > [//] sie wollte schnell weglaufen, dass sie zusammengestoßen sind, ne [: nicht] ?  
 (and she [//] he is [//] she wanted to get away fast (so?) that they bumped into one another, right?)

\*XYZ: er < hatte > [\*] sie ja im [: in dem] Grunde genommen dadurch aufgehalten .  
 (he basically had kept her (from getting away) that way.)

This sequence of events is not entirely logical, since Charlie's foiling the girl's attempt to get away did not precede their bumping into one another, but was concurrent with it or even the effect of it.

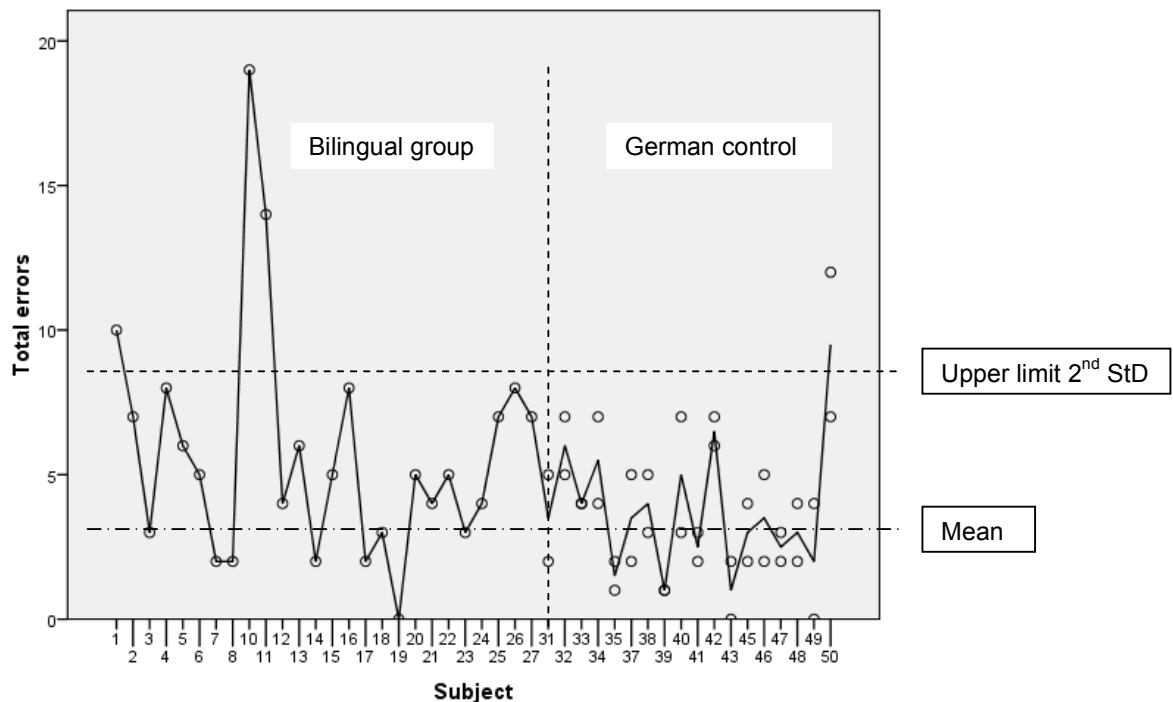


Figure 11: Total errors German film retelling

To conclude, we again examine individuals' performance on this measure. As we had seen previously in respect of the other tasks, there is again much variability between individuals in both groups on this measure (cf. Figure 11 and Table 33). The scores of the German control cluster around the 3-4 error mark and would show a normal curve except for one individual, S50, one of whose scores by far exceeds the upper 2<sup>nd</sup> StD. The majority of S50's errors come from the morpho-syntactic area, and he is often the only control producing a particular type of error (passives, verb past tense forms, regular confusion of dative and accusative).

We have no explanation for this performance, other than taking it as evidence for the existence of relatively poor native speakers (to which every teacher of German as a native language will testify). On the other hand, there are five individuals who, even on a task such as this requiring the integration of a lot of different demands, manage to produce a near-perfect retelling. It is presumably owing to these demands that the error rate is not lower overall, as one might naïvely have expected from a group of "normal", non-attrited native-speakers.

<b>Number of total errors</b>	<b>Bilingual group (N = 26)</b>		<b>German control (N = 36)</b>	
0-1	1 (3.8%)	S19	5 (13.9%)	
2-3	7 (26.9%)		13 (36.1%)	
4-5	7 (26.9%)		11 (30.6%)	
6-7	5 (19.2%)		6 (16.7%)	
8-9	3 (11.5%)	S4, S16, S26	—	
10-11	1 (3.8%)	S1	—	
12-13	1 (3.8%)	S11	1 (2.8%)	S50/2
14-15	—		—	
16-17	—		—	
18-19	1 (3.8%)	S10	—	

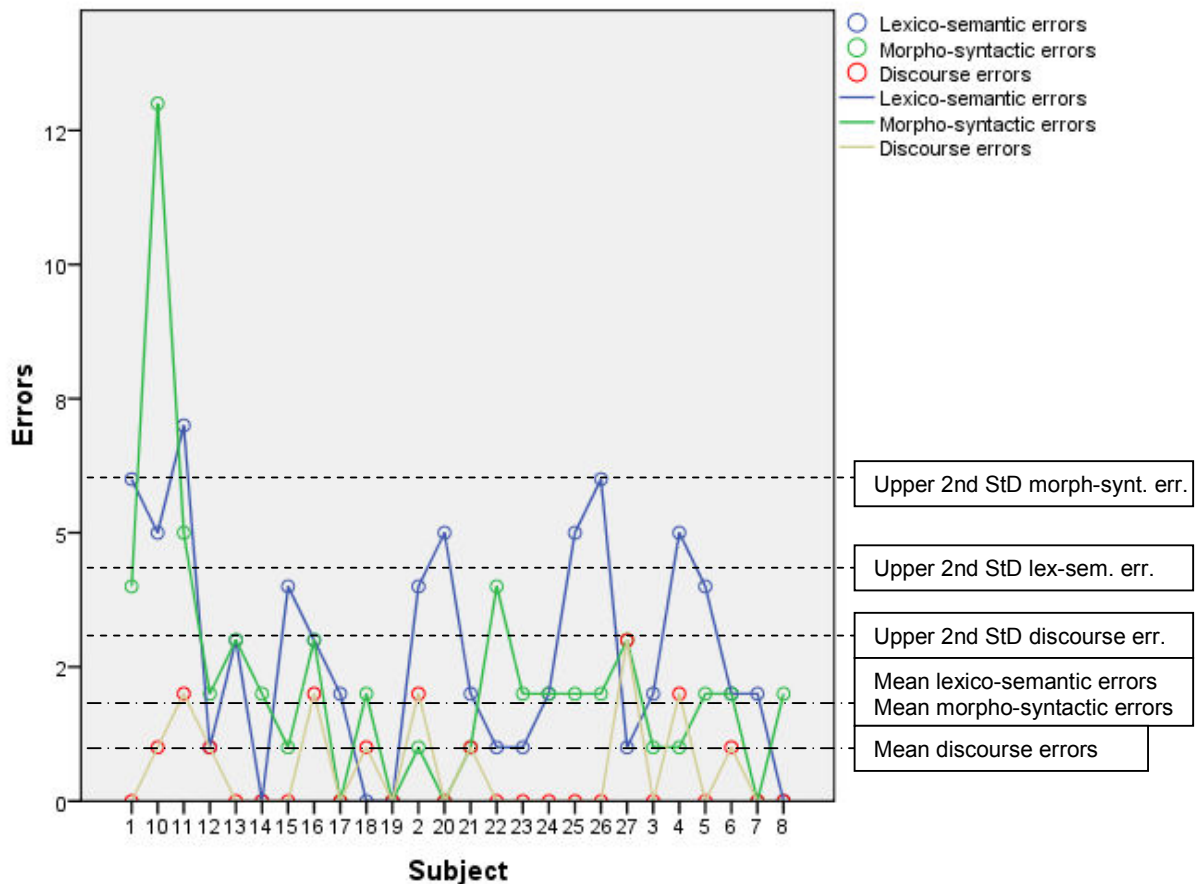
**Table 33: Total errors German film retelling by individual**

On the part of the bilingual group, most scores are also distributed around the 3-4 error mark; however, there is less homogeneity. There is only one participant, S19, who produces a near-flawless retelling, and quite a few more with higher error rates. S10 and S11 had already been earmarked as potential cases of L1 attrition on the basis of the fluency and lexical diversity measures.



S10 ranks first for the morpho-syntactic measure, owing to word order and other syntactic transfers (13 errors), but has a tied third place (5 errors; rank 7.5) on lexico-semantic measures. S11 comes first on lexico-semantic measures (7 errors), and second on morpho-syntax, owing to lexical interferences and agreement errors (6 errors).

High up on both measures (second and third place, 6 and 4 errors, respectively) is S1; again, transfer is playing a big part. S4 and S26's profile is similar to S11's; S16's to S10's, but at lower error rates. S20 and S25, whom we had mentioned under the fluency measures above, have the same error rate as S10 in the lexico-semantic area, but much lower morpho-syntactic errors (S25: 2 errors; S20: 0 errors).



**Figure 12: Errors bilingual group film retelling German by type**

The error analysis has so far given the clearest evidence of the impact of that "other" language in attrition situations. While there is some overlap in the types of errors made, the expected types

almost exclusively occur in the bilingual group, albeit to different degrees for different individuals. On the other hand, some individuals in the control group perform very poorly compared to a number of the bilinguals.

#### 4.3.2.3 Gem analysis

In this section, we turn to the gem analysis proper. Not all the gems analysed proved interesting, but some interesting areas for comparison emerged: reference, lexical diversity, and lexical retrieval/phrasing difficulties. Continuing the above discussion of gems in relation to code-switches and errors, we now present a sub-set for these areas (cf. Table 34).

<b>Gem tag</b>	<b>Description</b>	<b>Target</b>
1 shipyard	Charlie's first place of work	<i>(Schiffs-) Werft, Schiffbauer</i>
2 trustworthy	description of the letter; description of Charlie's character	<i>Empfehlung (~sschreiben); ehrenhaft/ehrenwert, vertrauenswürdig</i>
4 wedge	first job	<i>Keil</i>
6 stuck	position/location of the wedge	<i>feststecken, das Schiff festhalten</i>
8 exwedge	removal of the wedge	<i>herausschlagen</i>
13 girl	description of the girl	<i>die junge Frau, das junge Mädchen</i>
17 witness	description of the lady witnessing the theft	<i>Zeugin</i>
29 unconscious	everybody's state after the accident	<i>ohnmächtig, bewusstlos, benommen</i>
33 hit	knocking the policeman out again	<i>über die Rübe/ den Kopf hauen/ schlagen, eins übergeben/ überziehen, ihm eine verpassen</i>
35 grass	naming the grass verge	<i>am Weges-/ Straßenrand</i>
36 dream	the fantasy of Charlie and the girl living together	<i>Traum (~szene, ~sequenz), Paradies, Schlaraffenland</i>
37 fruit	fruit growing in through windows	<i>zum Fenster hineinwachsen</i>
39 cow	the cow producing milk without being milked	<i>nicht melken brauchen, sich selber melken</i>
40 jug	the container for the milk	<i>(Milch-) Krug, Kanne, Topf, Messbecher</i>
41 loitering	the reason for the second policeman to object	<i>herumlungern</i>
43 Minna	the police van	<i>Polizei-/Gefangenenwagen, -fahrzeug, -transporter; Grüne Minna</i>
44 scene	change of scene, getting in to dream sequence, return to reality	<i>eine neue Szene einblenden, Schnitt</i>

**Table 34: Gems analysed for lexical diversity and retrieval problems (German)**

With regard to lexical retrieval or phrasing, we had already given a few examples in the sections above, and several of the gems in Table 34 presenting explicit and implicit retrieval difficulties are also interesting from a lexical diversity point of view. "Pure" retrieval cases are **Gems 1** (shipyard) and **4** (wedge). S1 had a big problem remembering both these words as well as some more (*tray*), commenting with exasperation that she had started feeling quite confident about her German again after a trip to Germany (Example 29) (but now realised that it is still somewhat rusty):

- 29) \*XYZ:     und < ich &da > [/ -] ich hab mich so gut gefühlt in der Woche .  
                   (and I tho(ught) [/ -] I was feeling so good this week)
- \*XYZ:     und dachte, ich könnt(e) wieder so gut Deutsch !  
                   (and thought I spoke such good German again!)

S1 claimed not to know either *Werft* or *Keil*, even after being provided with them (Example 30), but instead retrieved the English equivalent *wharf/shipyard* or considered a related German word (*Keilriemen* – "V belt"/"transmission belt" in engines) (Example 31):

- 30) \*XYZ:     wie nennt man das in Deutsch ?  
                   (how do you say in German?)
- \*INV:     Werft .  
                   (shipyard .)
- \*XYZ:     Werft !  
                   (shipyard !)
- \*XYZ:     mit (ei)nem T@l !  
                   (with a T!)
- \*XYZ:     das Wort < kennt ich jetzt auch > [//] kannte ich auch net [: nicht] .  
                   (that word I also did not know now.)
- 31) \*XYZ:     ja, < a@s wharf@s > [\*] # sagt man auch.  
                   (yes, they also say "a wharf".)
- \*XYZ:     Keilriemen is(t) was ganz Anderes .  
                   ("Keilriemen" is something totally different.)

Similarly, S17 makes explicit reference to not knowing or remembering *wedge* and *trustworthy* (**Gem 2**). All three gems (shipyard, wedge, trustworthy) in fact seem to have presented difficulties for more bilinguals and some controls, judging by hesitation patterns, non-specific references (*da – "there"*, for *shipyard*, S26) and various circumlocutions/paraphrases.

Thus, for *wedge*, individuals in both groups produced solutions such as *ein dreieckiges Stück Holz* ("a triangular piece of wood"), *ein Stück Holz von einer bestimmten Länge/Form* ("a piece of wood of a particular length/shape"), and *ein Stück Holz, genau wie es ihm gezeigt wurde* ("a piece of wood exactly as he had been shown"). One control remembered *Kantholz* ("square timber") but could not think of the target word. It is not apparent why two controls chose to refer to the wedge as *ein keilförmiger Holzstab\** ("a wedge-shaped wooden staff\*"), or *ein Stück keilförmiges Holz* ("a wedge-shaped piece of wood"), both of which contain the target word, rather than simply *Keil*. *trustworthy* gave rise to paraphrases such as *dem man vertrauen kann* ("who may be trusted"), or was omitted in favour of the other description from the letter "honest" (*ehrlich*). This arose mainly with the bilinguals rather than the controls, for reasons to be discussed below under lexical diversity.

One of the reasons for why these items caused problems is probably the fact that they are relatively infrequent (number of occurrences in the Web-CELEX corpus at <http://celex.mpi.nl/> (Baayen, Piepenbrock et al. 1995): *vertrauenswürdig* – 12 (however, the set with the root *Vertrauen* ("trust") is larger than for the other two), *Keil* – 20, (*Schiffs~*) *Werft* – 96).

Mainly, retrieval problems arise with whole words, but not exclusively so. For example, one bilingual attempted to describe the manner in which Charlie, the girl and the policeman found themselves on the ground having been thrown out of the police van, presumably something like "rolled up" or "piled up" but abandoned the search having provided the first part of the verb compound *zusammen-* ("... up") including the prefix for the past tense participle *ge-* (Example 32):

32) \*XYZ: ähm@fp alle drei bewusstlos im [: in dem] Haufen &zusammenge^ äh@fp +...  
(erm all three unconscious er ... up in a heap)

\*XYZ: na ja .  
(anyway.)

One bilingual, S3, had a problem coming up with a German equivalent for the offence of *loitering* (**Gem 41**) causing her to code-switch. This is probably related to the fact that "loitering" is a word that would (subjectively) require some effort to learn, but would then be remembered well. The phrase *No loitering* which S3 referred to, can also be found on frequent signs around Ireland, reinforcing the word and therefore possibly causing it to compete in the mind with the rather infrequent (WebCelex count – 2) direct German equivalent (*he*)*rumlungern*. Additionally, there is no direct correspondence for the offence in German, the closest possibly being *Hausfriedensbruch* ("breach of the peace"), *Erregung öffentlichen Ärgernisses* (legally "indecent behaviour", but the words themselves could imply causing annoyances for different reasons), or, if phrased as a prohibition: *Betreteten verboten* ("No entry", literally "Entry prohibited").

The other bilinguals worked around the problem by stating that the policeman objected to their *sitzen* ("sitting") on the grass/lawn/meadow. This was also the reason given by most controls, some of whom instead used the word in question *herumlungern* ("they must not loiter"). Additionally, however, the controls usually spelled out why they must not sit or loiter there – because it is a chic area and because they are *Vagabunden*, *Landstreicher*, *Penner*, *Rumtreiber*, *Gesindel* or *abgerissene Gestalten* (all meaning "vagabonds", "tramps"), displaying a great command of terms for itinerant people.

**Gems 37** (fruit) and **39** (cow) caused difficulty in coming up with a satisfactory solution for participants in both groups. This is probably connected to the fact that both relate to improbable scenes, which some people found hard to perceive (i.e. they did not necessarily take in the detail) and/or to formulate. **Gem 37** presents an improbable scenario in that the branches of a tree and vine literally grow in through the window/door, which is a complex phrase containing a verb (*to grow*) that acts like a motion verb; **Gem 39**, because the cow does not require milking but dispenses her milk automatically into a jug.

For **Gem 37**, the solutions are fairly uniform across the two groups. Some participants chose to refer to the unusual growth pattern of the fruit plants: *zum Fenster/zur Tür hineinwachsen* ("grow in through the window/the door"), or *ins Zimmer/Haus hineinwachsen* ("to grow into the room/house"). This was usually accompanied by retracings on the function words and the "destination" of the growth. Several used different prepositions for this phrase: *durch die Tür*

("through the door") which is acceptable, and *in das Fenster* ("into the window"), which probably is a contamination or mixing of phrases. Others used verbs which are incompatible with their subject: *der Baum hängt ins Zimmer* ("the tree hangs into the room") instead of *die Äste* ("the branches") for which *to hang* this is a normal verb. Others again avoided the problem by choosing to keep the plants outside in the garden by stating they were *vor dem Fenster/der Tür* ("outside the window/door"), or simply *im Garten* ("in the garden"), or by making no reference at all to their location.

Similarly, **Gem 39** (cow) gave rise to any number of different phrases noting either the fact that the cow did not require milking/milked itself; the process of the milk *spurting* into the jug; or the jug "filling itself as if by magic" (*fällt sich wie von Zauberhand*). It is worth noting that, apart from the expressive reference to magic just given, most bilinguals used simple expressions such as *von allein, von selbst*, ("by itself"), *automatisch* ("automatically") or *einfach* ("simply") for the absence of human action, whereas the controls used a wider array of phrases: *auf Kommando* ("on command"), *quasi auf Knopfdruck* ("on pressing a button, so to speak").

The last example already hints at a pattern relating to lexical diversity and reference, found in many of the gems. The quantitative analysis of this feature presented above had not found any group differences, only a non-significant trend for the bilinguals' scores to be somewhat lower. However, comparing the solutions proposed for particular items yields several trends which we will now turn to.

The first trend is for the bilingual group to be fairly homogeneous in how individuals refer to certain items, and for the control group to produce a greater number of types. This is partially expected since there are more participants in the control group than the bilingual group on this task. However, where differences are very pronounced, they deserve attention. This trend holds for most items in Table 34. For example, on **Gem 29** (unconscious), the bilinguals used just two of the possible ways to refer to somebody being unconscious or dazed: *ohnmächtig* and *bewusstlos*, whereas the controls additionally used *benommen, nicht bei Bewusstsein, nicht ganz da*. One bilingual presumably intended to use *nicht bei Bewusstsein* but ended up saying *nicht bei Sinnen* ("not in their right mind"/"out of their senses"), which is semantically related but different.

Interestingly, some participants in both groups evidently believe that one may be "a little bit ~" or "partially unconscious".

Similarly, when referring to the dream painted by Charlie for the girl (**Gem 36**), the bilinguals use *Traum* ("dream") and related words, *Fantasie* ("fantasy"), *Paradies* ("paradise") and derivations as well as *Schlaraffenland* ("land of plenty"/"~ of milk and honey"). The controls used these too, but additionally use *Wunschvorstellung* ("imagining a desirable future"), *Tagtraum* ("day dream"), *heile Welt* ("(the idea of) an ideal world") and *Garten Eden* ("Garden of Eden"). Surprisingly, given that this is not an idiom in German but appears to be derived from English, two controls name the place *wo Milch und Honig fließen* ("where milk and honey flow").

In relation to **Gem 33** (hit), there are very many idiomatic ways to refer to the policeman being knocked out. The bilinguals came up with six, the controls with 16 different expressions; and one of the bilinguals' solutions (from S11) is not quite correct and marked by hesitations and repetitions during production ("*gibt ihm einen Stoß\* mit seinem eigenen Schlagstock*" – "pushes\* him with his own baton").

The second trend is for the bilingual group to use concrete references, and for the controls to use more abstract language. **Gem 6**, describing where the wedge that Charlie knocked out was placed, created considerable difficulty for most people, probably again because few people would frequently have to describe such situations. On this occasion, the bilinguals exclusively used verbal constructions of the type *the wedge holds/keeps/supports a ship/beam/other pieces of wood*, or *the wedge is stuck/wedged in under some wood*.

While a means also used by the controls, some individuals in that group did not refer to the location but used an abstract expression referring to the purpose of the wedge in that place, usually when stating how the wedge was removed: *löst/schlägt den Keil aus einer Konstruktion/Halterung/Verankerung/Sperre, die ein Schiff festhält/...* ("removes/knocks out the wedge from a construction/support which keeps/... a ship in place"), or *das war ein haltender Keil/eine Stütze* ("that was a supporting wedge/a support").

This trend is also present in **Gem 2** (trustworthy), in naming the letter Charlie brought to his new place of work. Most bilinguals call it simply *Brief* ("letter") or a sort of reference *Empfehlung/Referenz* (seven and six respectively), four using a more formal word for letter

(*Schreiben*). In contrast, the controls almost invariably specified that it was a letter of reference, often in lieu of spelling out the text of the letter, which we had already mentioned above. When they did use *Brief*, they subsequently (bar two) rephrased it as *Empfehlung/~sschreiben*.

A related phenomenon is the third trend we noted: the bilinguals tended to use more generic, unmarked (and more frequent) terms whereas the controls used more specific vocabulary. For example, in **Gem 35** (grass), the most likely way to refer to the place where Charlie and the girl sat down (apart from *unter einem Baum* ("under a tree")/*vor einem Haus* ("in front of a house") which were used with equal frequency in both groups) was *ins Gras/auf ein Stückchen Gras/einen Grasflecken* ("onto the/a piece of grass"). The controls, on the other hand, used *am Straßenrand/am Wegesrand* ("at the side of the road", three bilinguals vs. nine controls) with greater frequency, plus additionally *auf dem Seitenstreifen/Grünstreifen* (five controls) which are terms usually associated with road markings. Two controls had them sit in the front garden, and if they referred to the grass, it was as often as *Rasen* ("lawn").

This trend for controls to deploy more specific and/or more technical vocabulary is also noticeable in **Gem 44** (scene) with most bilinguals simply referring to the change in scene with *die nächste Szene ist ...* ("the next scene is ..."), *und dann geht es zurück in die Wirklichkeit* ("and then it goes back to reality"). The controls, on the other hand, deploy terms from film production in greater measure (of course, *Szene* itself is borrowed from that area): *als Nächstes wird eingeblendet* ("next they show", from *Blende* - "aperture" on a camera), *Schnitt* ("cut"), *ein Schwenk* ("pan" (to the next scene/picture – *Bild*)).

The trends may also combine: for **Gem 40** (jug), most bilinguals used *Krug* ("jug", seven), just one each mentioned a *Glas* ("glass"), *Kanne* (another word for "jug", usually one with a long spout). These were also used by the controls, but these also used *Becher/Tasse* ("mug"), (*Milch-*) *Pott/Topf* (pot"), *Karaffe* ("carafe") and *Gefäß/Behälter* ("vessel"), some being more specific, and others more abstract than "jug".

The trend for specific reference is particularly noticeable in how people are referred to. In relation to **Gem 17** (witness), the lady who had observed the girl stealing the bread is called just that by almost everybody on the first occurrence: *die Frau/Dame, die gesehen/beobachtet hat, wie das Mädchen das Brot gestohlen hat*. On the second occurrence, however, the bilinguals stay with



*Dame/Frau* ("lady"/"woman"), whereas several controls switch to calling her *Zeugin* ("witness"). Also, some controls refer to the lady as *Passantin* ("passer-by").

Ascribing and specifying specific roles/jobs/ranks can also be observed in **Gem 13** (girl) which some controls later on called *Delinquentin* or *Täterin* ("the delinquent"), or in relation to the people inside the prison van: *Leute* ("people") on the one hand, vs. *Passagiere* ("passengers"), *Straffällige* ("criminals"), *Verhaftete/Gefangene* ("arrestees"/"prisoners") on the other. The policeman is referred to by participants in both groups as *Polizist* ("policeman"), *Wachmann/Wachmeister* ("security guard"), *Schutzmann* ("constable"), but only the controls use *Polizeibeamter*, the official term for a police officer. (We had already discussed the use of *der Officer* above.)

**Gem 17** (witness) and **Gem 8** (exwedge) yield a fourth trend – the use of more colloquial or colourful language for effect by the bilinguals. While essentially the same solutions were used for the removal of the wedge: *heraus-/los-/weg-/freischlagen/~hauen* ("to knock out"), the bilinguals favour the more colloquial lexeme *hauen*, and the controls *schlagen*. The witness, on the other hand, is given an expletive (*die alte Schachtel* - literally "the old box"; *die dumme Kuh* – "the stupid cow") by two of the bilinguals. A third use is probably a mistake and completely inappropriate since it refers to gay people (*die alte Schwuchtel*) (perhaps *die alte Schachtel* was the intended target).

In this context, and finally, it is worth pointing out **Gem 43** (Minna) which exemplifies both the trends for a greater number of word types in the control group, and for colourful language in the bilingual group. In relation to the former, the most common lexemes to refer to the prison van in both groups are (*Polizei-/Polizisten-) Wagen/Auto* ("police van"). *Gefangenenwagen/Gefängniswagen* ("arrest/prison van") is also used. The controls additionally use different lexemes for *van* in both compounds, such as *Transporter* ("van", "truck") and *Fahrzeug* ("vehicle"), as well as for *prison*: *Knast*. Further, they supply *Mannschaftswagen* ("personnel carrier"), *Einsatzwagen* and *Streifenwagen* (both "squad" or "patrol car" and not quite correct here) which were not mentioned by the bilinguals.

Only the bilinguals, on the other hand, referred to the police car (or at least tried to refer to it) using slang terms from different cultural contexts. This created difficulties, maybe because of the clash of language, for one bilingual who chose to call the van *Black Maria*, but translated the phrase into German, at the first attempt produced *Schwarze Emma* ("Black Emma") and at the second *Schwarze Mary* ("Black Mary", as in the cocktail). When asked about the German equivalent *Grüne Minna*, he

was unable to retrieve it. This, however, was provided by four of the bilinguals (and none of the controls). A further participant used the American/Irish term *paddywagon*, germanised as *Paddy-Wagen*, and subsequently explained it was a *Polizeiwagen*.

#### **4.4 Summary**

The results of the three tasks indicate that the bilingual group is able to perform comparably to the German control group overall since few significant differences were detected on the chosen measures. The only measures resulting in a significant group difference were one fluency marker (repetitions) and some types of errors (lexical interferences, semantic transfer, word order).

However, given the small sample sizes on some of the individual conditions of FiCA2 and FiCA3, the absence of a significant effect for these conditions cannot be seen as a conclusive result.

Moreover, on each task, there is a non-significant trend for the bilingual group to score a little lower, take a little longer, pause (or repeat) more, have more problems retrieving lexical items, use less diverse/more prototypical language and in many domains demonstrate crosslinguistic influence.

Thus, with regard to the four hypotheses put forward at the beginning of the chapter, we find partial support and the need for modification.

a) The bilingual group will achieve lower scores in and/or took longer to complete the C-test.

This hypothesis is not supported at the group level; however, there are some individuals who fall below the native-speaker range on either one or both of the measures. Further, the qualitative analysis of item solutions highlighted differences in the level of difficulty of some items, often apparently connected with retrieval problems, crosslinguistic influence or insufficient monitoring. There is evidence for difficulties affecting all linguistic levels.

b) The bilingual group will achieve lower scores in the verbal fluency tasks.

This hypothesis is not supported since the individual results for each condition fail to reach significance. However, combined scores for each task show that more bilinguals are represented in the lower score bands, and more controls in the higher bands. Again, some bilinguals perform

rather poorly on this test. The particular items named by the two groups partially reflect differences in the environments they find themselves in, and the bilingual group displays signs of some level of difficulty in retrieving particular items, resulting in failed, partial or incorrect retrieval.

c) The bilingual group will demonstrate lower lexical diversity in the verbal fluency tasks/the film retelling.

This hypothesis was confirmed with regard to the verbal fluency task, where the bilinguals were shown to use more prototypical vocabulary, but not in relation to the film retelling, where both groups are comparable. However, the analysis of "gems" contained in the film retellings shows that, for several of the gems, the bilingual group uses less diversified, less specialist language, with the exception of some examples of "colourful" language.

d) The bilingual group will display a higher rate of errors, code-switches and hesitation phenomena in the film retelling.

This hypothesis was confirmed. The bilingual group does have a higher rate of lexico-semantic and word-order errors, which translates into a significant difference on errors overall. It also has more code-switches, which it used for different purposes than the control group. Amongst the hesitation phenomena, only repetitions yielded a significant difference, but the bilingual group was found to have many more lexical retrieval difficulties. The gem analysis provided further evidence for differences between the groups, for example adoption of L2 norms in language use by the bilingual group, and trends relating to lexical diversity mentioned under c).

These results are consistent with previous studies into L1 attrition with comparable groups of bilinguals in three areas. First, despite the perception, not least by the bilinguals themselves, that L1 attrition is a "ubiquitous" phenomenon (Seliger 1991: 227), its existence has been exceedingly hard to prove in adult bilinguals, and, when attested, effects were small (Schoenmakers-Klein Gunnewiek 1998; cited in Köpke 2004: 1341; Köpke and Schmid 2004). Second, while effects may potentially be found at all linguistic levels, the lexico-semantic level is most affected (Köpke and Nespoulous 2001; Köpke 2002). Finally, crosslinguistic influence has been found to account

for the majority of, though not all, effects (Sharwood Smith 1983; Sharwood Smith 1989), which is also supported by our data.

An important result is the great variability between individuals in both groups, increasing as online task demands increase (the control group's performance on the C-test is relatively homogeneous). While this partially explains the absence of any significant group differences (cf. below), it also highlights the fact that there is a continuum of native speakers who may differ in their mother tongue capacities for a number of reasons, of which relative input deprivation and the impact of other language systems is just one. Some of the specific effects found also point to continuity in the groups' behaviour, since often there is just a distributional difference on items which might be seen as borderline/infrequent, reinforcing the idea of attrition in the area of grammar following the direction of "drift" (Clyne 1992: 19).

Our findings in the main also follow the predictions for L1 attrition arising out of both the Activation Threshold Hypothesis, and DMM, discussed previously in Chapter 2. To start with the former, certain features of the bilingual data, such as disfluency and L2 interference, may result from the proposed (inter)action of activation thresholds in bilingual systems in an L2 environment, in which the entire language system is assumed to become less active due to infrequent L1 use, or to be inhibited through regular L2 use.

Thus, retrieval difficulties can be brought about by too high an activation threshold of the L1 overall and/or of particular language items. Such items include "rare", infrequent ones that one might not use in daily life and whose activation threshold therefore is higher to start with. As the gem analysis for the film retelling showed, retrieval difficulties occurred more often with such "rarer" items, such as "wedge" or "tray". "Rare" cases are also those where the participants were required to retell accounts of fantastic or improbable events or situations. Similarly, the observed trend for reduced lexical diversity may also be explained by a higher L1 activation threshold for such items, with only the more prototypical and frequent items being capable of being recalled.

Interferences of various types may be traced back to the inhibitory effect of frequent L2 use. The accidental code-switches we found, for example, most likely resulted from the presence of a more available L2 competitor whose activation threshold was below that of the German target and which

thus "succeeded" in "interloping". It is interesting that cognates lent themselves to this inadvertent use of the L2.

There were also several instances of a participant being unable to recall an item in German when it was clearly present in English, particularly when it was an item well anchored in Irish reality (e.g. "loitering") and/or a fairly infrequent and probably unrehearsed word in German. In these cases the English language lexical item apparently "eclipsed" the German one, which plausibly may be the result of an unfavourable constellation of activation thresholds for these items in German and in English. Finally, the syntactic interferences discussed in relation to the film retelling may also be seen as instances where the English sentence structure is more readily available.

Thus, our data may be read as tentative evidence for the proposition that the activation level of languages, and of items within them, may change in response to the communicative specificity of an L2 environment, and that, under specific circumstances, the L2 activation threshold may fall below that of L1, causing specific effects both in language processing, and possibly longer-term, as hypothesized in Chapter 2. While we have shown the effect of this particular constellation of activation thresholds on L1 processing by way of the above examples, the opposite would hold for the L2 and will not be discussed further.

If particular linguistic features and behaviours may be interpreted within the Activation threshold hypothesis, then DMM may be drawn on to provide an explanation for the absence of clear group differences. First, this may simply result from the particular constellation of initial L1 proficiency and the maintenance effort required and expendable in an L2 environment as proposed in DMM. According to its (negative) growth model, small changes in L1 proficiency significantly decrease the maintenance effort required to keep the language at the new level. Furthermore, it is proposed that at maximal levels of proficiency the curve is almost flat, so lack of maintenance takes relatively longer to manifest itself than at lower levels of proficiency. It is therefore possible that even reduced L1 input is still sufficient to keep the L1 at a merely slightly reduced level. We will consider the role of input, and its predictive power for our data, in Section 6.1.2 below.

Second, the absence of larger effects may not necessarily mean that no attrition is taking place. In fact, the heterogeneity of the bilingual group, which on most measures exceeds that of the control group, is indirect evidence that something is happening. In recent work adopting a dynamic

systems approach to bilingualism, it has been argued that large effects are not the only way in which changes may express themselves, and that, due to the multifarious and multidirectional interaction of a multitude of factors, small effects are as important (de Bot 2008: 174). Both de Bot (ibid.) and Herdina and Jessner (Herdina and Jessner 2002), the originators of the Dynamic Model of Multilingualism, argue that large changes in language systems will be preceded by "increased scatter" (Herdina and Jessner *ibid.*: 96), and this has been borne out in computer simulations of attrition (Meara 2004) as discussed in Chapter 2.

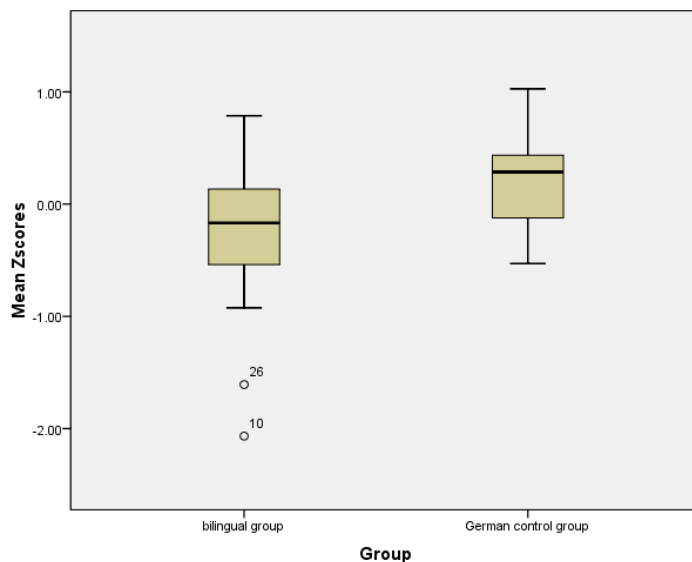
An additional explanation is the possibility that group comparisons may disguise differences found between individuals. Therefore, we felt it would be useful to gauge each person's overall L1 proficiency, in addition to reviewing the results for each test separately, in order to place them on a continuum relative to one another. While each test was designed to look at slightly different aspects of the same construct, we feel such a summary is warranted since together the tests attempted to produce a fairly comprehensive assessment of L1 proficiency. However, there are areas which could possibly have been explored with further tests or additional analyses, so by necessity the summary scores are relative to the current test battery and should not be viewed as an absolute expression of a particular person's proficiency.

For this analysis, we excluded participants who were missing data for 50% or more of the variables summarized, although the overall picture is unaffected by this precautionary measure. Since all the tests are scored differently, we first needed to convert all relevant scores to standardized measures (*z*-scores) to allow comparison. *z*-scores are used to identify the location of a participant's score relative to all other participants' (some argue, the overall population's) in a unit-neutral way by relating all scores to a mean of zero and an StD of 1 (Field 2005). The further away from 0 a participant's score on either side, the less representative it is. The *z*-score is a useful tool for identifying outliers (Field *ibid.*: 76) and for comparing data across different tasks that use different measurements.

The measures to be compared were C-test scores, C-test time, FiCA score, Charlie "raw" VOCD, Charlie hesitation index, Charlie error rate. Having converted these to *z*-scores, we then produced a mean score giving all variables equal value. Since there is a different number of variables for each task, we considered weighting the variables to give each task the same value; however, we decided

against that option since in reality the tasks were not equal in either duration or complexity. Another option would have been to choose just one variable for each task, but this would have given undue weight to some, while ignoring other, equally important features of participants' performance.

The summary scores thus produced yield very interesting results. Although we decided to average rather than sum the *z*-scores, any tendencies present in the individual tasks come out more strongly in the summary scores. There is now a much clearer difference between the two groups, which reaches significance at the .05 level ( $t(41) = -2.277, p = .014$  one-tailed). Figure 13 graphically represents the result of this comparison showing that the groups now overlap only partially. The groups' means are at  $-.195$  (StD  $0.654$ , range  $-2.09$ - $0.79$ ) for the bilingual group, and  $0.205$  (StD  $0.419$ , range  $-.53$ - $1.03$ ) for the control group. The dispersion of scores is such that the majority of the bilinguals' scores are below 0 (14/25), and the majority of the controls' – above (8/18).



**Figure 13: German groups' performance across all three German tasks**

The graph also illustrates that some individuals in the bilingual group fall outside the "norm", which we had previously defined as the range of the performance of the control group. Specifically, seven individuals, S10, S26, S1, S8, S7, S25 and S27, have a mean *z*-score below any of the controls, but the last four do come relatively close to the control with the lowest overall score, S40 (mean *z*-score  $-0.53$ ). S10, S26 and S1 (ranks 43-41, mean *z*-score  $-2.07$ ,  $-1.61$  and  $-0.92$

respectively) have been mentioned in several context as being possible cases of attrition, and this analysis confirms that their reduced performance relative to the rest of the group is consistent across several tasks and measures.

For example, S10 has the second-lowest C-test score of all (with S1), took almost the longest, has the highest number of errors and by far the highest number of hesitation markers in the film retelling. S1 took longer on the C-test, has the second-highest number of errors and a high rate of hesitation markers, while her VOCD is quite high. Finally, S26 has the lowest C-test score of all, the second-highest time and is similar to S1 in errors and hesitation markers.

At the other end of the range, we find some very proficient native speakers. While none of the bilinguals reach a score higher than the controls' range (S46, rank 1, mean  $z$ -score 1.03; S41, rank 2, mean  $z$ -score 0.85, three bilinguals, S6, S19 and S21 rank 3<sup>rd</sup>, 4<sup>th</sup> and 6<sup>th</sup> respectively (mean  $z$ -scores 0.79, 0.70 and 0.53 respectively); S16 occupies rank 9 (mean  $z$ -score 0.41), S24 rank 12 (mean  $z$ -score 0.35). We had come across S6 as the person with the highest FiCA score of all, but she has also got a very high C-test and lexical diversity score with a medium number of hesitations and errors. S19 excels on the C-test which he also solved very fast, and has no errors and few hesitations. S21 is comparable to S6, but instead of the extremely high FiCA score, she has the highest VOCD. We shall discuss these participants' profiles further in the section where we will compare proficiencies across the two languages.

It would appear, therefore, that the L1 of our participants is affected to varying degrees by the experience of living in an L2 environment. While on each individual measure differences are small, they produce a tangible group difference when combined, with a small sub-group showing some degree of L1 attrition.

In the next subsections, we compare the findings from the tests with the participants' perceptions of their own proficiency past and present, while the factors shaping L1 proficiency in the L2 context will be considered in Chapter 6.



## 4.5 Discussion

### 4.5.1 Self-reported L1 proficiency at time of testing (can-do scales)

All participants were asked to optionally fill out a can-do questionnaire (cf. Appendix for the English-language version of the questionnaire). For each statement of 43 statements relating in varying measure to the four language skills listening (8 statements), reading (7 statements), speaking (17 statements) and writing (11 statements), participants rated themselves on a 5-point Likert scale ranging from "1 = I cannot do that" to "5 = I can do that without any difficulty" (a description for each point was provided). The ratings were converted into percentage scores for each skill. 21 bilinguals, and 17 German controls returned the questionnaire, but only complete entries were included in the calculations below.

<b>Can-do</b>	<b>Bilingual group (N = 18)</b>	<b>German controls (N = 17)</b>
<b>Listening</b>		
Mean	97.78	97.94
Median	100.00	100.00
StD	3.82	3.77
Range	88-100	88-100
<b>Reading</b>		
Mean	93.65	97.65
Median	100.00	100.00
StD	10.07	3.53
Range	71-100	89-100
<b>Speaking</b>		
Mean	96.60	97.72
Median	99.41	100.00
StD	5.72	4.39
Range	79-100	82-100
<b>Writing</b>		
Mean	91.82	93.90
Median	97.27	96.36
StD	12.63	7.71
Range	55-100	71-100

**Table 35: Results of can-do statements German (percentage scores)**

On the basis of these self-ratings, there are no significant differences between the two groups on any of the sub-skills or averages of each skill (cf. Table 35 above), with the exception of one statement within the written proficiency section in favour of the control group: "I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points." ( $U = 131.000, p = .048$  one-tailed).

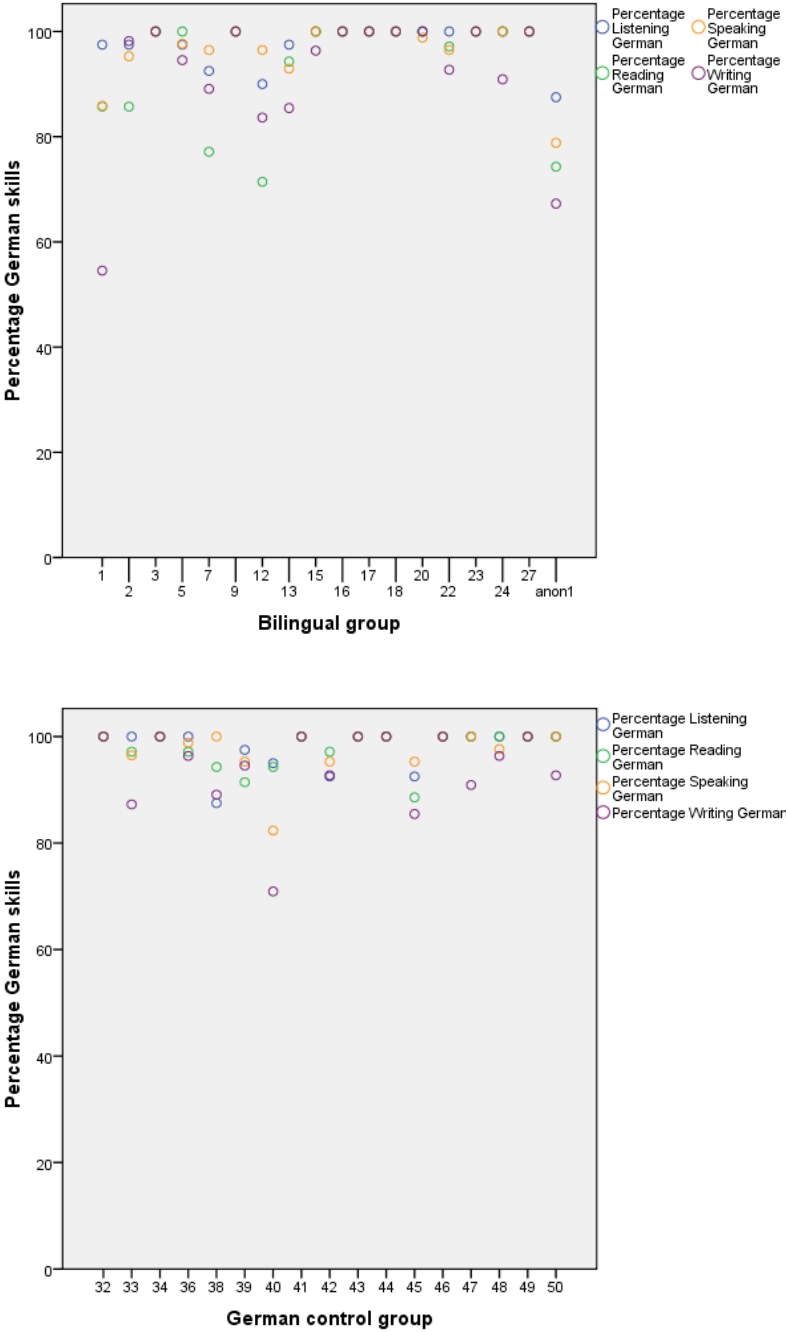
In contrast with more global self-ratings elicited from the bilingual group during the interview in response to Question 45 of the sociolinguistic questionnaire (5-point Likert scales ranging from minimal skill to native-like skill for each of listening, reading, speaking and writing), where all bilingual participants rated all their skills as native-like, except for three participants (11.1%), who rated themselves as advanced for writing, this table provides a more differentiated picture.

For the bilingual group, it suggests a hierarchy in the "complexity", and thus "vulnerability" of skills, which is reflected in decreasing self-ratings for Listening > Speaking > Reading > Writing. Such a hierarchy has been suggested in the acquisition and attrition literature (Yukawa 1997; Ecke 2004): the two oral skills listening and speaking are primary to the two literacy-related skills reading and writing and have been hypothesized to be easier to acquire and maintain. Receptive skills require less computational effort than productive skills and again ought to be more stable.

The bilingual participants apparently feel that some of their skills have suffered somewhat compared to the controls. In the control group, on the other hand, only writing, the productive literacy skill, has a lower average and greater range, while the figures for listening, reading and speaking are virtually identical (the magnitude of difference lying after the decimal point) at the high point of the spectrum, suggesting that these skills are overlearned in many participants as might be expected in native speakers (cf. McLaughlin 1995; de Bot 1998; de Bot 2008 in relation to skill routinisation and the threshold hypothesis with regard to knowledge stability).

Figure 14 displays the distribution of percentage scores for all four skills in both groups. In line with the notion of overlearned skills, scores cluster at or near the maximum. However, the figure also points to writing being the least proficient skill for some participants in both groups, and to a greater dispersion of scores in the bilingual group. Going back to Table 35 above, we again observe, similar to what we have seen before, a non-significant trend for slightly lower averages, larger StDs and/or larger ranges in the bilingual group. The trend for lower averages is also

observable in the individual can-do statements. For example, all controls rate themselves "5" ("I can do that without any problems") for ten statements out of 43; the bilinguals only for four.



**Figure 14: Distribution of skills scores by participant**

To test to what extent self-rated proficiency corresponds to objective proficiency scores, we correlated the averages for the four skills with the two measures elicited on the German C-test

(scores and time taken) and the *z*-scores, for the two groups (cf. Table 36 for the results of the bilingual group).

		Listening	Reading	Speaking	Writing	C-test scores	C-test time	<i>z</i> -score
Listening German	PCorr	1.000	.931**	.407	.429	.197	-.441	.405
	Sig. (2-tailed)		.000	.083	.086	.420	.059	.085
	N	19.000	19	19	17	19	19	19
Reading German	PCorr	.931**	1.000	.521*	.542*	.157	-.488*	.379
	Sig. (2-tailed)	.000		.022	.025	.520	.034	.109
	N	19	19.000	19	17	19	19	19
Speaking German	PCorr	.407	.521*	1.000	.912**	.506*	-.829**	.367
	Sig. (2-tailed)	.083	.022		.000	.027	.000	.122
	N	19	19	19.000	17	19	19	19
Writing German	PCorr	.429	.542*	.912**	1.000	.555*	-.823**	.200
	Sig. (2-tailed)	.086	.025	.000		.021	.000	.442
	N	17	17	17	17.000	17	17	17
C-test scores	PCorr	.197	.157	.506*	.555*	1.000	-.682**	.455*
	Sig. (2-tailed)	.420	.520	.027	.021		.000	.017
	N	19	19	19	17	27.000	27	27
C-test time	PCorr	-.441	-.488*	-.829**	-.823**	-.682**	1.000	-.564**
	Sig. (2-tailed)	.059	.034	.000	.000	.000		.002
	N	19	19	19	17	27	27.000	27
<i>z</i> -score	PCorr	.405	.379	.367	.200	.455*	-.564**	1.000
	Sig. (2-tailed)	.085	.109	.122	.442	.017	.002	
	N	19	19	19	17	27	27.000	27.000

**Table 36: Correlations between German can-do statements and C-test for bilingual group**

PCorr = Pearson correlation

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

Regarding the relationship of the four skills with one another, we observe strong significant correlations between the receptive skills listening and reading on the one hand ( $r = .931$  in the bilingual group and  $.685$  in the control group), and the productive skills on the other ( $r = .912$  in the bilingual group and  $.854$  in the control group; all *ps* at  $.01$  level). There are also significant

medium-sized effects between reading and writing (the two literacy skills) ( $r = .542$  bilingual group,  $r = .608$  control group) and reading and speaking ( $r = .521$  bilingual group,  $r = .526$  control group;  $p$ s mainly at .05 level except reading-writing controls,  $p = .01$ ). The control group also has a significant correlation between self-attested listening and writing skills ( $r = .528$ ,  $p = .03$ ).

Interestingly, only the two productive skills speaking and writing correlate significantly with the C-test scores, at the .05 level, with a medium effect of  $r = .506$  (speaking bilingual group)/.581 (speaking control group) and  $r = .555$  (writing bilingual group)/.586 (writing control group) respectively, while there is no effect for the receptive skills. None of the skill self-ratings correlate with the  $z$ -scores, though those do correlate with the C-test score and time, as would have been expected. For the German controls, the  $z$ -scores also correlate with the productive skills ( $r = .523$  and .489, at the .05 level).

Regarding the relationship between the self-rated skills and the time it took to solve the C-test, an interesting difference between the groups emerges: There is a highly significant negative relationship between the productive skills and C-test time ( $r = -.829$  and  $-.832$  respectively,  $p$ s at the .01 level), as well as between reading and C-test time ( $r = -.488$ ,  $p$  at .05 level) in the bilingual group (listening just fails to reach significance), indicating that the longer people took to complete the C-test, the lower their self-ratings. The control group, on the other hand, manifests no significant correlations between C-test time and any of the skills.

This is a very interesting result. Language proficiency can be construed as the coming together of three components: fluency, accuracy and complexity (Larsen-Freeman 2009; Skehan 2009), the first of which is an online, processing dimension, while the latter two relate more to the knowledge built up in the language. The C-test being an untimed task, the fluency dimension would be particularly noticeable in the time measure, whereas all three would bear on the C-test scores. It would thus appear that the self-ratings of the bilingual participants, particularly those relating to reading and speaking, reflect the fluency dimension of oral and written production (possibly the aforementioned difficulties in lexical retrieval), while the controls' do not, presumably because fluency is not at issue. Consequently, if all controls are, and see themselves as fluent, then differences in the time it took to solve the task are due to other reasons.

The picture that is beginning to emerge from this study, and indeed from earlier studies, is that the dimension most affected in attrition is the fluency dimension, though accuracy and complexity are ultimately liable to be affected too, as was apparent from the results of all three tasks. In the untimed C-test, the fluency dimension is of lesser relevance for proficiency scores, and indeed these were very similar for both groups with the exception of some individuals. It should, however, play a greater role in the "online tasks" FiCA and the film-retelling. The FiCA scores, as we remember, were somewhat depressed for several bilingual participants compared to the control group. In the film retelling, only repetitions were found to be significantly higher in the bilingual group, but as we discussed, there may be a trade-off between repetition and pausing (disfluency) behaviour. Nevertheless, retrieval problems are the most common problem isolated across all three tasks.

Thus, it would appear that the bilingual participants home in on these specific difficulties in their assessment of their L1 proficiency. We may hypothesize - without being able to develop this thought further here, and without wanting to overstate the case -, that for some people, the very notion of the possibility of the L1 deteriorating, and the slightest sign thereof, may have a profound impact on their self-perception as native speakers. The next subsection explores the question of self-perceived changes in participants' L1 proficiency further.

#### **4.5.2 Self-reported L1 proficiency over time**

During the interview, bilingual participants were asked to state whether their L1 proficiency had changed since emigration, and if so, in which direction (Question 39 of the sociolinguistic questionnaire). Participants also supplied a rating for their proficiency level prior to emigration, and at the time of testing (Questions 40 and 41). These questions were asked after the testing session, so that performance on the tasks would be unbiased.

The bilinguals overwhelmingly rated their pre-emigration proficiency in German as "fully nativelike" (89%); the remainder (three individuals) as either "very good" or "good" (cf. Table 37). At time of testing, only two-thirds still rated their proficiency at the highest point at the scale. 22.2% felt that their proficiency was still very good (but not as good as it had been); and one (S10)

felt her proficiency had dropped from the maximum to a mere "satisfactory". The two individuals who had rated their pre-migration proficiency as "good", still gave it the same rating. Nevertheless, a Wilcoxon Signed Rank test reveals that the difference in the ratings of L1 proficiency pre- and post-migration is significant ( $T = 1$ ,  $z = -2.11$ ,  $p = .031$  one-tailed,  $r = -.29$ ).

<b>L1 proficiency</b>	<b>Prior to migration</b>		<b>At time of testing</b>	
Sufficient	0	0%	1	3.7%
Good	2	7.4%	2	7.4%
Very good	1	3.7%	6	22.2%
(Near-) Native	24	88.9%	18	66.7%

**Table 37: Self-ratings L1 proficiency**

When asked explicitly whether their proficiency had changed (Question 39), 49% felt it was still the same, and 51% said it had disimproved. Given that the majority of participants still rated their proficiency as native-like, the scale of negative changes is evidently relatively small and partially lies within the bounds of nativeness. This is borne out in some participants' comments, who emphasise that L1 is just "a little bit worse", or who do not see the characteristics of their current speech constituting a disimprovement. For example, some of the participants who say that their proficiency is the same, go on to list areas of difficulty.

This presumably reflects individually differing conceptions of proficiency, and probably explains why one participant states that her German is now actually better, despite the fact that she describes lexical retrieval difficulties ("the words are not always present"), the need to translate phrases from English to German when speaking, or to find paraphrases. The improvement she sees in the area of speaking with awareness as a result of making an effort. Several participants also state that their German was now "back to normal" since it had become more affordable and easier to keep telephone and other contact, often via the internet. Other factors in improved L1 proficiency that were mentioned were starting a new job or training as a translator/interpreter, or having children. A trip to Germany also helps to "relearn" the language which otherwise has got "rusty" and requires some warming-up (like "muscles").

Several participants point out that the disimprovement they have noticed is not across the board, but relates to specific domains where they are lacking specialist vocabulary, or where the development of their German simply has not kept up with their development as people (re hobbies, professional vocabulary). One participant states that she feels that her German is still the German of an 18-year old. In general, participants notice having missed out on new developments in the German language since they left; and as a result sometimes find it easier to express themselves in English, particularly where precise and complex language is required. Strictly speaking, this latter area does not come under "attrition" but under "non-acquisition", but it is a consequence of the same circumstances that may also lead to attrition.

In response to Question 39, participants also describe changes relating to different linguistic levels (cf. Table 38 below). Mirroring our findings of most problems occurring in the lexical area, many participants comment on specific difficulties they have encountered, such as retrieving words and certain types of expressions, for example humorous, literary and indeed "foul" language, with facility, which impacts on overall fluency and pragmatic effectiveness.

Several participants feel that their lexicon is not just less accessible but has actually shrunk in size. Some participants also indicate that their knowledge of words and expressions may have been impacted, when they report being unsure of the particular shape of a word in relation to their spelling and their morphemic make-up, for example the end suffix in *Therapeut* ("therapist"). This example hints at another source of difficulty: interferences and the need to translate from English or find other compensatory means for forms that are (temporarily) unavailable. One participant, for example, admits to resorting to innovation, i.e. to creating novel words in German in lieu of those that she would have liked to use.

As Table 38 shows, the self-professed problems are not restricted to the lexical area, since participants also report noticing syntactic transfer in their speech, having to consciously apply rules for forming grammatically correct sentences, rather than knowing "intuitively" what is correct, and making grammatical mistakes; however, these are mentioned less often than changes in the lexical area.



<b>Changes in L1 German</b>	<b>Mentions (N (%))</b>
<b>Phonology</b>	
L2 accent in German	1 (3.7%)
<b>Lexicon</b>	
Smaller size of lexicon	6 (22.2%)
Speed and precision of lexical retrieval/fluency	17 (63.0%)
Insecurity in choice and precise shape of words/expressions	5 (18.5%)
Insecurity in relation to spelling	4 (14.8%)
Interferences, innovations	8 (29.6%)
<b>Morpho-syntax</b>	
Loss of sensitivity for grammatical distinctions, accuracy	1 (3.7%)
Loss of automaticity of rule application	1 (3.7%)
Syntactic transfer	5 (18.5%)
<b>Pragmatics</b>	
Appropriacy of formal/informal address	2 (7.4%)
Facility for being witty	1 (3.7%)
Facility for partaking in complex discourse	2 (7.4%)
Facility for using expressive means	1 (3.7%)

**Table 38: Reported changes in L1 proficiency**

In the area of pragmatics, in addition to the changes discussed above, two participants mention being unsure of when to use the formal address "Sie" and the informal address "du" appropriately, another area which is more than likely influenced by L2. Only one participant, S10, whom we have encountered as the most likely candidate for L1 attrition, reports that her accent in German has become foreign. Although native accent does seem to be relatively stable, perceptible and imperceptible changes have been reported (Major 1992; Yukawa 1997; de Leeuw, Schmid et al. 2010).

Overall then, the self-reported changes to the L1 suggest the same pattern as was observed in the proficiency data. On the basis of the combined proficiency and self-report data, we can thus surmise that the L1 proficiency of our bilingual participants has indeed been affected by their experience of living abroad and speaking another language in daily life, but that some effects which may have been thought of as attrition also occur, perhaps to a lesser extent, in non-attributed

native speakers. S1, S10 and S26 are the most clear-cut cases of those effects. However, the magnitude of change for the most part is too small to register as performance that is significantly different from non-attrited native speakers, which in turn may be a result of dynamic interactions between various variables.

## 5 L2 attainment - between-group study English

The comparison of the bilingual participants' performance on the English language tests with that of the Irish control group serves to indicate the degree of English language acquisition (research question 2). It is worth reiterating that due to the cross-sectional nature of this study, we have no direct access to data relating to the level of the bilinguals' L2 proficiency prior to emigration, so instead must rely on self-report.

All bilinguals had some knowledge of English prior to emigration, having studied the subject in school. In response to Question 37 in the sociolinguistic questionnaire, most participants, almost half, rated their pre-migration English proficiency as good (44.4%; cf. Table 39). A sizeable proportion stated that it was very good (25.9%), and the remainder stated that their proficiency was lower. Nobody had near-native proficiency at that stage. Since then, proficiency at the group level has improved to a minimum of a "good" rating, with two-thirds now rating their proficiency as native-like. None of the participants felt that their L2 proficiency had decreased, and the improvement registers as a large positive change (Wilcoxon Signed Rank test:  $T = 0$ ,  $z = -4.53$ ,  $p = .000$  one-tailed,  $r = -.62$ ).

<b>L2 proficiency</b>	<b>Prior to migration</b>		<b>At time of testing</b>	
Very low	3	11.1%	-	-
Low	1	3.7%	-	-
Sufficient	4	14.8%	-	-
Good	12	44.4%	3	11.1%
Very good	7	25.9%	6	22.2%
(Near-) Native	-	-	18	66.7%

**Table 39: Self-ratings L2 proficiency**

Incidentally, the figures for proficiency at time of testing are identical to the (group) self-ratings for German at the time of testing, except that one bilingual, S10, rates her German proficiency lower, as pointed out above.

As a baseline against which to assess the degree of second language acquisition, we deployed a control group of matched native speakers of English, as set out in Chapter 3, who also have at least

a good level of proficiency in German. The degree of second-language attainment is operationalised as the proximity of participants' performance to or its overlap with the Irish controls' native-speaker range. We shall investigate the same features and hypotheses as formulated in the German section: score ranges and time taken to complete tasks, lexical diversity, errors, code-switching behaviour and hesitations.

## 5.1 C-test English

### 5.1.1 Test results

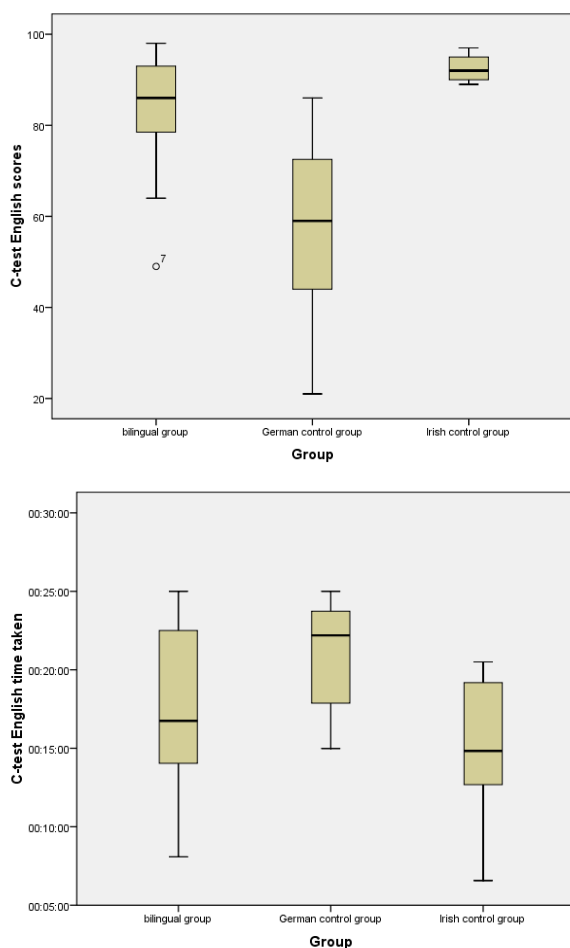
The English C-test, which was completed by all three groups, was scored according to the same procedure as outlined above for the German C-test. The results are given in Table 40 (cf. also Figure 15). Significant differences in the C-test scores are revealed (Kruskal-Wallis:  $H = 37.29$ , 2df,  $p < .001$ ): between the bilingual group and the Irish control group (Mann-Whitney with Bonferroni correction:  $U = 128$ ,  $p < .017$ ), but also between the German and Irish control groups (Mann-Whitney:  $U = 0$ ,  $p < .0003$ ), and the bilingual group and the German control group (Mann-Whitney:  $U = 49$ ,  $p < .0003$ ).

C-test English	Bilingual group (N = 27)	Irish controls (N = 18)	German controls (N = 20)
Mean scores (ex 100)	84.3*B	92.7	57.4***B
Median	86	92	59
Range	49-98	89-97	21-86
StD	11.5	2.8	18.1
Mean time taken (min:sec)	17:20	14:55	21:11***
Median	16:45	14:50	22:12
Range	8:05-25:00	6:34-20:30	14:59-25:00
StD	5:09	4:10	3:01

**Table 40: English C-test mean scores and time taken**

\*B Results are significant at  $p < .017$  (Bonferroni corrected  $p$ ).

\*\*\*B Results are significant at  $p < .0003$  (Bonferroni corrected  $p$ ).



**Figure 15: Box plots of English C-test scores and times for all three groups**

A slightly different picture presents itself with respect to the time it took participants to complete the test: the difference between the bilingual group and the Irish control group does not reach significance, while there are significant differences between the two control groups on the one hand, and between the bilingual group and the German control group on the other (One-Way ANOVA: Welch  $F(2, 39) = 14.82, p < .001$ , Games-Howell).

This means that the bilingual group has not overall achieved a similar level of proficiency in English to that of the Irish control group, but it has a significantly higher level than the German control group. Moreover, it comes close to the Irish control group in terms of speed of solving the test, while the German control group is significantly slower.

Similarly to the findings on the German C-test, the bilingual group's range of scores is much larger than that of the Irish control group: 49-98 as opposed to 89-97, which is not unusual in comparisons of L2 learners and native speakers. Nevertheless, twelve of the bilingual participants

(ex 27; 44%) achieve scores within the English native-speaker range, which satisfies our criterion of native-like achievement. Moreover, 16 participants (59%), including ten of the twelve participants with native-like scores, complete the test in a comparable amount of time (Table 41). (It may be worth re-emphasising that participants were not instructed to solve the test as fast as possible, but at their own speed.)

Thus, even though the bilingual group's scores are significantly different from the Irish controls', a considerable number of individuals achieve a native-like level of L2 proficiency as measured on this test. In stark contrast, none of the German controls achieve a score within the English native-speaker range, while just under half (nine participants ex 20; 45%) complete the test in comparable time.

<b>C-test English</b>	<b>Irish controls</b>	<b>S16</b>	<b>S10</b>	<b>S21</b>	<b>S11</b>	<b>S5</b>
Mean score (range)	92.7 (89-97)	<b>98</b>	<b>96</b>	<b>96</b>	<b>95</b>	<b>94</b>
Mean test time (range) (min:sec)	11:34 (7:03-17:25)	<b>08:05</b>	<b>14:05</b>	<b>12:16</b>	<b>09:30</b>	<b>13:45</b>

<b>continued</b>	<b>S14</b>	<b>S17</b>	<b>S22</b>	<b>S9</b>	<b>S25</b>	<b>S27</b>	<b>S23</b>
Mean score (range)	<b>94</b>	<b>94</b>	<b>92</b>	<b>91</b>	<b>91</b>	<b>91</b>	<b>90</b>
Mean test time (range)	<b>14:07</b>	<b>15:42</b>	19:35	<b>09:35</b>	<b>17:15</b>	<b>17:01</b>	22:50

<b>continued</b>	<b>S3</b>	<b>S12</b>	<b>S19</b>	<b>S20</b>	<b>S15</b>	<b>S18</b>	<b>S18</b>
Mean score (range)	88	86	86	86	81	64	64
Mean test time (range)	<b>16:45</b>	<b>13:26</b>	<b>09:11</b>	<b>14:31</b>	<b>15:37</b>	<b>14:00</b>	<b>14:00</b>

**Table 41: Individual English C-test scores and test times\***

\*Results given in bold fall within the native-speaker range.

The best achievers are naturally those who can integrate accuracy and speed in their test performance. On the basis of Table 41, S16 and S11 stand out as the most proficient individuals who achieved a very high score, and managed to do so in a very short time (only two Irish controls, with a score of 95, are faster than them). S10, S21, S5, S14 and S17 perform at a similar level, but take considerably longer. S9 and S19 come close to the "best performers" in terms of time, but their score is at the lower end or just outside of the native-speaker range respectively.

At the other end of the scale, the lowest bilingual score of 49 has been identified as an outlier (S7, Figure 15), with the second and third lowest scores being as high as 64 and 68 respectively (S18 and S24). In the German control group, on the other hand, while all scores are below the range of the Irish control group, 40% of scores (eight participants ex 20) also fall below the lowest bilingual score.

We venture to suggest, and will return to this suggestion when discussing the later tests (cf. external ratings), that S18's score does not reflect her true level of proficiency, since for personality and health reasons she felt that she simply could not do this test. Impressionistically, she is a very fluent L2 user who uses English with great facility, but she gave up on several of the items claiming she did not know them. She did not enjoy any of the tasks, but found this test particularly hard.

Similarly, S26, who has a score of 72 and who comes across as a very authentic English speaker, warned that he had learnt English primarily in its spoken form, and felt under pressure and anxious from having to complete a written test. Learning a language in its written form naturally ought to impact on one's ability to spell it – which we disregarded in scoring the test, and which was one of the difficulties raised by S26 –, but has been argued to play a role in advancing language proficiency in general.

Correlating C-test score and time for the bilingual group again yields a significant result ( $r = -.612$ ,  $p = .001$ ), indicating that people who were faster at solving the test also had a higher result. However, given the picture presenting itself in Table 41 and the explained variance of 37%, factors apart from the level of proficiency do apparently play a role in determining the speed with which the test was solved, as we had already discussed in the corresponding section for the German C-test above.

### **5.1.2 Item analysis**

In the following section, we shall again focus on the comparison between bilingual group and the relevant, Irish, control group, leaving aside the performance of the German control group. The analysis of the success with which individual items were solved in both groups confirms the two

features of the bilingual group's performance already identified in the preceding section: that of partial overlap as well as difference.

Both difference and overlap are apparent from the figures in Table 42, which shows that a much smaller number of items were solved correctly by all bilinguals than by all controls. Consequently, the figures for the lower bands are higher, with most items solved with a success rate in the 90-99% band, but in general follow the distributional pattern of the control group, each band roughly corresponding to the next higher band of the control group. There was one item, *flout*, which caused severe problems for the bilingual group, as well as the control group, and was solved correctly by only three bilinguals (11.1%). We shall return to this item below.

<b>Number of items solved correctly by % of participants</b>	<b>Bilingual group</b>	<b>Irish controls</b>
all participants	18	53
90-99%	28	20
80-89%	21	15
70-79%	18	5
60-69%	7	3
50-59%	4	4
40-49%	3	–
30-39%	–	–
20-29%	–	–
10-19%	1	–
<b>Total number of items</b>	<b>100</b>	<b>100</b>

**Table 42: Breakdown of number of correct item solutions by percentage of participants**

There is again considerable overlap of items between groups – 17 of the bilingual group's 18 error-free items did not present a problem for the controls either, and a further 28 resulted in an error for fewer than 10% of participants in either group. 19 more items were problematic for up to 15% of participants, and the remaining 36 items presented problems for 16% and more in either group (mostly the bilingual group). The distribution of these items across the five texts is represented in Table 43 below.



Still, the relative success with which items were solved in either group varies significantly between groups (Wilcoxon Signed-Rank Test,  $z = -6.612$ ,  $p = .000$ ,  $r = -.47$ ). There were 20 ties, i.e. both groups were experiencing similar levels of difficulties on these items; 71 items on which the control group outperformed the bilingual group; and 9 items on which the bilingual participants performed better than the control group. However, as the analysis of item overlap by bands indicated, the relative level of difficulty of many items is similar across groups.

Text 1	Text 2	Text 3	Text 4	Text 5	Overall
4	10	6	13	3	36

**Table 43: Problematic items per text across both groups**

Table 43 highlights the fact that the five texts were of uneven difficulty. The two hardest texts were the second and the fourth text; the former was experienced as very difficult by both groups, whereas the latter presented a greater problem for the bilingual group. The third text was somewhat easier overall, but it contained the most difficult item, **Item 19**, *flout*.

Item	Error type*	Bilingual group % (Number) correct ex 27	**	Irish controls % (Number) correct ex 18
<b>Text 1</b>				
1) <i>reflection</i>	GR	78% (21)	<	94% (17)
2) <i>predictions</i>	LEX/GR	63% (17)	<	89% (16)
<b>Text 2</b>				
3) <i>commercial</i>	LEX	85% (23)	=	89% (16)
4) <i>pursued</i> +	CL	74% (20)	<	89% (16)
5) <i>such</i> +		74% (20)	=	72% (13)
6) <i>help</i> +		78% (21)	<	83% (15)
7) <i>realize</i> +		63% (17)	=	61% (11)
8) <i>value</i> +		59% (16)	<	78% (14)
9) <i>assets</i>		67% (18)	<	72% (13)
10) <i>generate</i>	GR	41% (11)	<	50% (9)
11) <i>ploughed</i>	LEX/GR	78% (21)	<	94% (17)
12) <i>set</i> +	LEX/GR	48% (13)	<	61% (11)
13) <i>support</i>		74% (20)	<	83% (15)

<b>Item</b>	<b>Error type*</b>	<b>Bilingual group</b> % (Number) correct ex 27	<b>**</b>	<b>Irish controls</b> % (Number) correct ex 18
<b>Text 3</b>				
14) <i>vessels</i>	GR	85% (23)	<	94% (17)
15) <i>towed</i>	LEX/GR	78% (21)	<	100% (18)
16) <i>plans +</i>	CL	81% (22)	<	100% (18)
17) <i>shelved</i>		52% (14)	<	94% (17)
18) <i>deemed</i>	LEX	67% (18)	<	100% (18)
19) <i>flout</i>	LEX	11% (3)	=	50% (9)
20) <i>stored</i>	LEX	85% (23)	>	56% (10)
<b>Text 4</b>				
21) <i>superficiality</i>	LEX/GR	74% (20)	<	94% (17)
22) <i>women's</i>	LEX	74% (20)	<	100% (18)
23) <i>boy</i>	LEX	63% (17)	<	89% (16)
24) <i>reckon</i>	LEX	56% (15)	<	83% (15)
25) <i>one</i>	LEX	63% (17)	<	89% (16)
26) <i>easily</i>	GR	82% (22)	<	100% (18)
27) <i>cover</i>	LEX	70% (19)	<	78% (14)
28) <i>head +</i>	LEX	74% (20)	<	100% (18)
29) <i>shoulders</i>	(CL)	70% (19)	<	100% (18)
30) <i>smiling</i>	GR	78% (21)	<	89% (16)
31) <i>airbrushed</i>	LEX	78% (21)	<	100% (18)
32) <i>optionally</i>	GR	41% (11)	<	89% (16)
<b>Text 5</b>				
33) <i>see</i>	LEX	74% (20)	>	50% (9)
34) <i>walk</i>	LEX	78% (21)	<	100% (18)

**Table 44: Selected items from English C-test for detailed analysis**

\* LEX = lexico-semantic error, GR = morpho-syntactic error, CL = cluster

\*\* < GC outperforming BG, = tie, > BG outperforming GC

Table 44 above presents a selection of the more problematic items, occasionally in combination with some items that evidently contributed to the formation of clusters of difficulty. The English C-test (gapped texts and solutions) may be found in the Appendix. The relevant items are put in bold face and are indexed with the number from the table above.

Table 44 not only confirms the relative level of difficulty of each text, but also the two trends already mentioned above: partial convergence or overlap between the groups in the level of

difficulty, and the type of difficulty on the one hand, which very often translated into similar solutions as we will see below, but also divergence or difference with regard to other items that presented no or minimal difficulty for the control group. Again, we find that difficulties are not restricted to the lexical and semantic realms, but concern grammar and syntax as well, which is indirect proof of the C-test functioning as an integrative test.

To start with the latter, **Items 1, 2, 10, 11, 12, 13, 14** and **30** are examples of both groups using the wrong part of speech or not agreeing subjects and verbs despite evidence to the contrary. The errors made are the same, but the rate is usually higher in the bilingual group. For example, for **Item 1**, *reflection*, one control and five bilinguals provide the plural form despite the indirect article in front of the noun, which might indicate lack of monitoring.

Of course, this is an easy mistake to make given that the article is a rather short word, making it hard to notice. Also, the sentence starts with a plural demonstrative and verb, *these are*, which refer to *expectations* in the previous sentence and may plausibly create the expectation of a following plural object. Further, the item does not recur, unlike **Item 14**, *vessels*, which is the theme of the third text. Nevertheless, one control and four bilinguals provide the singular for this item, although there are several references to "ships" throughout the text: *two former US navy ships, were expected, were being towed, dismantle them, the ships*. The item occurs early on in the text and is perhaps located too far away from the markers highlighting plurality. Several people did in fact correct their solution after having read more of the text; failure to do so again indicates insufficient monitoring.

Not readily explained are **Items 12** and **13**, for which several individuals provide the third person singular (*set*: two controls, one bilingual; *support*: three controls, two bilinguals), although the subject for *set*, *these Commercial Policy Guidelines*, has two salient markers of plurality, the determiner, and plural -s, and that for *support* – *activities* – a salient orthographic change in the plural. Thus, while for Item 14 one might argue that participants simply forgot the ending, for Items 1, 12 and 13 they actually supplied unnecessary endings, which suggests more than a spelling error.

Dolan (1999), in the introduction to his Dictionary of Hiberno-English, notes that "[s]peakers of Hiberno-English often retain the singular form [of verbs] for plural subjects" (ibid.: xxvf.), which

would provide a suitable explanation for Items 12 and 13, provided the feature was used deliberately, but not Item 1, for which, however, we pointed to other sources above. If participants used the singular verb with the plural nouns consciously, then the appropriacy of so doing in a formal written text created by somebody else, as opposed to self-created speech, might be questioned.

The other items resulting in errors in the morpho-syntactic area concern the use of wrong parts of speech for the given syntactic context: **Item 2**, *predictions*, was variably rendered as *predictive/predicated* (two controls, two bilinguals) or *predicting* (five bilinguals), despite the preceding article, which is not combinable with any of the suggestions, and the *of* after the gap, which only works with *predictive* and, in some contexts, with *predictable* (*the most/least predictable one of a number of x*).

**Item 11**, (*to be*) *ploughed*, generated *plough/ploughing* (one control, two bilinguals; it also resulted in a gap indicating a lexical problem for four bilinguals); **Item 30**, the adjective *smiling*, was rendered as the noun *smile* (following the preposition *of* and article *a*; two controls, six bilinguals), although the next segment is the adjectival phrase *heavily made-up*, and although both controls and two of the bilinguals used the correct adjectival target *airbrushed*, **Item 31**, in the next gap. Four bilinguals left a gap for Item 31, and one supplied an invented word, *airbrouse*, mirroring his use of *smile*.

Finally, **Item 10**, *generate*, was rendered as *general* by eight controls and 14 bilinguals, i.e. as an adjective instead of a verb. This appears to have been brought on by the preceding large cluster of difficult gaps (Items 4-9), which made syntactic analysis of this portion of text difficult for some participants, and which we shall discuss further below. Since seven controls and ten bilinguals actually do correctly provide a verb for the slot of the preceding *realize*, **Item 6**, with which *generate* is joined by *and*, we have to surmise that *and* was analysed as connecting *income* with the preceding noun *assets*. Possibly, this reading of the text is promoted by interference from the common collocation *general income*, which appears to have overridden other considerations. There is another very interesting instance of an over-extension of a collocation in the data, to which we shall return below.

One type of part-of-speech/syntactic violation occurs almost exclusively in the bilingual group – using an adjective instead of an adverb, e.g. in **Items 26**, *easily* (five bilinguals) and **32**, *optionally* (one control, twelve bilinguals). We have no ready explanation for this. It is possible that this error is simply an oversight or results from genuine confusion. The bilinguals, for example, may be influenced by German, which distinguishes between adjectives and adverbs by the latter being ending-less. The control's use of *optional*, on the other hand, is probably not. However Berry (1971), in his (not so recent) book of common mistakes in English, which in the first instance appears to have been written for native speakers in third-level education, has a section on "Errors Arising from the Confusion of Adjectives and Adverbs" (ibid.: 68-70), which lists *easy-easily* as item 133. Similarly, several internet fora have sections spelling out the differences between the two parts of speech, indicating that it may present a problem for native speakers and learners of English alike.

**Item 15**, *towed*, and **Item 21**, *superficiality*, are further examples of syntactic violations by the bilingual group, for which the underlying cause is probably a lexical gap. **Item 15** was solved correctly by all controls, but two bilinguals left a gap, one put *together*, and three used *to go/to guide*, violating the requirement for a past participle in a progressive passive construction, as well as the requirement for gap solutions to be single words.

For **Item 21**, two bilinguals used the adjective of the same root which clashes with the preceding article and following preposition: *something about the superficiality of*. The remaining solutions that contributed to the error count were *superficialness* (one control, two bilinguals), which is only listed in some references of English, usually with a cross-reference to *superficiality*; it does not occur in either the Advanced Learner's Dictionary (ALD; Hornby 1995) or the Duden-Oxford Großwörterbuch Deutsch-Englisch, Englisch-Deutsch (Duden-Oxford; Scholze-Stubenrecht and Sykes 1990). One bilingual left a gap, and one each used a host of either existing or made-up, but equally non-fitting words: *superfiction*, *superfigur*, *superfiality*, *superficency*, *superfinery*.

Lexical problems account for the remainder of the items listed above, sometimes compounded by participants not homing in on the theme of a text sufficiently. A case in point is **Item 34**, *walk*, which is not a difficult word, and the theme of walking and driving is reiterated in the last sentence: "lifts to someone unable to walk or who doesn't have a car". Nevertheless, four bilinguals

put *wait*, the thematic link of which with *drive* is unclear, and two put *way*, which may be due to crosslinguistic transfer of *Weg* ("way", or "distance" in this context).

As an aside, another example of possible crosslinguistic transfer is **Item 22**, *women's*, for which six bilinguals produce *women*, and one *woman*, probably as a result of the German compound *Frauenzeitschrift* ("women's magazine"). Here, the possessive is not obvious since the ending *-en* also stands for plural.

Disregarding the overall or local theme occasionally gave rise to clusters of problem gaps, the solutions to which individually are common words, for example, in the second text. This text, as we pointed out above, was experienced as the hardest, presumably because of the moderately technical subject matter; however, anecdotally, the expressions used in the text are commonly enough found in daily newspapers and in people's everyday speech. The text states its genre and theme quite unambiguously in the opening sentences and restates it at the end: "the BBC's commercial activities", "Commercial Policy Guidelines". Nevertheless, for **Item 3**, *commercial*, two controls and four bilinguals do not identify the "obvious" target, but provide *commentary/commentating*, presumably motivated by the fact that it is a text about the BBC, or leave a gap.

But even some of those who do supply *commercial* have trouble with **Item 7**, *realize*, **Item 8**, *value*, and **Item 9**, *assets* (we already discussed Item 10, *generate*). Our analysis shows that these items represent a cluster of difficult items: people either solve all gaps correctly (seven in either group), or make several errors across the four gaps, on average 1.6 in the control group, and 1.8 in the bilingual group. Not solving these items also increases chances of an error for Item 10 for the majority of people in both groups (seven ex eleven controls, and 13 ex 20 bilinguals; only two controls and three bilinguals have an error on Item 10 while solving the cluster correctly). For some participants, Items 3-10 become a "super-cluster" of difficulty (two controls, three bilinguals with five or six errors each).

Two more clusters presented a problem for the bilingual group only. **Items 28 and 29**, *head (and) shoulders (shot)*, was solved correctly by all controls, but eight bilinguals could not think of a fitting determiner, although the topic again is quite clearly set as "Magazines". Interestingly, all of these but one solved the preceding **Item 27**, *cover* (image) correctly, which should have kept them

on the photographic track. Item 27 presents a problem for four controls and eight bilinguals, who variably propose *cosmopolitan*, *common*, *corporate*, *corny* and *corner*, all more or less plausible within the context. One control and three bilinguals leave a gap.

The other cluster concerns **Items 16-18**, *plans (have been) shelved (after being) deemed (to) ...* This presented minimal problems for the controls, but the collocation *to shelve plans* (Items 16 + 17) apparently was not available to 13 bilinguals. Three of them (and one control) suggested *shed*, eight left a gap and one commented "I know what they mean but I can't think of it." and wrote *discarded*. Item 18 literally drew a blank for ten bilinguals.

There are several more examples of lexical problems in both groups, not surprisingly usually with a higher percentage of error in the bilingual group (e.g. **Item 24**, *reckon*), but there are three very interesting phenomena causing the control group to perform worse than one might have expected. The first concerns collocations and has been labelled "Too smart for their own good", the second is related and will be referred to as the "Small words phenomenon", and the third one is "The hardest word".

The "too smart for their own good" phenomenon occurs with **Item 4**, *pursued*, and **Item 20**, *stored*. In both cases participants pick a ready solution without interrogating the wider context. For **Item 4**, having established that the text deals with commerce, one control and two bilinguals go for *purchased*, although it is unclear how one would *purchase something improperly*. One control uses *purposed*, three bilinguals use a gap, one makes up a word, and one uses *purpose* which again does not work syntactically.

**Item 20** represents an even clearer case of this phenomenon, with eight controls and two bilinguals proposing *stowed* instead of *stored*. While *stowed* is a word particularly used in a nautical context, it does not make sense in a text that talks about the problem of returning contaminated ships. Admittedly, *stored* does sound a bit strange too. This is one of the items on which the bilinguals outperformed the controls, presumably because the collocation *to stow (something in) a ship* was not known by everybody.

The "small words" phenomenon (**Items 5-6, 25 and 33**) describes attempts, primarily by the controls, to come up with more complex words than are necessary in the context. Discussing **Item 33**, *see (lower turnouts)*, should suffice to illustrate that point. Just half of the controls, and three

quarters of bilinguals, come up with the simple target word, the remainder propose, variably, *surpass, served, set, sustain, score, supply, surprise* and even *some* (one bilingual; this again causes a syntactic problem) – any word, it would appear, but the obvious one. Three in either group leave a gap.

**Item 23**, *(oh) boy*, probably also falls into this category but is complicated by the fact that it is an Americanism, a fact to which all participants were alerted. This item results either in gaps (one control, three bilinguals) or in more or less successful attempts to capture the exclamation, such as *(but oh) but* (one control, five bilinguals), *oh baby/babe* (two bilinguals), as well as the Irishisms *oh blimey* (one control, one bilingual) and (non-existent) *oh begosh* (one bilingual).

Finally, the "hardest word", as we pointed out above, is **Item 19**, *flout*, which just half of the controls (nine) and three bilinguals solved correctly. Two controls misspell the target as *flought*, but eight controls and four bilinguals use *flaunt* which according to a multitude of linguistic sources has, since the 1940s, become a frequent contamination, despite the fact that the two words *flout* and *flaunt* have almost diametrically opposed meanings. According to several print editions of reputable dictionaries, this use still has to be considered incorrect.

The bilinguals had greater difficulty with this item, twelve simply leaving a gap, and the remainder using various inappropriate solutions (*flounder/flaunder, flag(s), flaw, float*). Thus, in the case of the controls, the cause for error is habitual confusion, whereas for most bilinguals it is a lexical gap. Nevertheless, the incorrect use of *flaunt* by four bilinguals probably and ironically has to be seen as a sign of acquisition. Similarly, **Item 25**, *one*, classed under the "small word" problem for the controls, presented a lexical problem for the bilinguals, some of whom clearly did not know what the preceding word *cobble (together)* meant (gaps: two controls, seven bilinguals; one bilingual each *other, on*).

Nevertheless, in discussing the problems encountered by participants in both groups, we must not forget that in fact the majority of people actually solve most items correctly. In the control group, to reiterate, the lowest success rate for any individual item is 50%, in the bilingual group, there are just four items that are solved with less success. As we have seen, and with the exception of *flout*, and possibly *superficiality* in the bilingual group, it is likely that most people actually know the



words that were required in the gaps but occasionally – controls and bilinguals alike – failed to make sense of the wider context.

Summarising the analyses presented in the preceding two sections, we may conclude that while, on the basis of scores averages and the magnitude of problem experienced on individual items, the performance of the bilingual group differs from that of the controls, there is a great deal of continuity between the two groups. Twelve bilinguals achieve a native-like score on this test, ten of these do so within the control group's time frame, and more come very close. Even participants who achieve lower scores mostly solve the test relatively fluently. Bilinguals and controls alike stumble over the same gaps for the same reasons and propose remarkably similar solutions, including the use of illicit native uses. All of these features point to acquisition, for some individuals to a very high level.

The exception to this overall impression of successful acquisition is the adverb-adjective confusion, which probably and apparently constitutes a larger problem for learners than native speakers, and some lexical items/collocations which are not as well established in all bilinguals' vocabularies. There are also several individuals who achieve relatively low scores. As we suggested at the beginning of this section, some of these are likely to be more proficient than their score suggests (S18, S26), while others (S7, S24) are indeed weaker as confirmed by the item analysis, and might simply be poorer language learners.

Typically, these bilinguals propose the types of solutions not found in the control group, committing "typical" learner errors. However, in many ways they do perform similarly to the rest of the participants, again pointing to acquisition rather than lack thereof. Given the significant difference between the scores achieved by the bilingual group and the German control group, it is an interesting speculation, but one for which we are unable to provide an answer, what level of L2 proficiency these participants might have achieved without the extended LOR in the L2 environment.

## 5.2 Verbal fluency tasks English

### 5.2.1 Test results

As we had done with the German FiCAs, we analysed the difference between the conditions on each task based on the responses of the Irish control group to check whether results could be combined. This comparison returned a significant difference for FiCA1, the category naming task ( $t(33) = 2.38, p = .023$ ). Although there is no apparent reason for this difference (in German, the two conditions had yielded similar results), *Professions* was significantly more productive than *Furniture*. There are no significant differences between conditions on FiCA2 or FiCA3 for the control group.

FiCA	Stimulus	Bilingual group					Irish controls				
		N	Mean	Me- dian	Range	StD	N	Mean	Me- dian	Range	StD
FiCA1 – category naming	<i>Professions</i>	9	19.9	20	11-28	4.6	17	23.2	24	13-34	6.3
	<i>Furniture</i>	17	16.0*	16	9-22	3.4	18	18.7	17	12-31	4.9
FiCA2 – letter fluency	<i>F</i>	12	17.0	16	11-23	3.8	15	18.2	17	11-27	4.7
	<i>A</i>	8	14.6	14.5	11-23	4.2	9	19.6	19	10-32	7.2
	<i>S</i>	7	19.7	18	9-27	6.2	9	20.9	24	8-28	7.4
FiCA3 – word listing by domain	<i>School</i>	8	24.4	24	16-31	4.5	13	27.1	26	18-37	6.6
	<i>Kitchen</i>	3	29.7	31	27-31	2.3	3	30.3	35	21-35	8.1
	<i>Street</i>	9	17.7*	20	8-21	4.3	12	21.9	20.5	15-34	5.7
	<i>Church</i>	6	23.3	24	18-28	3.8	6	26.2	26.5	19-32	6.1
<b>Scaled total score</b>		<b>25</b>	<b>66.28**</b>	<b>67.35</b>	<b>41-82</b>	<b>9.3</b>	<b>32</b>	<b>74.83</b>	<b>70.38</b>	<b>44- 112</b>	<b>17.0</b>

**Table 45: Mean scores for bilingual group and Irish control for all English FiCA conditions**

\* Significant at the .05 level.

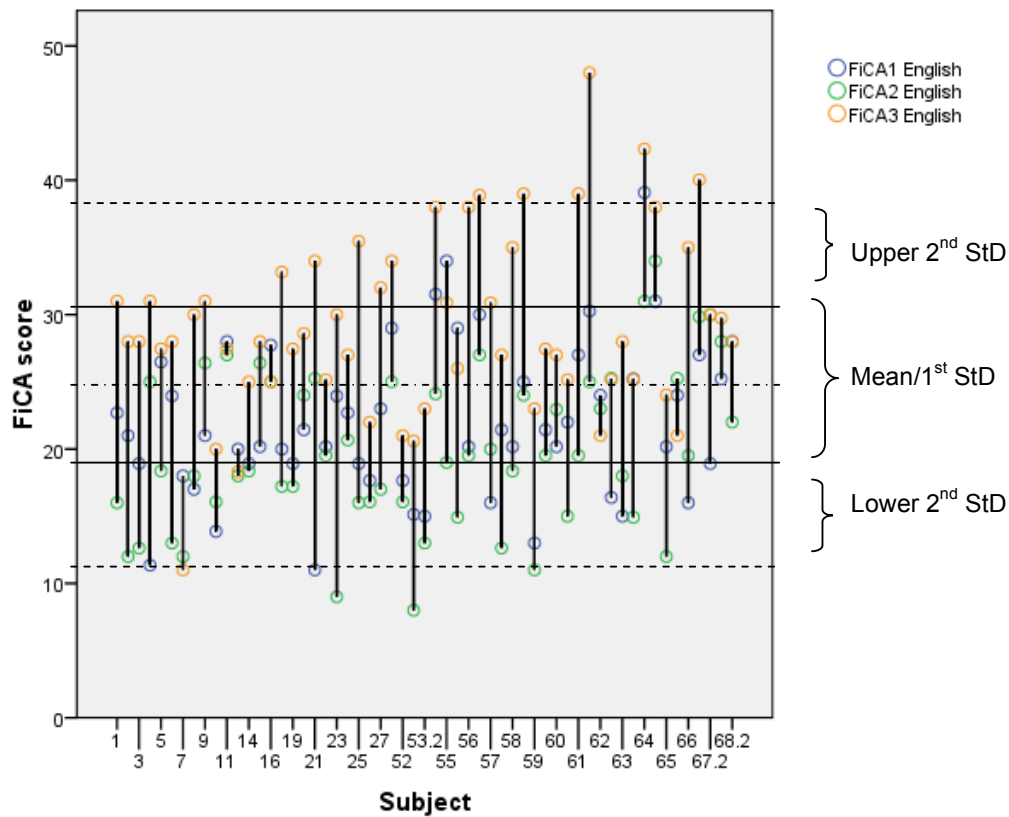
\*\* Significant at the .01 level.

A cursory glance at the group means (cf. Table 45) reveals that the means differences between bilingual group and control group are generally larger than on the German FiCA, and the difference concerning FiCA1 *Furniture*:  $t(33) = -1.90, p$  (one-tailed) = .033 and FiCA3 *Street* ( $t(19) = -1.87, p$  (one-tailed) = .039 is significant at the .05 level. However, most conditions fail to reach significance, and so the groups appear to be more similar than dissimilar (all  $ps$  one-tailed:

FiCA1 *Profession*:  $t(24) = -1.41, p = .086$ ; FiCA2 Letter *F*:  $t(25) = -0.71, p = .242$ ; FiCA2 Letter *A*:  $U = 22, p = .094$ ; FiCA2 Letter *S*:  $t(14) = -0.35, p = .365$ ; FiCA3 *School*:  $t(19) = -1.02, p = .159$ ; FiCA3 *Kitchen*:  $U = 3, p = .300$ ; FiCA3 *Church*:  $t(8.3) = -0.97, p = .181$ .

For some conditions, the absence of a significant effect may be a result of the relatively small numbers. Thus, should a difference between the groups exist, we were unable to demonstrate it on the basis of some of the individual conditions. However, after scaling the relevant conditions following the procedure described above in German FiCA, and combining them to summary scores, a *t*-test does return a significant difference ( $t(49.7) = -2.43, p = .01$  one-tailed), pointing to less fluent lexical processing and/or lower proficiency in English by the bilinguals.

As a trend we observe that the score ranges for the control group not only occupy higher bands, but they are generally larger than the bilinguals', as are the StDs. This points to heterogeneity of performance on the part of the Irish control group – some individuals appear to be much more verbose than others.



**Figure 16: Scores across three FiCA tasks for bilinguals (cases 1-27, N = 25) and German controls (cases 51+, N = 32)**

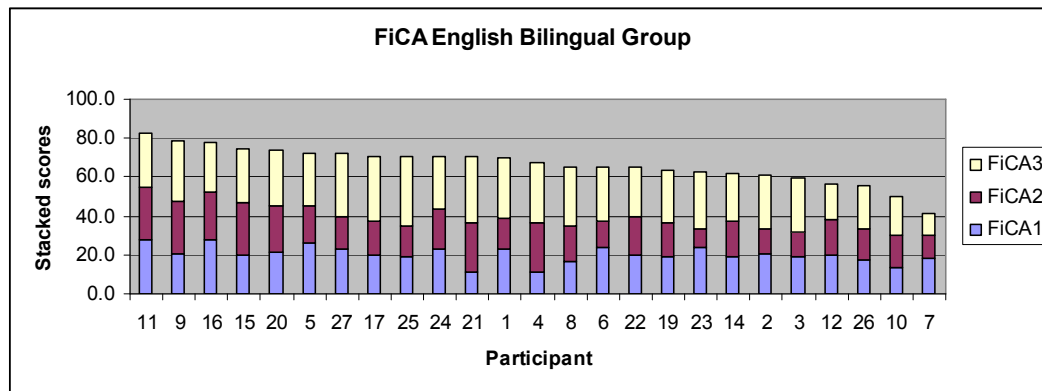
For a comparison of participants' performance, we again plot the scaled scores for each task. Figure 16 graphically represents the scores achieved by each individual across all three tasks after scaling, confirming the impression that the bilingual group mainly occupies the lower range of the scale, and that there are large inter-individual differences amongst the controls. This is also borne out in the figures presented in Table 46.

<b>FiCA German</b>	<b>Bilingual group</b>	<b>Irish control</b>
Number	27	34
Average score across three tests	n/a	24.94
Range of 1 <sup>st</sup> StD	n/a	18.3-31.6
Range of upper 2 <sup>nd</sup> StD	n/a	31.7-38.3
Range of lower 2 <sup>nd</sup> StD	n/a	11.6-18.2
Number with lowest value below lower 2 <sup>nd</sup> StD	6 (22%)	6 (18%)
Number with lowest value below lower 1st StD	13 (48%)	11 (32%)
Consistently low scores	4 (15%)	2 (6%)
Number with highest value above upper 2 <sup>nd</sup> StD	0 (0%)	7 (21%)
Number with highest value above 1st StD	4 (15%)	7 (21%)
Consistently high scores	7 (26%)	15 (44%)

**Table 46: Comparison of participant scores relative to control group's mean across three tasks**

Thus, over two-thirds of bilinguals (19 individuals, 70%) have their lowest score below or within the 2<sup>nd</sup> standard deviation, compared to 17 controls (50%). There are just four bilinguals (15%), whose highest score falls above the 1<sup>st</sup> StD, and none in the highest range, compared to 14 controls (42%), two of whom are very fluent individuals with all scores above or at the very top of the 1<sup>st</sup> StD. The total number of bilinguals with consistently high scores (values around mean and above) is half that of the controls (seven bilinguals (26%) vs. 15 controls (44%)). However, there are just four bilinguals (15%) and two controls (6%) with consistently low scores (below average). Moreover, there is not a single bilingual, whose performance could be classed as being

categorically different from that of any of the controls; and the lowest score overall (8 on FiCA2) in fact belongs to a control.



**Figure 17: Scores across three FiCA tasks for bilingual group (descending order)**

Figure 17 combines the scaled scores for the bilingual group for all three tests, and presents them in descending order by overall score. Participant S7 has the lowest score by far (41), which tallies with his performance on the C-test, followed by S10 (50). Three controls, S53 and S59, occupy roughly the same area with total scores of 44/51 and 47 respectively. At the other end of the scale, controls S61 and S64 are unmatched with 103 and 113/103 respectively. S64's performance is a very interesting case in terms of strategy use in verbal fluency tasks. While most participants usefully relied on semantic associations for FiCA1 and FiCA3, S64 started off in FiCA1 (*Professions*) by listing two or three professions following the letters of the alphabet.

In contrast with the wide range of total scores in the control group, the scores for the bilingual group are relatively homogeneous. However, while there is a trend for FiCA3 to be the most productive task, and FiCA2 the least productive one, this is not the case for all participants. Pronounced inter-individual differences between different categories are also present in the control group.

### 5.2.2 Item analysis

In this section, we again provide a detailed qualitative analysis of FiCA1, the category naming task (stimulus *Professions*). In preparing the data, plural and gendered forms were merged with their

standard form and are therefore not counted as separate types. The results of the type-token and common-contents analyses (cf. Table 47) show that the controls have a much higher token count, as well as a higher type count, but this may in part be explained by the difference in participant numbers for this stimulus.

<b>FiCA1 English</b>	<b>Bilingual group</b>	<b>Irish controls</b>
Category	<i>Professions</i>	
Number	10	17
Types	116	186
Tokens	184	379
Type-token ratio	0.63	0.49
Common contents	53%	33%
Basic occupations (N tokens (%))	297 (90.5%)*	209 (78.3%)*

**Table 47: Item analysis of FiCA1 Professions English**

\*\*\* The difference is significant at  $p = .000$ .

Over half of the bilingual group's types (52%) are also used by the control group, including all but one (*plumber*) of the bilinguals' top 10/top 25 items. (The top 10 items were named by at least four bilinguals/five controls, and the top 25 were named by at least two bilinguals/four controls.) Amongst both groups' top 10 are *teacher*, *doctor*, *nurse*, *baker* and *accountant*, amongst the top 25 – *butcher*, *gardener*, *engineer*, *dentist*, *carpenter* and *painter*, providing evidence for cross-lingual continuity, since these items were also amongst the top 10/25 in the German condition. The absence of *plumber* from the controls' list is interesting since it was in the top 25 for both groups in German.

The bilinguals' high-ranking *secretary*, *shopkeeper*, *artist*, *decorator*, *fireman*, *politician*, *bricklayer*, *computer programmer*, *electrician*, *farmer* and *policeman* do not make the control group's top 25, but all of them, bar *decorator*, are named by more than one person. All but three (*musician*, *author*, *plasterer*) of the remaining top 25 of the control group – *lecturer*, *builder*, *solicitor*, *waiter*, *banker*, *barrister*, *chef*, *civil servant*, *physiotherapist*, *professor* – occur further down the bilinguals' list.

Impressionistically, the differences between the groups' verbalisations appear smaller than had been the case with German FiCA1. Certain sets of jobs which we had found in the bilingual data, but which were not as widely represented in the German controls' data, such as trades (cf. top 25), professions and jobs connected to rural living (*farmer, gardener, conservationist, horticulturalist, landscaper*) and self-employment (*shopkeeper, restaurant owner*) as well as those reflecting job opportunities of the recent past (*builder, computer expert*), now appear in similar measure in both groups. Both groups also have different types of *drivers*. Interestingly, language-related jobs were only mentioned by the controls this time (*translator*, three mentions; *interpreter*, two mentions). Further, both groups name a few professions from the field of sports (*jockey, soccer trainer, tennis coach*) and politics (*politician*, three bilinguals, two controls; *foreign minister*), which were absent from the German set.

As on the German FiCA, there is a trend for the bilingual group to name more trades than the controls (40% vs. 33%). These in turn name more white-collar/specialist jobs ("the professions" – medical, health, judicial; 51% vs. 56%) and artistic jobs (*singer, film director, conductor, various musicians, artist* etc.; 3% vs. 8%). However, this difference fails to reach significance ( $t(559) = -.372, p = .710$ ). Similarly, despite differences in the type-token ratios, which seem to indicate that the bilingual group has a higher degree of lexical diversity, an analysis of basic level/prototypical vs. specialist jobs yields a non-significant result ( $t(383.418) = -1.820, p = .308$ ).

While the degree of lexical diversity is thus similar between the groups, they do differ somewhat in the verbalisation strategies used. A few controls extensively use associative sets reflecting subdivisions of professions or reporting structures, for example in relation to music or film production (*film director, producer, actor, stage hand, cameraman*) or clergy (*priest, pastor, minister, rabbi*), several types of *chartered, financial* or *banking professional*, management or catering/hospitality (*chef, sous-chef, front-of-house*). However, there is a good level of agreement on some tokens: 45 items are mentioned by at least three controls, and the top 2 (*doctor* and *teacher*) are mentioned by 16 ex 17 participants (94%).

In the bilingual group, on the other hand, and to a lesser degree in the control group, some participants apparently go out of their way to produce "novel" or interesting professions, such as *box manufacturer* and *nuclear power station attendant* (bilingual group), notably S7, whose scores

are low outliers on two conditions, but who at least on this condition does not appear to "pay" for his creativity with fewer verbalisations. These items also include a few somewhat borderline "jobs" in terms of "paid work": *burglar* (one mention in each group), *pickpocket* (one bilingual), *bank robber* (one bilingual) and *body-snatcher* (one control); as well as *liar* (one bilingual, possibly as in *a professional liar*) and *student* (one control; presumably accidentally as a result of listing the different participants in the educational process).

Finally, unlike the corresponding German FiCA condition, there are no signs of lexical retrieval difficulties with the bilinguals on this task. Any fluency-related problems appear to or were declared to stem from not being able to think of more concepts rather than the corresponding words. Also, there were no errors in the professions and jobs that were produced.

In summary, the combined evidence from the analysis of the English verbal fluency tasks paints a similar picture to the C-test results. At the group level, there are some differences between the groups but these are restricted to two conditions out of nine, and the overall score. The case for lack of difference is naturally stronger for FiCA1 and FiCA2 which had more participants in each condition. Moreover, many individuals perform well within the native speaker range on one or several of the tasks. In both contents and ease of access, as shown by the item analysis of one condition, *Professions*, both groups are remarkably similar. Again, we may conclude that second language acquisition, as measured by this test, has taken place, and that most bilinguals have attained a degree of verbal fluency approximating that of the control group.

### **5.3 Film retelling English**

The transcripts analysed in this section were prepared and coded in an analogous manner to the German film retellings, and several linguistically aware and/or trained native speakers were consulted in questions of accuracy and appropriacy of the retellings. No transcripts had to be excluded from the analysis, but from one control only Charlie 1 had been elicited. For a further eight controls only the monologic part was transcribed and marked up in full, since the bulk of the analyses is based on the monologues, and sufficient comparative material had been collected for the qualitative analysis. A total of 27 retellings from the bilingual group, and 35 retellings from the control group (cf. Table 48) were analysed for the same features previously outlined in Chapter 4:



1. sample length (time and tokens), lexical diversity (VOCD-D), pauses, repetitions and retracings on the basis of the narratives (monologues),
2. errors, code-switches and "gems" on the basis of the complete retellings.

<b>Film retelling English</b>	<b>Bilingual group</b>	<b>Irish controls</b>
Number of narratives/complete files, Charlie 1	12/12	18/11
Number of narratives/complete files, Charlie 2	15/15	17/9
Total number of narratives/complete files	27/27	35/20

**Table 48: Number of English film retellings**

### **5.3.1 Narratives**

#### **5.3.1.1 Sample length**

We again start by comparing the length of the narratives between the two groups for both parts of Charlie (cf. Table 49). For both retellings, the bilingual group took longer to complete the narratives, but the difference is less pronounced for Charlie 2, and the controls' median for Charlie 2 is in fact slightly higher than that of the bilinguals. The difference is significant for Charlie 1 at the .05 level ( $t(28) = 2.656, p = .013$ ) but not for Charlie 2 ( $U = 122.000, p = .845$ ).

<b>Film retelling English</b>	<b>Bilingual group</b>	<b>Irish controls</b>	<b>Bilingual group</b>	<b>Irish controls</b>
	<b>Charlie 1</b>		<b>Charlie 2</b>	
Number	12	18	15	17
Mean length of narratives (min:sec)	04:32	03:09	04:12	03:59
Median	04:33	03:13	04:09	04:14
StD	01:47.73	01:03.29	01:16.55	00:59.04
Range	02:18-9:15	01:08-05:25	02:00-06:47	01:55-05:35
Mean tokens	721.08	551.44	670.93	763.24
Median	646	586	594	744
StD	321.96	195.24	248.38	210.39
Range	450-1616	161-911	307-1199	341-1097

**Table 49: Length of English film retellings**

This picture is mirrored in the token count for Charlie 1, but not for Charlie 2, for which the controls have a higher token count than the bilinguals, while the range is fairly similar. The difference between the two groups is not significant for either part of Charlie (Charlie 1:  $t(28) = 1.801, p = .082$ ; Charlie 2:  $t(30) = 1.138, p = .264$ ). Time taken and the token count again correlate highly for both parts (Charlie 1:  $r = .890$ ; Charlie 2:  $r = .779$ ; both correlations are significant at the .01 level).

As the StDs and ranges for both measures reveal, there are large differences between individuals in both groups, probably again relating to fluency, personality and effort variables. Several participants' narratives are extremely succinct, requiring extensive elicitation of detail after the monologues. For example, control S55's narrative is just 161 tokens (22 utterances) long, and the whole episode of Charlie being asked to find a wedge by the foreman, knocking it out from under a supporting beam and thereby launching a half-finished ship is summarised as in Example 1:

- 1) \*XYZ:        he # wrecked < everything > [=! laughing] !  
           \*XYZ:        and sent the ship out to sea .  
           \*XYZ:        when it was n(o)t finished yet .

Other participants retold the stories with careful attention to detail and put plot and characters in a broader context, for example by spelling out the characters' reasons for their actions. S11, an outlier within the bilingual group, who produced 1616 tokens for Charlie 1, almost twice as many as the closest bilingual (S6, 879 tokens) and ten times as many as S55 above, and who "causes" the bilingual group's retellings for Charlie 1 to be longer rather than shorter than those for Charlie 2, uses 26 utterances just for that one "wedge episode", commenting on Charlie's character along the way (Example 2):

- 2) \*XYZ:        < you can > [/] you can tell that he (i)s supposed to be a bit simple .  
           \*XYZ:        because # < it (i)s > [/] it (i)s very difficult to mistake # one thing for another if  
                           you were shown it five minutes before .  
           [...]  
           \*XYZ:        because he (i)s quite aware that he did something stupid once again .

Similarly, S17 and S25 (also outliers in the bilingual group with the highest and second highest token count for Charlie 2 respectively), and the controls S56 (highest token count amongst the controls for both parts), S67 and S61 (second and third highest on Charlie 2), comment on the characters' motives and states of mind. S59 notices that the song playing in the background in Charlie 2 is "Smile though your heart is breaking", and relates that to the story; S67 comments on the health risks involved in pulling up a person who has fallen and now lies unconsciously on the road. S25 has several utterances situating the story culturally, wondering, for example, whether he correctly calls the policemen *coppers* since the film is set in America, and the term is British.

In general, it is worth noting that several controls have a markedly higher token count on the second part of Charlie, which also has a bearing on the lexical diversity scores discussed below. Impressionistically, this is not just related to the fact that Charlie 2 objectively is a longer sequence, as we pointed out above, but also to participants taking a more detailed approach the second time round, probably partially prompted by the investigator's follow-up questions after the first part. This had not been observed to the same extent on the German retellings. We shall review some more examples of how the narratives were embellished in the subsections on lexical diversity and the gem analysis below.

### **5.3.1.2 Lexical diversity**

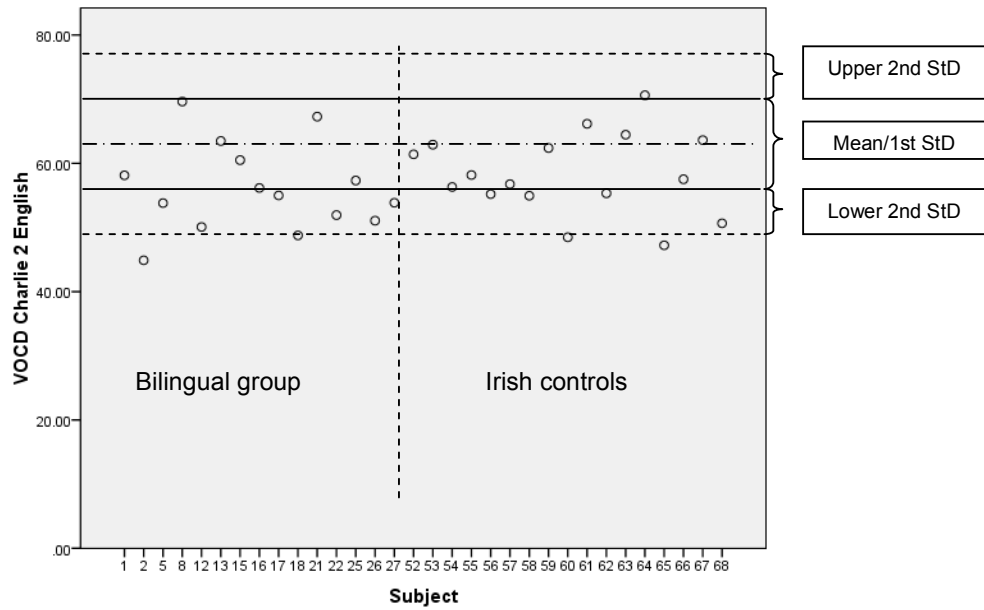
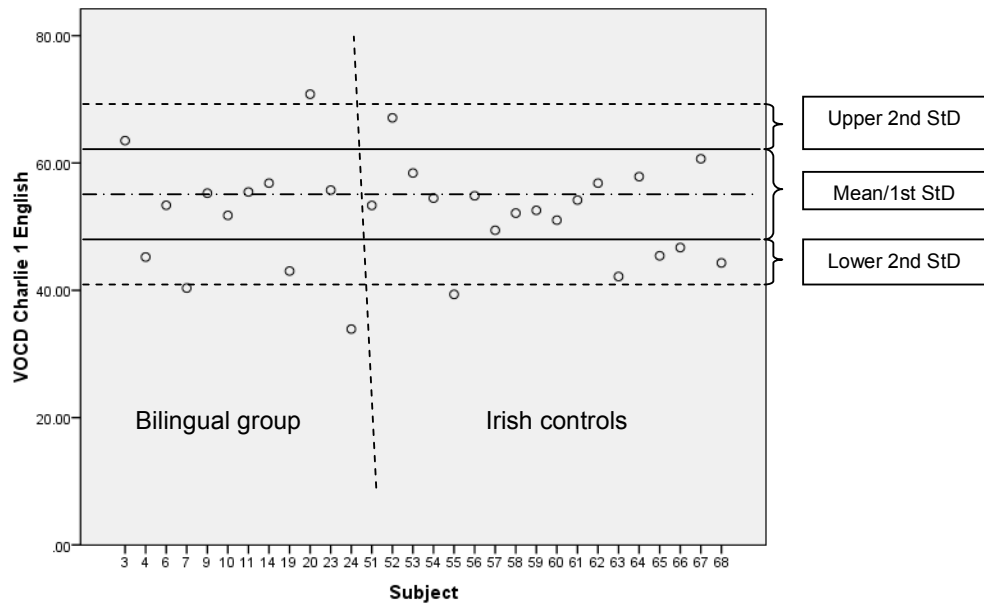
Turning now to the measure of lexical diversity, we again computed VOCD-D twice – as "raw" D, excluding hesitation words and retracings, and standard D (cf. Table 50). Apart from Charlie 1, where the bilingual group has a larger range compared to Charlie 2, the means, medians, StDs and ranges are very similar for both groups and parts. Indeed, the comparison between groups proves to be non-significant for both parts of Charlie on both measures (raw VOCD-D: Charlie 1:  $t(28) = -0.361$ ,  $p = .721$ , Charlie 2:  $t(30) = -1.270$ ,  $p = .214$ ; standard VOCD-D: Charlie 1:  $t(28) = -0.053$ ,  $p = .958$ , Charlie 2:  $t(30) = -0.956$ ,  $p = .345$ ). However, a comparison of the control group's data across both tasks returns a highly significant difference (raw VOCD-D:  $t(16) = -3.277$ ,  $p = .005$ ; standard VOCD-D:  $t(16) = -3.376$ ,  $p = .004$ ), confirming the differences we had noticed when comparing the token counts. This means that we cannot pool the data for this measure.

<b>Film retelling English</b>	<b>Bilingual group</b>	<b>Irish controls</b>	<b>Bilingual group</b>	<b>Irish controls</b>
	<b>Charlie 1</b>		<b>Charlie 2</b>	
Number of narratives	12	18	15	17
Mean VOCD-D, excluding hesitation words and retracings	52.10	52.27	56.12	58.36
Median	54.31	52.95	54.99	57.52
StD	10.17	6.93	6.85	6.33
Range	33.90-70.81	39.36-67.10	44.88-69.63	47.23-70.61
Mean standard VOCD-D	57.10	55.93	59.14	62.26
Median	56.92	57.11	58.87	62.54
StD	10.33	7.44	7.63	6.26
Range	42.13-71.87	43.25-67.81	43.68-72.77	50.43-72.18

**Table 50: VOCD-D values for English film retellings**

The absence of any significant differences between the groups implies that the bilingual ("learner") group is performing on a par with the native-speaker controls and manages to narrate the stories with a similar degree of lexical diversity overall. This may be seen as evidence of successful second-language acquisition on the part of the bilingual group. All VOCD-D values are lower than for the German retellings but since this is the case for both bilinguals and controls, we assume that that may partially be the result of differences in the structures of the particular languages used, and possibly crosslinguistically different narrative styles. The differences in the range of values are also lower than for German, creating the impression of a more homogeneous performance by both groups.

This impression of relative homogeneity within groups and of the essentially overlapping performance of groups is confirmed by Figure 18 which shows the distribution of lexical diversity scores by individuals. A clear tendency becomes apparent – there are very few high scores, with more scores clustering in the lower regions of the bands. For example, there is just one participant, S20, a bilingual who has a score above the 2<sup>nd</sup> StD (Charlie 1), and just one participant in each group for both parts of Charlie (Charlie 1: S3 and S52; Charlie 2: S8 and S64) who have a score within the 2<sup>nd</sup> StD (cf. Table 51).



**Figure 18: Lexical diversity scores (standard D) for English film retelling**

	<b>Bilingual group</b>	<b>Irish controls</b>
Above upper 2 <sup>nd</sup> StD	S20	-
Within upper 2 <sup>nd</sup> StD	S3, S8	S52, S64
Within lower 2 <sup>nd</sup> StD	S4, S6, S12, S17, S19, S22, S26, S27	S56, S58, S62, S63, S65, S66, S68 (x 2)
Below lower 2 <sup>nd</sup> StD	S2, S7, S18, S24	S55, S60, S65

**Table 51: Distribution of individual VOCD-D (standard D) scores**

As Table 51 shows, in each group there are several individuals with scores in or below the 2<sup>nd</sup> StD (Charlie 1: two bilinguals (S4, S19) within, two bilinguals (S7, S24) below, four controls (S66, S65, S68, S63) within, one control (S55) outside; Charlie 2: six bilinguals (S17, S27, S5, S22, S26, S12) within, two (S19, S2) below, four controls (S62, S56, S58, S68) within and two controls (S60, S65) below). Three of the bilinguals with the lowest values, S7, S18 and S24, also had the lowest C-test scores, and S7 additionally the lowest total FiCA score. However, their lower performance on this variable is put in perspective by the fact that it is matched by that of native speakers, so cannot be seen as an indication of lower attainment on the basis of this task.

At the other extreme, one bilingual, S20, actually outperforms all other participants, and two more, S3 and S8, have a high score on a par with the control group. Again, these results point to successful acquisition on the part of the bilingual group. Of course, lexical diversity is just one aspect of the quality of a retelling, so this score needs to be related to other measures to either confirm or disconfirm whether an individual speaker is an excellent or a less proficient speaker.

### **5.3.1.3 Fluency**

As we had done with the German film retellings before, we again looked at various fluency/hesitation measures (cf. Table 52), which were computed using the relevant CLAN commands and combined in an analogous manner to arrive at the three relevant indices – the pause index, the retracing index, and the overall fluency index.

Table 52 below shows that, again, the two groups perform very comparably on all measures. While overall there is a tendency for the bilingual group to have slightly higher averages than the control group, it actually outperforms the control group on some features (all retracing measures for Charlie 2). Only two of the fluency measures – silent pauses and repetitions, both for Charlie 1 only – reach significance (one-tailed test: silent pauses:  $t(28) = 1.844, p = .038$ ; repetitions:  $t(28) = 1.740, p = .047$ ; we refrain from reporting the non-significant results for the other features). When pooling the repetition data for both Charlies, the difference becomes non-significant ( $U = 359.500, p = .055$  one-tailed), unlike what we had seen on the German retellings.

Film retelling English	Bilingual group	Irish controls	Bilingual group	Irish controls
	Charlie 1		Charlie 2	
Number of narratives	12	18	15	17
Mean silent pauses (%)	1.84*	1.34*	2.31	1.63
Median	1.84	1.11	1.71	1.64
StD	0.68	0.77	1.42	0.53
Range	0.73-3.00	0.18-3.03	0.93-6.56	0.53-2.62
Mean filled pauses (%)	4.94	4.24	3.75	3.76
Median	4.84	3.91	3.26	3.57
StD	2.39	1.95	2.03	1.61
Range	1.63-9.50	0.17-8.80	1.35-8.52	2.05-8.04
Mean pause index (%)	6.79	5.57	5.62	5.39
Median	6.64	5.20	5.5	5
StD	2.27	2.04	1.94	1.67
Range	3.44-10.54	0.82-10.09	2.87-10.23	2.99-9.11
Mean repetitions (%)	1.94*	1.16*	1.57	1.16
Median	1.96	0.76	1.52	0.73
StD	1.32	1.12	1.64	1.02
Range	0.18-4.22	0.00-3.65	0.00-6.13	0.32-4.04
Mean retracings (%)	1.59	1.39	1.64	2.13
Median	1.57	1.71	1.26	2.05
StD	1.05	0.68	0.94	0.74
Range	0.41-3.41	0.26-2.66	0.36-3.25	0.84-3.21
Mean retraced material (%)	4.67	4.26	4.69	5.14
Median	3.92	4.42	4.23	4.26
StD	3.39	2.00	2.66	2.09
Range	0.41-11.87	0.79-7.23	1.10-10.09	2.06-10.23
Mean retracing index (%)	3.53	2.89	3.21	3.28
Median	3.35	2.79	2.58	2.91
StD	2.27	1.34	2.09	1.44
Range	0.61-7.52	0.52-5.48	0.74-7.72	1.22-6.02
Mean fluency index (%)	10.32	8.46	8.84	8.67
Median	11.19	7.89	8.09	7.43
StD	4.04	3.04	3.62	3.12
Range	4.05-18.06	2.48-15.57	4.41-17.96	4.27-17.46

**Table 52: Fluency measures for English film retellings**

\* Group differences are significant at the .05 level.

We were unable to pool the silent pause measure for both Charlies since a repeated measures test for the control group showed them to be significantly different; however, given the between-group result for Charlie 2 ( $U = 92.000, p = .095$  one-tailed) it is possible that again the difference would disappear overall.

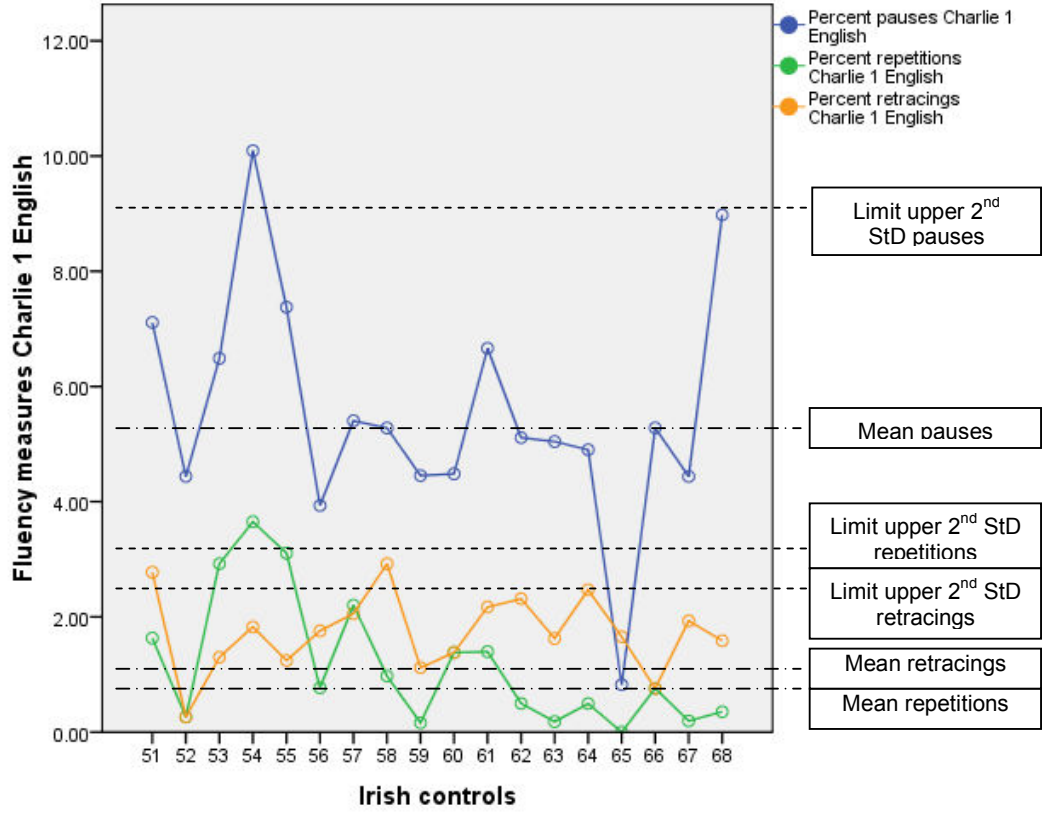
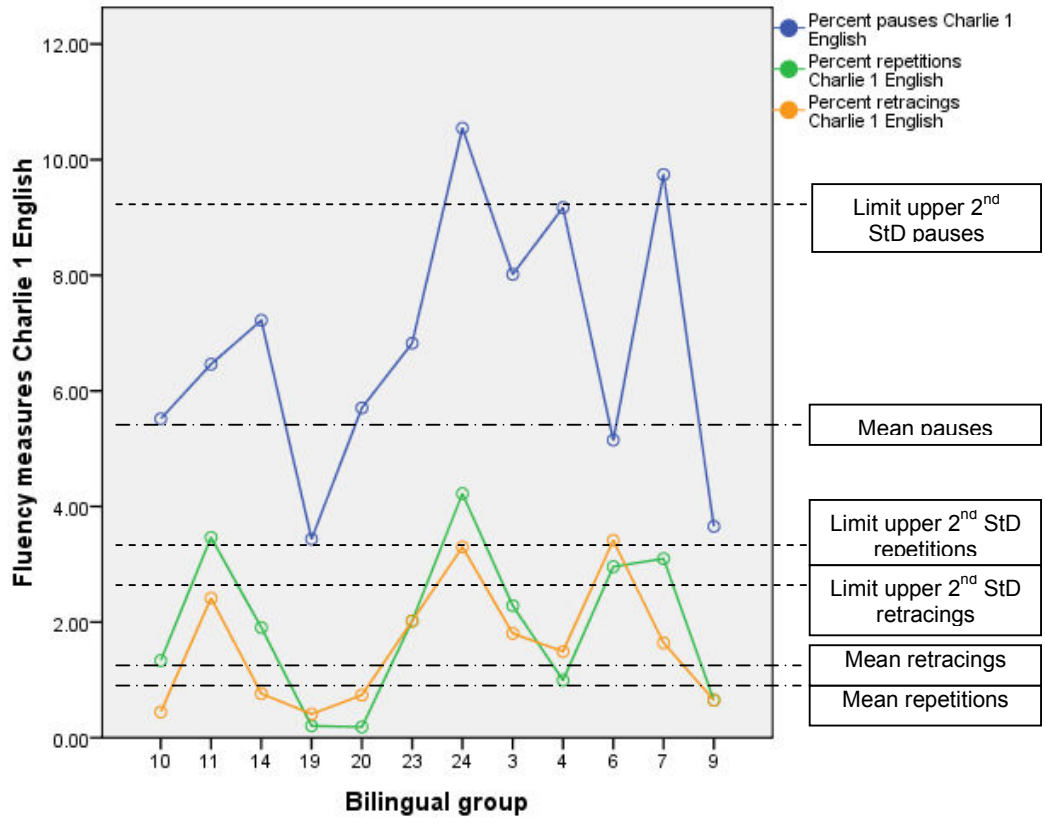
When combining the various features into the pause index and the retracing index respectively, the groups again appear to be performing similarly. Thus, we may surmise that at the overall group level the bilingual group may not be distinguished from native speakers with regard to the fluency of their speech, apart from some bilinguals narrating Charlie 1 who use more silent pauses and repeat words more often. Word repetition was the single feature on which the bilingual group differed from the German control group in the German film retellings, where we had argued that they may be traded-off for silent pauses. Since both features show up as significantly different, the bilinguals narrating Charlie 1 appear indeed to be somewhat more hesitant/less fluent in their production than the control group.

Since the control group's performance was not comparable on both parts of Charlie for all features (silent pauses and retracings), we provide separate graphs for Charlie 1 and 2 for the three principal fluency features (pauses, repetitions and retracings, cf. Figure 19 and Table 53 below).

Similarly to our results for the German retellings, there is a great deal of inter-individual variability, regardless of task. Table 53 lists those individuals (bilinguals: S1-S27, controls: S51-68) who on one or more measures fall within the upper 2<sup>nd</sup> StD (bold italics) or above (bold). S24 exceeds the 2<sup>nd</sup> StD on all measures, followed by S18 and S25 who are above the 2<sup>nd</sup> StD on two measures (pauses/repetitions, and repetitions/retracings respectively). S25 also has a high percentage of pauses, while S18's percentage of retracings is very low. S7 and S3 have a similar profile to S18 with slightly lower pause and repetition values (one or both within the 2<sup>nd</sup> StD).

S24, S18 and S25 thus fall outside, while S7 and S3 are at the very fringe of the native-speaker range on at least two fluency measures. S7, S18, and S24 had already been noted with low values on VOCD-D, as well as the C-test; S7 and S25 also had low FiCA values, providing converging evidence that these participants may be somewhat less proficient. S3, on the other hand, had a very high value of lexical diversity, thus in her case the high percent of pauses and repetitions merely indicates that her speech is carefully planned.





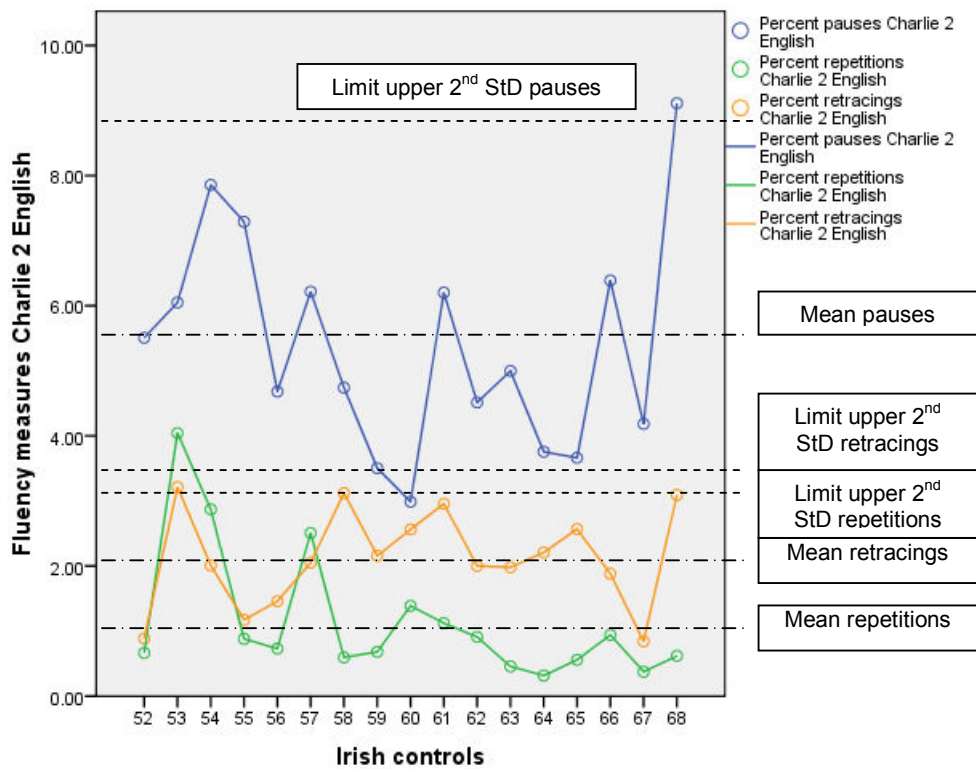
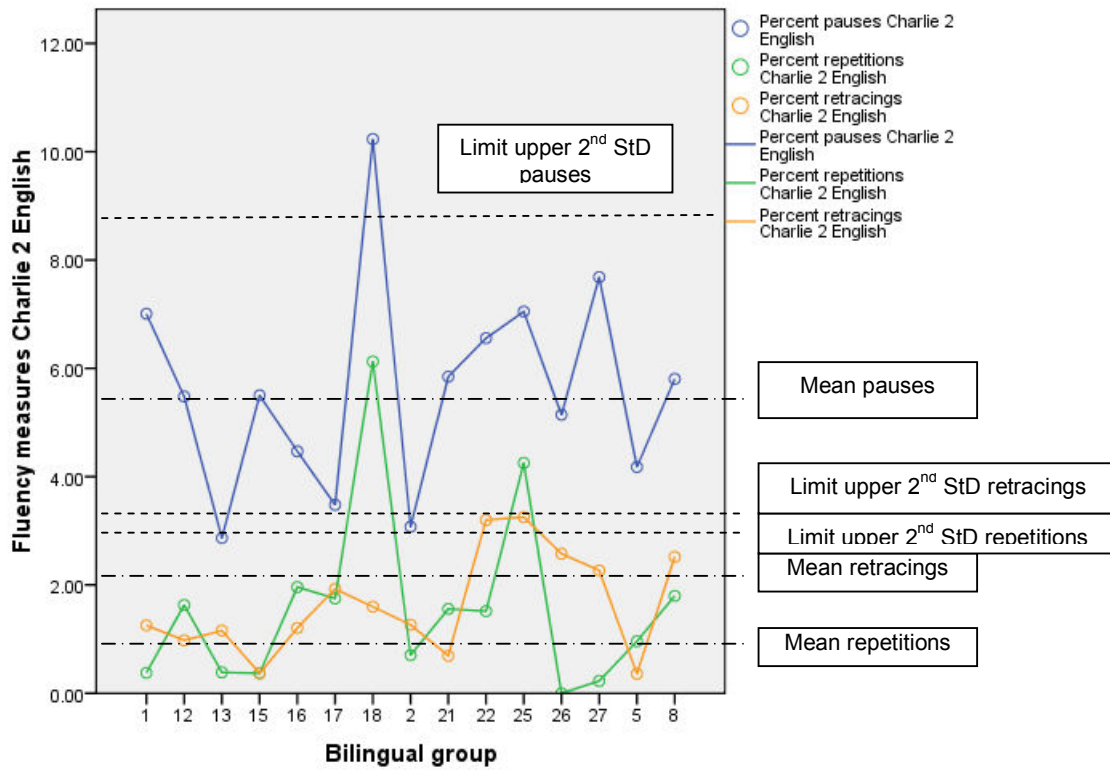


Figure 19: Fluency measures English film retelling vis-à-vis native-speaker baseline

<b>Charlie 1</b>	<b>Pauses</b> Mean: 5.57% Upper limit 2 <sup>nd</sup> StD: 9.65%	<b>Repetitions</b> Mean: 1.16% Upper limit 2 <sup>nd</sup> StD: 3.40%	<b>Retracings</b> Mean: 1.39 Upper limit 2 <sup>nd</sup> StD: 2.75%
<b>Participant</b>			
S24	<b>10.54%</b>	<b>4.22%</b>	<b>3.30%</b>
S7	<b>9.74%</b>	<b>3.10%</b>	1.64%
S4	<b>9.17%</b>	0.99%	1.49%
S3	<b>8.02%</b>	<b>2.28%</b>	1.80%
S11	6.46%	<b>3.47%</b>	<b>2.41%</b>
S54	<b>10.09%</b>	<b>3.65%</b>	1.83%
S68	<b>8.98%</b>	0.35%	1.58%
S55	7.38%	<b>3.11%</b>	1.24%
S53	6.49%	<b>2.92%</b>	1.30%
S58	5.28%	0.97%	<b>2.92%</b>
S51	7.11%	1.63%	<b>2.77%</b>

<b>Charlie 2</b>	<b>Pauses</b> Mean: 5.39% Upper limit 2 <sup>nd</sup> StD: 8.73%	<b>Repetitions</b> Mean: 1.16% Upper limit 2 <sup>nd</sup> StD: 3.20%	<b>Retracings</b> Mean: 2.13 Upper limit 2 <sup>nd</sup> StD: 3.61%
<b>Participant</b>			
S18	<b>10.23%</b>	<b>6.13%</b>	1.60%
S27	<b>7.68%</b>	0.23%	2.27%
S25	<b>7.05%</b>	<b>4.25%</b>	<b>3.25%</b>
S22	6.56%	1.52%	<b>3.20%</b>
S68	<b>9.11%</b>	0.62%	<b>3.09%</b>
S54	<b>7.86%</b>	<b>2.87%</b>	2.01%
S55	<b>7.29%</b>	0.88%	1.17%
S53	6.05%	<b>4.04%</b>	<b>3.21%</b>
S58	4.74%	0.59%	<b>3.12%</b>
S61	6.20%	1.12%	<b>2.96%</b>

**Table 53: Individual results on three fluency measures, English film retelling**

There are two controls, S54 and S68, who show a similar profile on both Charlies: S54 pauses and repeats a lot, while her retracings are average; and S68 pauses a lot and on Charlie 2 uses a lot of retracings but very rarely repeats words. S53 also has a higher level of repetitions (both Charlies) and retracings (Charlie 2) with pauses in the above average normal range, and his profile is matched by S11 (bilingual).

There are several more controls, and one bilingual who have higher values on one of the features: S22 (retracings); S27 (pauses); S58 (retracings, both Charlies); S55 (repetitions, Charlie 1; pauses, Charlie 2); S51 (retracings, Charlie 1), and S61 (retracings, Charlie 2). It is interesting to note on the basis of the control group's data that individual "profiles" of fluency features appear to be relatively stable within a person across different samples, possibly reflecting personality features and preferences in speaking, for example, whether or not to use filled pauses. S55's case seems to confirm the link between pausing and repetitions mentioned in the German retellings.

Thus, the analysis of the fluency profiles of individuals at the high end of the scale confirms that several bilinguals have very high values relative to the control group; the difference between their values and those of the controls is, however, not categorical but one of the degree, since several controls also have values in the same region. Where a (possibly categorical) difference does exist, it is in the number of features on which an individual has a high value – in the control group it is a maximum of two (S53, S54, S68), but usually just one feature (S51, S55, S58, S61). For the bilinguals, on the other hand, the high values range from three (S24, S25) through two (S3, S7, S11, S18) to one (S4, S22 and S27). The presence of several such features would seem to indicate a lower degree of fluency compared to the controls, even if the values themselves are more or less matched by the controls.

At the other end of the scale, there are several bilingual individuals on a par with, and in some cases outperforming the controls in terms of consistency (cf. Table 54). S19 and S13's values are all either within or below the 2<sup>nd</sup> StD, or very close to it. (There is a ceiling effect with regard to the low scores on some of the measures so that it is impossible to achieve scores outside the lower 2<sup>nd</sup> StD on these). These participants can be considered very fluent speakers all around. Interestingly, there are no matching controls with three very low values.

S9, S20, S2 and S15 are also quite fluent since they have two very low values each (below, within or very close to lower 2<sup>nd</sup> StD), like several of the controls (S67 both Charlies; S65, S52 Charlie 1; S64 Charlie 2). The remainder of the bilinguals and controls in Table 54 (S65 and S52 come up again) has one "fluent" value. We had already discussed S55 with regard to high values above – while she has high pause and repetition percentages on Charlie 2, her retracings are within the lower 2<sup>nd</sup> StD.

<b>Charlie 1</b>	<b>Pauses</b> Mean: 5.57% Lower limit 2 <sup>nd</sup> StD: 1.49%	<b>Repetitions</b> Mean: 1.16% Lower limit 2 <sup>nd</sup> StD: n/a	<b>Retracings</b> Mean: 1.39 Lower limit 2 <sup>nd</sup> StD: 0.03%
<b>Participant</b>			
S19	<b>3.44%</b>	<b>0.20%</b>	<b>0.41%</b>
S9	<b>3.66%</b>	0.65%	<b>0.65%</b>
S20	5.70%	<b>0.18%</b>	<b>0.74%</b>
S10	5.52%	1.33%	<b>0.44%</b>
S65	<b>0.82%</b>	<b>0.00%</b>	1.66%
S52	4.44%	<b>0.26%</b>	<b>0.26%</b>
S59	4.45%	<b>0.16%</b>	1.12%
S63	5.04%	<b>0.18%</b>	1.62%
S67	4.44%	<b>0.19%</b>	1.93%
S66	5.28%	0.76%	<b>0.76%</b>

<b>Charlie 2</b>	<b>Pauses</b> Mean: 5.39% Lower limit 2 <sup>nd</sup> StD: 2.05%	<b>Repetitions</b> Mean: 1.16% Lower limit 2 <sup>nd</sup> StD: n/a	<b>Retracings</b> Mean: 2.13 Lower limit 2 <sup>nd</sup> StD: 0.65%
<b>Participant</b>			
S13	<b>2.87%</b>	<b>0.39%</b>	<b>1.16%</b>
S2	<b>3.07%</b>	0.70%	<b>1.26%</b>
S15	5.50%	<b>0.37%</b>	<b>0.37%</b>
S5	4.18%	0.96%	<b>0.36%</b>
S17	<b>3.48%</b>	1.75%	1.93%
S26	5.14%	<b>0.00%</b>	2.58%
S27	7.68%	<b>0.23%</b>	2.27%
S67	4.18%	<b>0.37%</b>	<b>0.84%</b>
S52	5.51%	0.66%	<b>0.88%</b>
S60	<b>2.99%</b>	1.39%	2.65%
S64	<b>3.76%</b>	<b>0.32%</b>	2.21%
S59	<b>3.50%</b>	0.68%	2.16%
S65	<b>3.66%</b>	0.56%	1.64%

**Table 54: Individual results on three fluency measures, English film retelling, continued**

We will refrain from spelling out individual "fluency profiles" since this would not add substantially to the present discussion. Overall then, the differences between controls and bilinguals are rather small, and few bilinguals, sometimes apparently as a result of lower L2 proficiency, are less fluent than the controls. Indeed, several bilinguals apparently are just as fluent.

### 5.3.1.4 Measures combined

Last in this subsection, we again correlated the various fluency measures with sample length (time taken and total tokens) and the lexical diversity measure (raw VOCD-D). The results resemble those from the analogous comparison in the German section in the two length measures correlating highly with one another, as mentioned above, and in the retracing measures correlating amongst themselves.

<b>Charlie English</b>		<b>Time taken</b>	<b>Total tokens</b>	<b>Raw VOCD-D</b>	<b>% pauses</b>	<b>% reps</b>	<b>% retrace</b>	<b>% retr mater</b>
Time taken	Pearson Correlation	1.000	.845**	.266*	.177	.423**	.385**	.539**
	Sig. (2-tailed)		.000	.036	.169	.001	.002	.000
	N	62.000	62	62	62	62	62	58
Total tokens	Pearson Correlation	.845**	1.000	.185	-.070	.254*	.372**	.470**
	Sig. (2-tailed)	.000		.149	.586	.047	.003	.000
	N	62	62.000	62	62	62	62	58
VOCD-D excluding hesitation words & retracings	Pearson Correlation	.266*	.185	1.000	.096	.128	.113	.145
	Sig. (2-tailed)	.036	.149		.457	.322	.381	.276
	N	62	62	62.000	62	62	62	58
Percent pauses	Pearson Correlation	.177	-.070	.096	1.000	.578**	.204	.339**
	Sig. (2-tailed)	.169	.586	.457		.000	.112	.009
	N	62	62	62	62.000	62	62	58
Percent repetitions	Pearson Correlation	.423**	.254*	.128	.578**	1.000	.298*	.670**
	Sig. (2-tailed)	.001	.047	.322	.000		.018	.000
	N	62	62	62	62	62.000	62	58
Percent retracings	Pearson Correlation	.385**	.372**	.113	.204	.298*	1.000	.819**
	Sig. (2-tailed)	.002	.003	.381	.112	.018		.000
	N	62	62	62	62	62	62.000	58
Percent retraced material	Pearson Correlation	.539**	.470**	.145	.339**	.670**	.819**	1.000
	Sig. (2-tailed)	.000	.000	.276	.009	.000	.000	
	N	58	58	58	58	58	58	58.000

**Table 55: Correlations between sample length, lexical diversity and fluency measures**

\*\* Correlation is significant at the 0.01 level (2-tailed).

\* Correlation is significant at the 0.05 level (2-tailed).

However, pauses do not now correlate with time taken or retracings, which seems a bit counter-intuitive. The (still non-significant) relationships between pauses and tokens/lexical diversity have reversed their signs, thus indicating that the more tokens a person had, the fewer pauses they made, and the more pauses they made, the higher the degree of lexical diversity. The latter seems plausible, the former less so. Lexical diversity, on the other hand, still does not correlate with the majority of features, but the relationship with the time taken to tell the story has become significant, indicating that those people who took longer to narrate the story had a higher degree of lexical diversity. Part of this effect we would have expected to come from pauses but this is not borne out in the data.

### 5.3.2 Scenes

#### 5.3.2.1 Code-switching

For this section, we again coded all instances of code-switches for subsequent analysis with the relevant CLAN command, as outlined in the discussion of the German film retellings. The number of code-switches is much smaller overall, and their distribution differs from that in the German narratives. There are no instances of quotes, self-talk or reference to actors; however, there are isolated instances of lexical gaps and of interferences, interestingly in both groups. The most frequent type of code-switch are interjections and discourse particles, such as *Ach ja!* ("Oh yes!") or *also* ("well"), in line with findings in the literature (Olshtain and Blum-Kulka 1989; Maschler 1994; both cited in Callahan 2004).

Type of code-switch	Bilingual group (N = 27)		Irish control (N = 35)	
	Instances	Individuals	Instances	Individuals
Lexical gaps	2	S22, S24	1	S57
Interferences	1	S18	1	S56
Interjections/discourse particles	12	S2, S7, S8, S24, S26	0	—
<b>Total</b>	<b>15</b>	<b>7</b>	<b>2</b>	<b>2</b>

**Table 56: Code-switches film retelling English**

The lexical gaps concern the items *barrel*, *land of milk and honey*, and *kiosk* or similar (Examples 3 and 4, bilingual group; Example 5, control group). In Example 3, the participant was talking about Charlie wanting to put his coat and hat down on a barrel, as he had been told to do by the foreman, and made the lexical gap explicit. The code-switch is again marked @s.

- 3) \*XYZ: and so Charlie put it back onto erm@fp erm@fp +//.  
 \*XYZ: ein@s Fass@s ?  
 (a barrel ?)  
 \*XYZ: I do (not) know .
- 4) \*XYZ: it (i)s a < sup of > [//] sort of < Schlaraffenland@s > ["] .  
 (it is a sort of "land of milk and honey" .)
- 5) \*XYZ: he [-] there was just a kind of erm@fp a shop # or a counter , or a +...  
 \*XYZ: whatever you call those xxx .  
 \*XYZ: an Imbiss@s thing !  
 (a snack bar thing !)

A related case is the loan translation into English employed by a bilingual to fill a lexical gap (*meat loaf*), for which he first produced *meat bake* and then *fake rabbit* (German: *Falscher Hase*, traditional name of the dish).

In Example 4, the participant went for the German equivalent of the same item that some of the German controls had imported from English into German, thereby indicating that the English expression is quite current. This switch is prosodically flagged, meaning that it was conscious – presumably since the German phrase was more available, but it is possible that the participant deliberately chose what to her seemed to be a more fitting expression (cf. also Example 5). This participant had started off her narrative in German, having misunderstood the instructions, which may have resulted in her German being more active than it otherwise might have been.

In Example 5, the control used a term that is culturally quite specific for a German context, after a longer lexical search which subjectively failed to turn up a fitting lexeme ("whatever you call those"). The word chosen instead is, however, not entirely suitable either since *Imbiss* refers to stalls or kiosks offering food.



Such occurrences are puzzling in the case of the controls since they were speaking in their mother tongue, which all controls were dominant in. Furthermore, the narratives were elicited as one of the tasks in the middle of a longer English-only section, which was clearly flagged as such. This example highlights the fact that the "other language" was nevertheless available enough in some participants to be consciously drawn on, as in Examples 4 and 5, or to interfere, as in Example 6 (control group), which we attribute to the impact of the bilingual experimental setting.

- 6) \*XYZ:       takes off his hat .  
\*XYZ:       and gets into the police # < &wag > [//] +//.  
\*XYZ:       erm@fp Wagen@s !  
              (erm car !)  
\*XYZ:       +, erm@fp police car .

In this example, *Wagen* ("car", "van") is phonetically very similar to *wagon*, which incidentally is part of the "colourful" denomination for a police car, *paddywagon*, which is employed by several participants.

The single case of an interference in the bilingual group that is not related to interjections or discourse markers concerns *ganz* ("very", "totally"), which intruded and was "caught" (i.e. the phrase was retraced) when the participant tried to explain the girl's attitude towards Charlie (she was playing "very flirtatiously" with a flower she had picked).

In the German retellings, the sub-group of interjections was attested with a few occurrences of *and*, used to link utterances, and the interjection *well*. These were used by participants who, in line with their results on other measures, have fairly active English (S1, S3, S10, S11). In contrast, the use of German interjections in English occurs mostly with participants whose English appears to be somewhat less proficient or fluent. All the relevant participants (S2, S7, S18, S24, S26), apart from S8, were mentioned in the previous subsections in relation with lower lexical diversity and/or higher hesitation percentages.

The participant who uses German interjections most frequently is S2, with once occurrence each of *also* ("well", Example 7), *ach* ("oh!"/ "ok!") and *ne* (*nicht*) (question tag), and two of *ach ja!* ("ah yes!", Example 8). *also* is also used by S26, *ach* by S7 (Example 9), *ach ja!* and *ne* (*nicht*) by S8 (Example 10). S8 also at some points says *Hä?*, which is used to express that one has not

understood. S24, on the other hand, twice in the process of considering what happened next and/or deciding how to phrase this, says *Ja!* (yes), which seems to indicate that she has reached a conclusion and is ready to continue.

While some of these expressions are similar in English - *also* being a "false friend" because of its different meaning in English, which may nevertheless very occasionally occur in the same context as German *also*, and *yes* often being actualised as *yeah* in speaking -, the phonetics in these instances are quite unequivocally German.

7) \*XYZ: +< and er@fp then < this lovely couple > [//] also@s husband and wife came out.

8) \*XYZ: +< another plant ?

\*XYZ: ach@s ja@s, grapes +//.

9) \*XYZ: +" ach@s, he saw it that way .

10) \*XYZ: and < he > [/] well, # he landed [...] on the lap of that coloured woman .

\*XYZ: not very skinny # coloured woman, ne@s [: nicht@s] ?

To summarise, our results in this section are in line with other studies that also find interjections to be a very frequent type of code-switch, as we pointed out above. The role of language proficiency in the frequency of occurrence is ambiguous, as we saw from the comparison of the German and English data. The almost complete absence of code-switches to fill lexical gaps in the English data, on the other hand, is in contrast with the German film retellings where they had a greater share and affected more individuals. We take this as an indication that a) the bilinguals by and large have acquired the relevant lexical items in English, and that b) they do not experience lexical access problems to the same degree as they had in German.

### **5.3.2.2 Errors and idiomatic language use**

As with the German film retellings, all errors were tagged according to the scheme reproduced in the Appendix and checked with linguistically aware/trained native speakers, before being combined into the figures given in Table 57 below.

Three things stand out in comparison with the German data. First, the overall error rate is lower, by about one error per person in the bilingual group and two in the control group. We checked whether this was a result of a more lenient approach by the native-speaker raters; however, adding in borderline cases which are excluded from the total error count does not change the picture substantially.

Second, the proportions of error types are different. While the order of frequency is still the same in the bilingual group, morpho-syntactic errors are relatively less important. These two findings are probably related to English being a more weakly inflected language than German.

Third, the control group has again a larger proportion and number of anaphoric reference errors (albeit at a magnitude of less than one error per person) compared to the bilinguals, who seem to be slightly better able to take into account the listener's perspective. However, unlike the other categories (cf. below), the group difference for discourse errors is not significant ( $U = 439.500, p = .291$  one-tailed).

There are highly significant differences between the groups on all other error categories (total errors:  $U = 187.000, p = .000$  one-tailed; lexico-semantic errors:  $U = 153.000, p = .000$  one-tailed; morpho-syntactic errors:  $U = 320.000, p = .005$  one-tailed). Within lexico-semantic errors, all error types - lexical and semantic errors ( $U = 152.000, p = .000$  one-tailed) and, in contrast with the German data, function words - contribute to the difference ( $U = 361.000, p = .013$  one-tailed). Within morpho-syntactic errors, only form errors yield a significant difference ( $U = 364.000, p = .017$  one-tailed), but not syntactic or word order errors.

These results are a little surprising. Although we are comparing a group of learners with native speakers, we would have expected the differences to be less pronounced in light of the relatively high self-ratings of L2 proficiency presented at the beginning of this chapter section, and indeed of some independent ratings as well, as we will see below. This high level of proficiency probably explains the absence of significant effects in syntax and word order, since syntactic/word order transfer is usually associated with lower levels. Indeed, the few errors in this area that do occur are almost exclusively produced by those bilinguals, who overall appear to be a bit less proficient, though not exclusively so.

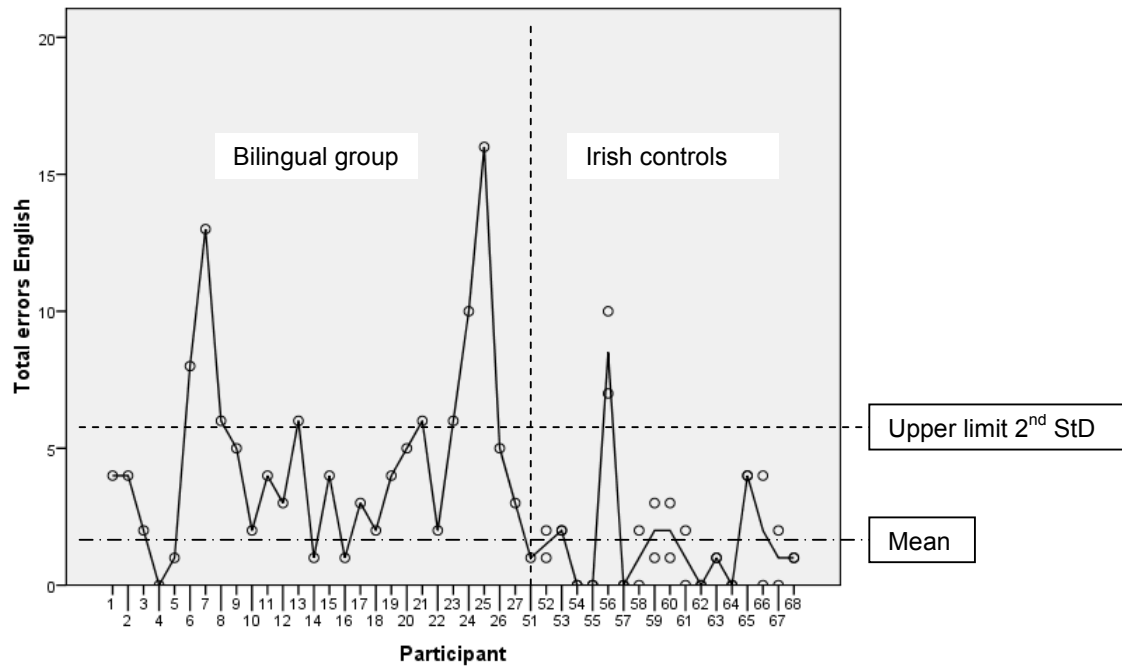
	<b>Bilingual group (N = 27)</b>	<b>Irish controls (N = 35)</b>
Total number errors	126	56
Mean total errors	4.67***	1.60***
Median	4.00	1.00
StD	3.65	2.19
Range	0-16	0-10
Mean lexico-semantic errors	3.15***	0.83***
Median	3.00	0.00
StD	2.66	1.60
Range	0-13	0-7
Percent of total errors	67.46%	51.79%
Mean morpho-syntactic errors	1.19**	0.29**
Median	0.00	0.00
StD	1.69	0.62
Range	0-7	0-2
Percent of total errors	25.40%	17.86%
Mean discourse errors	0.33	0.49
Median	0.00	0.00
StD	0.68	0.85
Range	0-3	0-3
Percent of total errors	7.14%	30.36%

**Table 57: Errors film retelling English**

\*\* Group difference is significant at  $p = .01$ .

\*\*\* Group difference is significant at  $p = .001$ .

Figure 20, Figure 21 and Table 58 represent the error scores by individual for total errors, and errors subdivided by types, plus the percentage distribution of individual error scores. We decided to give total error scores rather than percentages relative to tokens as for the German retellings, since the number of errors compared to the number of tokens is very small, so that individual differences would not really become apparent. We feel justified in doing so, since there is no significant correlation between the number of tokens and the number of errors for the bilinguals, while for the controls the correlation only reaches significance on the one-tailed test (bilingual group:  $r = .178$ ,  $p = .187$ , control group:  $r = .306$ ,  $p = .037$  one-tailed).



**Figure 20: Total errors English film retelling**

Number of total errors	Bilingual group (N = 27)		Irish controls (N = 35)	
0-1	4 (14.8%)	S4, S5, S14, S16	22 (62.9%)	
2-3	7 (25.9%)		8 (22.9%)	
4-5	8 (29.6%)		3 (8.6%)	S65, S65/2, S66
6-7	4 (14.8%)	S8, S13, S21, S23	1 (2.9%)	S56/2
8-9	1 (3.7%)	S6	—	
10-11	1 (3.7%)	S24	1 (2.9%)	S56
12-13	1 (3.7%)	S7	—	
14-15	—		—	
16-17	1 (3.7%)	S25	—	

**Table 58: Total errors English film retelling by individual**

To start our discussion of individuals' performance, we briefly outline the extreme cases and how sample length may have impacted on the error count. The vast majority of the controls have no more than one error, regardless of sample length, which ranges from 161 to 1069. Four bilinguals (sample range 450-823) match this level. A smaller group of controls and roughly as many bilinguals have up to three errors, with sample lengths of 308-981 and 307-1172 respectively.

At the other end of the scale, one control and one bilingual each have errors in a range far beyond the nearest comparison participant. Both have a high token count (S56: 911, S25: 1199), suggesting that their error count might have been lower if they had spoken less; however, the nearest control is actually S56's other production with a higher token count of 109. S25 can be compared to S17 and S11 who at similar or indeed higher token counts (1172, 1616 respectively) have far lower error rates (3 and 4). Thus, even after proportional adjustment (perhaps to the group mean), S56 and S25 would still have a relatively high count.

S65 and S66, who occupy the next range, have average and very low token counts respectively (S65: 604/896; S66: 264), as does the nearest bilingual (S7: 523). If the relationship between number of tokens and errors holds, then these participants have a higher error rate. However, as this quick survey shows the relationship between both measures is neither straightforward nor particularly strong. Furthermore, on the basis of the control group data we can surmise that the number of errors a participant is likely to make is relatively stable across different productions, almost irrespective of its length.

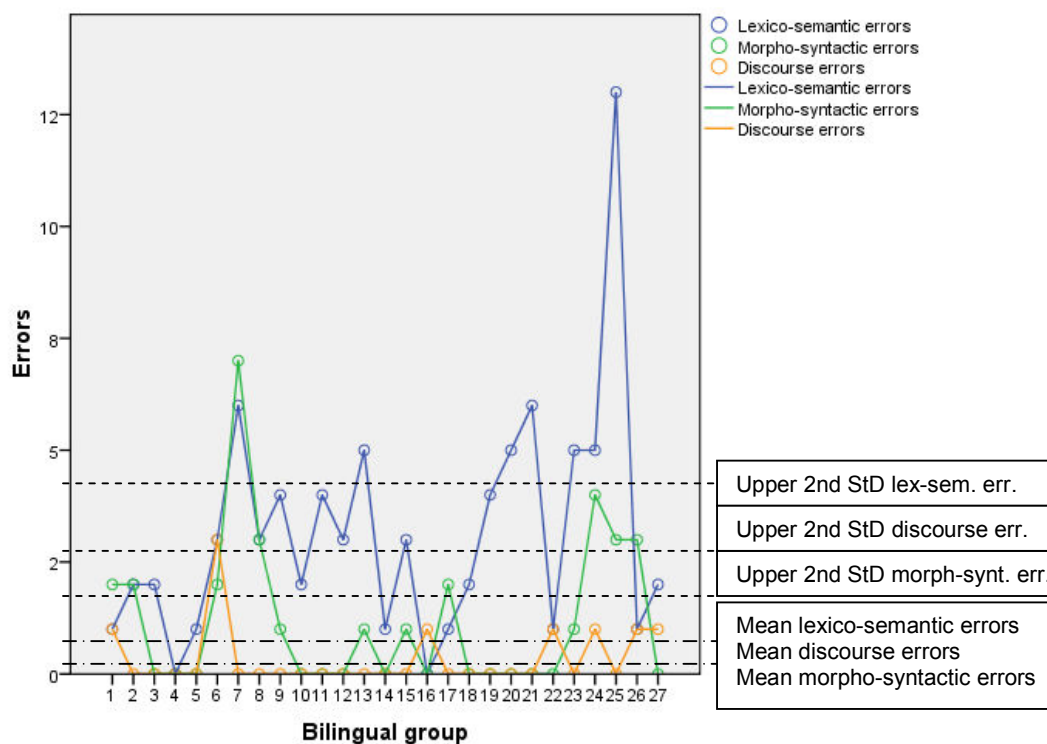


Figure 21: Errors bilingual group film retelling English by type

In view of the distribution of error types in Figure 21, we may, additionally detect different profiles: S25 is the person with the highest overall error count, which is almost entirely made up by lexico-semantic errors. S7, on the other hand, has an almost equally high error count for both lexico-semantic and morpho-syntactic errors, followed by S24, who has the same profile at a slightly lower level. In other words, S7 and S24 are slightly less proficient than other bilinguals in the morpho-syntactic domain, as well as the lexico-semantic one, and commit a number of the errors we will refer to as "learner errors", as do S25 and S26. In contrast, S21 and S13, who also have a high overall error count, have (almost) no errors in the morpho-syntactic domain.

Thus, starting our discussion with the morpho-syntactic field, S24 has three syntactic errors, one being a case of syntactic transfer concerning object placement (Example 11), one representing a strange idiosyncratic way (the construction occurs twice) of denoting location (Example 12), and the third concerning word order in indirect questions (Example 13). In Example 11, the participant uses the order direct object-indirect object, which in English requires the placement of a preposition in front of the indirect object, while in German this is not necessary because of case marking on both objects. A similar case is an utterance produced by S26 who places a temporal phrase before the direct object (*cooks him in the meantime a steak*), as would be required in German. In Example 12, the intended phrase was simply *There was a little kiosk*, which is actually the preferred construction in both languages.

11) \*XYZ:      and er@fp < she was telling that the shop owner > [\*] .

12) \*XYZ:      and er@fp < the > [/ -] < it > [\*] was < a > [/] a little kiosk there .

13) \*XYZ:      and and I don't know < what did the policeman > [\*] .

In Example 13, the participant inverted subject and verb after the interrogative pronoun, possibly in line with the requirements for English and German direct questions. However, this construction would not be correct in German either as in a sub-clause the verb goes to the end. For a direct question in English, the clause is in any case missing the main verb (or rather, the participant uses

the main verb in place of an auxiliary). Different, but related is S8's production of a direct question without an auxiliary (Example 14), in line with German syntactic rules.

- 14) \*XYZ: and they come back to # reality .  
\*XYZ: and < who they see > [\*] passing by ?

An interesting case comes from S6 who uses an Irishism for *let* (*left* instead of *let*) in the phrase *he was let go* but follows it with the gerund (*he was left going*) rather than the infinitive. This phrase was used twice, so it seems to be an established, or "fossilised", form in this participant's vocabulary, of which there are several more examples in the data (cf. below). S6 also uses German negation structure, deploying *no* as a (quasi) negative article, in connection with an apparent lexical retrieval problem (*silent movie*, Example 15). Interestingly, the opposite, i.e. using the English structure with negated verb rather than negated object, had occurred in the German data set.

- 15) \*XYZ: well , < you can (no)t > [/] you can (no)t hear him talking .  
\*XYZ: +< < it (i)s > [/] < it (i)s erm@fp no sound film > [\*] .

In contrast with such "learner errors", the very few occurrences of syntactic problems in the control group are of a different calibre. S56, S60 and S67 use a fairly complex construction, in which the subject is parsed incorrectly or constraints on maintaining subject identity are violated. In the first case (Example 16, S56), the subject of the main clause (*he*) is understood to be the same as that of the (subject-less) gerundial construction. This leads to an implausible, even funny interpretation (Charlie being a good-looking girl). Presumably, two different subjects were intended in the gerundial construction (her being good-looking, and him having a bad reputation). Example 17 (S67) is analogous – *she* is being set upon rather than *the policeman* (so *her* needs to be moved out of object position into subject position).

- 16) \*XYZ: and obviously, being < &good > [//] a good\_looking girl and already < having > [\*] a bad reputation, he decides he (wi)ll take the rap .



- 17) \*XYZ: and then a policeman comes running after her .  
 \*XYZ: < being er@fp set upon her by > [\*] er@fp the baker and another passerby .

- 18) \*XYZ: and he picks up the first one .  
 \*XYZ: and it looks nothing like what he < was > [\*] < first > [=! laughing] described as .

In Example 18, *he* (Charlie) appears as the grammatical subject in a passive construction, which should in fact be the wedge (*one*); removing the *as* and (preferably) changing the tense would solve this problem (*and it looks nothing like what he had first been described*; or *and it looks nothing like what had been described to him first*). It is interesting to note that we had come across similar problems with complex phrases in the case of the German controls.

The sub-category of form errors provides a few more examples of "typical" learner errors from the bilingual group, often on verbs, such as *what they were sawing* instead of *seeing* and *he then stolen* instead of *stole* (S24), *he throwed it away* instead of *threw* (S7); non-agreement of verb and subject (singular/plural) (S2, S7); choice of tense/mode (*she stands up* instead of the Irishism *she is standing up* "she is standing", S17); using incorrect conjunctions and pronouns (*what causes her to cry* for *which* (S25), and *the girl which he met* for *whom* (S13)); problems with conditionals (*ach, wouldn't it be nice if we would have a house like that* (S2)); and confusion arising in connection with supplying genitive endings for an irregular noun (Example 19, S25).

- 19) \*XYZ: erm@fp there (i)s erm@fp a scene .  
 \*XYZ: he sits on a < black's # women > [/] black < women's > [\*] lap .

However, there are a few instances of controls not agreeing subject and verb either, which might mean that this kind of error may alternatively have to do with a lack of control or monitoring, rather than with lack of knowledge, as we had already suggested in relation to the German data. The fact that many of these above examples are drawn from possibly somewhat less proficient bilinguals does not necessarily mean that these do not know the structures in question, but that their resources for monitoring are taken up by other processes. Most certainly a case of insufficient control is the recurrence of gender mismatches (using a personal pronoun with the wrong gender

for the character in question) in both groups, which had already been attested in the German narratives for both bilinguals and controls.

There seem to be, nevertheless, a few cases of fossilised forms in what otherwise are very proficient productions. Thus, S19 produces the false friend *tablet* for *tray* (German *Tablett* "tray") twice, and provides three different occurrences/forms for *passerby/to pass by*, presumably in analogy with *bystander: a bypasser, the lady-bypasser, children who were bypassing*. S13 provides *they got unconscious* (German *ohnmächtig werden* "to become unconscious") instead of *they were knocked unconscious*. These belong to the area of lexico-semantic errors, which, as we pointed out earlier, are the most frequent error source in both groups.

In both groups, there are cases of incorrect lexical choices. Some of these seem to be simple lapses of attention/slips of the tongue, for example, S51 stating that *the shop is looking at Charlie through the window*; while others apparently are genuine malapropisms. Thus, two controls (S52, S65) and one bilingual (S18) have Charlie *reminisce* when referring to his making up the fantasy about the girl and himself living together (**Gem 36**, dream). Further, one control (S56) refers to the police van (**Gem 43**, Minna) as *carriage*, and one bilingual (S25) as *cart*.

S56, the control with the highest number of errors of any of the controls (10 and 7 on Charlie 1 and 2 respectively), primarily because of her idiosyncratic lexical choices (7 and 6 lexico-semantic errors respectively), has the cow *spurt off milk*, and later, *spit milk into a jug* (**Gem 39**, cow) and has the boat, which is *not yet fully developed, collapse into the sea*. According to her (very lively and enjoyable) retelling, the policeman is *stuck* to Charlie at the end of the scene, and the newsagent did not *get on the money*.

However, this is rather unusual for the controls. In contrast, lexical and semantic problems occur with almost all bilinguals, where we additionally find, for example, Charlie *clapping* rather than *slapping the girl's face* (S2), Charlie, the girl and the policeman *lying in unison* on the street, rather than *in a heap* (S21), the girl *begging* rather than *pleading with the policeman* (S7), the girl giving Charlie a *hint*, rather than *warning* him (S27).

The groups do concur with regard to some of the gems already mentioned (cf. Table 59), which caused lexical or semantic problems for most participants in both groups. Before presenting these in more detail, we briefly want to outline four additional problem areas, almost exclusively

concerning the bilinguals: use of functions words; use of prepositional and phrasal verbs; semantic transfers, including false friends, which we already touched upon above; and idioms.

<b>Gem tag</b>	<b>Description</b>
11 determined	"determined to go back to jail"
32 baton	naming the policeman's baton/truncheon
36 dream	the fantasy of Charlie and the girl living together
39 cow	the cow producing milk without being milked
43 Minna	the police van
44 scene	change of scene, getting into dream sequence, return to reality

**Table 59: Gems resulting in lexico-semantic errors**

Starting with the first, function word errors concern almost exclusively the use of prepositions, apart from a few cases of article or pronoun errors. Sometimes, incorrect prepositions are used, such as in *a telephone at the wall* (S19), *sitting on the table* (S24) or *on the front door* (S12), *something similar like this one* (S24), *having a go for it* (S25), and *being scared for being arrested* (S23). At other times, prepositions are not supplied where needed, for example, in *knocking the glass of the window* (S6) or *wiping his hands off the curtains* (S25).

Related to these problems is the area of prepositional and phrasal verbs, where again some participants use incorrect prepositions for the intended meaning, while others use the wrong verb or unnecessary prepositions. Thus, we find Charlie *knocking* or *pulling the wedge off* instead of *out* (S14, S24), *the wife showing off the husband* instead of *seeing him off* (S8), *the girl picking up a flower/an orange* rather than *picking it* (S18, S25), the policeman *putting* Charlie away instead of *taking him away* (S24) – naturally, he will eventually put him away, but that is not what the participant intended to say –, *taking Charlie* rather than *taking him outside* (S20), and Charlie *showing up a wedge* rather than simply *showing it* (S19).

Turning now to semantic transfers, S18 has the policeman *wear* his baton, as opposed to *carry* (both *tragen* in German), and in S1's narrative Charlie and the girl *make* rather than *have a rest* (German: *Pause machen*). S25, who, partially as a result of his very long narrative, has the highest number of errors within the bilingual group (16), the majority of which (13) are in the lexico-

semantic area, also has several instances of transfers. For example, he refers to the baton (**Gem 32**) as *stock* (German *Stock* "stick") and provides *freshly married* (German *frisch verheiratet*) instead of *newly wed*. This interference had occurred in the opposite direction in the German data. He also initially has the girl *deck the table* (German *den Tisch decken*), which he noticed and corrected to *set*.

S9 produces a construction, which presumably has the same source as the German controls' (mis-) understanding of **Gem 11** (determined) discussed above: the fact that one German verb (*bestimmt sein*) may be rendered in English with two different lexemes (*to be determined, to be destined*), the first one implying volition, the second, outside forces (Example 20).

20) \*XYZ: and erm@fp the owner of the ship yard takes him on to the job .

\*XYZ: but from the very first moment you see that it (i)s < determined > [\*] to go wrong .

One interesting case of semantic transfer occurs in relation to **Gem 44** (scene, cf. Table 59), when S10 states that *we blend into a scene* ("we switch to a scene"; German *eine Szene wird eingeblendet*). By and large, controls and bilinguals come up with similar solutions for this gem, mostly referring to the sequence of scenes, deploying, as we had already seen in relation to the German controls (but not the bilinguals), terminology from film-making: *next/final scene/thing, the scene changes/switches over/back to, the shot is framed/moves over to, it cuts to/back to*, etc. However, a few controls use such terms in somewhat strange ways (Example 21, S59).

21) \*XYZ: and < they &b > [//] I think they both kind of, you know, < fade off into this > [\*] imaginary world .

*Fading in and out of scenes* are of course bona fide filmmaking terms, but in this example, different processes and meanings appear to be bundled together: Presumably, the scene fades into the new one, rather than Charlie and the girl, who, however, would be the correct subjects for verbs implying movement away/off. Similarly, for S60, *it splits into a fantasy scene*, while for S52 *Charlie has a scene*, which is plausible if Charlie is seen as the director but not as the protagonist. In S59, they (Charlie and the girl) *zoom back out*, which is not really what happens, and in S65, we

*flash back to the two of them in reality*, which is also not quite correct, since the dream was about a potential, not an actual future.

A final case of semantic transfer from German to English in the bilingual group was the use of an idiom in the context of Charlie eating as much as he can in the cafeteria (**Example 22**, S20).

- 22) \*XYZ:     so Mister Chaplin decides to go into er@fp one of the er@fp self\_service cafeterias.  
          \*XYZ:     and er@fp < has n(o)t got enough hands > [\*] to take everything that he would like to take .

The phrase *not to have enough hand* (German *nicht genug Hände haben*, "need help to", "not managing") is probably clear enough, from the phrase *there are not enough hands for ...*, but constitutes a coinage or a novel use in this context.

The other idioms we noted all occurred in the context of **Gem 36** (dream). Three bilinguals (S12, S13 and S15) declare that *the bubble burst* when Charlie and the girl return to reality after the dream sequence. This is a plausible use since the idiom refers to dreams not having worked out; however, its use requires that one has attempted to put the dream into practice, which to do our protagonists have not yet had a chance. Similarly, S18 initially states that Charlie and the girl are *back to square one* but subsequently corrects the phrase to *return to reality*.

This subsection has provided evidence for differences between the two groups, some of which are not just one of degree. Some of the errors discussed, both form and lexico-semantic errors, do appear to stem from a knowledge deficit on the part of some bilinguals and would be indicative of somewhat lower levels of L2 proficiency ("learner errors"). Moreover, some errors persist even in highly proficient bilinguals, such as fossilised forms and errors of lexical choice. Some, though not all, of these errors can be traced back to an influence from L1.

Such errors do not occur in the control group at all. However, some controls do use incorrect words and grammatical structures, albeit usually more sophisticated ones, again confirming that there are differences in native speakers' linguistic knowledge also. Furthermore, certain errors in both groups can be seen to result from the complexities of spoken production (speech errors), and thus cannot be directly related to the level of L1 or L2 proficiency.

Despite these differences at the group level, the overall number of errors in the bilingual group is still relatively small; in fact, the mean of total errors is actually lower than the German mean, as is the control group's. Assuming that this is due to typological differences, as discussed above, we may adjust the bilingual group's mean on the basis of the control groups', with the result that it is actually only one error higher than the German mean errors, and this difference is due to the lexico-semantic errors, not to discourse or morpho-syntactic errors. Moreover, a sizeable proportion of individuals have error rates well within the native-speaker range, and only two fall outside the native-speaker baseline. Several more appear to be a bit weaker, but their level is matched by one control.

### 5.3.2.3 *Gem analysis*

In conclusion of our discussion of the English film retellings, we again present an analysis of selected gems, some of which have already been tabled in the subsection on errors. Such an analysis affords a complementary perspective on the bilinguals' level of L2 acquisition which is not based on deficit. As we had done in the analogous German section, we focus here on issues of lexical diversity, retrieval and idiomatic or target-like language use, having introduced an additional gem, **Gem 47** (idiom) (cf. Appendix), to tag such uses.

<b>Gem tag</b>	<b>Description</b>
5 ship	object being built
13 girl	description of girl
14 steal	stealing the bread
16 collide	Charlie and girl colliding
17 witness	description of the lady witnessing the theft
33 hit	knocking the policeman out again
35 grass	naming the grass verge
41 loitering	the reason for the second policeman to object
42 arrest	Charlie/the girl being arrested
46 retrieval	explicit and implicit lexical retrieval difficulties
47 idiom	idiomatic uses of English, "Irishisms"

**Table 60: Gems analysed for lexical diversity, idiomatic use and retrieval difficulties (English)**

The three broad trends we had identified in relation to the German gems were distributional differences in the frequency of use of certain words, a certain number of retrieval difficulties, and lower lexical diversity in the bilingual group, which used more prototypical or concrete words, except for the use of expressive language. There is limited evidence for these differences in the English data; the overriding impression is instead that of target-like language use in terms of the frequency of use and the range of expressions, and the very low incidence of retrieval/lexical knowledge problems, some of which also occur in the control group.

To start with, with regard to **Gem 13** (girl) we found that in German the controls preferred the expression *young woman* while the bilinguals were divided between that and *young girl*. By analysing the same gem in English, it seems quite clear that this use was in fact a result of the adoption and transfer of L2 norms by some of the bilinguals: in English, the preferred use by far by both controls and by bilinguals is *the young girl* (eight bilinguals, twelve controls; compared to three/two mentions of *young woman*, and three/two mentions of *young lady*).

In relation to **Gem 5** (ship), we had seen that again the German controls had a preference for *ship* over *boat*, unlike the bilinguals. In English, both groups use *ship* more often (eight bilinguals, 14 controls, compared to three/five mentions of *boat*). However, the impression that the distinction between both terms is less strict in English than in German is confirmed by the fact that two controls use both terms alternately.

**Gem 32** (baton) was predominantly rendered as either *baton* or *truncheon*. The proportion of *truncheons* was higher in the control group; however, this seems to be a regional difference since it was almost exclusively used by the Northerners amongst the controls, while that region is not represented by the bilinguals. Three bilinguals and one control used a more general word, *stick*, which seemed to result either from lexical gaps or retrieval difficulties. Just one of the bilinguals (S21) used *stick* without trying to find a more specific term, while the remainder all paused extensively, finally settling for *stick*.

In the case of the control, S66, this solution was accompanied by a giggle, which seemed to indicate that she was well aware that this was not the optimal word. Bilingual participant S26, after an unsuccessful search process (Example 23) and apparently despite his better knowledge, asserts that "we call it a stick" ("we" presumably being "we in Ireland", as in a previous utterance).

Alternatively, he may have used the phrase in the sense of "We'll call it a stick" or "Let's call it a stick", to indicate that the negotiation of labels was complete. Bilingual S18, on the other hand, is experiencing a tip-of-the-tongue moment since she does recall a fragment of the word (Example 24). Thus, bilinguals and controls "agree" both on the default and non-default solutions for this item.

23) \*XYZ: and he gives him one with the erm@fp < stick > [//] or +//.

\*XYZ: whatever .

\*XYZ: we call it a stick .

24) \*XYZ: by erm@fp hitting him # on the head with his stick # he was < wearing > [\*] .

\*XYZ: there (i)s a special name for the stick .

\*XYZ: but # I can (no)t remember that .

\*XYZ: erm@fp it (i)s starting with < b@l > [/] b@l I think .

A gem where we do find a small distributional difference is **Gem 35** (grass), in combination with **Gem 41** (loitering). Both controls and bilinguals frequently name the place where Charlie and the girl sit down as *on/in the grass* or *on the grass verge*, but the bilinguals prefer the former, and the controls the latter (bilinguals: five and two mentions, controls: three and six mentions, plus one *grassy knoll*). In other words, the controls are more specific about the place. *Verge* was an unsolved retrieval problem for S25.

Additionally, the bilinguals mainly use concrete references, such as *under a tree/on the side of the road* (six mentions each) and, most popularly, *outside a house/bungalow* (twelve mentions). All of these were also used by the controls (one, two and seven mentions respectively), who, however, more often than the bilinguals abstractly refer to the area where the house is located rather than specifying Charlie's physical location in relation to the house: *suburbia, nice/rich/well-to-do suburban/residential area/neighbourhood* (bilinguals: five mentions; controls: twelve mentions).

In the German narratives, **Gem 41** (loitering) had caused some retrieval problems and a code-switch, and led to less diversified explanations in the bilingual group. In the English narratives, the



controls also mention the fact that Charlie and the girl are *layabouts*, *dropouts*, *vagrants* or *tramps*, slightly more often than the bilinguals (one vs. four mentions). This difference may not mean too much since the sample size is also bigger.

On the other hand, controls and bilinguals alike produce the target *loitering* without too much hesitation, except for S25, who nevertheless manages to come up with it. One control (S56), on the other hand, uses *trespassing*, but seems to query it in her mind, since she returns to it later saying *trespassing with intent*. It is not clear whether this is ultimately the term she wants since the phrase is plausible if unusual in an Irish context where it normally occurs as *loitering with intent*.

**Gem 14** (steal) is an example where we find a distributional difference reflecting greater lexical diversity/colourful language use on the part of the bilingual group. Although both groups overwhelmingly use the same "default" lexemes (*steal*: nine bilinguals, sixteen controls; followed by *take*: three bilinguals, two controls), as on most of the other gems analysed, the bilingual group uses a wider range of colloquial lexemes: *rob* (two bilinguals, one control), *grab* (two bilinguals), *pinch*, *snatch* and *nick* (one bilingual each).

Sometimes the choice of solutions is complicated by the fact that the scene or facts to be described are rather unlikely, as we had already seen in the case of **Gem 39** (cow) in the German retellings. In English, we find the same wide range of perspectives on the cow scene, with minor differences in distribution between the two groups. The subject of the relevant utterance is either Charlie who *milks* or *does not milk the cow* (the controls point this latter fact out more often) or *pulls on the udder*; or the milk *which runs/flows/goes/shoots into the jug* or *comes out* or *appears*; or the cow which variably *gives/drains/lets off/squirts milk* or *fills the jug with milk*. We had already mentioned some of the verbs used (*spitting* and *spurting off*) in the error analysis.

Some bilinguals and controls avoid the problem by simply stating that you *get milk from the cow*, *have milk* or *the cow supplies/provides milk*; and both groups are united in using the image of *milk on tap* (more or less successfully), or *the automatic/self-milking cow*. Notably absent are retrieval problems or errors in using the lexeme *milking*, which in German had constituted a problem for some bilinguals, except for S53 (a control) who said he did not know the technical term for "pulling (on/at) the udder(s)". *Udder* did represent a phonetic problem for S18, who was not sure of how the word was pronounced.

There are several more gems, for which there is a great range of possible solutions, particularly if one desires to be expressive, which is a common intention of many of the controls and bilinguals: Gem 16 (collide), Gem 33 (hit), Gem 42 (arrest), Gem 17 (witness) and Gem 43 (Minna). Sometimes this desire gives rise to more or less acceptable solutions.

For example, for **Gem 16** (collide), the most popular phrase in both groups is *bump into* (four bilinguals, eight controls), followed by *run (smack) into* (two bilinguals, four controls). Both groups also use *crash into* and *collide with*, while the controls specify that *the girl knocks Charlie over/down* more often (one bilingual, five controls). Two bilinguals use *knock into*, and one more *bang into*, which is still ok, but the line of acceptability is probably crossed with S7's *they bang together*, which seems to be another case of transfer (German *zusammenknallen*).

Similarly, for **Gem 33** (hit), the groups agree on the most frequent ways to phrase this scene, by using *hitting* or *knocking the policeman on the head* (five bilinguals, three controls; four bilinguals, five controls). Additionally, there are *hits him a bump* (controls), *whacks him over the head* (both groups), *clicks/clocks/clonks him over the head* (both groups), *bashes him* (controls), all of which involve a "hitting" verb. Though some may be more unusual than others, these were all accepted by the native speakers consulted.

An alternative phrasing uses *give* + a "hitting" noun, which is used more often by the bilinguals, though not exclusively so: *gives him a little tap* (two bilinguals, one control), *gives him a bonk* (one control), *gives him a little bang* (three bilinguals)/*a flake/a little knock/a good bash* (one bilingual each), and finally, and more borderline, *gives him one/another smack* (which recurs as *smash* in relation to **Gem 8** (exwedge)). The construction with *give* + verb might be a case of transfer, though presumably not in the case of the controls, so it would appear that while both forms are possible in English, the active verb one is preferred in English.

In relation to **Gem 42** (arrest) most participants in either group use *to arrest* and *to take away* for the actual arrest (as passive or active; 13 bilinguals, 20 controls), and *put into the van* for the transfer to prison (ten bilinguals, eight controls), and there is also great agreement on other alternative or more expressive ways to phrase these meanings: *is led/escorted/taken/thrown/pushed/loaded/bundled into the van*, *grab/nab him*, *take him into custody* and *take/pull/drag/march/cart off/away*. However, the controls seem to have a slightly greater range of

(active) expressions, since they also use *chuck* or *pack into the van*, and *haul/whip Charlie off/away*, as well as *the slammer* for prison. One of the bilinguals presumably inadvertently uses *accost* instead of *apprehend*, and there are two examples in either group of Charlie simply being *taken*, which did not find too much favour with the native speakers consulted.

In a similar vein, the witness from **Gem 17** is initially referred to as *the woman* or *lady who saw what happened* (ten bilinguals, twelve controls). Two controls and one bilingual call her *a/the witness*, one each *a passer-by*, and interestingly, one bilingual refers to her as a *housewife*, and another as *a lady out shopping*. Several people qualify the basic term further, most commonly with *fine/rich/well-dressed/wealthy looking* (six bilinguals, three controls), but the controls also use *posh* and *hoity-toity*. One bilingual and one control call her *nasty*, and one control *interfering*; while another control and one bilingual take this one step further by referring to her as an *auld biddy* ("an interfering, reactionary, gossiping busybody of mature years" (White Lennon 2006: 13)) and *a well-dressed rich bitch* respectively.

The last example is reminiscent of the tendency we had seen during the German film retellings for the bilinguals to be inclined to use colourful language. Also in analogy to what we had seen there, for **Gem 43** (Minna) most participants in either group use police van or police car (13 bilinguals, 20 controls), but there is also frequent mention of the *Black Maria* (again called *Black Mary* by S25) (four bilinguals, eight controls) and the *paddywagon* (four each). This gem combines other trends already mentioned, with the controls having a few extra labels (*police wagon*, *prison transport*, *jailer van* and *truck*), and the bilinguals producing one slightly off *squad car* (like in the German narratives) and, surprisingly, a *station wagon*.

To conclude, the discussion of gems just presented serves to highlight the fact that many of the bilinguals have indeed acquired native-like ways for referring to the items discussed, in the lexical and grammatical choices overall, and the preference for some forms over others, though this does not apply across the board. So, in the gems analysed, we find considerable overlap, very few "aberrations" in terms of incorrect or inappropriate choices, some of which also occur in the control group, and little crosslinguistic influence.

We have already discussed the main instances of lexical retrieval problems, to which we add just one more problem with naming the *ramp* in the shipyard (S6, bilingual), and expressing insecurity

about the words *belching* (S26, bilingual), the *foreman* and the *shipyard* itself (S61, S66, controls). When taken together, these problems boil down to three controls (S53, S61 and S66) and five bilinguals (S6, S18, S24, S25 and S26) explicitly stating a retrieval problem or lexical gap. Two of these, S25 and S18, have several instances (three and four, respectively). In S25's case, he only remembers an approximation of the sounds for one of these (*verge*), but for the other two (*set the table*, *loitering*) he actually finds the solution himself. S18 claimed she did not know the correct name for *police van*, the correct pronunciation for *udder* and produced the starting sounds for *reminiscing* and *baton*. It would seem, however, that this is a psychological rather than a proficiency or knowledge question, which is partially confirmed by the external ratings we will discuss in the next subsection, and the participant's overall approach to the testing, which we had already briefly referred to in relation to the C-test.

Overall, then, lexical retrieval/knowledge is not an issue for the bilingual group overall, which ties in with the impression of considerable fluency gleaned from the hesitation measures. This stands in contrast with several cases of retrieval difficulties in the German retellings, which caused a certain degree of exasperation. Furthermore, several of the lexemes causing a problem in German were unproblematic for the bilinguals in English: *to milk*, the *wedge*, the *shipyard*, *to loiter*, the *foreman*, confirming the impression of fluency and a high level of acquisition with little L1 influence noticeable in the L2.

As a final piece in the puzzle of the bilinguals' L2 attainment level on the basis of the film retellings, we analysed features that would be considered idiomatic (Hiberno-) English. The following brief sketch is partially based on Dolan's Dictionary of Hiberno-English (Dolan 1999), and partially on features the researcher stumbled upon: morpho-syntactic peculiarities on the one hand, and collocational/discourse ones on the other. First, typical features of Hiberno-English grammar are using singular verbs with plural subjects (Example 25), using *after* + verb+ing instead of perfect (Example 26), and using clauses without a finite verb (Example 27).

25) \*XYZ: and there is other people in there that are all on the way to the police station. (S15)

\*XYZ: and there is all sorts of beautiful cakes and tarts and lovely things. (S9)

\*XYZ: and there is apples growing outside window. (S5; also S2, S17 ...)

- 26) \*XYZ: what he is after doing. (S6)
- 27) \*XYZ: and the cop still lying there. (S8)
- \*XYZ: and Charlie looking after it. (S11)
- \*XYZ: and behind them the policeman. (S26; also S7 ...)

Moreover, it is quite common to use *your* or *this* instead of the indefinite article in narratives (Example 28), as is the ubiquitous use (not just in our set, but generally) of adjectives such as *little/wee*, *beautiful/lovely* and *grand* (cf. Example 25 above; and also S1, S2, S6, S9, S10, S12, S18). This presumably provides the reason for using diminutives like *wifey/girlie* (S5, S14), although in Ireland the suffix *-een* is more commonly added for the purpose. We already pointed out the use of *leave* + verb instead of *let*. Not as common in our set, but still profoundly authentic is the use of *inside* as an adverb to be followed by a preposition (Example 29), and the use of *standing up* and *sitting down* for *standing* and *sitting*, already referred to.

- 28) \*XYZ: there was your man, Charlie. (S6, also S5)
- 29) \*XYZ: and she sees some bread inside in the van. (S10)

Apart from the above-mentioned *paddywagon*, there are other culture-specific references, such as the *guard* for policeman (S7, S18, S22), and *a total knacker* (S5), defined as "(pejor.) a person involved in shady deals, [...] (pejor.) a member of the travelling community" (Dolan 1999: 152).

Last, but not least we should mention some figures of speech that are very common in Irish discourse and are also made extensive use of by the bilinguals: *thanks be to God*, *be away with the fairies* (S1), *what d' you call it* (S5, S6 ...); *being the (gentleman) he is* (S1, S8, S10, S12); *to be delighted (with himself)* (S9, S17, S22); *she's all (dressed-up/smartly dressed)* (S5, S17, S21); *(chatting) away at her/(chopping/hacking/sawing) away at it* (S5, S25); and S25's "modern" use of *like* (an Americanism rather than an Irishism, and prevalent in youth speak).

All of these features are also abundantly attested in the control group's narratives, and thus provide another piece of support for the claim that many bilinguals have indeed developed native-like language skills. In the final subsection, we will check this claim against some external proficiency ratings of the film retelling task, as well as bring together the various English-language tasks and features investigated.

#### **5.4 Summary**

Across the three tasks analysed in this chapter, we find that the bilingual group performs comparably to the Irish control group on some measures, while on others there are significant differences between the groups. Thus, the bilingual group has significantly lower C-test scores, at comparable test times; significantly lower scores on three of ten FiCA measures (FiCA total, FiCA1 *Professions*, FiCA3 *Street*); a significantly higher percentage of silent pauses and repetitions on Charlie 1; and significantly more lexico-semantic, function-word and form errors.

However, the bilingual group performed very similarly to the Irish control group on the fluency and lexical diversity measures of both the verbal fluency task and the film retelling, which was confirmed by the qualitative analyses. The bilingual group's performance is marked by the relative absence of retrieval/production difficulties and transfer from L1 to L2. (Again, we cannot rule out that, due to small sample sizes on some of the individual conditions of FiCA2 and FiCA3, the actual differences may be larger than we have been able to show.) On the verbal fluency task, the control group's performance is more heterogeneous than the bilinguals'.

Moreover, there is a considerable number of individuals in the bilingual group, who fall within the native-speaker range on all or most measures, while the group differences in the main may be traced back to the same individuals, who seem to have achieved a somewhat lower level of L2 proficiency than the rest of the group, as we will discuss further in relation to the summary scores below. Nevertheless, the bilingual group, including those individuals, performs far better than the German control group on the C-test.

At the beginning of Chapter 4, we had formulated four hypotheses relating to the bilinguals' performance vis-à-vis the relevant control group, which we will review next:

a) The bilingual group will achieve lower scores in and/or took longer to complete the C-test.

This hypothesis is supported overall; however, a substantial number of bilinguals has scores within the native-speaker range (12/27, 44%), and a larger number completes the test within that range (16/27, 59%). The qualitative analysis of item solutions showed a considerable degree of overlap in item solutions at a somewhat lower level of accuracy. Problems occurred at all linguistic levels, but were mainly related to lexical gaps or retrieval problems and insufficient monitoring of the wider context. Crosslinguistic influence from L1 to L2 did not seem to play a part.

b) The bilingual group will achieve lower scores in the verbal fluency tasks.

This hypothesis is supported in relation to two of the FiCA conditions, and the overall scaled score, while the groups perform comparably on the other conditions. The significant differences can be connected to few very fluent controls, who outperform the rest of the control group. The remainder of the controls largely overlap with the bilinguals. The bilingual group displays a similar level of lexical diversity; the particular items named by the two groups reflect similar experiences and environments, although there is a non-significant trend for the controls to name more white-collar and artistic jobs. There is no evidence of lexical retrieval difficulties in the bilingual group.

c) The bilingual group will demonstrate lower lexical diversity in the verbal fluency tasks/the film retelling.

This hypothesis was not confirmed with regard to either task. Moreover, the gem analysis shows that, in the main, the bilingual group approximates native-like use in relation to preferential lexical choice and the range of expressions used.

d) The bilingual group will display a higher rate of errors, code-switches and hesitation phenomena in the film retelling.

This hypothesis was partially confirmed. The bilingual group does have a higher rate of lexico-semantic, function word and form errors, resulting in a significant difference in the total error rate. However, the gem analysis provided evidence for wide-spread idiomatic uses, including that of features of Hiberno-English, in the bilingual group. There are very few cases of code-switches

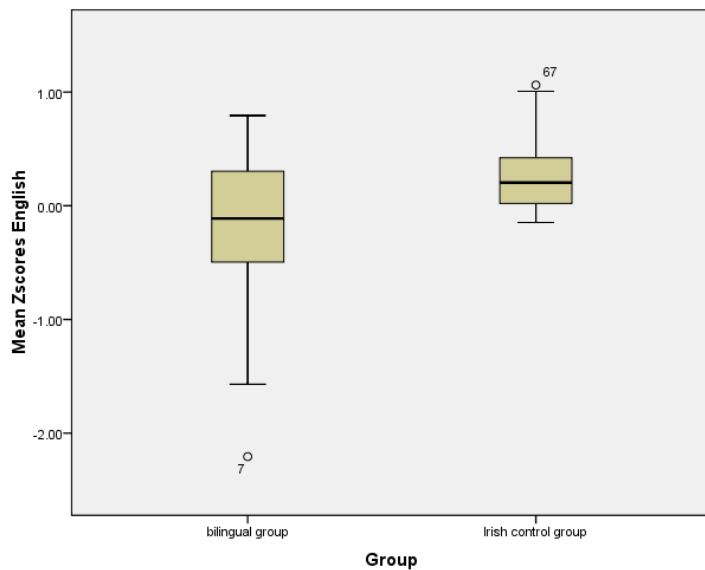
apart from some interjections, which, like the majority of the form errors, were produced by the same somewhat less proficient individuals. Only one part of Charlie yielded a significant difference in some hesitation phenomena, namely silent pauses and repetitions, but the overall impression is that of considerable fluency and a high level of attainment.

This last statement sums up the overall performance of the bilingual group well: while the group does display some differences with the control group on measures testing breadth of linguistic knowledge and accuracy (the C-test, the error measure of the film retelling), it performs similarly to the native speakers on measures focusing on fluency, lexical diversity and idiomatic language use (FiCA, the Charlie fluency index, the gems). Given the difference between the bilingual group and the German control group on the C-test, one of the knowledge/accuracy measures, which is larger than that between the bilingual and the Irish control group, evidence points to successful L2 acquisition on the part of the bilingual group.

Throughout the discussion we had noted that some very low and high scores could often be traced back to the same individuals. Therefore, we again computed summary scores to be able to assess each individual's overall L2 proficiency relative to the other participants. As in the German data set, the summary scores are mean  $z$ -scores based on the unweighted C test score and time taken, the FiCA score, raw VOCD-D, fluency and error indices for the film retelling.

Figure 22 graphically represents the results of the group comparison. As the box plot shows, the bilingual group's mean is lower than the control's (-0.192 and 0.287 respectively), its range is much larger, and thus there is only partial overlap between the groups. This difference is significant at  $t(41.184) = -3.154, p = .003$ . The controls generally occupy the positive side of 0 (15/18), which is a higher proportion than had been the case with the German  $z$ -scores. While its overall performance looks relatively homogeneous, there are two controls, who are distinguished from the rest of the group by their much higher scores: S67 and S64, with scores of 1.062 and 1.007 respectively, the next highest control's score being 0.671 from S54.





**Figure 22: Bilingual and Irish control groups' performance across all three English tasks**

However, 13 (ex 27) of the bilinguals also have positive  $z$ -scores, and more than half (15 ex 27) have their scores within the native-speaker range, thus indicating native-like attainment across the three tasks. The highest-ranking of these, S16 (0.793) actually comes in third place overall, S11 (0.499) in fifth, thus outperforming 15 native speakers, or 83% of the native control group. S9 (0.466), S5 (0.451) and S20 (0.444) occupy ranks seven to nine, still performing better than 14 individuals, or 78% of the control group, which would appear to be a very strong testimony to the level of achievement of these bilinguals. For comparison, the lowest-ranking controls, S63 and S53, have  $z$ -scores of  $-0.0147$  and  $-0.091$  respectively. Bilinguals S21, S14 and S17, in ranks 12, 14 and 16 (of 45), can also be seen as highly proficient L2 learners. Since late bilinguals are often characterized as unable to acquire a high level of L2 proficiency, this is surely a very positive result.

The twelve bilinguals, whose score falls below the native-speaker range, on the other hand, have attained a level of proficiency, which is somewhat or in some respects lower than that of native speakers, as gauged by these tests. The lowest scores by far are those of S7 ( $-2.206$ ) and S24 ( $-1.570$ ); S18 follows in third lowest place, but her score is very close to the next participant, S6, with all other scores rising gradually from thereon.

The  $z$ -scores of five of the individuals, S1, S7, S8, S25 and S26, are also below the German control group's range, while S2, S4, S6, S13, S18, S23 and S24's German scores were within that range. In contrast, S10 and S27, also with non-native German scores, have a  $z$ -score above 0 on the English tests.

In summary, then, there is a substantial number of individuals in the bilingual group, who do not correspond to the stereotype of the unsuccessful late L2 learner, and even those who evidently are less proficient have acquired some features approximating native-like use, and use the language fluently and largely accurately.

In the next section, we will present and discuss evidence to complement the overall proficiency scores just presented. First, we will relate them to self-ratings of L2 proficiency elicited from the participants, and second to independent ratings commissioned from two linguistically trained native speakers.

## **5.5 Discussion**

### **5.5.1 Self-reported L2 proficiency at time of testing (can-do scales)**

In addition to the proficiency questions on the background and sociolinguistic questionnaires, participants optionally responded to 43 can-do statements in relation to both English and German (21 bilinguals, 14 Irish controls).

In contrast with the German results, the bilingual group rates itself lower than the Irish control group on their English-language skills (cf. Table 61). That difference is significant for all four skills, but is much more pronounced for the literacy skills, which are significantly different at the .001 level, while the two spoken skills are different at the .05 level (listening:  $U = 80.500$ ,  $p = .025$  one-tailed, reading:  $U = 39.500$  one-tailed,  $p = .000$ , speaking:  $U = 78.000$ ,  $p = .016$  one-tailed, writing:  $U = 45.500$ ,  $p = .000$  one-tailed). The hierarchy of skills – Listening > Speaking > Reading > Writing – is the same as in the German data and follows the order of computational effort, which has been proposed for these skills. The control group's average scores vary minutely, again pointing to overlearned skills, but the two literacy skills are also rated a fraction lower.

<b>Can-do</b>	<b>Bilingual group (N = 21)</b>	<b>Irish controls (N = 14)</b>
<b>Listening</b>		
Mean	95.95*	99.38*
Median	97.50	100
StD	6.91	1.55
Range	75-100	95-100
<b>Reading</b>		
Mean	89.80***	99.80***
Median	91.43	100
StD	12.21	0.76
Range	49-100	97-100
<b>Speaking</b>		
Mean	94.59*	98.74*
Median	97.65	100
StD	7.53	1.93
Range	71-100	94-100
<b>Writing</b>		
Mean	88.73***	98.32***
Median	92.73	100
StD	10.98	3.36
Range	65-100	91-100

**Table 61: Results of can-do statements English (percentage scores)**

\* Significant at  $p < .05$  one-tailed.

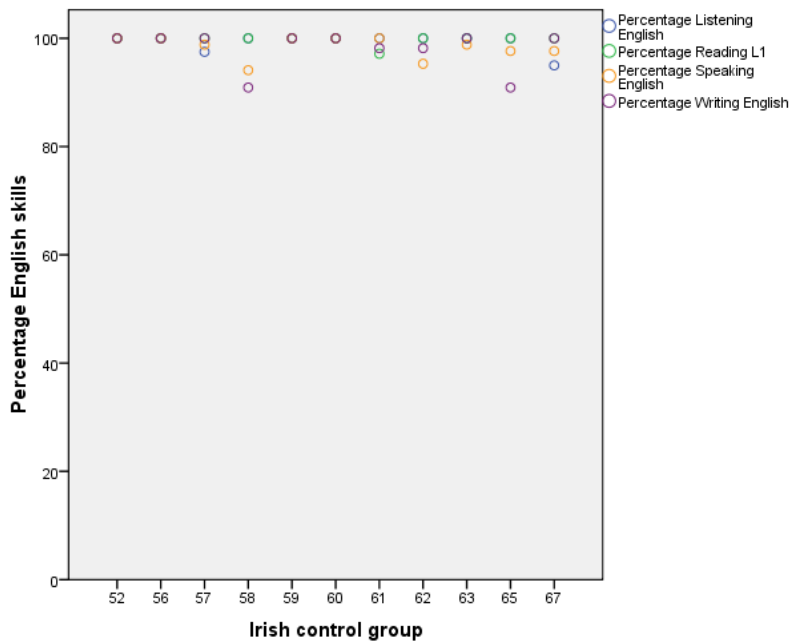
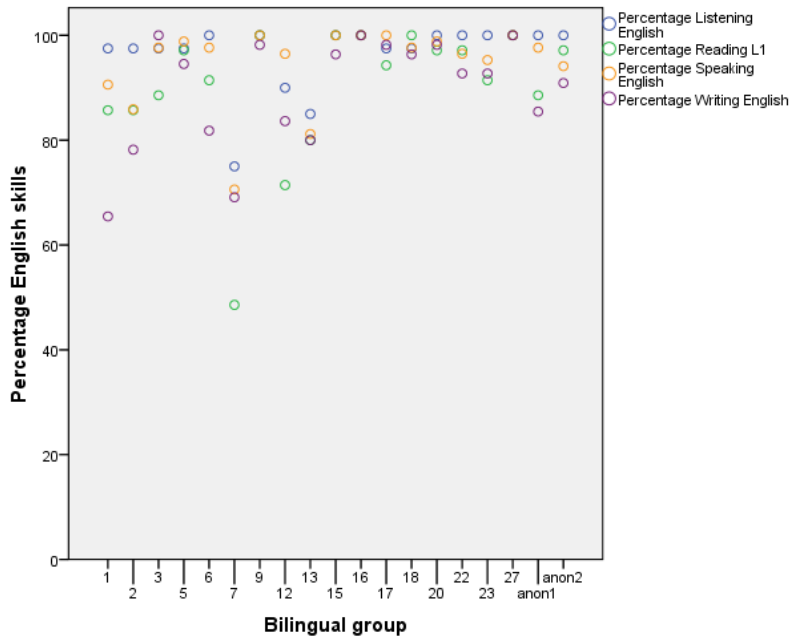
\*\*\* Significant at  $p < .001$  one-tailed.

A total of 21 out of the 43 individual statements returned a significant difference (Table 62), in contrast with German, where only one statement received lower ratings by the bilingual group compared to the control. Interestingly, while in relation to the degree of difference (i.e. the level of skill overall), we saw a dissociation between oral and literacy-related skills, it is the productive skills vis-à-vis the receptive ones that caused greater difficulty in terms of the number of statements, i.e. the sub-skills.

Despite the significant difference with the Irish control group, the bilinguals' overall ratings are much more like the self-ratings by the Irish control group than the German control group's, whose self-ratings range from 64.56 for listening to 57.65 for writing skills (cf. Figure 23).

<b>Language skill</b>	<b>Can-do statement</b>
<b>Listening</b>	<p>I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided that I have some time to get familiar with the accent.</p> <p>I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar.</p>
<b>Reading</b>	<p>I can understand long and complex factual and literary texts, appreciating distinctions of style.</p> <p>I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.</p> <p>I can understand specialised articles and longer technical instructions, even when they do not relate to my field.</p>
<b>Speaking</b>	<p>I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.</p> <p>I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms.</p> <p>I can present clear, detailed descriptions on a wide range of subjects related to my field of interest.</p> <p>I can narrate a story or relate the plot of a book or film and describe my reactions.</p> <p>I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.</p> <p>I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.</p> <p>I can express myself fluently and convey finer shades of meaning precisely.</p>
<b>Writing</b>	<p>I can select style appropriate to the reader in mind.</p> <p>I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view.</p> <p>I can write simple connected text on topics which are familiar or of personal interest.</p> <p>I can express myself in clear, well-structured text, expressing points of view at some length.</p> <p>I can write clear, detailed text on a wide range of subjects related to my interests.</p> <p>I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points.</p> <p>I can write letters highlighting the personal significance of events and experiences.</p> <p>I can write summaries and reviews of professional or literary works.</p> <p>I can write about complex subjects in a letter, an essay or a report, underlying what I consider to be the salient issues.</p>

**Table 62: Can-do statements English, resulting in significant group differences**



**Figure 23: Distribution of skills scores by participant**

Figure 23 shows the dispersion of scores for each skill within the bilingual group vis-à-vis the Irish control group, highlighting both the clustering of scores at or near the maximum within both groups, but also pronounced inter-individual differences within the bilingual group. S7 stands out as the participants who rated himself lowest with regard to reading and overall. S1 also gave herself a very low score for writing, followed by S12 with three low scores within reading.

We again checked the level of correspondence between self-ratings and objective proficiency measures by correlating them with the C-test scores, time taken, and the z-scores (cf. Table 63). All self-ratings correlate very highly with the proficiency measures, except for listening and reading skills, which play a lesser role in the time it took to solve the C-test (listening almost reaches significance, reading is significant at the .05 level). This reduced role for the receptive skills was also evident in the German data. The self-ratings also correlate very highly with one another, pointing to relatively balanced skills, as do the proficiency measures, as was to be expected.

		Listening	Reading	Speaking	Writing	C-test score	C-test time	z-score
Listening English	PCorr	1.000	.870**	.865**	.663**	.658**	-.429	.743**
	Sig. (2-tailed)		.000	.000	.003	.002	.067	.000
	N	19.000	19	18	18	19	19	19
Reading English	PCorr	.870**	1.000	.853**	.731**	.643**	-.526*	.717**
	Sig. (2-tailed)	.000		.000	.001	.003	.021	.001
	N	19	19.000	18	18	19	19	19
Speaking English	PCorr	.865**	.853**	1.000	.782**	.692**	-.661**	.789**
	Sig. (2-tailed)	.000	.000		.000	.001	.003	.000
	N	18	18	18.000	17	18	18	18
Writing English	PCorr	.663**	.731**	.782**	1.000	.642**	-.767**	.733**
	Sig. (2-tailed)	.003	.001	.000		.004	.000	.001
	N	18	18	17	18.000	18	18	18
C-test scores	PCorr	.658**	.643**	.692**	.642**	1.000	-.606**	.881**
	Sig. (2-tailed)	.002	.003	.001	.004		.001	.000
	N	19	19	18	18	27.000	27	27
C-test time	PCorr	-.429	-.526*	-.661**	-.767**	-.606**	1.000	-.738**
	Sig. (2-tailed)	.067	.021	.003	.000	.001		.000
	N	19	19	18	18	27	27.000	27
z-scores	PCorr	.743**	.717**	.789**	.733**	.881**	-.738**	1.000
	Sig. (2-tailed)	.000	.001	.000	.001	.000	.000	
	N	19	19	18	18	27	27	27.000

**Table 63: Correlations between English can-do statements and C-test for bilingual group**

PCorr = Pearson Correlation

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\* . Correlation is significant at the 0.01 level (2-tailed).

In contrast, we find few meaningful interactions in the control group's data. Speaking and writing correlate ( $r = .677, p = .011$ ), as do  $z$ -scores with C-test scores and time ( $r = .549, p = .018$  and  $r = -.671, p = .002$ ). These make sense. The only other significant correlations were between listening and C-test time ( $r = .763, p = .004$ ), and listening and  $z$ -scores ( $r = -.764, p = .004$ ), thus those with higher listening skills take longer on the C-text and have a lower overall score. The only explanation we can offer is that the control group's self-ratings for listening were almost exclusively at 100%, so there could be a ceiling effect despite actual differences in performance.

### **5.5.2 External proficiency ratings**

While this study was set up in such a way as to provide a fairly comprehensive picture of proficiency in the two languages in question, we decided that it would be useful to get some independent measures of the level of English proficiency, since it is possible that some proficiency aspects that do play a role in how a second language learner is perceived are not captured as well in our measures (accent springs to mind).

We commissioned two linguistically trained native speakers to listen to the monologues of the film retelling and to provide a global rating for each bilingual participant, following the CEF proficiency levels. We did not disguise the fact that the participants were L2 learners, and did not ask them to rate native speakers, although some comparative samples were provided for reference if they wished to listen to them.

The ratings paint a very clear picture: there was a single participant who achieved a borderline rating of B2-C1 on the basis of her retelling (S24, 3.7%); all other scores are at least C1. Overall, four participants (15%) were rated C1, ten (37%) C1-C2, and twelve (44%) C2 (cf. Table 64), alongside comments such as "difficult to distinguish from native speaker", "excellent all rounder". These results correlate very highly with the  $z$ -scores ( $r = .742, p = .000$ ), and also with the English C-test scores achieved by the participants ( $r = .520, p = .005$ ). The latter slightly lower correlation may be explained by the fact that the global rating provided by the C-test is based on written materials and therefore cannot take into account features such as accent or fluency.

Reference level	N (%)	Participant
C2	12 (44%)	S1, S3, S4, S5, S9, S11, S14, S15, S16, S17, S19, S20
C1-C2	10 (37%)	S8, S10, S12, S18, S21, S22, S23, S25, S26, S27
C1	4 (15%)	S2, S6, S7, S13
B2-C1	1 (4%)	S24

**Table 64: Proficiency ratings by external raters by participant**

These features were, however, commented on by the raters where they were found noteworthy. For example, seven participants were described as having acquired distinctly Irish pronunciation ("significant Irish/Hiberno-English influence on phonology"; "sounds like native speaker phonology"), and two more as having acquired some features, such as post-vocalic /r/ and Irish vowel values. Another participant has "definitely non-Irish, though not necessarily non-native, phonology". Just six participants were described as "evidently" or "identifiably non-native" or "non-Irish" with regard to their phonology. Similarly, three participants' high level of fluency was noted ("very proficient vocabulary and fluency of speech"), alongside five others, who appeared a bit hesitant, but this was assumed to be partially connected to the contents of the story.

Apart from comments relating to the level of command of grammar and lexicon ("native-like grammatical and lexical control", "very colloquial, native-like vocabulary"), the raters also picked out certain features, such as the use of colloquialisms and idiomatic phrases, as well as some errors, which we already discussed in the error/gem analyses of the film retelling.

One final comment relates to S18, the participant whom we mentioned at the beginning of this chapter as someone who for a number of reasons experienced the experimental setting as challenging, and who on several occasions, notably in the C-test, but also the film retelling, claimed not to know particular words or phrases. One of the raters had this to say about her:

"Irish influences on phonology, e.g. fricative /t/ ("shot"). Very correct use of grammar ("Charlie has to remind her as to who he is"). I think she's more proficient than she thinks she is, more than once she mentioned that she couldn't remember the appropriate word where another simpler word would have sufficed."



In conclusion of this subsection then, we can be confident that our results do describe the level of L2 achievement by our participants fairly accurately, so we are dealing with a group of bilinguals that has become very proficient in their second language.

### **5.5.3 Self-reported L2 proficiency over time**

To round off the discussion, we point to a few themes arising out of the sociolinguistic questionnaire/interview. Several questions dealt directly with the participants learning of foreign languages: Questions 37/38 asked about their proficiency level in the past and now, Questions 34 and 36 enquired about their motivation for and difficulties in learning languages, and Question 48 asked what in their opinion promotes successful language learning.

We already presented the figures for the self-ratings of L2 proficiency prior to migration and at the time of testing in the introduction to this chapter. The relevant question did not yield much additional information, apart from one participant relating that she went through a "silent period" during which she noted down expression she noticed, and another talking about realizing how her range of language domains expanded, and the opportunities and challenges associated with that.

Most participants initially learnt languages in school, while two participants (S8, S26) are more or less self-taught as far as English is concerned. Some enjoyed language subjects and went on to do language courses after school, others realised the opportunities for friendships, job opportunities, travelling and ultimately settling abroad implicit in knowing other languages. However, even those who did have an interest in learning languages found certain aspects of instructed language learning boring or difficult – formal/grammar/rote learning, the absence of practical application and the lack of progress.

Instead, when asked about what helped them acquire English to their current level, the participants overwhelmingly pointed to their experience of living abroad, and the opportunities afforded by that fact for speaking and discovering the language in meaningful contexts, for listening and copying others (the "parrot system", as one participant called it), be that in person by listening to the radio and the television, watching films and news in the original, as well as reading books. All these

activities constitute language acquisition/maintenance effort becoming quite effortless in the appropriate language environment, as we had proposed in Chapter 2.

Several participants mention the role of significant others in providing feedback, input and opportunities for language use, and more than half highlight the need to be prepared to take risks by trying the language out and taking the initiative.

Many agree that having a basis in the language prior to experiencing immersion is helpful and speeds up the progress, but the case of the two essentially self-tutored participants would seem to indicate that it is not a necessary prerequisite.

Thus, to summarise, the bilingual participants in this study represent a group of fairly to very highly proficient late L2 learners, who seem to have benefited from a rich linguistic environment and sufficient input. This result is contrary to the view that ultimate attainment is exclusively constrained by age of acquisition, and nativelikeness is unachievable for adult learners, as discussed in the Introduction. Instead, it confirms that native-like attainment is possible, as has been suggested in several recent studies on late bilinguals, for example, by White and Genesee (1996), Bongaerts (1999), Birdsong and Molis (2001), Flege and MacKay (2004) and Birdsong (2006).

Birdsong (*ibid.*: 20) argues that "nativelikeness in late L2A is not typical, but neither is it exceedingly rare" under conditions of "sufficient LoR and contact with natives". Similarly, Bongaerts, Mennen and van der Slik (2000: 298) conclude from their analysis "that a combination of input, motivational, and instructional factors may compensate for the neurological disadvantages of a late start". Birdsong and Molis (2001) further suggest that the incidence of native-like attainment may depend on the particular L1-L2 pairing, with greater typological proximity, as in our case, possibly fostering L2 acquisition. Whether some of our participants show "narrow" or "broad" nativelikeness (Hyltenstam and Abrahamsson 2000), further analyses may reveal.

## **6 Bilingual profiles – within-group study**

This chapter provides a synopsis of results from the within-group study in relation to research questions 3 and 4. The within-group study ultimately aims to shed light on what factors determine the interrelationship between the two (and more) language systems in the bilingual mind. We first look at some of the predictor variables that have been suggested in the literature, biographical variables, language use, language attitudes and linguistic aptitude. As a first step in assessing whether L2 proficiency may be an independent predictor variable, we then give an overview over the bilingual profiles of the participants. The discussion serves to set the scene for the future direction of this research and research into first-language attrition more generally.

### **6.1 Predictor variables**

This section expands two conference presentations on the impact of language use and aptitude as predictor variables for L1 and L2 proficiency presented at AILA 2008 and ISB7 2009 respectively (Opitz 2008; 2009; slides can be made available), by taking into account additional predictor variables. For each of the variable groups corresponding to the four subsections below, we ran correlations with the bilingual group's  $z$ -scores.

#### **6.1.1 Biographical factors**

We first consider several factors that are often proposed as playing a part in attrition: length of residence (LOR), age at time of testing; gender, education and the number of languages known by the bilingual; nationality, first language of partner and children living with the family (cf. Table 65 below).

For the German  $z$ -scores, significant correlations were obtained for LOR, education level and nationality. (For the sake of brevity, we do not replicate the  $r$  values for the correlations given in the tables.) The direction of the effects mean that German is maintained better by people who have stayed in Ireland for shorter periods of time, who have a higher level of education and who have kept their German nationality (very few people have changed their passport, since double citizenship only became possible quite recently).

		Length of residence (LOR)	Age at time of testing	Gender	Education level	Number languages	Nationality	L1 of partner	Children
Mean z-scores	Pcorr	-.441*	-.280	.268	.524**	.307	.585**	.184	.191
	Sig. 2-t	.021	.158	.177	.005	.119	.001	.358	.341
German	N	27	27	27	27	27	27	27	27
Mean z-scores	Pcorr	.051	-.169	.445*	.170	.429*	-.056	-.597**	-.306
	Sig. 2-t	.802	.401	.020	.398	.026	.781	.001	.120
English	N	27	27	27	27	27	27	27	27

**Table 65: Correlations between bilinguals' z-scores and biographical predictor variables**

Pcorr = Pearson correlation

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

The English results, on the other hand, correlate significantly with gender, the number of languages known and the L1 of the partner. In other words, the English scores are higher for females, whose partner is Irish and who know more languages.

It is interesting to note which combinations do not result in significant correlations. For example, it seems to make no difference to one's level of L1 maintenance *per se* whether one is male or female, whether the partner also speaks German, or whether there are children to speak German with. Age at time of testing is not relevant for either L1 maintenance or the level of L2 proficiency. (The latter could be related to the fact that all bilinguals are more or less "stable" bilinguals (Ng and Wigglesworth 2007).) Similarly, length of residence does not play a role (anymore) in L2 attainment. Also, education level is less relevant for L2 acquisition in naturalistic settings, while having a partner to speak the language with makes a difference.

There are also significant interactions amongst the biographical variables. The main correlations are, not surprisingly, between age at time of testing and length of residence ( $r = .764, p = .000$ ) and the number of children ( $r = .452, p = .018$ ) – people who are older have longer LORs and are more likely to have children.

Further, again not surprisingly, there is a significant negative interaction between education and age as well as LOR ( $r = .572, p = .002$  and  $r = .597, p = .001$  respectively) – older people have received less education. Education also correlates with gender and the number of languages known

( $r = .476, p = .012$  and  $r = .396, p = .041$  respectively) – females are more highly educated and know more languages. The negative correlation between number of languages known and children (those without children know more languages), and the positive one between languages and gender (females know more languages) follow from the interactions just mentioned.

An interesting fact is that the males in the sample tend to have German partners ( $r = .571, p = .002$ ). This variable, like the presence of children, is likely to be linked to language use, which we will consider next.

### **6.1.2 Language use**

As discussed in the Chapter 2, language contact is often seen as a major predictor for language maintenance. For both the Activation Threshold Hypothesis and DMM, it is the prime factor in bringing about or combating language attrition. Thus, the Activation Threshold Hypothesis holds that if a language is not used, its activation threshold, and that of items within it, rises and may become too high for fluent access. If another language is used instead, it serves to inhibit the first language, and that of items within it, again potentially leading to higher activation thresholds, disfluency and crosslinguistic influence.

Within DMM, language exposure/use is the most important part of language maintenance effort. If there is insufficient maintenance effort, a language will attrite. If language effort is expended on acquiring or using another language or other languages, the L2 user may reach psychological and practical limits preventing him or her from supporting the first language through continued use.

On the basis of the responses to the questionnaires, we computed two language use variables (cf. Table 66). The first, "Average L1 frequency", takes into account questions relating to how often German is used by the bilinguals overall, ranging from "never" (lowest score) to "all the time" (highest score). The second index, "Average bilingual use", relates to the amount/frequency with which either English or German is used for particular purposes. It ranges from "English only" (lowest score) to "German only" (highest score).

		Average L1 frequency	Average bilingual use
Mean <i>z</i> -scores	Pcorr	.545**	.452*
German	Sig. (2-tailed)	.003	.018
	N	27	27
Mean <i>z</i> -scores	Pcorr	-.167	-.567**
English	Sig. (2-tailed)	.406	.002
	N	27	27

**Table 66: Correlations between bilinguals' *z*-scores and language use**

Pcorr = Pearson correlation

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

The correlations show a strong positive correlation between the overall amount of German use and the German *z*-scores. Thus, a person who uses German more has maintained German better. Similarly, the bilingual index shows that those people who use more German have a higher level of German-language proficiency, and those who use more English have higher proficiency in English. However, use of German has no significant negative impact on English-language proficiency, as one might have ventured.

While this is in line with expectations, these effects have not always been proven. We thus find support for the role ascribed to this factor within the context of our two explanatory models.

### **6.1.3 Language attitudes**

As discussed in the literature, attitudes are a multi-componential phenomenon, and intuitively it would make sense to distinguish attitudes relating to questions of identity from attitudes towards other languages and cultures. However, in analysing the responses, it became apparent that questions may have been answered in a similar manner for different reasons.

For example, the statement "I would like my children to know both German and English" might have been given a rating of "I totally agree" because the speaker would like to pass on that part of his/her identity, or it might be related to an instrumental motivation – the consideration that the children will have better chances in the job market if they know "foreign" languages. In fact, it is

quite likely that different motivations come together in rating such statements, as the interviews show anecdotally.

Therefore, we decided to work from the ratings as they are for the time being. We proceeded in a similar manner to the language use data just discussed to produce attitude indices (cf. Table 67).

Index 1, "Orientation towards English and other L2s" combines statements that probed people's disposition towards other languages. The lowest score corresponds to the strongest L2 orientation.

Index 2, "Bilingual index", is a combination of statements that try to assess the value placed on knowing more than one language. This is similar, but not identical to a foreign language orientation, which might be interpreted as an orientation away from the L1. Again, the lowest score indicates the strongest agreement with the relevant statements.

Finally, Index 3, "Orientation towards German", combines several statements, which were conceived of as giving an indication of a person's identification. However, as we just pointed out, the responses in some cases may be related to different motivations. Here, the highest score corresponds to the highest level of agreement, in order to allow combining indices 1 and 3. This would be plausible if one argued that an orientation towards other languages implies an orientation away from the mother tongue, but as our preliminary results show, the issue is more complicated than that.

		Index 1: Orientation towards English and other L2s	Index 2: Bilingual orientation	Index 3: Orientation towards German
Mean z-scores	Pcorr	-.027	-.526*	.245
German	Sig. (2-tailed)	.914	.021	.218
	N	19	19	27
Mean z-scores	Pcorr	-.471*	-.036	-.540**
English	Sig. (2-tailed)	.042	.884	.004
	N	19	19	27

**Table 67: Correlations between bilinguals' z-scores and attitude**

Pcorr = Pearson correlation

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

Index 1 and 3 both correlate highly with the English scores, meaning that the more a person values English and foreign languages, the higher their attainment. However, the opposite is not true, as the lack of correlation with the German scores shows. The German scores, on the other hand, correlate with Index 2, indicating that the higher a person values bilingualism, the higher their level of L1 maintenance.

This is a very interesting and slightly unexpected result in relation to the German data. It would have been more plausible if an orientation towards German rather than bilingualism was important in L1 maintenance. However, perhaps the ratings on Index 2 reflect the realities of living in an integrative bilingual situation, where the exclusive use of the mother tongue is neither possible nor desirable, so the value placed on keeping both languages active has replaced an exclusive orientation towards German.

#### **6.1.4 Linguistic aptitude**

The aptitude test was taken by all three groups, so we first give the results for each sub-test and group separately (cf. Table 68 below). The norms for each sub-test, and therefore the range of the first StD differ, so scores mean different things for each sub-test. We therefore refrained from combining them.

As we already pointed out in the Chapter 3, the scores of the two German groups are lower, in the case of LAT B significantly lower, than those of the Irish control group on the two sub-tests that use English as the language of presentation. This would seem to indicate that the level of proficiency achieved in English has some part to play in the success with which the tests were solved, although, as we argued, the actual level of proficiency strictly needed to be able to do the tasks, is much lower than the level of attainment of our bilingual participants in the very least.

Notwithstanding this proviso, we decided to run the relevant correlations, though in this case, even more than is usually the case with correlations, we cannot draw any conclusion about the direction of cause and effect.



LAT	Bilingual group	German controls	Irish controls
<b>D aural recognition</b> (range of 1 <sup>st</sup> StD: 70-84)			
Mean	77.8	82.3	83.5
Median	80	85	80
StD	11.47	10.57	7.326
Range	50-95	55-95	65-90
<b>B visual memory</b> (range of 1 <sup>st</sup> StD: 43-73)			
Mean	58.0	48.2	74.6***
Median	66	43	82.5
StD	28.11	23.55	11.94
Range	14-100	12-90	60-95
<b>C rule inference</b> (range of 1 <sup>st</sup> StD: 60-69)			
Mean	65.6	62.8	77.0
Median	70	65	74
StD	16.43	18.81	17.12
Range	35-100	20-90	40-98

**Table 68: Mean aptitude scores for all three groups**

		Aptitude D Aural recognition	Aptitude B Visual memory	Aptitude C Rule inference
Mean z-scores	Pcorr	-.210	.477*	.094
German	Sig. (2-tailed)	.314	.018	.661
	N	25	24	24
Mean z-scores	Pcorr	.107	.584**	.541**
English	Sig. (2-tailed)	.610	.003	.006
	N	25	24	24

**Table 69: Correlations between bilinguals' z-scores and linguistic aptitude**

Pcorr = Pearson correlation

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

The results prove interesting. While on the one sub-test not employing English, but only the artificial test language, we do not see a significant effect with either proficiency score, the English scores do correlate very highly with the other two sub-test. As discussed, this may mean that higher English language proficiency made it easier to do the task, or that those with high aptitude have

achieved a high level of L2 proficiency. It is possible that both are true. The latter interpretation does make sense in a naturalistic learning environment where input is not pre-structured. On the other hand, participants had a certain level of L2 proficiency before they emigrated, so perhaps they already had sufficient access to the language, diminishing the role for aptitude.

However, as discussed before, linguistic aptitude is usually only considered for its role in L2 acquisition, so the fact that the word-learning task LAT B correlates with the level of L1 maintenance is interesting. This could indicate that those participants who have higher verbal skills find it easier to maintain their mother tongue, which, given the abundance of lexical retrieval difficulties in the German data, seems plausible.

## **6.2 Profiles**

The brief analysis of bilingual participants' language profiles ("bilingual profiles", subsection 3) in this section follows the procedure discussed in Opitz (2010), which presented a detailed analysis of individual profiles based on the C-test and FiCA tasks as well as considering task demands on performance. Following a cluster analysis on the level of language proficiency in L1 and L2 for each of the two tasks, we found that participants could be divided into bands that differed along the dimensions of size and direction of the L1/L2 score differential. In other words, we identified three main bilingual profiles:

- a) equal or balanced performance in L1 and L2 ( $L1 = L2$ )
- b) slightly better performance in L1 or L2 ( $L1 > L2$  or  $L1 < L2$ ) and
- c) markedly better performance in L1 or L2 ( $L1 \gg \gg L2$  and  $L1 \gg L2$  or  $L1 \ll L2$  and  $L1 \ll \ll L2$ ).

In addition to these relative proficiency levels, the absolute L1 proficiency level (relative to the German control group) also turned out to be an important factor. We identified four such levels which corresponded to the position of a score either above the upper 2<sup>nd</sup> StD of the control group's mean (outstanding), within the 1<sup>st</sup> and upper 2<sup>nd</sup> StD (high), within the lower 2<sup>nd</sup> StD (medium) and at the lower end or below the native-speaker range (low L1 proficiency).

These criteria can also be used to identify a person's absolute L2 proficiency level. The first three levels represent bilinguals who perform comparably to the control groups and who can thus be said

to have maintained their German at native-like levels or who have attained native-like proficiency as measured on our tests in English. The fourth level, on the other hand, represents bilinguals who perform at levels below the native-speaker range as established by the controls, either as a result of L1 attrition or lower L2 acquisition.

The results presented in the above paper showed that bilinguals' profiles were relatively stable across the two tasks, and that most differences could plausibly be attributed to the different task demands of FiCA and C-test. Thus, we felt justified in performing a cluster analysis (hierarchical cluster analysis using average linkage between groups) on the overall German and English scores (mean *z*-scores) for the bilingual group. Since *z*-scores are standardized scores, there was no need to adjust one set of scores by scaling.

The analysis yielded 13 relevant clusters (cf. Table 70 and Figure 24), which suggested some small modification to the labels for the proficiency levels compared to the bands just discussed:

- i) very high/outstanding L1/L2 proficiency – upper end of or above the 2<sup>nd</sup> StD of the relevant control group's mean,
- ii) high L1/L2 proficiency – around and above the relevant control group's mean (principally the upper 1<sup>st</sup> StD and parts of the upper 2<sup>nd</sup> StD),
- iii) medium L1/L2 proficiency – below the relevant control group's mean, including the lower 2<sup>nd</sup> StD and
- iv) lower L1/L2 proficiency – below the native-speaker range/lower 2<sup>nd</sup> StD.

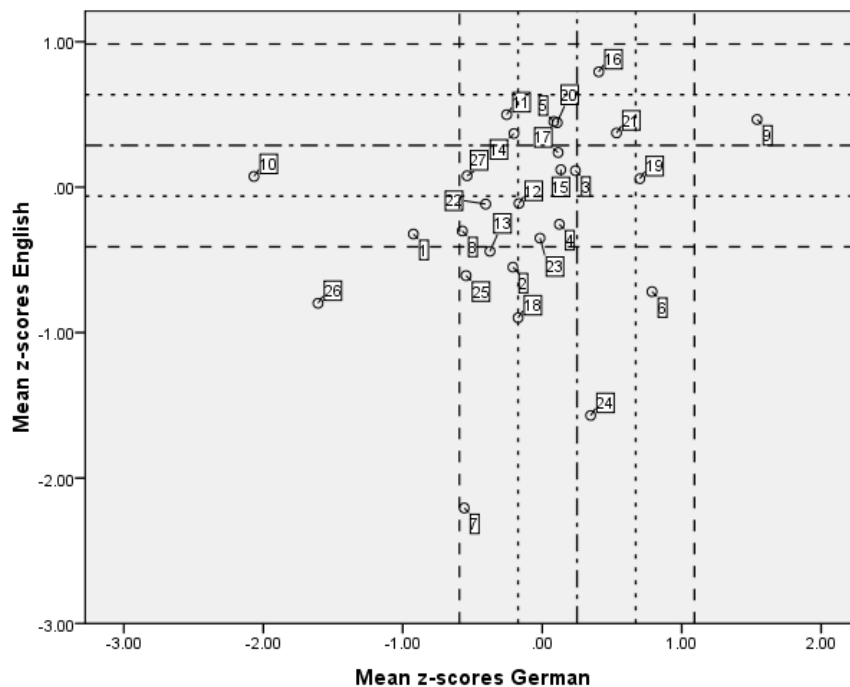
The last level has been subdivided further (low level) where differences within the band were large; and some clusters contain sub-clusters.

There is a large range of profiles, with few members in each (sub-) cluster. The L1 dominant profile accounts for nine cases, or a third of participants, and is the profile one might have expected for late bilinguals. However, just three of these, S7, S24 and S6, have a strongly L1 dominant profile (cf. Figure 24). S7 has the lowest score for English, but his German is not very good either, while S24 and S6 have good German but are less proficient in English. S7 and S24 are the individuals whose L2 attainment is much lower than the rest of the bilinguals. S9 seems to be particularly good in German, with a high score in English as well, but her German score has to be considered with caution since we only have the C-test data for this participant in German.

Cluster number	Profile*	Participants	N = 27 (percent)
1	HH L1 > H L2	S9	1 (3.7%)
2	H L1 >> LL L2	S24	1 (3.7%)
3	H L1 >> L L2	S6	1 (3.7%)
4	H L1 > M L2	S19	1 (3.7%)
5	H L1 = H L2	S16, S21	2 (7.4%)
6	M L1 > L L2	S4, S23 S2, S18	4 (14.8%)
7	M L1 = M L2	S3, S15, S17 S12, S22, S27	6 (22.2%)
8	M L1 < H L2	S5, S11, S14, S20	4 (14.8%)
9	L L1 >> LL L2	S7	1 (3.7%)
10	L L1 = L L2	S13, S25 S1, S8	4 (14.8%)
11	LL L1 << M L2	S10	1 (3.7%)
13	LL L1 < L L2	S26	1 (3.7%)

**Table 70: Clusters of bilingual profiles based on German/English z-scores**

\* HH = very high, H = high, M = medium, L = lower, LL = low level of proficiency



**Figure 24: Scatter plot of bilingual profiles**

The dot-dash lines represent the mean of the control groups, the dotted lines the range of the 1<sup>st</sup> StD, and the dashed lines the range of the 2<sup>nd</sup> StD. The native-speaker range is slightly above the limits of the lower 2<sup>nd</sup> StD.

The majority of the participants (12/27, 44%) has a balanced profile across both languages, albeit at different absolute levels of proficiency. S16, who has the highest English score of any of the participants, is a very good all-rounder; S21 is somewhat better in German and a little less proficient in English, resembling S19, but all three of them still have very good German language skills. Six more individuals within the balanced group have good proficiency in both languages, but S3, S15 and S17 have maintained their German slightly better than S12, S22 or S27.

S1's German score, on the other hand, is below the native-speaker range, meaning that she may suffer a degree of L1 attrition, which matches the impression formed on the basis of the detailed discussion of her German data. Her attainment of English is also in the lower range. In this respect, she is similar to S8, and also S13 and S25.

The remaining six individuals (22%) have an L2-dominant profile, again at different absolute proficiency levels. S26 stands out as having low scores in both languages, part of which needs to be attributed to his anxiety in relation to the C-test, which we referred to above. S10 is the individual with the lowest score overall in German, while her English is good. S10, S26 and S1 thus are people who can be called L1 attriters with some justification.

On the other hand, S5, S11, S14 and S20 have maintained a good level of German, but they also have achieved a high level of L2 proficiency, which exceeds their current German proficiency level. These, along with S16, S21 and S9, may be seen as very proficient adult bilinguals.

To summarise, while the expected L1-dominant profile does account for a sizeable share of cases, the proportion of balanced cases outweighs those, and there are also several individuals with L2 being stronger than L1. Some participants have high skills across both languages, others are not very good in either. Furthermore, some cases point to L1 attrition, others to L1 maintenance, others again to more or less successful L2 acquisition, ranging from "typical L2 learners" to highly proficient bilinguals.

Given the variety of profiles present in our data, including the profiles for the participants that do seem to exhibit a degree of L1 attrition, we conclude that the level of L2 proficiency is not by itself a reliable predictor variable for the degree of L1 attrition, pending further analyses. Instead, individuals' profiles are evidently influenced by, probably a combination of, factors, such as those discussed in the first section of this chapter.

### **6.3 Summary and discussion**

As a result of our decision to focus on the between-group studies in this thesis, the exposition of the within-group data in this chapter has been very brief and by no means comprehensive. In relation to predictor variables for L1 attrition and L2 acquisition, the results show that some predictors are strong enough to result in significant correlations by themselves, while others fail to do so on their own, but possibly play a part due to their strong relationships with other factors. The findings are similar to Köpke's (2007) who concludes that

"many of these factors are closely linked, (b) some of them may be considered cluster factors which are more influential and more strongly interrelated than others, and (c) none of the factors may alone be considered responsible for attrition." (Köpke *ibid.*: 9)

These results are in line with dynamic approaches to bilingualism that argue that single-factor analyses of linguistic data, and a preoccupation with large effects are ultimately doomed to failure given the complete interconnectedness of the language system(s), and the multi-directional interaction of a host of different factors (e.g. de Bot 2007).

The proposed interaction of factors probably also explains the large number of different profile types relative to the small number of participants tested. In view of the ongoing debate about adult bilinguals' capacity to acquire a second language to a high level, however, it is an interesting finding that the number of participants who have balanced or L2-dominant profiles at relatively high absolute proficiency levels actually outweighs that of less proficient bilinguals.

Similarly, it is heartening to know that few participants show reduced L1 proficiency/L1 attrition, despite high levels of L2 attainment, in view of the fact that some researchers see the beginnings of L1 attrition firmly in the context of high L2 proficiency (Seliger and Vago 1991; Cook 2002; de Groot 2002). Our findings would seem to support the idea of adults' bilingual capacity.

By necessity, the findings presented in this section are preliminary. We already proposed that investigating the impact of clusters of factors appears to present a fruitful avenue for continuing this particular piece of research. Further, we intend to extend the profiles analysis to the individual level, exploring possible commonalities between people with particular profiles, and the combination of factors that have led to extreme profiles at either end of the range.

## 7 Conclusions

This thesis set out to explore the outcomes of the parallel processes of L1 attrition and L2 acquisition in an L2 setting. Taking a multi-competence perspective of linguistic development and bringing together different subfields of linguistics, we were able to show that L1 attrition does indeed not take place in a vacuum but interrelates in various complex ways with the process of L2 acquisition. Our findings are consistent with the two models we invoked in Chapter 2 – Paradis's Activation Threshold Hypothesis and Herdina and Jessner's Dynamic Model of Multilingualism – in order to predict and account for those two processes and their outcomes.

Our findings confirm that L1 attrition is a real process, albeit one that appears to proceed slowly and may be overestimated in its extent by those affected. Few participants display consistent signs of L1 attrition to mark them out as L1 attriters overall. However, there is evidence of small-scale effects in most participants, indicating increased variability, or "scatter", which has been hypothesised (within DMM) and modelled to precede an actual deterioration in proficiency (Jessner 2003; Meara 2004). Moreover, in line with DMM's proposed attrition curves, we would not expect to see large effects in a language that had been acquired to a maximal level if language maintenance effort is stepped up sufficiently. The required maintenance effort, however, is substantially reduced following even a small reduction in proficiency.

There is a large role for the L2, in that most effects found in the bilingual group can be accounted for by crosslinguistic influence. On the other hand, some difficulties appear to be simply due to the (temporary) unavailability of the lexical item or grammatical structure in question, hypothesised to be due to insufficiently low activation thresholds. This in the main confirms previous findings, and is indirect confirmation of the interrelatedness of the two processes in question, as well as the interconnectedness of the language systems, be that at the processing or the representational level.

Regardless of the scale of these effects, they serve to counter the myth that L1 proficiency is immutable, along with the idea of "the perfect native speaker", as the results of the control groups show. The variability in the bilingual group's, and, to a lesser extent, the control groups', performance serves to support the view that bilinguals/L2 users are "competent but specific speaker-hearer[s]" (Grosjean 2008: 9).

With regard to L2 acquisition, we have been able to show that a considerable number of the bilinguals perform similarly to the Irish control group on each measure, and overall. This lends support to the idea that late learners are capable of benefiting from a rich linguistic environment and sufficient input: "a combination of input, motivational, and instructional factors may compensate for the neurological disadvantages of a late start" (Bongaerts, Mennen et al. 2000: 298), in contrast with the common view of late learners as a "general failure" (Ellis 1994: 107).

The importance of the learning context, i.e. immersion, becomes clear from the qualitative analysis of participant responses. Although it is sometimes claimed that adult learners have to rely on explicit learning mechanism for second-language acquisition, (e.g. DeKeyser 2000), the fact that none of the bilinguals attended any language courses once in the L2 environment would suggest that implicit L2 acquisition is at least a possibility.

In the same way that many of the bilinguals experienced L2 influence in their L1 performance, L1 influence on L2 (transfer) was also attested in some of the less proficient bilinguals. The occurrence of L1-L2 transfer is not per se noteworthy, other than to say that it is conspicuous by its relative absence, and that "reverse transfer" (L2->L1), which was only "rediscovered" as part of the development of the multi-competence view (Cook 2003: 1) is far more prevalent in the data.

Since transfer is often associated with lower proficiency levels, its absence indirectly testifies to the level of L2 attainment by the bilinguals. Moreover, some cases of transfer appeared to be cases of fossilised forms in the sense that they stood out as forms from an earlier stage of development in the bilinguals in question who overall had achieved a higher level of L2 proficiency (Han 2004: 21).

With regard to factors potentially impacting on the degree of L1 maintenance and/or L2 acquisition, some, like language use, attitudes, and maybe to a lesser degree aptitude, seem to be relevant for both, reinforcing the idea of the interrelatedness of the two processes. With other, particularly biographical, factors, such as length of residence, education etc., however, there seems to be a dissociation, in that some are important for L1 maintenance but not L2 acquisition, and vice versa. Interestingly, age at time of testing does not appear to play a role in either.

The bilingual profiles provide evidence for the complex interactions between all the potential variables by way of very diverse outcomes. Although some profile types and groups of similar



participants could be identified, there is a great deal of variability in the ultimate degree of L1 attrition and L2 acquisition, which requires further investigation, as does the question of proficiency-related interactions between the languages in the bi-/multilingual mind.

### **7.1 *Limitations of the present study and directions for future research***

The applicability of the data presented in this thesis to wider contexts is limited by a number of factors. Before discussing these in turn, we would like to mention a regret which resulted from an early decision to refer to our participants by using the S code followed by a number. While this served the purpose of anonymising the data (which we were obliged to do since we had undertaken to treat all information confidentially) and for pre-empting any biases in scoring the materials, it is a very impersonal way of reference. Given the intensity of contact we had with the participants, and the wealth of information they shared, there was no need for this degree of formality.

A better alternative would have been to use code-names, but unfortunately the researcher displayed an insufficient level of creativity on this front. This solution would presumably also have made it easier for the readers of this work who may have difficulties distinguishing participants on the basis of very similar labels.

The first limitation we would like to mention is the fact that the two languages investigated are typologically quite close. It is possible that different effects will obtain with different combinations of languages. However, as explained in the early part of the thesis, the study utilised instruments which were developed as part of a test battery that is made available to researchers undertaking L1 attrition research. This will make it possible to draw inferences on the basis of a larger number of participants and across different language combinations, improving the applicability of the findings.

Second, although not strictly a case study, the research is limited by its relatively small scale, which may impact on the generalisability of the results. This is more relevant for some tests in the battery where participant numbers were smaller than on others (e.g. FiCA). However, the number of respondents to the sociolinguistic questionnaire and some other instruments were large enough for the type of analyses they were subjected to, so this is not an absolute limitation.

Third, any statement we made in relation to the L1 and L2 proficiency of our participants is by definition valid only in relation to the tasks they were subjected to. In other words, the tasks were designed to measure particular constructs of proficiency rather than an abstract notion of proficiency. However, we took care to devise tasks with a high level of validity and combined the results for a more comprehensive assessment of participants' proficiency.

Fourth, and related to the third point, while we consider being able to make reliable statements concerning our participants' overall proficiency, the three proficiency components of complexity, accuracy and fluency we set out at the beginning were not tested to the same extent. We have several measures for global proficiency (C-test, total z-scores, L2 native-speaker ratings), accuracy (C-test, error rate) and fluency (FiCA, various hesitation measures), but only one for complexity (lexical diversity; this was measured in two different contexts). However, we were unable to devise a suitable measure of grammatical complexity (apart from what the C-test tests). MLU analyses provided with CLAN are only suitable at lower levels of proficiency. This is a point we would like to address in future research.

Finally, despite our subscribing to a dynamic perspective on bilingualism across the lifespan, the data presented so far can only provide a snapshot, mostly of the current levels of L1 and L2 proficiency of the participants, apart from the self-assessed changes in proficiency. However, during the study much more material was collected, which we did not have space to consider here, but which ultimately should help to shed some light on the dynamic changes not just in L1 and the L2 of the host country, but also other languages known by the bilingual in questions.

In addition to the future research avenues outlined in the previous sections, such as the qualitative analysis of the bilingual profiles, a more detailed and qualitatively different look at the predictor variables, and, of course, the analysis of the parts of the test battery which were not included in this thesis, this type of research should ultimately try and answer not just the questions of whether, to what degree and why/under what circumstances L1 attrition and L2 acquisition are likely to occur, but also become more proactive in making recommendations – at an individual and a societal/political level – concerning how to promote additive bilingualism – which also implies redressing the relative importance that is given to the processes of L2 acquisition and L1 maintenance in L2 settings.

## References

- Abrahamsson, N. and K. Hyltenstam (2008). "The Robustness of Aptitude Effects in Near-Native Second Language Acquisition." Studies in Second Language Acquisition 30(04): 481-509.
- Abrahamsson, N. and K. Hyltenstam (2008). The Role of Language Learning Aptitude for Near-Native L2 Attainment.
- Abrahamsson, N. and K. Hyltenstam (2009). "Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny." Language Learning 59(2): 249-306.
- Alderson, C. J. (2002). Testing Proficiency and Achievement: Principles and Practice. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 15-30.
- ALTE (1990). Association of Language Testers in Europe.
- Altenberg, E. P. (1991). Assessing First Language Vulnerability to Attrition. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 189-206.
- Ammerlaan, T. (1996). "You Get a Bit Wobbly..." - Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition. Nijmegen, Katholieke Universiteit Nijmegen.
- Ammerlaan, T., M. Hulsen, et al. (2001). Language Maintenance, Shift, and Loss: Work in Progress. Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. T. Ammerlaan, M. Hulsen, H. Strating and K. Yagmur. Münster/New York/München/Berlin, Waxmann: 1-13.
- Ammerlaan, T., M. Hulsen, et al., Eds. (2001). Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. Münster/New York/München/Berlin, Waxmann.
- Andersen, R. W. (1982). Determining the Linguistic Attributes of Language Attrition. The Loss of Language Skills. R. D. Lambert and B. F. Freed. Rowley, MA, Newbury House: 83-117.
- Anderson, J. R. (1983). The Architecture of Cognition. Cambridge, MA, Harvard University Press.
- Baayen, R. H., R. Piepenbrock, et al. (1995). The Celex Lexical Database. Philadelphia, University of Pennsylvania, Philadelphia Linguistic Data Consortium.
- Baker, C. (1996). Foundations of Bilingual Education and Bilingualism. Clevedon, Philadelphia, Adelaide, Multilingual Matters.
- Baker, C. (2001). Foundations of Bilingual Education and Bilingualism. Clevedon, Multilingual Matters.
- Beausoleil, N., R. Fortin, et al. (2003). "Unconstrained Oral Naming Performance in Right- and Left-Hemisphere-Damaged Individuals: When Education Overrides the Lesion." Aphasiology 17(2): 143 - 158.
- Benton, A. L., P. J. Eslinger, et al. (1985). Normative Observations on Neuropsychological Test Performances in Old Age. Studies in Neuropsychology. Selected Papers of Arthur Benton. L. Costa and O. Spreen. New York, Oxford, Oxford University Press: 319-326.
- Berry, T. E. (1971). The Most Common Mistakes in English Usage, McGraw-Hill.
- Bialystok, E. (2001). "Metalinguistic Aspects of Bilingual Processing." Annual Review of Applied Linguistics 21(1): 169-181.
- Biber, D. (1988). Variation Across Speech and Writing. Cambridge, Cambridge University Press.
- Birdsong, D. (1992). "Ultimate Attainment in Second Language Acquisition." Language 68(4): 706-755.
- Birdsong, D. (2006). "Age and Second Language Acquisition and Processing: A Selective Overview." Language Learning 56(s1): 9-49.
- Birdsong, D. and M. Molis (2001). "On the Evidence for Maturational Constraints in Second-Language Acquisition." Journal of Memory and Language 44(2): 235-249.
- Bloomfield, L. (1933). Language. New York, Holt, Rinehart and Winston.
- Blumenfeld, H. K. and V. Marian (2007). "Constraints on Parallel Activation in Bilingual Spoken Language Processing: Examining Proficiency and Lexical Status Using Eye-Tracking." Language and Cognitive Processes 22(5): 633-660.
- Bolonyai, A. (1999). The Hidden Dimension of Language Contact: The Case of Hungarian-English Bilingual Children. Columbia SC, University of South Carolina.
- Bongaerts, T. (1999). Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners. Second Language Acquisition and the Critical Period Hypothesis. D. Birdsong. Mahwah, New York, Lawrence Erlbaum: 33-159.
- Bongaerts, T., S. Mennen, et al. (2000). "Authenticity of Pronunciation in Naturalistic Second Language Acquisition: The Case of Very Advanced Late Learners of Dutch as a Second Language." Studia Linguistica 54(2): 298-308.
- Boyd, S. (1986). Using the Present to Predict the Future in Language Contact: The Case of Immigrant Minority Languages in Sweden. Language Attrition in Progress. B. Weltens, K. de Bot and T. van Els. Dordrecht, Foris: 99-115.
- Broeder, P., G. Extra, et al., Eds. (1988). Processes in the Developing Lexicon. Final report on the European Science Foundation's Additional Activity on Second language Acquisition by Adult Immigrants. Strasbourg, Tilburg and Göteborg.

- Brown, J. D. (2001). Using Surveys in Language Programmes. Cambridge, Cambridge University Press.
- Bullock, A. and S. Trombley (1999). The New Fontana Dictionary of Modern Thought. London, Harper-Collins.
- Bylund, E., N. Abrahamsson, et al. (2010). "The Role of Language Aptitude in First Language Attrition: The Case of Prepubescent Attriters." Applied Linguistics 31(3): 443-464.
- Callahan, L. (2004). Spanish/English Code-Switching in a Written Corpus. Amsterdam, John Benjamins.
- Carroll, J. B. (1981). Twenty-five Years of Research in Foreign Language Aptitude. Individual Differences & Universals in Language Learning Aptitude K. C. Diller. Rowley, Mass, Newbury House.
- Carroll, J. B. and S. Sapon (1959). Modern Language Aptitude Test: Form A. New York, Psychological Corporation.
- Cenoz, J. and F. Genesee, Eds. (1998). Beyond Bilingualism: Multilingualism and Multilingual Education. Multilingual Matters: 110. Clevedon, Multilingual Matters.
- Cenoz, J., B. Hufeisen, et al., Eds. (2001). Cross-linguistic Influence in Third Language Acquisition. Psycholinguistic Perspectives. Bilingual Education and Bilingualism 31. Clevedon, Multilingual Matters.
- Cenoz, J. and U. Jessner, Eds. (2000). English in Europe: The Acquisition of a Third Language. Bilingual Education and Bilingualism 13. Clevedon, Multilingual Matters.
- Cherciov, M. (to appear). The Role of Attitude and Aptitude in First Language Attrition: Preliminary Results from a Study of Adult Bilinguals Living in Canada. Actes du colloque: Les XXIIes Journées de linguistique. Québec City, 2008.
- Clark, J. L. D. (1982). Measurement Considerations in Language Attrition Research. The Loss of Language Skills. R. D. Lambert and B. F. Freed. Rowley, Massachusetts, Newbury House: 138-154.
- Clyne, M. (1987). "Constraints on Code Switching: how Universal are they?" Linguistics 25: 739-764.
- Clyne, M. (1992). Linguistic and Sociolinguistic Aspects of Language Contact, Maintenance and Loss. Towards a Multifacet Theory. Maintenance and Loss of Minority Languages. W. Fase, K. Jaspaert and S. Kroon. Amsterdam, John Benjamins: 17-36.
- Clyne, M., C. Rossi Hunt, et al. (2004). "Learning a Community Language as a Third Language." The International Journal of Multilingualism 1(1): 33-52.
- Clyne, M. G. (1995). The German Language in a Changing Europe. Cambridge, Cambridge University Press.
- Cochran, B. P., J. L. McDonald, et al. (1999). "Too Smart for Their Own Good: The Disadvantage of a Superior Processing Capacity for Adult Language Learners,." Journal of Memory and Language 41(1): 30-58.
- Collentine, J. and B. F. Freed, Eds. (2004). Learning Context and its Effects on Second Language Acquisition. = Studies in Second Language Acquisition 26. Cambridge, Cambridge University Press.
- Cook, V. (1995). Multi-Competence and Effects of Age. The Age Factor in Language Acquisition. D. Singleton and Z. Lengyel. Clevedon, Multilingual Matters: 51-66.
- Cook, V. (1997). The Consequences of Bilingualism for Cognitive Processing. Tutorials in Bilingualism. Psycholinguistic Perspectives. A. de Groot and J. F. Kroll. Mahwah, New Jersey, Lawrence Erlbaum: 279-299.
- Cook, V. (2002). Background to the L2 User. Portraits of the L2 User. V. Cook. Clevedon, Multilingual Matters: 1-28.
- Cook, V. (2003). Introduction: The Changing L1 in the L2 User's Mind. Effects of the Second Language on the First. V. Cook. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters: 1-18.
- Coolican, H. (1999). Research Methods and Statistics in Psychology. London, Hodder & Stoughton.
- Costa, A., Á. Colomé, et al. (2004). "Another Look at Cross-Language Competition in Bilingual Speech Production: Lexical and Phonological Factors." Bilingualism: Language and Cognition 6(03): 167-179.
- Daller, H., J. Treffers-Daller, et al. (2002). The Development of a Turkish C-Test. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 187-199.
- Davies, A. (1986). Language Loss and Symbolic Gain: The Meaning of Institutional Maintenance. Language Attrition in Progress. B. Weltens, K. de Bot and T. van Els. Dordrecht, Foris: 117-127.
- Davies, A. (2004). The Native Spaker in Applied Linguistics. The Handbook of Applied Linguistics. A. Davies and C. Elder. Oxford, Blackwell. 17: 431-450.
- de Bot, K. (1998). The Psycholinguistics of Language Loss. Bilingualism and Migration. G. Extra and L. Verhoeven. Berlin/New York, Mouton de Gruyter: 345-361.
- de Bot, K. (2002) "Language Attrition: The Next Phase." Keynote address at International Conference on L1 Attrition, Amsterdam, August 2002.
- de Bot, K. (2007). Dynamic Systems Theory, Lifespan Development and Language Attrition. Language Attrition. Theoretical Perspectives. B. Köpke, M. S. Schmid, M. Keijzer and S. Dostert. Amsterdam, John Benjamins. 33: 53-68.

- de Bot, K. (2008). "Introduction: Second Language Development as a Dynamic Process." The Modern Language Journal **92**(2): 166-178.
- de Bot, K. and M. Clyne (1989). "Language Reversion Revisited." Studies in Second Language Acquisition **11**: 167-177.
- de Bot, K. and M. Clyne (1994). "A 16-year Longitudinal Study of Language Attrition In Dutch Immigrants in Australia." Journal of Multilingual and Multicultural Development **15**(1): 17-28.
- de Bot, K., P. Gommans, et al. (1991). L1 Loss in an L2 Environment: Dutch Immigrants in France. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 87-98.
- de Bot, K. and M. Hulsen (2002). Language Attrition: Tests, Self-Assessments and Perceptions. Portraits of the L2 User. V. Cook. Clevedon/Buffalo/Toronto/Sydney, Multilingual Matters: 253-274.
- de Bot, K. and T. Lintsen (1986). Foreign-Language Proficiency in the Elderly. Language Attrition in Progress. B. Weltens, K. de Bot and T. J. M. van Els. Dordrecht, Foris. **2**: 131-141.
- de Bot, K. and S. Makoni (2005). Language and Aging in Multilingual Contexts. Clevedon, Multilingual Matters Ltd.
- de Bot, K. and R. W. Schrauf, Eds. (2009). Language Development over the Lifespan. New York, Routledge.
- de Bot, K. and S. Stössel (1999). Finding residual lexical knowledge: The 'savings' approach to testing vocabulary. Paper presented at the American Association for Applied Linguistics Conference, Stanford, CT.
- de Bot, K. and B. Weltens (1995). "Foreign Language Attrition." Annual Review of Applied Linguistics **15**: 151-164.
- de Groot, A. (2002). Lexical Representation and Lexical Processing in the L2 User. Portraits of the L2 User. V. Cook. Clevedon, Multilingual Matters: 29-63.
- de Groot, A. and J. F. Kroll, Eds. (1997). Tutorials in Bilingualism. Psycholinguistic Perspectives. Mahwah, New Jersey, Lawrence Erlbaum.
- de Leeuw, E., M. S. Schmid, et al. (2010). "The Effects of Contact on Native Language Pronunciation in an L2 Migrant Setting." Bilingualism: Language and Cognition **13**(Special Issue 01): 33-40.
- DeKeyser, R. M. (2000). "The Robustness of Critical Period Effects in Second Language Acquisition." Studies in Second Language Acquisition **22**(04): 499-533.
- DeKeyser, R. M., I. Alfi-Shabtay, et al. (2010). "Cross-Linguistic Evidence for the Nature of Age Effects in Second Language Acquisition." Applied Psycholinguistics **31**(03): 413-438.
- Dewaele, J.-M. and A. Pavlenko. (2001-2003). "Web Questionnaire on Bilingualism and Emotion."
- Dolan, T. P. (1999). A Dictionary of Hiberno-English. Dublin, Gill and Macmillan.
- Dostert, S. (2009). Multilingualism, L1 Attrition and the Concept of 'Native Speaker'. Department of English Language and Linguistics. Düsseldorf, Heinrich-Heine Universität Düsseldorf. **PhD**: 292.
- Duden-Redaktion. (n/a, 26/7/2010). "<http://www.duden-suche.de>."
- Ecke, P. (2004). "Language Attrition and Theories of Forgetting: A Cross-Disciplinary Review." International Journal of Bilingualism **8**(3): 321-345.
- El Aissati, A. and A. Schaufeli (1999). Language Maintenance and Loss: Evidence from Language Perception and Production. Bilingualism and Migration. G. Extra and L. Verhoeven. Berlin New York, Mouton de Gruyter: 363-377.
- Ellis, N. C. (2002). "Frequency Effects in Language Processing. A Review with Implications for Theories of Implicit and Explicit Language Acquisition." SSLA **24**: 143-188.
- Ellis, R. (1994). The Study of Second Language Acquisition. Oxford, Oxford University Press.
- Ellis, R. (2000). The Study of Second Language Acquisition. Oxford, Oxford University Press.
- European Commission (2005). Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. A Common Agenda for Integration Framework for the Integration of Third-Country Nationals in the European Union.
- Extra, G. and L. Verhoeven, Eds. (1999). Bilingualism and Migration. Studies on Language Acquisition; 14. Berlin, New York, Mouton de Gruyter.
- Fase, W., K. Jaspaert, et al. (1992). Maintenance and Loss of Minority Languages. Introductory Remarks. Maintenance and Loss of Minority Languages. W. Fase, K. Jaspaert and S. Kroon. Amsterdam, John Benjamins: 3-13.
- Field, A. (2005). Discovering Statistics Using SPSS. London, Sage.
- Finkbeiner, M., J. Almeida, et al. (2006). "Lexical selection in bilingual speech production does not involve language suppression." Journal of Experimental Psychology: Learning, Memory, and Cognition **32**(5): 1075-1089.
- Fishman, J. (1972). Language Maintenance and Language Shift as a Field of Inquiry: Revisited. Language in Sociocultural Change. J. Fishman. Stanford, Stanford University Press: 76-132.
- Fishman, J. A. and R. L. Cooper (1969). "Alternative Measures of Bilingualism." Journal of Verbal Learning and Verbal Behaviour **8**(2): 276-282.
- Flege, J. E. and I. R. A. MacKay (2004). "Perceiving Vowels in a Second Language." Studies in Second Language Acquisition **26**: 1-34.

- Frenck-Mestre, C. (1993). "Use of Orthographic Redundancies and Word Identification Speed in Bilinguals." Journal of Psycholinguistic Research **22**(4): 397-410.
- García Mayo, M. d. P. and M. L. García Lecumberri, Eds. (2003). Age and the Acquisition of English as a Foreign Language. Second Language Acquisition. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters.
- Gardner, R. and P. MacIntyre (1992). "A Student's Contributions to Second Language Learning. Part 1: Cognitive Variables." Language Teaching **25**(211-220).
- Gass, S. M., C. Madden, et al. (1989). Introduction. Variation in Second Language Acquisition. Volume II: Psycholinguistic Issues. S. M. Gass, C. Madden, D. Preston and L. Selinker. Clevedon, Philadelphia, Multilingual Matters. **2**: 3-10.
- Gass, S. M. and L. Selinker (2008). Second Language Acquisition : an Introductory Course. New York London, Routledge.
- Ginsberg, R. B. (1986). Issues in the Analysis of Language Loss: Methodology of the Language Skills Attrition Project. Language Attrition in Progress. B. Weltens, K. de Bot and T. van Els. Dordrecht, Foris Publications: 19-36.
- Gollan, T. H., R. I. Montoya, et al. (2002). "Semantic and letter fluency in Spanish-English bilinguals." Neuropsychology **16**(4): 562-576.
- Goral, M. (2004). "First-Language Decline in Healthy Aging: Implications for Attrition in Bilingualism." Journal of Neurolinguistics **17**(1): 31-52.
- Goulet, P., Y. Joannette, et al. (1997). "Word fluency after a right-hemisphere lesion." Neuropsychologia **35**(12): 1565-1570.
- Green, D. W. (1998). "Mental Control of the Bilingual Lexico-Semantic System." Bilingualism: Language and Cognition **1**(02): 67-81.
- Green, D. W. (2003). Exploring the Convergence Hypothesis: Processing and Representational Implications. EuroSLA. Edinburgh.
- Green, D. W. (2003). Exploring the Convergence Hypothesis: Processing and Representational Implications. EuroSLA.
- Grosjean, F. (1989). "Neurolinguists, Beware! The Bilingual is not Two Monolinguals in One Person." Brain and Language **36**(1): 3-15.
- Grosjean, F. (1998). "Studying Bilinguals: Methodological and Conceptual Issues." Bilingualism: Language and Cognition **1**: 131-149.
- Grosjean, F. (2001). The Bilingual's Language Modes. One Mind, two Languages: Bilingual Language Processing. J. Nicol, Blackwell. **1**: 1-22.
- Grosjean, F. (2008). Studying Bilinguals. Oxford, Oxford University Press.
- Grotjahn, R. (2000). C-test. Routledge Encyclopedia of Language Teaching and Learning. M. Byram. London, Routledge: 158-159.
- Grotjahn, R., Ed. (2006). Der C-Test: Theorie, Empirie, Anwendungen/The C-Test: Theory, Empirical Research, Applications. Language Testing and Evaluation. Frankfurt am Main, Peter Lang.
- Grotjahn, R., C. Klein-Braley, et al. (2002). C-Tests: an Overview. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 93-114.
- Grotjahn, R. and B. Stemmer (2002). C-Tests and Language Processing. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 115-130.
- Gubrium, J. F. and J. A. Holstein (2002). Handbook of Interview Research: Context & Method Thousand Oaks, Calif. ; London : , Sage Publications.
- Gumperz, J. J. (1982). Discourse Strategies. Cambridge & New York, Cambridge University Press.
- Gürel, A. (2002). Linguistic Characteristics of Second Language Acquisition and First Language Attrition: Overt versus Null Pronouns. Montreal, Canada, McGill University. **Unpublished PhD dissertation**.
- Han, Z. (2004). Fossilization in Adult Second Language Acquisition. Clevedon, Multilingual Matters.
- Han, Z. and T. Odlin (2006). Chapter 1. Introduction. Studies in Fossilization in Second Language Acquisition. Z. Han and T. Odlin. Clevedon, Multilingual Matters. **14**: 1-20.
- Hansen, L., Ed. (1999). Second Language Attrition in Japanese Contexts. New York, Oxford University Press.
- Hansen, L. (2001). "Language Attrition: The Fate of the Start." Annual Review of Applied Linguistics **21**: 60-73.
- Hansen, L., Y. Umeda, et al. (2000). Forgotten but not Gone: The Use of the Savings Paradigm in the Recovery of L2 Japanese and L2 Korean Lexical Knowledge. Paper presented at the American Association for Applied Linguistics Conference, Vancouver, B.C.
- Harres, A. (1989). ""Being a Good German": A Case Study Analysis of Language Retention and Loss among German Migrants in North Queensland." Journal of Multilingual and Multicultural Development **10**(5): 383-399.

- Harris, R. J. and E. M. McGhee Nelson (1992). Bilingualism: Not the Exception Anymore. Cognitive Processing in Bilinguals. R. J. Harris. Amsterdam, London, New York, Tokyo, North-Holland/Elsevier Science Publishers: 3-14.
- Harrison, J. E., P. Buxton, et al. (2000). "Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability." British Journal of Clinical Psychology **39**: 181-191.
- Hasselmo, N. (1961). American Swedish, Harvard University. **Unpublished Ph.D. thesis**.
- Haugen, E. (1953). The Norwegian Language in America: A Study in Bilingual Behavior. Philadelphia, Bloomington, University of Philadelphia Press, Indiana University Press 1964.
- Hegde, M. N. and D. E. Hartman (1979). "Factors affecting Judgments of Fluency: II. Word Repetitions." Journal of Fluency Disorders **4**(1): 13-22.
- Herdina, P. and U. Jessner (2002). A Dynamic Model of Multilingualism. Perspectives of Change in Psycholinguistics. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters.
- Herwig, A. (2001). Plurilingual Lexical Organisation: Evidence from Lexical Processing in L1-L2-L3-L4 Translation. Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives. J. Cenoz, B. Hufstein and U. Jessner. Clevedon, Multilingual Matters.
- Hiller-Foti, I. (1985). "Erhebung über den Verlust der Muttersprache deutscher Immigranten in Sizilien." Rassegna Italiana di Linguistica Applicata **17**(1): 107-110.
- Hilton, H. (2007). The 'Pausological' Interface between Language Knowledge and Production Skill. EuroSLA 17. Newcastle, England.
- Holden, L. (2005). Keeping an open mind on intelligence. The Irish Times. Dublin: 1 (page 15).
- Hornby, A. S. (1995). Advanced Learner's Dictionary of Current English. Oxford, Oxford University Press.
- Howell, P. and S. Sackin (2001). "Function Word Repetitions Emerge When Speakers Are Operantly Conditioned to Reduce Frequency of Silent Pauses." Journal of Psycholinguistic Research **30**(5): 457-474.
- Hulsen, M. (2000). Language Loss and Language Processing. Three Generations of Dutch Migrants in New Zealand. Nijmegen, Katholieke Universiteit Nijmegen: 261.
- Hulsen, M., K. de Bot, et al. (2001). The Interaction between Language Shift and Language Processing: Some First Data on Three Generations of Dutch Immigrants in New Zealand. Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. Münster/New York/München/Berlin, Waxmann: 153-168.
- Hutz, M. (2004). Is there a Natural Process of Decay? A Longitudinal Study of Language Attrition. Language Attrition: Interdisciplinary perspectives on methodological issues. M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar. Amsterdam, John Benjamins: 189-206.
- Hyltenstam, K. (1992). Non-Native Features of Near-Native Speakers: On the Ultimate Attainment of Childhood L2 Learners. Cognitive Processing in Bilinguals. R. J. Harris. Amsterdam, London, New York, Tokyo, North-Holland. **83**: 351-368.
- Hyltenstam, K. and N. Abrahamsson (2000). "Who Can Become Native-like in a Second Language? All, Some, or None? On the maturational constraints controversy in second language acquisition." Studia Linguistica **54**(2): 150-166.
- Hyltenstam, K. and L. K. Obler (1989). Bilingualism Across the Lifespan: An Introduction. Bilingualism Across the Lifespan. Cambridge, Cambridge University Press: 1-12.
- Hyltenstam, K. and C. Stroud (1989). Bilingualism in Alzheimer's Dementia: Two Case Studies. Bilingualism Across the Life-Span. K. Hyltenstam and L. K. Obler. Cambridge, Cambridge University Press: 202-226.
- Hyltenstam, K. and C. Stroud (1993). Second Language Regression in Alzheimer's Dementia. Progression and Regression in Language. K. Hyltenstam and Å. Viberg. Cambridge, Cambridge University Press.
- Hyltenstam, K. and Å. Viberg, Eds. (1993). Progression and Regression in Language. Sociocultural, Neuropsychological, and Linguistic Perspectives. Cambridge, Cambridge University Press.
- Jakobson, R. (1941). Kindersprache, Aphasie und allgemeine Lautgesetze. Uppsala, Almqvist.
- Jaspaert, K. and S. Kroon (1989). "Social Determinants of Language Loss." Review of Applied Linguistics, I.T.L. **83/84**: 75-98.
- Jaspaert, K. and S. Kroon (1992). From the Typewriter of A.L. A Case study in Language Loss. Maintenance and Loss of Minority Languages. W. Fase, K. Jaspaert and S. Kroon. Amsterdam, John Benjamins: 138-147.
- Jaspaert, K., S. Kroon, et al. (1986). Points of Reference in First-Language Loss Research. Language Attrition in Progress. B. Weltens, K. de Bot and T. van Els. Dordrecht, Foris: 37-49.
- Jessner, U. (2003). A Dynamic Approach to Language Attrition in Multilingual Systems. Effects of the Second Language on the First. V. Cook. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters: 234-246.
- Jessner, U. (2008). "A DST Model of Multilingualism and the Role of Metalinguistic Awareness." The Modern Language Journal **92**(2): 270-283.

- Jia, G. and D. Aaronson (1999). Age Differences in Second Language Acquisition: the Dominant Language Switch and Maintenance Hypothesis. Proceedings of the 23rd Boston University Conference on Language Development. A. Greenhill, H. Littlefield and C. Tano. Somerville, MA, Cascadilla Press. **23**: 301-312.
- Jiménez Jiménez, A. F. (2004). A Sociocultural Approach to Language Attrition. First Language Attrition: Interdisciplinary Perspectives on Methodological Issues. M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar. Amsterdam/Philadelphia, John Benjamins. **28**: 61-80.
- Joanette, Y., P. Goulet, et al. (1988). "Impaired Word Naming in Right-Brain-Damaged Right-Handers: Error Types and Time-Course Analyses." Brain and Language **34**: 54-64.
- Jordens, P., K. de Bot, et al. (1989). "Linguistic Aspects of Regression in German Case Marking." Studies of Second Language Acquisition **11**: 179-204.
- Kadt, E. d. (2001). "You Still Speak German?" Teenage Language Skills in a German-speaking community. Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. T. Ammerlaan, M. Hulsen, H. Strating and K. Yagmur. Münster/New York/München/Berlin, Waxmann. **2**: 61-75.
- Karmiloff-Smith, A. (1986). "From Meta-Processes to Conscious Access: Evidence from Metalinguistic and Repair Data." Cognition **23**: 95-147.
- Karmiloff-Smith, A. (1992). Beyond Modularity: A Developmental Perspective on Cognitive Science. Cambridge, MA, MIT Press.
- Kaufman, D. H. (1991). First Language Attrition as a Creative Interplay Between Two Languages, State University of New York at Stony Brook: 203.
- Kecskes, I. and T. Papp (2003). How to Demonstrate the Conceptual Effect of L2 on L1? Methods and Techniques. Effects of the Second Language on the First. V. Cook. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters: 247-265.
- Keijzer, M. (2007). Last in First out? An Investigation of the Regression Hypothesis in Dutch Emigrants in Anglophone Canada. English Department. Amsterdam, Vrije Universiteit. **PhD**: 429.
- Keijzer, M. (2010). "The Regression Hypothesis as a Framework for First Language Attrition." Bilingualism: Language and Cognition **13**(Special Issue 01): 9-18.
- Klein-Braley, C. (1985). "A cloze-up on the C-Test: a study in the construct validation of authentic tests." Language Testing **2**(1): 76-104.
- Klein-Braley, C. (1997). "C-Tests in the Context of Reduced Redundancy Testing: An Appraisal." Language Testing **14**(1): 47-84.
- Klein-Braley, C. (2002). Psycholinguistics of C-Test Taking. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 131-142.
- Köpke, B. (2002). Activation Thresholds and Non-Pathological First Language Attrition. Advances in the Neurolinguistics of Bilingualism. Essays in Honor of Michel Paradis. F. Fabbro. Udine, Forum: 119-142.
- Köpke, B. (2004). Attrition is Not a Unitary Phenomenon. On Different Possible Outcomes of Language Contact Situations. Bilingual Socialization and Bilingual Language Acquisition: Proceedings from the Second International Symposium on Bilingualism. A. M. L. Suarez, F. Ramallo and X.-P. Rodriguez-Yanez. Vigo, Servicio de Publicaciones da Universidade de Vigo.: 1331-1347.
- Köpke, B. (2004). "Neurolinguistic Aspects of Attrition." Journal of Neurolinguistics **17**(1): 3-30.
- Köpke, B. (2007). Language Attrition at the Crossroads of Brain, Mind, and Society. Language Attrition: Theoretical Perspectives. B. Köpke, M. S. Schmid, M. Keijzer and S. Dostert. Amsterdam, John Benjamins. **33**: 9-37.
- Köpke, B. and J.-L. Nespoulos (2001). First Language Attrition in Production Skills and Metalinguistic Abilities in German-English and German-French Bilinguals. Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. R. W. Andersen, M. Hulsen, H. Strating and K. Yagmur. Münster/New York/München/Berlin: 221-234.
- Köpke, B. and M. S. Schmid (2004). Language Attrition: The Next Phase. Language Attrition: Interdisciplinary Perspectives on Methodological Issues. M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar. Amsterdam, John Benjamins. **28**: 1-43.
- Köpke, B., M. S. Schmid, et al., Eds. (2007). Language Attrition. Theoretical Perspectives. Studies in Bilingualism. Amsterdam/Philadelphia, John Benjamins.
- Kroll, J. F. and A. M. B. De Groot (1997). Lexical and Conceptual Memory in the Bilingual: Mapping Form to Meaning in Two Languages. Tutorials in Bilingualism: Psycholinguistic Perspectives. A. M. B. De Groot and J. F. Kroll. Mahwah, New Jersey, Lawrence Erlbaum: 169-190.
- Kroll, J. F. and E. Stewart (1994). "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations." Journal of Memory and Language **33**(2): 149-174.
- Lacy, M. A., P. A. Gore, et al. (1996). "Verbal Fluency Task Equivalence." The Clinical Neuropsychologist **10**(3): 305 - 308.



- Lambert, R. D. and B. F. Freed, Eds. (1982). The Loss of Language Skills. Rowley, Massachusetts, Newbury House.
- Lambert, W. E. (1955). "Measurement of the Linguistic Dominance of Bilinguals." The Journal of Abnormal and Social Psychology **50**(2): 197-200.
- Lambert, W. E., J. Havelka, et al. (1959). "Linguistic Manifestations of Bilingualism." The American Journal of Psychology **72**(1): 77-82.
- Langdon, H. W., E. H. Wiig, et al. (2005). "Dual-Dimension Naming-Speed and Language-Dominance Ratings by Bilingual Hispanic Adults." Bilingual Research Journal **29**(2): 319-336.
- Larsen-Freeman, D. (2006). "The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English." Applied Linguistics **27**(4): 590-619.
- Larsen-Freeman, D. (2007). "On the Complementarity of Chaos/Complexity Theory and Dynamic Systems Theory in Understanding the Second Language Acquisition Process." Bilingualism: Language and Cognition **10**(01): 35-37.
- Larsen-Freeman, D. (2009). "Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition." Applied Linguistics **30**(4): 579-589.
- Larsen-Freeman, D. and L. Cameron (2008). Complex Systems and Applied Linguistics. Oxford, Oxford University Press.
- Larsen-Freeman, D. and M. H. Long (1991). An Introduction to Second Language Acquisition Research. London New York, Longman.
- Leyen, I. A. (1984). Native Language Attrition: A Study of Vocabulary Decline. Austin, University of Texas at Austin: 214.
- Linck, J. A., J. F. Kroll, et al. (2009). "Losing Access to the Native Language while Immersed in a Second Language: Evidence for the Role of Inhibition in Second-Language Learning." Psychological Science **20**(12): 1507-1515.
- Littlewood, D. (2004). Second Language Learning. The Handbook of Applied Linguistics. A. Davies and C. Elder. Malden, MA, Blackwell: 501-524.
- Lupien, S., A. Roch Lecours, et al. (1994). "Basal Cortisol Levels and Cognitive Deficits in Human Aging." The Journal of Neuroscience **14**(5): 2893-2903.
- Mackey, W. F. (1962). "The Description of Bilingualism." Canadian Journal of Linguistics **7**: 51-85.
- MacWhinney, B. (1997). Second Language Acquisition and the Competition Model. Tutorials in Bilingualism. Psycholinguistic Perspectives. A. de Groot and J. F. Kroll. Mahwah, New Jersey, Lawrence Erlbaum: 113-142.
- MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ, Lawrence Erlbaum Associates.
- MacWhinney, B. (2009). The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Part 1: The CHAT Transcription Format. Mahwah, NJ, Lawrence Erlbaum Associates. **1**.
- MacWhinney, B. (2009). The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Volume 2: Transcription Format and Programs. Mahwah, NJ, Lawrence Erlbaum Associates. **2**.
- Mägiste, E. (1979). "The Competing Language Systems of the Multilingual: A Developmental Study of Decoding and Encoding Processes." Journal of Verbal Learning and Verbal Behavior **18**: 79-89.
- Major, R. C. (1992). "Losing English as a First Language." The Modern Language Journal **76**(2): 190-208.
- Major, R. C. (1993). Sociolinguistic Factors in Loss and Acquisition of Phonology. Progression and Regression in Language. Sociocultural, Neuropsychological, and Linguistic Perspectives. K. Hyltenstam and Å. Viberg. Cambridge, Cambridge University Press: 463-478.
- Major, R. C. (2002). The Phonology of the L2 User. Portraits of the L2 User. V. Cook. Clevedon, Multilingual Matters: 67-92.
- Malvern, D. D., B. J. Richards, et al. (2004). Lexical Diversity and Language Development: Quantification and Assessment. Houndmills, NH, Palgrave Macmillan.
- Marian, V. (2008). Bilingual Research Methods. An Introduction to Bilingualism. Principles and Processes. J. Altarriba and R. R. Heredia. New York, London, Taylor and Francis: 13-37.
- Marian, V. and M. Spivey (2003). "Bilingual and Monolingual Processing of Competing Lexical Items." Applied Psycholinguistics **24**(02): 173-193.
- Marian, V. and M. Spivey (2003). "Competing Activation in Bilingual Language Processing: Within- and Between-Language Competition." Bilingualism: Language and Cognition **6**(02): 97-115.
- Maschler, Y. (1994). "Metalinguaging and Discourse Markers in Bilingual Conversation." Language in Society **23**(3): 325-366.
- McCarthy, P. M. and S. Jarvis (2007). "vocd: A Theoretical and Empirical Evaluation." Language Testing **24**(4): 459-488.
- McCarthy, P. M. and S. Jarvis (2010) "MTLD, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment." Behavior Research Methods.
- McKee, G., D. Malvern, et al. (2000). "Measuring Vocabulary Diversity using Dedicated Software." Literary & Linguistic Computing **15**: 323-337.

- McLaughlin, B. (1995). "Aptitude from an Information-Processing Perspective." Language Testing **12**: 370-387.
- Meara, P. (2003). Modelling Vocabulary Loss. Eurosla 13, Edinburgh.
- Meara, P. (2004). "Modelling Vocabulary Loss." Applied Linguistics **25**(2): 137-155.
- Meara, P. (2005). LLAMA Language Aptitude Tests. Swansea, UK, Lognostics.
- Meara, P. (2005). LLAMA Language Aptitude Tests: The Manual. Swansea, UK, Lognostics.
- Meara, P., J. Milton, et al. (2001). Language Aptitude Tests. Newbury, Express Publishing.
- Michael, E. B. and T. H. Gollan (2005). Being and Becoming Bilingual. Individual Differences and Consequences for Language Production. Handbook of Bilingualism. Psycholinguistic Approaches. J. F. Kroll and A. de Groot. Oxford, Oxford University Press: 389-407.
- Mitchell, M. L. and J. M. Jolley (2009). Research Design Explained. Belmont, California; London, Thomson Wadsworth.
- Myers Scotton, C. and J. Jake (2000). "Four types of morphemes: Evidence from aphasia, codeswitching and second language acquisition." Linguistics **8**(6): 1053-1100.
- Nakamizu, E. (2003). "Language Contact between Portuguese and Japanese. Functions of Code-Switching in the Speech of Brazilians living in Japan." Bulletin of Portuguese/Japanese Studies **6**: 73-91.
- Ng, B. C. and G. Wigglesworth (2007). Bilingualism: An Advanced Resource Book. London, New York, Routledge.
- Nunan, D. (1992). Research Methods in Language Learning. Cambridge, Cambridge University Press.
- O'Rourke, B. (2002). Metalinguistic Knowledge in Instructed Second Language Acquisition: A Theoretical Model and its Pedagogical Application in Computer-Mediated Communication. Centre for Language and Communication Studies. Dublin, University of Dublin, Trinity College: 311.
- Obler, L. K. and N. R. Mahecha (1991). First Language Loss in Bilingual and Polyglot Aphasics. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 53-66.
- Okimura-Bichard, F. (1985). "Mother tongue maintenance and second language learning: A case of Japanese children." Language learning(35): 63-89.
- Olshtain, E. and M. Barzilay (1991). Lexical Retrieval Difficulties in Adult Language Attrition. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 139-150.
- Olshtain, E. and S. Blum-Kulka (1989). Happy Hebrish: Mixing and switching in American Israeli family interactions. Variations in Second Language Acquisition: Discourse and Pragmatics. S. M. Gass, S. Madden, D. R. Preston and L. Selinker. Philadelphia, Multilingual Matters: 37-59.
- Opitz, C. (2005). "L1 Attrition in Adult Bilinguals in the Context of L2 Acquisition." International Journal of Bilingualism **8**(3): 395-398.
- Opitz, C. (2008). L1 Attrition in Adult Bilinguals: The Impact of L2 Proficiency, Language Use and Linguistic Aptitude. AILA. Essen.
- Opitz, C. (2009). The Role of Linguistic Aptitude and Language Use for Adult L2 Acquisition and L1 Attrition. ISB7. Utrecht.
- Opitz, C. (2010). "L1 Attrition and L2 Acquisition: Global Language Proficiency and Language Dominance in Adult Bilinguals." EUROSLA Yearbook **10**: 248-281.
- Pallier, C., S. Dehaene, et al. (2003). "Brain Imaging of Language Plasticity in Adopted Adults: Can a Second Language Replace the First?" Cerebral Cortex **13**(2): 144-161.
- Pan, B. A. and J. B. Gleason (1986). "The Study of Language Loss: Models and Hypotheses for an Emerging Discipline." Applied Psycholinguistics **7**: 193-206.
- Paradis, M. (1993). "Linguistic, Psycholinguistic, and Neurolinguistic Aspects of "Interference" in Bilingual Speakers: The Activation Threshold Hypothesis." International Journal of Psycholinguistics **9**(2): 133-145.
- Paradis, M. (1994). Neurolinguistic Aspects of Implicit and Explicit Memory: Implications for Bilingualism and SLA. Implicit and explicit language learning. N. C. Ellis. London, Academic Press: 393-419.
- Paradis, M. (2001). An Integrated Neurolinguistic Theory of Bilingualism (1976-2000). LACUS Forum XXVII. Speaking and Comprehending. R. M. Brend, A. K. Melby and A. R. Lommel: 5-15.
- Paradis, M. (2007). L1 Attrition Features Predicted by a Neurolinguistic Theory of Bilingualism. Language Attrition. Theoretical Perspectives. B. Köpke, M. S. Schmid, M. Keijzer and S. Dostert. Amsterdam, John Benjamins. **33**: 121-133.
- Pavlenko, A. (2003). 'I Feel Clumsy Speaking Russian': L2 influence on L1 in Narratives of Russian L2 Users of English. Effects of the Second Language on the First. V. Cook. Clevedon, Multilingual Matters: 32-61.
- Perdue, C., Ed. (1993). Adult Language Acquisition. Cross-Linguistic Perspectives: Field Methods. Cambridge, Cambridge University Press.
- Perdue, C., Ed. (1993). Adult Language Acquisition. Cross-Linguistic Perspectives: The Results. Cambridge, Cambridge University Press.

- Pérez-Vidal, C. and M. Juan-Garau (2009). The Effect of Study Abroad (SA) on Written Performance. Eurosla Yearbook. L. Roberts, D. Véronique, A. Nilsson and M. Tellier. Amsterdam, John Benjamins. **9**: 269–295.
- Poplack, S. (1980). "Sometimes I'll Start a Sentence in Spanish y Termino en Español": Toward a Typology of Code-Switching." Linguistics **18**(7/8): 581-618
- Raatz, U. (2002). C-Tests and Intelligence. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag. **31**: 169-178.
- Raatz, U. and C. Klein-Braley (1981). The C-Test - a Modification of the Cloze Procedure. Practice and Problems in Language Testing. T. Culhane, C. Klein-Braley and D. K. Stevenson. Essex: 113-148.
- Raatz, U. and C. Klein-Braley (2002). Introduction to Language Testing and to C-Tests. University Language Testing and the C-Test. J. A. Coleman, R. Grotjahn and U. Raatz. Bochum, AKS-Verlag: 75-91.
- Ribbert, A. and F. Kuiken (2010). "L2-induced changes in the L1 of Germans living in the Netherlands." Bilingualism: Language and Cognition **13**(Special Issue 01): 41-48.
- Roberts, P. M. and G. Le Dorze (1997). "Semantic Organization, Strategy Use, and Productivity in Bilingual Semantic Verbal Fluency." Brain and Language **59**(3): 412-449.
- Robinson, P. (2007). "Task Complexity, Theory of Mind, and Intentional Reasoning: Effects on L2 Speech Production, Interaction, Uptake and Perceptions of Task Difficulty." IRAL - International Review of Applied Linguistics in Language Teaching **45**(3): 193-213.
- Robinson, P. and R. Gilabert (2007). "Task Complexity, the Cognition Hypothesis and Second Language Learning and Performance." IRAL **45**(3): 161-176.
- Rohrer, D., J. T. Wixted, et al. (1995). "Retrieval from Semantic Memory and its Implications for Alzheimer's Disease." Journal of Experimental Psychology: Learning, Memory, and Cognition **21**(5): 1127-1139.
- Rosen, W. G. (1980). "Verbal Fluency in Aging and Dementia." Journal of Clinical Neuropsychology **2**: 135-146.
- Rosselli, M., A. Ardila, et al. (2000). "Verbal Fluency and Repetition Skills in Healthy Older Spanish-English Bilinguals." Applied Neuropsychology **7**(1): 17 - 24.
- Sabourin, L. (1988). Processus lexico-sémantiques et hemisphere droit: etude de la fluence verbale a partir de critres formels et semantiques. Département de Psychologie. Montréal, Université de Montréal. **MSc**.
- Schmid, M. and B. Köpke (2007). Bilingualism and Attrition. Language Attrition. Theoretical Perspectives. B. Köpke, M. S. Schmid, M. Keijzer and S. Dostert. Amsterdam, John Benjamins. **33**: 1-7.
- Schmid, M. S. (2001). Language Use and Language Loss of German-Jewish Refugees. Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages. T. Ammerlaan, M. Hulsen, H. Strating and K. Yagmur. Münster/New York/München/Berlin, Waxmann: 267-280.
- Schmid, M. S. (2004). The Language Attrition Test Battery. A Research Manual. Amsterdam: 61.
- Schmid, M. S. (2004). A New Blueprint for Language Attrition Research. Language Attrition: Interdisciplinary Perspectives on Methodological Issues. M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar. Amsterdam, John Benjamins. **28**: 348-363.
- Schmid, M. S. (2007). The Role of L1 Use for L1 Attrition. Language Attrition. Theoretical Perspectives. B. Köpke, M. S. Schmid, M. Keijzer and S. Dostert. Amsterdam/Philadelphia, Benjamins. **33**: 135-153.
- Schmid, M. S. (2009). L1 Attrition across the Lifespan. Language Development over the Lifespan. K. de Bot and R. W. Schrauf. New York, Routledge: 171-188.
- Schmid, M. S. (2010). "Languages at Play: The Relevance of L1 Attrition to the Study of Bilingualism." Bilingualism: Language and Cognition **13**(Special Issue 01): 1-7.
- Schmid, M. S. and K. de Bot (2004). Language Attrition. The Handbook of Applied Linguistics. A. Davies and C. Elder. Oxford, Blackwell. **17**: 210-234.
- Schmid, M. S., B. Köpke, et al., Eds. (2004). First Language Attrition: Interdisciplinary Perspectives on Methodological Issues. Studies in Bilingualism. Amsterdam/Philadelphia, John Benjamins.
- Schmitt, E. (2000). The Lost Word: Language Attrition among Children. Proceedings of the 24th Annual Boston University Conference on Language Development. C. S. Howell, S. A. Fish and T. Keith-Lucas. Somerville, MA, Cascadilla Press. **2**: 645-656.
- Schoenmakers-Klein Gunnewiek, M. (1998). Taalverlies door taalcontact? Een onderzoek bij Portugese migranten. Tilburg, Tilburg University Press.
- Scholze-Stubenrecht, W. and J. Sykes, Eds. (1990). Duden-Oxford Großwörterbuch English: Englisch-Deutsch, Deutsch-Englisch. Mannheim, Wien, Zürich, Dudenredaktion, Oxford University Press.
- Segalowitz, N. (1991). "Does Advanced Skill in a Second Language Reduce Automaticity in the First Language?" Language Learning **41**(1): 59-83.
- Seliger, H. W. (1991). Language attrition, reduced redundancy, and creativity. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 227-240.

- Seliger, H. W. and R. M. Vago, Eds. (1991). First Language Attrition. Cambridge, Cambridge University Press.
- Seliger, H. W. and R. M. Vago (1991). The Study of First Language Attrition: An Overview. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 3-15.
- Sharwood Smith, M. (1983). On Explaining Language Loss. Language Development at the Crossroads. S. W. Felix and H. Wode. Tübingen, Gunter Narr: 49-89.
- Sharwood Smith, M. (1983). "On First Language Loss in the Second Language Acquirer: Problems of Transfer." Language Transfer in Language Learning: 222-231.
- Sharwood Smith, M. (1989). Crosslinguistic Influences in Language Loss. Bilingualism Across the Lifespan. K. Hyltenstam and L. K. Obler. Cambridge, Cambridge University Press: 185-201.
- Sharwood Smith, M. and P. van Buren (1991). First Language Attrition and the Parameter Setting Model. First Language Attrition. H. W. Seliger and R. M. Vago. Cambridge, Cambridge University Press: 17-30.
- Silverman, S. and N. B. Ratner (2002). "Measuring Lexical Diversity in Children who Stutter: Application of *vocd*." Journal of Fluency Disorders 27(4): 289-304.
- Singleton, D. (1989). Language Acquisition: The Age Factor. Clevedon, Multilingual Matters.
- Singleton, D. (2003). Critical Period or General Age Factor(s)? Age and the Acquisition of English as a Foreign Language. M. d. P. García Mayo and M. L. García Lecumberri. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters. 4: 3-22.
- Singleton, D. and D. Little (1991). "The second language lexicon: some evidence from university-level learners of French and German." Second Language Research 7(1): 61-81.
- Singleton, D. and L. Ryan (2004). Language Acquisition: The Age Factor. Clevedon, Buffalo, Toronto, Multilingual Matters.
- Skehan, P. (1998). A Cognitive Approach to Language Learning. Oxford, Oxford University Press.
- Skehan, P. (2009). "Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis." Applied Linguistics 30(4): 510-532.
- Soares, C. and F. Grosjean (1984). "Bilinguals in a Monolingual and a Bilingual Speech Mode: The Effect on Lexical Access." Memory and Cognition 12: 380-386.
- Sorace, A. (2000). Differential Effects of Attrition in the L1 Syntax of Near-native L2 speakers. Proceedings of the 24th Annual Boston University Conference on Language Development. C. S. Howell, S. A. Fish and T. Keith-Lucas. Somerville, MA, Cascadilla Press. 2: 719-725.
- Spolsky, B. (1995). "Prognostication and Language Aptitude Testing. 1925-62." Language Testing 12: 321-340.
- Spolsky, B., B. Sigurd, et al. (1968). "Preliminary Studies for the Development of Techniques for Testing Overall Second Language Proficiency." Language Learning Upshur, J. A. and J. Fata (eds.) **Problems in Foreign Language Testing**(Special Issue 3): 79-98.
- Takona, J. P. (2002). Educational Research: Principles and Practice. Lincoln, NE, Writers Club Press.
- Treffers-Daller, J. (1997). Variability in Code-switching Styles: Turkish-German Code-switching Patterns. Code-switching Worldwide. R. Jacobson. Berlin, Mouton de Gruyter: 177-197.
- Troyer, A. K., M. Moscovitch, et al. (1997). "Clustering and Switching as Two Components of Verbal Fluency: Evidence from Younger and Older Healthy Adults." Neuropsychology 11(1): 138-146.
- Ullman, M. (2003). A Neurocognitive Perspective on Second Language Acquisition and Processing: The Declarative/Procedural Model. Eurosla 13, Edinburgh, unpublished.
- Valdés, G. and R. A. Figueroa (1994). Bilingualism and Testing: A Special Case of Bias. Norwood, NJ, Ablex.
- van Els, T. (1986). An Overview of European Research on Language Attrition. Language Attrition in Progress. B. Weltens, K. de Bot and T. van Els. Dordrecht, Foris Publications: 3-18.
- van Hout, R. and S. Strömquist (1993). 8.3 The Influence of Socio-Biographical Factors. Adult Language Acquisition. Cross-linguistic perspectives. C. Perdue. Cambridge, Cambridge University Press. 1: 164-172.
- van Ness, S. (1988). Pennsylvania German in West Virginia: Language Variation and Language Attrition, State University of New York at Albany: 254.
- Ventureyra, V. A. G., C. Pallier, et al. (2004). "The loss of first language phonetic perception in adopted Koreans." Journal of Neurolinguistics 17(1): 79-91.
- Victori, M. and E. Tragant (2003). Learner Strategies: A Cross-Sectional and Longitudinal Study of Primary and High-school EFL Learners. Age and the Acquisition of English as a Foreign Language. M. d. P. García Mayo and M. L. García Lecumberri. Clevedon, Buffalo, Toronto, Sydney, Multilingual Matters. 4: 182-209.
- Waas, M. (1996). Language Attrition Downunder. Frankfurt, Peter Lang.
- Wahrig, G., U. Hermann, et al., Eds. (1986). Deutsches Wörterbuch. Gütersloh/München, Bertelsmann Lexikon Verlag.
- Weinreich, U. (1953). Languages in Contact. New York, Linguistic Circle of New York.

- Weltens, B., K. de Bot, et al., Eds. (1986). Language Attrition in Progress. Studies on Language Acquisition; 2. Dordrecht, Foris.
- White, L. and F. Genesee (1996). How Native is Near-Native? The Issue of Ultimate Attainment in Adult Second Language Acquisition. Second Language Research, **12**: 233-265.
- White Lennon, B. (2006). Best of Irish Festive Cooking. O'Brien Press.
- Wray, A., K. Trott, et al. (1998). Projects in Linguistics. A Practical Guide to Researching Language. London, Arnold.
- Yağmur, K. (1997). First Language Attrition among Turkish Speakers in Sydney. Tilburg, Tilburg University Press.
- Yağmur, K., K. De Bot, et al. (1999). "Language Attrition, Language Shift and Ethnolinguistic Vitality of Turkish in Australia." Journal of Multilingual and Multicultural Development **20**(1): 51 - 69.
- Yeni-Komshian, G. H., J. E. Flege, et al. (2000). "Pronunciation Proficiency in the First and Second Languages of Korean-English Bilinguals." Bilingualism: Language and Cognition **3**(2): 131-149.
- Yukawa, E. (1997). Language Attrition from the Psycholinguistic Perspective: A Literature Review. Rapporter om tvåspråkighet. Stockholm, Centre for Research on Bilingualism, Stockholm University: 66.