

Measuring Variation in Perception of Acceptability: A Magnitude Estimation Investigation of Netherlands and Belgian Dutch

Laurie A. Stowe^a, Robert Hartsuiker^b, Magda Devos^b, Jack Hoeksema^a

^aUniversity of Groningen, ^bGhent University

Abstract We report a study of the perceived acceptability of three constructions with variable forms: the form of the adjectives in definite neuter noun phrases, the production of interruptions of the verbal cluster and the position of verbal particles in verb clusters. These constructions differ in frequency between northern standard Dutch used in the Netherlands and the standard variant used in Belgium. The perception of acceptability followed the frequency of production.

1. Introduction

In this paper, the perception of syntactic variation along the North-South axis in Dutch was investigated. Much has been written on North-South variation, using corpora (e.g. Nieuwborg 1968, De Cubber 1973, Grondelaers, Speelman and Carbonez 2001), collections of production data from tape recordings or elicitation (e.g. Pauwels, 1953; De Vriend-De Man, 1969; Stroop, 1970; De Schutter, 1974; Gerritsen, 1991; Barbiers, Bennis, De Vogelaer, Devos and Van der Ham, 2005). These studies thus focus on production of various structures and preference differences as reflected in usage. Relatively little work has been done on the issue of whether perception of these structures show parallel preferences by eliciting informant judgments, with the exception of Haeseryn (1990).

In the present paper, we use the technique of magnitude estimation to gain a more refined understanding of North-South differences in perception of acceptability in a number of areas that have been identified in the literature as areas of relatively significant syntactic variation, namely (1) adjective-noun agreement in noun phrases; (2) interruption by nonverbal material in verb clusters; and (3) splitting of particle verbs in verb clusters. Since we were primarily interested in the basic distinction between Southern (= Belgian) and Northern (=Netherlandic) Dutch, we avoided manipulating variables that would distinguish various regions within Belgium from one another, or regions within the Netherlands. Additionally since the Dutch spoken in the southern provinces of the Netherlands shares a number of characteristics with Belgian Dutch, we decided not to use participants from those provinces in our experiments, in order to keep the comparison simple and binary. Thus, whenever we speak of “North-South variation,” we actually mean the differences that are found between Belgian Dutch and the standard Dutch of the western, eastern and northern provinces of the Netherlands.

2. Syntactic Variation between Northern and Southern Dutch

Before turning to the experiments, it will be useful to take a brief look at the three syntactic variables that we chose to study.

2.1 Adjective noun agreement

Adjective-noun agreement in Dutch is simple on the morphological level. There are only two forms of the adjective, the short or uninflected form consisting of just the stem of the adjective, and the long or inflected form in which a schwa has been added to the stem (in words ending in a vowel, this schwa is automatically deleted). The syntactic conditioning of agreement is a bit more complex, since it depends on two factors: nominal gender and type of determiner. Standard Northern Dutch only has two genders, neuter and common gender (because the original Germanic feminine and masculine genders merged in the early modern period). Standard Southern Dutch maintains a distinction between masculine and feminine in its system of pronominal reference, but not in the inflectional system. Moreover, the gender distinction has completely disappeared in the plural, where all nouns are treated as common gender, as will be evident from the examples in (1) below. In Northern Dutch, singular neuter nouns with an indefinite determiner (such as *een* 'a, one' or *geen* 'no') are preceded by short forms, while all other combinations of noun and determiner require the long form of the adjective. Compare:

- | | | | |
|-----|----|------------------|---------------------------------------|
| (1) | a. | het oude huis | [singular; neuter; definite] |
| | | the old house | |
| | b. | een oud huis | [singular; neuter; indefinite] |
| | | an old house | |
| | c. | de oude man | [singular; common gender; definite] |
| | | the old man | |
| | d. | een oude man | [singular; common gender; indefinite] |
| | | an old man | |
| | e. | de oude huizen | [plural; definite] |
| | | the old houses | |
| | f. | geen oude huizen | [plural; indefinite] |
| | | no old houses | |

As noted in Haeseryn et al. (1997, section 6.4.1.3), southern Dutch sometimes employs the short form in definite neuter noun phrases such as

- (2) a. het bruin paard
the brown horse
b. het groot huis
the large house

These forms are judged to be ungrammatical in Northern Dutch, while they are apparently grammatical in Southern Dutch production. How acceptable they are in perception is another matter. In production the stem form occurs fairly infrequently.

2.2 Interruptions within the verbal cluster

The second and third constructions investigated here have to do with the order of elements in the verb cluster. The first concerns interruption effects in the verb cluster, in which non-verbal elements appear within the cluster. Interruption effects have been studied by a great many linguists. In Dutch, as in German, the main verb and its auxiliary verbs tend to form an uninterrupted sequence at the end of a subordinate clause, e.g. Evers (1975), Den Besten (1989), Haeseryn (1990). An example can be seen in (3); the verbal cluster is indicated by italics.

- (3) omdat de dames de brieven in hun tas *hadden gestopt*
because the ladies the letters in their bags *had put*
'because the ladies had put the letters in their bags'

Alternative word orders, in which nonverbal elements appear in the middle of the verb cluster occasionally appear, especially in the usage of Southern speakers (Koelmans 1965, Den Besten and Edmondson 1983, Haegeman and Van Riemsdijk 1986), a phenomenon referred to in the generative literature as 'Verb Projection Raising' and elsewhere as 'interruption of the verb cluster.' Evers (2003) writes in this connection of 'cluster creepers'. Compare:

- (4) omdat de dames de brieven *hadden* in hun tas *gestopt*
because the ladies the letters *had* in their bags *put*
'id'

Koelmans (1965) noted that nonverbal items that form an idiomatic unit (e.g. *keep an eye on* in English or its Dutch version *in het oog houden*) with the main verb are far more likely

to occur in the verb cluster than nonidiomatic nonverbal items. Koelmans' observation was based on corpus research; it is one goal of our study to see if first, the difference between Northern and Southern variants of Dutch is reflected in the acceptability judgments of native speakers as well and second if the effect of idiomaticity can be seen in perception and if so it affects the Southern group only.

2.3. Splitting of particle verbs

Closely related to the previous construction is the splitting of particle verbs. Standard Dutch (both northern and southern) typically allows two options for particle verbs in the verb cluster: the particle may either appear inside the cluster, typically adjoined to the main verb, or with another verbal element intervening, leading to a splitting of the verb-particle combination.

- (5) ...hoe het team de werkzaamheden zou *indelen*.
 ...how the team the tasks would apportion
 '...how the team would apportion the tasks.'
- (6) ...hoe het team de werkzaamheden *in* zou *delen*
 'id'
- (7) ...hoe het team de werkzaamheden *in* zou willen *delen*
 ...how the team the tasks would want-to apportion
 '...how the team would want to apportion the tasks'

Verb-particle combinations are actually special cases of the verbal idioms mentioned above in connection with interruptions of the verbal cluster. What is different about them is that particles are always acceptable as cluster creepers, both in the North and in the South. What we would like to know, however, is whether there are significant differences between sentences like (6) and sentences like (7) in terms of preference or relative acceptability. There are issues as to the relative acceptability of the possible positions of the particle (cf. Bennis 1992, Koopman & Szabolcsi 2000). Our primary concern in this study is whether there is a difference between the Northern and Southern participants with regard to the acceptability of the split version. De Cubber (1973) noted that non-adjacent position is less common in Southern Dutch. However, we also manipulate the distance between the particle and its verb by varying the size of the verbal cluster from two to three elements, which may serve to accentuate differences in preference between the split and non-split versions of the construction.

3. Investigating Variation in Perception

Let us emphasize at this point that the constructions which we are studying never differ to the extent that only one version of the construction is used in the South and only the other in the North. Both inflected and uninflected versions of the adjective are possible in the Southern variant, but the uninflected form is more common than in Northern Dutch. Likewise, interruption of the verbal cluster, particularly in idiomatic verb phrases, is more common in southern Dutch, although still not particularly frequent. Separation of the particle from its verb occurs fairly frequently in both variants, but is claimed to be more common in Northern Dutch (De Cubber, 1973). For this reason, binary grammaticality judgments are probably not sensitive enough to reveal the effects of interest: both will frequently be judged grammatical.

To investigate whether subtle differences in frequency in production are mirrored by differences in acceptability in perception, it is important to find a way to allow the informants to express their judgments even in cases where the perceived differences are small. N-point scales (e.g. acceptable from 1 to 20) such as a Likert-scale, may not be suitable for this, because they specify the number of distinctions in advance. Our experiments therefore used Magnitude Estimation (ME), a technique imported into linguistics from psychophysics (see Bard et al., 1996), in which participants give a relative judgment of each stimulus in comparison to a reference stimulus. Participants are usually free to choose an arbitrary value for the reference themselves. This technique has the advantage that the participants can add as many intermediate points as are necessary in contrast to a Likert scale. Importantly, the technique has been demonstrated to be sensitive to gradations in judgments of the acceptability of sentences, yielding data that is replicable and that is robust enough to show statistically significant differences (e.g., Bard et al., 1996; Keller & Alexopolou, 2001; Featherston, 2005a,b; see Sorace & Keller 2005 for an overview). As far as we are aware, our study is the first to use ME in order to compare acceptability judgments for speakers of two different variants of one language. We discuss group comparisons with this technique later.

4. Experimental Methods

4.1 Participants

The same Northern and Southern Dutch speakers participated in all three experiments described below, which were interleaved with each other and acted as distracters for each other. The group of Northern Dutch speakers consisted of 56 first year students at the

Faculty of Arts of the University of Groningen in Northeast Netherlands (50 F; 6 M; mean age = 21.4; s.d. = 4.71). For the Southern Dutch group, 56 students of psychology at the University of Gent in the northwest of Belgium (40 F; 16 M; mean age = 18.5; s.d. = 2.04) participated. These participants were also selected for inclusion from a larger group on the basis of being native speakers of Northern or Southern Dutch without extensive exposure to the other variant, based on a language experience questionnaire that they filled in. Some additional subjects were excluded to make the number of participants from each group on each list equal (see materials).

4.2 Materials

The experiments were interleaved in a single data collection and served as distracters for each other. For this reason, we will describe all three sets of materials together. Magnitude estimation encourages participants to judge relative acceptability, and it is important to include a number of very clearly ungrammatical items and very clearly acceptable items to calibrate the extremes of the scale. A description of fillers which served to achieve this will be provided at the end of this section, as will a description of the allocation of the materials and fillers to experimental lists.

4.2.1 Experiment 1: Adjective-Noun Agreement

We constructed 12 sentence pairs which contained a prepositional phrase with a definite neuter gender noun phrase, as in (8).⁶ This could appear with an adjective in the inflected form *-e*, which is standard in Northern Dutch, or without the inflection which is acceptable for a large number of adjective noun combinations in Southern Dutch. This variability between language variants is indicated by the question mark before the sentence. The sentences varied somewhat in structure, but had in common that the noun

⁶ *Inflection production pre-test.* A number of factors seem to affect the frequency of the uninflected form with definite neuter gender nouns in Southern Dutch, including the exact combinations of adjective and noun. The major goal of the current experiment was to investigate the degree to which the perception of acceptability mirrors the degree to which the forms are produced. For this reason, we chose to maximize the degree to which Southern speakers are likely to produce the actual sequences which we used in the experiment. Therefore, we conducted a pretest in which we presented a questionnaire with indefinite neuter gender noun phrases (e.g., *een belangrijk gesprek*), in which the uninflected form is correct, and asked participants to orally generate the definite form. They then filled in whether they had used the *-e* form or not. We used combinations for which at least 70% of our sample of 10 had generated the uninflected form. Note that this pretest, if anything, may underestimate the frequency of the uninflected form, since in a conscious generation task like this one the uninflected form might be underreported.

phrase containing the target adjective appeared in a prepositional phrase presented after the inflected verb.

- (8) a. Er werden tijdens het **belangrijke** gesprek een paar afspraken gemaakt
 b. ?Er werden tijdens het **belangrijk** gesprek een paar afspraken gemaakt
 Lit. There were during the important conversation a few agreements made
 'A few agreements were made during the important conversation.'
- (9) a. Er werden tijdens de **lange** vergadering een paar punten besproken
 b. *Er werden tijdens de **lang** vergadering een paar punten besproken
 lit. There were during the long meeting a couple-of points discussed
 'A couple of points were discussed during the long meeting.'

Participants may judge the uninflected forms to be grammatical because they do not notice the missing *-e*. To control for this, twelve sentence pairs as in (9) were constructed which contained a common gender rather than a neuter gender phrase. The control condition is also important in our between-groups comparison. The *-e* inflection is considered to be equally obligatory in these phrases in both Northern and Southern Dutch, which allowed us to assess the relative acceptability of the uninflected definite common gender form both within and across groups. Each *de* pair was matched in syntactic structure with a *het* pair in order to reduce differences in acceptability rating due to differences between constructions.

We predicted that for Northern speakers, there should be no difference in acceptability between the uninflected neuter- and common gender conditions (i.e. a main effect of inflection and no interaction with gender). For Southern speakers, the uninflected neuter gender condition should be more acceptable than the uninflected common gender condition (a significant interaction between gender and inflection).

4.2.2 Experiment 2: Nonverbal Interruptions in the Verb Cluster

The construction with nonverbal material inside the verbal cluster is archaic in Dutch, but particularly in idioms this word order type is still occasionally produced in Southern Dutch. In Northern Dutch, according to standard descriptions interruption by nonverbal material (with the exception of particles from particle verbs) is considered ungrammatical even in idiomatic phrases (Koelmans 1965, Haeseryn 1990). To test whether interruptions are relatively acceptable in idiomatic sentences, we constructed 12 sentence pairs containing

common idiomatic verb phrases (e.g., (10), such as *uit het oog verliezen* (lose track of or literally out of the eye lose).

- (10) a. Dat waren kwesties die de mannen *uit het oog* **hadden** *verloren*
 b. Dat waren kwesties die de mannen **hadden** *uit het oog* *verloren*
 Lit. Those were issues that the men had out the eye lost (lost track of)
 ‘Those were issues that the gentleman had lost track of.’
- (11) a. Dat waren brieven die de dames *in hun tas* **hadden** *gestopt*
 b. *Dat waren brieven die de dames **hadden** *in hun tas* *gestopt*
 lit Those were letters that the ladies had in their bags placed.
 ‘Those were letters that the ladies had put in their handbags.’

For each idiomatic item, a non-idiomatic sentence pair was constructed matched in sentence structure, like the example in (11), in order to test whether the asymmetry in acceptability between idiomatic and non-idiomatic sentences is also found in perceived acceptability. Based on Koelmans’ production data the version of these sentences with nonverbal interruptions are expected to be equally unacceptable regardless of idiomaticity for the Northern groups (main effect of interruption and no interaction with idiomaticity), but they are expected to be more acceptable for Southern dialect speakers, particularly in the case of the idiomatic sentences (interaction between idiomaticity and interruption).

4.2.3 Experiment 3: Splitting Particle Verbs

We constructed 24 sentence sets like those in (12) containing particle verb combinations like *indelen*, which means *apportion* or, syntactically more apposite, *carve up*. These sentences occurred in four versions: one version with the particle and verb together in clause-final position in an embedded clause (12a), another version with the particle fronted to a position before a single auxiliary or modal in the verb cluster (12b), and similar versions that contain two auxiliary and/or modal elements in the cluster (e.g. 12c and d). The occurrence of the particle both adjacent to the verb and separated from it occurs in both Northern Dutch and Southern Dutch, but the non-adjacent position appears to occur less frequently in Southern Dutch (De Cubber 1973). For the experimental sentences there was some variability in the structure of the main clause and transitivity of the embedded clause, but all sentences contained an embedded clause with a particle verb.

- (12) a. De chef vroeg hoe het team de werkzaamheden zou *indelen*.
 b. ?De chef vroeg hoe het team de werkzaamheden *in* zou *delen*
 Lit. The manager asked how the team the tasks would apportion
 ‘The manager asked how the team would apportion the tasks.’
 c. De chef vroeg hoe het team de werkzaamheden zou willen *indelen*
 d. ?De chef vroeg hoe het team de werkzaamheden *in* zou willen *delen*
 Lit. The manager asked how the team the tasks would want to apportion.
 ‘The manager asked how the team would want to apportion the tasks.’

In this case it was not possible to provide control conditions which both groups are hypothesized to find equally unacceptable as in the previous experiments, so our control condition is the form which both are hypothesized to find equally acceptable (the adjacent versions). If De Cubber’s (1973) production data is correct and is parallel to perceived acceptability, we expect to find that the Northern group finds all versions of these sentences equally acceptable. The Southern group on the other hand is expected to find both versions of the sentence with the particle not adjacent to the verb less acceptable. The variation in the complexity of the verbal cluster was included to probe the degree to which splitting the particle is accepted, since with a longer distance, unacceptability may be stronger.

4.2.4 Construction of Fillers and Lists

A number of sorts of filler sentences provided clear extreme values for the magnitude estimation scale, with the simple canonical intransitive and transitive sentences at the high end of the acceptability scale (14-15) and the missing-argument, extra-argument, and fronted-particle sentences at the low end of the scale (16-18). Sentences with and without *te* (19-20) were originally intended as part of an additional experiment, but due to design problems will not be discussed further here. Additionally, we used a set of sentences that provided two further points along the acceptability scale. To fill in the middle of the range, sentences were included containing a short noun phrase after a long adverbial phrase, in other words, sentences with an unacceptable ‘heavy NP shift’ (e.g., (21)). This theoretically provides a midway point between clearly grammatical and clearly ungrammatical. Twelve other sentences (e.g., (22) contained appropriate heavy NP shift (a long NP following a short adverbial). The number of items of each type of filler is provided following the examples given.

(14)	De puber heeft tegen zijn broer gejammerd	12
Lit.	The teenager had to his brother complained	
(15)	De tennisser heeft de finale bereikt	12
Lit.	The tennis player had the finals reached	
(16)	*De jongen heeft zijn moeder gejammerd	12
Lit.	The boy has his mother complained	
(17)	*De renner heeft in de wedstrijd bereikt	12
Lit.	The runner has in the contest reached	
(18)	* De oppas realiseerde dat weg het jongetje liep	12
Lit:	The babysitter realized that away the kid ran	
(19)	De man begon te werken	6
Lit:	The man began to work	
(20)	*De man begon werken	9
Lit:	The man began work	
(21)	?De jongen vond bij de school waar hij elke dag speelde het poesje	12
Lit.	The boy found next to the school where he every day played the kitten	
(22)	De man zag in de tuin een lange, slanke dame gekleed in witte zijde	12
	The man saw in the garden a tall slender woman dressed in white silk	

To prevent effects of having already seen and rated a nearly identical sentence, versions of the experimental sentences were allocated to the four lists using a Latin square design, so that subjects saw an equal number of each condition in an experiment (6) with no repeated items. When allocating experimental sentences to conditions on the lists, care was taken that the variable sentence structures were evenly spread across the conditions. All filler sentences appeared on each of the four lists in order to provide the same scale for magnitude estimation. Six blocks were made containing sentences of each type of filler sentence and one sentence of each condition of each experiment. Each block of sentences was then randomly ordered and the blocks were concatenated with each other to provide an order in which the sentences types were thoroughly mixed and were evenly spread across the list. A second version of each list was also created with the sentences in reversed order, giving eight versions of the experiment, each of which was rated by an equal number of the participants in the final analysis.⁷

⁷ The versions presented in Ghent and in Groningen were almost identical, except that one filler sentence contained the word *hesp* ('ham') in the Southern version, but *ham* ('ham') in the Northern version, and in another two filler sentences, the neuter determiner was used in the Southern version and the common gender determiner in the Northern version.

4.3 Procedure

The participants first read a written instruction concerning the tasks and then carried out two versions of the magnitude estimation task. In the first version they became acquainted with the concept of doing magnitude estimation on the basis of a short training using lines and estimating the difference in length relative to a reference line. They were then asked to carry out the same comparison but now on the basis of how much the sentences differ in acceptability from a reference sentence. The reference sentence contained an inappropriate heavy NP shift like the fillers which served as an intermediate point on the acceptability scale (i.e., example (21)). The participants gave a numerical weight to the reference sentence and then proceeded to estimate how far away from this reference other sentences were, on a scale including both less acceptable or more acceptable. The reference sentence was presented throughout the list after each group of ten sentences to help the participant to maintain the same relative scale, and subjects were instructed to give the reference the same rating each time they saw it. Participants who failed to do so were excluded from the analysis.

4.4 Data Analysis Common to All Three Experiments

Because the participants were free to generate their own scale of acceptability (some used 1-4000, some .001 to 1), we first normalized the data using Z-scores. In this calculation, a mean and standard deviation were calculated over the data from all four experiments for each subject, and each score was recoded as $(\text{Score} - \text{Mean})/\text{Standard deviation}$. By this means the extreme tails of each distribution become comparable, despite the actual numbers used for scoring.⁸

After recoding the data to Z-scores, the mean of the ratings for each condition were calculated for each subject and for each item. For each experiment, ANOVAs were carried out over the participant means (F_1) and item means (F_2). Effects will be regarded as significant if they are significant at alpha-level .05 in both the participant and item analysis, as both items and participants are random factors (Clark, 1973). For the sake of readability we do not report the full statistics when the F-ratio is below 1 and therefore

⁸ We used this procedure rather than the more common transformation to log scores (see Bard et al., 1996), because we found that the groups used sufficiently different scales that there were significant interactions between group and other factors for virtually any set of conditions when the log transformation procedure was used, even when the pattern of data was identical. Other experimenters have reported similar difficulties with between-groups comparisons (e.g., Featherstone, 2005b; Hopp, 2009). The Z-score transformation appears to be somewhat more reliable in this respect, although remnants of the scale differences can be seen in the figures below.

clearly not significant. In general, our interest is in the effects of the language variants used by the two groups, and we will therefore concentrate on interactions of other factors with group (although see footnote 3).

5. Results and Discussion

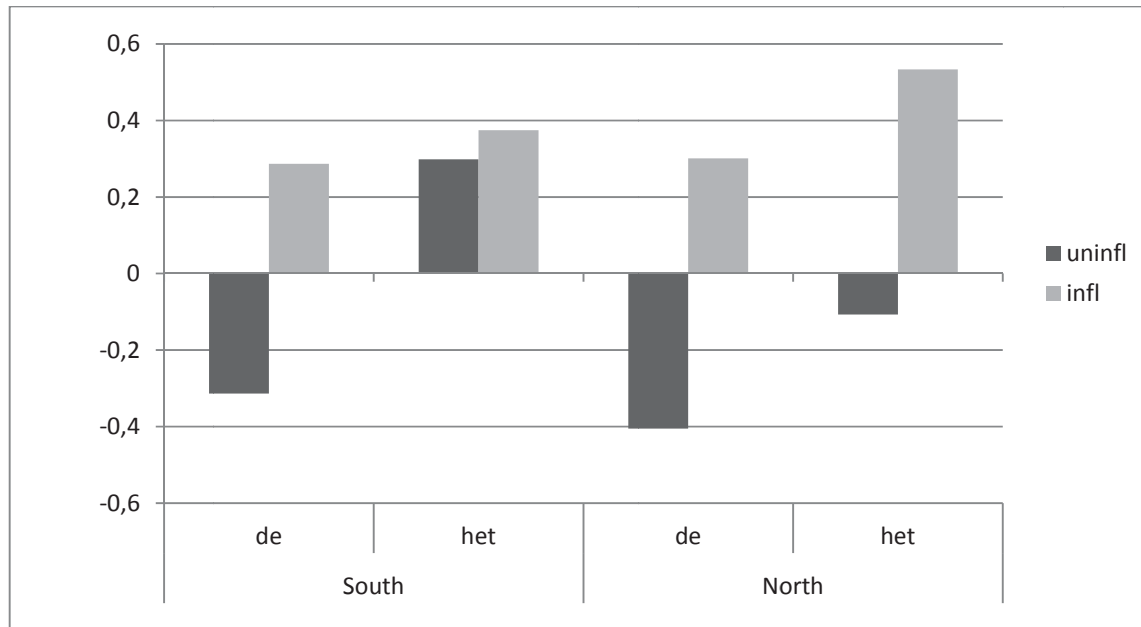
5.1 Experiment 1: Adjective-Noun Agreement: Results and Discussion

The results of the experiment in which we investigated the acceptability of uninflected adjective forms in neuter gender noun phrases are shown in Figure 1 below.

Both groups showed significant main effects of Gender [Southern: $F_1(1, 52) = 78.84$, $p < .001$; $F_2(1, 20) = 14.58$, $p = .001$; Northern: $F_1(1, 52) = 40.39$, $p < .001$; $F_2(1, 20) = 7.12$, $p < .05$] and Inflection [Southern: $F_1(1, 52) = 47.44$, $p < .001$; $F_2(1, 20) = 46.90$, $p < .001$; Northern: $F_1(1, 52) = 160.40$, $p < .001$; $F_2(1, 20) = 129.23$, $p < .001$]. However, only the Southern group showed a significant interaction between Gender and Inflection [Southern: $F_1(1, 52) = 31.41$, $p < .001$; $F_2(1, 20) = 28.76$, $p < .001$; Northern: F_1 and $F_2 < 1$]. An overall interaction between Group (Ghent/Groningen), Gender, and Inflection [$F_1(1, 104) = 13.76$, $p < .001$; $F_2(1, 20) = 8.48$, $p < .001$] suggests that the two groups do indeed differ in how they react to the inflectionless definite neuter.

Previous work (Haeseryn 1990) has suggested that the use of the uninflected form with neuter gender (*het*) words is relatively more frequent and acceptable in Southern than in Northern Dutch. The results of the current experiment clearly confirm this claim. The Southern participants find the uninflected form with *het* much more acceptable than with *de* ($p < .001$ for both subjects and items). Furthermore, the results suggest that the current generation of Southern speakers find the uninflected form approximately as acceptable as the inflected form showing virtually no difference in acceptability between the two variants (neither subjects or items significant). The uninflected form with common gender *de* is regarded as quite unacceptable relative to the inflected form by both Northern and Southern groups ($p < .001$ for both subjects and items), so the effect cannot be explained as readers missing the omitted *-e*.

Figure 1: Mean Z-score ratings for sentences containing inflected and uninflected adjectives in neuter gender (*het*) and common gender (*de*) definite noun phrases

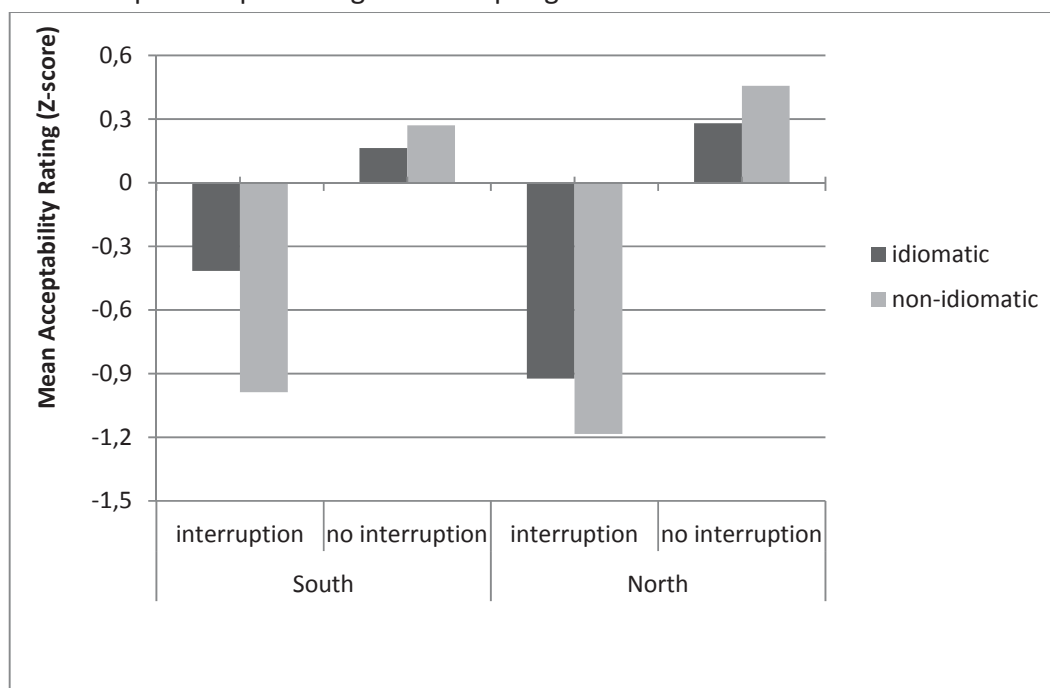


These results probably overestimate the acceptability of the omitted *-e*, since we chose lexical items that are produced frequently in this form. An interesting follow up to this study would be to systematically vary the production frequency for various combinations and see to what extent this correlates with perceived acceptability for each sequence. However, note that the inflected forms were judged to be equally acceptable. Since these were not the form which was frequently produced in the pre-test, they should have been judged to be less acceptable if the results are purely due to the frequency of individual combinations.

5.2 Experiment 2: Verb Cluster Interruptions: Results and Discussion

The results of the experiment with verb cluster interruptions in embedded clauses is shown in Figure 2 below.

Figure 2: Mean Z-score ratings for embedded clauses containing idiomatic and non-idiomatic phrases preceding or interrupting verbal clusters



The Southern group showed a significant Interruption x Idiomaticity interaction [$F_1(1, 52) = 53.54, p < .001$; $F_2(1, 20) = 17.76, p < .001$], as well as main effects of Interruption [$F_1(1, 52) = 189.70, p < .001$; $F_2(1, 20) = 133.14, p < .001$] and Idiomaticity [$F_1(1, 52) = 34.47, p < .001$; $F_2(1, 20) = 175.34, p < .05$]. The interaction reflects a significantly increased acceptability for the interruption in idiomatic verb phrases ($F_1(1, 52) = 74.055, p < .001$; $t_2(22) = 5.460, p < .001$), as well as a non-reliable tendency toward decreased acceptability for idiomatic sentences without intrusion ($F_1(1, 52) = 3.773, p < .058$; $t_2(22) = -711, p < .3$). The group from Groningen also showed a significant Interruption x Idiomaticity interaction [$F_1(1, 52) = 44.82, p < .001$; $F_2(1, 20) = 16.61, p = .001$] and a main effect of Interruption [$F_1(1, 52) = 822.52, p < .001$; $F_2(1, 20) = 696.79, p < .001$], but they showed no effect of idiomaticity [$F_1(1, 52) = 1.34, p > .25$; $F_2 < 1$]. As in the Southern group, idiomaticity modified the acceptability of sentences with intrusions ($F_1(1, 52) = 24.883, p < .001$; $t_2(22) = 6.023, p < .001$), and showed a tendency to do so as well for those without ($F_1(1, 52) = 15.43, p < .001$; $t_2(22) = -1.788, p = .088$). The two groups show essentially the same pattern, although it is somewhat more extreme for the Northern group.

In earlier forms of Dutch non-verbal material within the verbal cluster was grammatical regardless of idiomaticity, but it is currently normally regarded as

ungrammatical in Standard Netherlands Dutch and is at least heavily non-preferred in Southern Dutch. From corpus studies (e.g., Koelmans, 1965), it appears that interruptions of the verbal cluster in embedded clauses occur more commonly in Southern Dutch than in Northern Dutch, particularly in idioms, although it remains infrequent. Although the Northerners' ratings are more negative for both interruption conditions, it is difficult to be sure that this is meaningful. Overall, both groups show the same pattern, despite the interactions with Group. Additionally, the Northern group tends to show a more extreme scale in their judgments in general, even in the normalized Z-score ratings. This shows up for conditions which are almost certainly equally unacceptable in both variants, cf. uninflected adjectives in common gender noun phrases in Experiment 1 above.

What we can conclude is that the Northern group also finds the intrusion more acceptable for idiomatic phrases than for non-idiomatic phrases, just like the Southern group. Since interruptions have been reported to occur more frequently with idioms, we also predicted that acceptability might further be qualified by idiomaticity, particularly for the Southern group. As suggested by Koelmans (1965), idiomaticity modified the unacceptability of verbal intrusions. However, this increased acceptability is not limited to the Southern group, as shown by the interactions reported for both groups above. Interestingly, there is a trend in the data for idiomaticity to have the converse effect for both groups as well. The sentences containing idiomatic verb phrases were less acceptable than non-idiomatic sentences when the highly frequent continuous order was not employed, even though it is the frequent clearly grammatical order for this construction. This suggests that the drive toward continuity in idiomatic phrases is a strong one. However, the effects failed to reach significance in the item analysis which suggests that the effect may be limited to certain items.

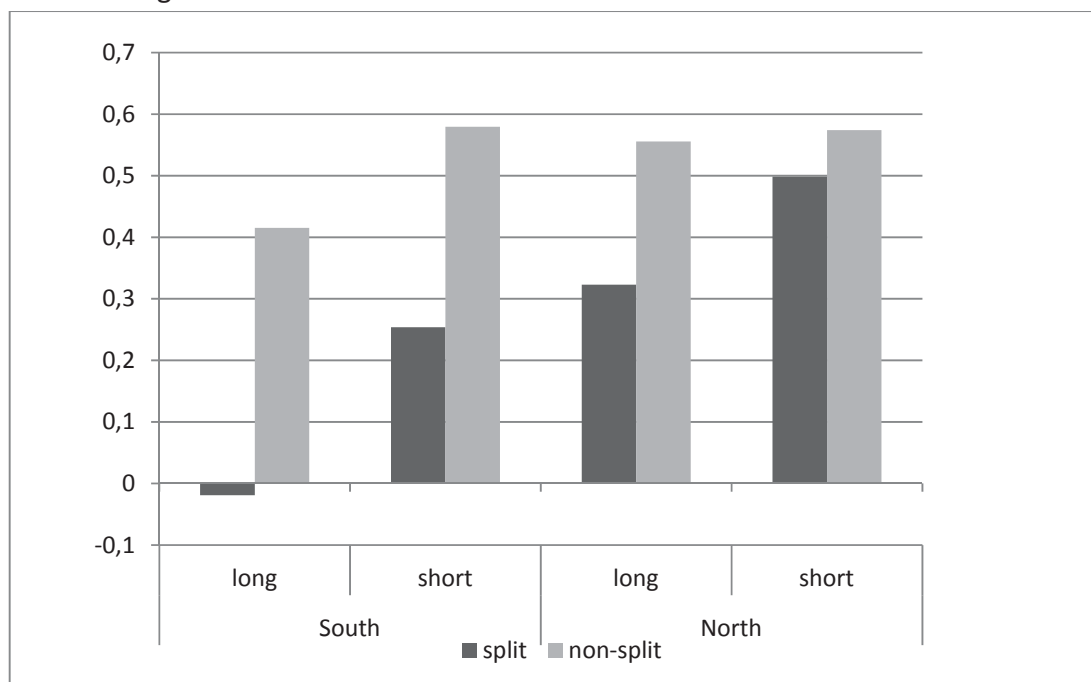
5.3 Experiment 3: Splitting Particle Verbs: Results and Discussion

Figure 3 displays the results of the third experiment, in which the acceptability of fronting a verbal particle to the position before auxiliary and modal verbs in embedded clauses was examined. We varied the number of verb cluster elements between the particle and verb as well, to further probe the degree to which splitting the verb and particle is less acceptable.

Analyses of the two groups separately do not illuminate the crucial interaction between Group and Particle Splitting [$F(1, 104) = 14.26$, $p < .001$; $F(1, 20) = 8.36$; $p < .01$], since both groups show similar patterns. However, taking the short versions alone, there is a clear interaction between Group and Splitting [$F(1, 104) = 19.860$, $p < .001$; $F(1, 20) = 4.606$, $p < .044$], which reflects the fact that only the Southern group shows an

effect of Particle Splitting when only one element intervenes between the verb and the particle [Southern: $F1(1,52) = 34.093$, $p < .001$; $F2(1, 20) = 9.826$, $p = .005$; Northern: $F1(1,52) = 2.162$, $p > 1$; $F2(1, 20) < 1$].

Figure 3 Mean Z-score ratings for sentences containing split and non-split particles, with short or long verbal clusters



Although non-split particle verbs are considered to be interruptions within the verb cluster, they are not considered less grammatical than the split variant in general. In fact, they seem to be the preferred form. This is confirmed by the main effect of splitting shown by both groups. It has been noted (De Cubber 1973) that splitting a particle from its verb within the verbal cluster is less common in Southern Dutch than in Northern Dutch. We investigated whether this pattern can also be seen in the perceived acceptability of these constructions. As predicted, the effect is more obvious for the Southern group than for the Northern group. This is confirmed by the significant interaction between Group and Particle Splitting reported above.⁹ The results demonstrate that in perception, as in production, the language variants differ. However, even though less common in the

⁹ This result cannot be attributed to the general tendency of the Northern group to assign a more extreme scale in their ratings, because in this case they are consistently more positive than the Southern group for the less preferred version. This interaction, then, can be accepted without further reservation.

south, they remain relatively acceptable. Note that most of the conditions in the table above are rated as comparable to the best alternatives in Experiments 1 and 2.

As an additional variable, we varied the length of the separation between the particle and its verb with an additional auxiliary or modal component. Additional distance diminished acceptability in both groups as can be seen in the significant interaction between Particle and Length [$F_1(1, 104) = 8.53, p < .001$; $F_2(1, 20) = 12.11, p < .01$]. This interaction bolsters the conclusion that in general the non-split version is preferred, even though the preference is less pronounced for the Northern speakers.

6. General Discussion and Conclusion

In this study, we used experimental methods to investigate three differences which have been claimed to exist between the Northern and southern Dutch language variants, the production of uninflected adjectives in definite neuter gender noun phrases, splitting of particle verbs, and nonverbal interruptions in the verbal cluster in embedded clauses. We have demonstrated that this sort of experiment can give insight into the relationship between frequency of various constructions in production and their perceived acceptability in comprehension.

The results support the hypothesis that when two constructions differ in frequency of production, they will also differ in perceived acceptability. For example, both Experiments 2 and 3 included a less acceptable variant, but one of those was relatively frequent for both variants and the other was very infrequent. This is clearly reflected in the acceptability ratings across the two groups. More importantly, we can indeed see that the two groups differ in how acceptable they find some forms. This is particularly clear in Experiments 1 and 3. The fact that we can see these differences between groups even in relatively subtle differences like the difference in particle splitting shows that this method has some scope for application in the future.

However future experiments need to be concerned with spurious interactions with group which may be produced due to group differences in the manner in which the magnitude estimation scale is applied. It is not clear why this is so problematic, but group differences have emerged in several other studies as well (Featherstone, 2005b; Hopp, 2009). For single group studies, the logarithmic scale is the most common manner of making the judgments comparable, but this leads to extreme interactions. Even use of Z-scores does not totally eliminate the issue. In future studies, it might be better to use a very extensive Likert-scale (e.g. 1 to 100; e.g. Hopp, 2009) with a number of examples for the extreme values and intervening values using fillers as in the current study, before initiating ratings for the target items. This may eliminate the group differences and make

the ratings more reliable. Replications of the current studies might confirm that the apparent group differences in Experiment 1 and 2 really exist. Despite these difficulties, elicitation of acceptability judgments seems suitable for further investigation of differences between language variants, particularly since it can be web-based, allowing large sets of data to be collected, adding to our arsenal of methods for investigating language variation.

7. References

- Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos, Margreet van der Ham. 2005. *Syntactic Atlas of the Dutch Dialects. Vol. 1: Pronouns, Agreement and Dependencies*. Amsterdam: Amsterdam University Press.
- Bard, Ellen Gurman, Robertson, Dan, & Sorace, Antonella (1996). Magnitude estimation of linguistic acceptability. *Language* 72(1), 32-68.
- Bennis, Hans (1992). 'Long Head Movement: The Position of Particles in the Verbal Cluster in Dutch.' In: *Linguistics in the Netherlands 1992*, ed. by R. Bok-Bennema and R. van Hout,, Amsterdam: John Benjamins, 37-47.
- Besten, Hans den (1989). *Studies in West-Germanic Syntax*. Doctoral dissertation, Catholic University of Brabant.
- Besten, Hans den and Jerrold A. Edmondson (1983). The Verbal Complex in Continental West Germanic. On the Formal Syntax of the Westgermania, ed. by Werner Abraham, Amsterdam/ Philadelphia: John Benjamins, 155-216.
- Clark, Herbert H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12 (X), 335-359.
- De Cubber, W. 1973, De splitsing van scheidbaar samengestelde werkwoorden in hedendaags proza, *Studia Germanica Gandensia* XIV, 39-63.
- Evers, Arnold (1975). *The transformational cycle in Dutch and German*. Bloomington: Indiana University Linguistics Club.
- Evers, Arnold (2003). Verbal Clusters and Cluster Creepers. In: *Verb Constructions in German and Dutch*, ed. by Pieter A.M. Seuren and Gerard Kempen, Amsterdam: John Benjamins, 43-89.
- Featherston, Sam. (2005a). Magnitude estimation and what it can do for your syntax: some wh-constraints in German. *Lingua* 115, 1525-1550.

- Featherston, S. (2005b). Universals and grammaticality: Wh-constraints in German and English. *Linguistics*, 43(4), 667-711.
- Gerritsen, Marinel (1991). Atlas van de Nederlandse dialectsyntaxis. Amsterdam: KNAW.
- Grondelaers, Stefan, Dirk Speelman and An Carbonez. 2001. Regionale variatie in de postverbale distributie van presentatief er.
- Haegeman, Liliane and Henk van Riemsdijk (1986) Verb Projection Raising, Scope, and the Typology of Rules Affecting Verbs. *Linguistic Inquiry* 17, 417-466.
- Haeseryn, Walter (1990). *Syntactische normen in het Nederlands. Een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep*. Dissertation, Catholic University of Nijmegen.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij and M.C. van den Toorn (1997). *Algemene Nederlandse Spraakkunst*, 2nd, revised edition, Groningen/Deurne: Martinus Nijhoff and Wolters/Plantyn.
- Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), 463-483.
- Keller, Frank, & Alexopoulou, Theodora (2001). Phonology competes with syntax: experimental evidence for the interaction of word order and accent placement in the realization of Information Structure. *Cognition*, 79, 301-372.
- Koelmans, L., 1965. Iets over de woordorde bij samengestelde predikaten in het Nederlands. *De Nieuwe Taalgids* 58, 156-165.
- Koopman, Hilda and Anna Szabolcsi (2000). *Verbal complexes*. Cambridge, MA: MIT-Press.
- Nieuwborg, E. (1968). *De distributie van het onderwerp en het lijdend voorwerp*. Antwerpen: Plantyn.
- Pauwels, A. (1953) De plaats van hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin. Leuven, Symons.
- Sorace, Antonella & Keller, Frank (2005). Gradience in linguistic data. *Lingua*, 115, 1497-1524.
- De Schutter, G. (1974). "Wezen vissen. Dialektgeografie van een constructie." *Taal en Tongval* 26, 70-85.
- Stroop, Jan (1970). 'Systeem in gesproken werkwoordsgroepen.' *Taal en Tongval* 22, 128-147. Reprinted in: J. Stroop, ed., *Nederlands dialectonderzoek*. Amsterdam: Huis aan de drie grachten, 1983, 247-264.
- De Vriendt-de Man, M.J. , 1969, *Frequentie van woorden en structuren in spontaan gesproken Nederlands*. Brussels: Didier.