

## **Intelligibility of written and spoken words in Germanic languages**

### **1. Research goal**

In this research project we want to assess how well secondary school pupils with various Germanic mother tongues are able to understand spoken and written words in Germanic languages that they are not familiar with. More specifically, we would like to answer the following research questions:

- Is there a difference between the intelligibility of words in their written and spoken form?
- What is the role of various linguistic factors (linguistic distance, word length, false friends, etc.) for word intelligibility?
- What is the relative intelligibility of native words compared to loan words?
- Is there a difference between informal, formal and neutral words?

### **2. Selection of words**

The point of departure in this research project is a database of formal and informal speech that we constructed in the spring of 2006. The informal speech was selected from the Corpus of Spoken Dutch and consists of conversations between friends and relatives in a homely atmosphere. The formal speech consists of Dutch and Swedish monologues in the European parliament. From both databases we took the 1500 most frequent words. As there was some overlap, the combined database included 2575 words. This database was translated into the following Germanic languages:

- Dutch
- Frisian
- High German
- Low German
- Danish
- Swedish
- Norwegian

In the database each word in each language was enriched with the following information:

- word class
- pronunciation
- origin (native word or loan word)
- in case of a loan word: language from which the word was borrowed, original language, and, whenever known, year of introduction into the language
- style (Corpus of Spoken Dutch = informal, European Parliament = formal, present in both databases = neutral)
- cognate (historically related with the corresponding words in the other languages) or non-cognate (not historically related)

For pragmatic reasons, the number of words to be tested in the present study had to be restricted. We decided to only use simple nouns (no proper names, no compounds). This

resulted in 815 words. The Dutch versions of these words were presented to Dutch secondary school pupils of 15, 16, 17 and 18 years of age. They were asked to indicate which words were unknown to them. 82 words were ticked by one or more participants. These were removed from the sample. From the remaining 733 words we made a random selection of 400 words. These are the test words in our study. The written versions of the words were available in the database of translations. To obtain the spoken versions, male, native speakers of the seven languages read the words in a professional audio studio.

### 3. Subjects

Subjects with the following mother tongues will participate in the study:

- **Danish** (from regions not adjacent to Germany or Sweden -> from the provinces of Fyn, Ribe, Vejle, Ringkøbing, Århus, Viborg, Nordjylland)
- **South Jutlandic** (from Danish Sønderjyllands Amt)
- **Norwegian** (from regions not adjacent to Sweden -> from the provinces of Møre og Romsdal, Aust-Agder, Vest-Agder, Rogaland, Hordaland, Sogn og Fjordane)
- **Swedish** (from regions not adjacent to Denmark or Norway -> from the provinces of Västernorrland, Gävleborg, Västmanland, Örebro, Östergötland, Jönköping, Kronoberg, Kalmar, Gotland, Blekinge, Stockholm, Uppsala, Södermanland, Västmanland)
- **Dutch** (from regions not adjacent to Germany -> from the provinces of Zeeland, Noord-Brabant, Utrecht, Flevoland, Zuid-Holland, Noord-holland)
- **Saxon** (from the Dutch provinces of Groningen, Drenthe, Overijssel)
- **Frisian** (from the Dutch province of Friesland)

The various subject groups should have comparable age and educational characteristics. We aim at pupils between 15 and 18 years of age taking a type of education allowing access to the university but without Latin and Greek as subject matter.

We are only interested in the performance of subjects who have (virtually) no knowledge of the test language. Therefore pupils of whom (one of) the parents have the test language as their mother tongue will be excluded from the data analysis (an exception is German, which is taught at schools in The Netherlands and Scandinavia). As an extra precaution, subjects will be asked to translate a number of non-cognates. Without knowledge of the test language these words should be unintelligible. Subjects with too many correct translations of non-cognates will also be removed from the analyses.

### 4. Task

The research is carried out via the Internet. Every subject is tested in one language. Each person is presented with 200 words, 100 in a spoken form and 100 in a written form. Half of the subjects start with the spoken words and the other half with the written words. This means that for each combination of languages (language of the subjects and language of the words presented) eight groups of subjects are needed to obtain data for all 400 words, both spoken and written, in the sample. The task is simple: subjects are asked to translate each word presented into their own language. The response time is the same for each

word, but the subject can also proceed to the following word by pressing the enter key. When the response time is finished, the next word is automatically presented.

To make it attractive for the pupils to participate in the test and to stimulate their efforts to do well, an extra reward will be given to the best-performing pupil in each group. Furthermore, all subjects make the chance of winning the price regardless of their performance. The internet program has been designed in such a way that the accuracy of the responses can be checked automatically and that the percentage of correct responses can be reported to the participant immediately after the test.

## 5. Design

The following combinations of subject groups and test languages will be tested:

Subject group	Test language						
	Danish	Norwegian	Swedish	Dutch	Frisian	High German	Low German
Danish		A	B	C	D	E	F
<b>Zuidjuts ???</b>						Q	R
Norwegian	G		H				
Swedish	I	J					
Dutch	K					L	M
Saxon						N	O
Frisian	P						

On the basis of these combinations we can answer the following questions:

- How well can Scandinavians understand each other's vocabulary? (A, B, G, H, I, J)
- What do Danes understand better: Low German or High German? (E, F)
- Do subjects with a South Jutlandic background have a better understanding of Low German than other Danes? (F, R). Possible comparison with High German (E, Q)
- Is it true that Danes and Frisian understand each other? (D, P) Comparison with Danes (C) and Danish by Dutch subjects (K)
- What do Dutch subjects understand better, Low German or High German? (L, M)
- Do subjects with a Saxon background have a better understanding of Low German than other Dutch subjects? (M, O). Possible comparison with High German (L, N)

For each combination of subject group and test language 40 subjects will be tested. As there are 18 combinations (A through P), in total 720 subjects will participate in this research project. The internet program automatically keeps track so that the subjects are neatly distributed over various test languages and test conditions.

A complete test includes the following elements:

1. Access page in English
2. Questions to determine from which country the subject originates
3. Short Introduction to the research
4. Personal questionnaire to select subjects afterwards
5. Instruction to test 1 (written or spoken part)
6. 5 test words
7. 100 words (written or spoken)
8. Instruction to test 2 (written or spoken part)
9. 5 test words
10. 100 words (written or spoken)
11. End: percentage of correct answers, thanks

The information in 2 and 4 is used to assign subjects to the correct test. All subjects are admitted to the test even if they do not meet the criteria (see section 3). A subject from the Dutch province of Limburg will be admitted, even though the province borders on Germany. A subject of 40 years of age will be admitted even though he is too old to participate. The responses of the subjects that do not meet the criteria will not be included in the data analysis.