

Varieties in contemporary Dutch: Combining the research possibilities offered by MIMORE and GrETEL

Liesbeth Augustinus

KU Leuven, Belgium
liesbeth@ccl.kuleuven.be

Ineke Schuurman

KU Leuven, Belgium
ineke@ccl.kuleuven.be

Sjef Barbiers

Meertens Institute Amsterdam, The Netherlands
sjef.barbiers@meertens.knaw.nl

MIMORE and GrETEL are two linguistic search engines. MIMORE enables linguists to query three related databases on morphosyntactic variation in Flemish and Dutch dialects: SAND, DiDDD, and GTRP. The data are based on interviews and elicitation. GrETEL enables linguists to consult syntactically annotated corpora (or treebanks) in a user-friendly way. For Dutch the CGN treebank (spoken Dutch, 1M words),¹ Lassy Small (written Dutch, 1M words) and SoNaR (500M words) are available. These data concern actual utterances in the ‘standard language’. But is this standard language the same in the entire Dutch-speaking region?

Some questions presenting themselves: Is there any evidence of a correlation with the dialect used in a specific area? Is there a stronger correlation with spoken language than with written language? Thus: can MIMORE be used to interpret findings in GrETEL? Are data found by means of elicitation (MIMORE) reflected in actual standard language use (GrETEL)?

In a combined educative CLARIN-NL use case on verb clusters we investigated those questions. In this case study, we investigated verb clusters of (1) a finite modal verb (e.g. *moeten* ‘must’), (2) a temporal auxiliary (e.g. *hebben* ‘have’) and (3) a main verb functioning as past participle (e.g. *gemaakt* ‘made’), as in:

De juf zegt dat hij zijn huiswerk morgen moet hebben gemaakt.

The teacher says that he his homework tomorrow must have made

‘The teacher says that he has to make his homework by tomorrow.’

¹ CGN covers both Dutch as spoken in Flanders (CGN-VL) and Dutch as spoken in the Netherlands (CGN-NL)

Depending on the dialect the order of the verbs in the sentence may differ, without changing the meaning of the sentence. Figure 1 shows the word order variation per region, based on data obtained from MIMORE, while Table 1 presents the figures of the variation found in GrETEL. In our presentation we will show how we obtained those figures and we will present a series of related findings.

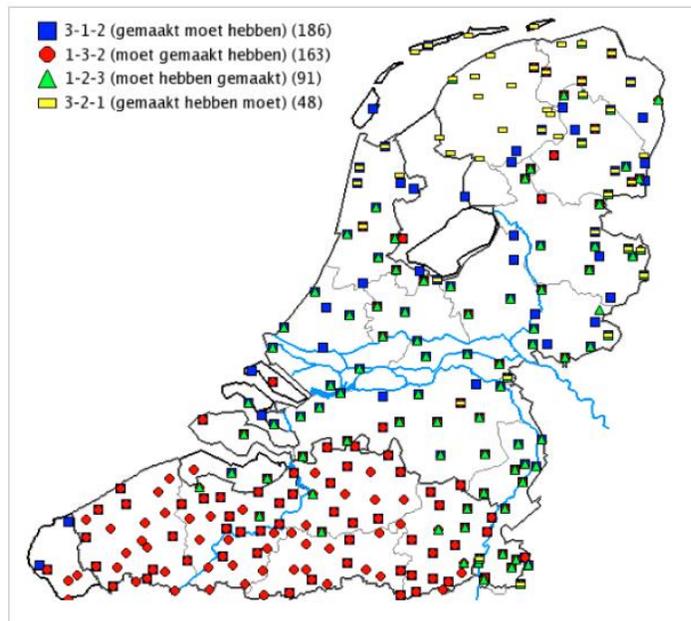


Figure 1: Frequency per region (MIMORE)

	CGN (NL + VL)	LASSY	SUM
1-2-3	19 + 7	66	92
1-3-2	0+3	2	5
3-1-2	14+3	33	50
3-2-1	0+0	0	0
2-1-3	0+0	0	0
2-3-1	0+0	0	0
SUM	33+13	101	147

Table 1: Frequency in the standard language (GrETEL)

References

MIMORE website: <http://www.meertens.knaw.nl/mimore/>

GrETEL website: <http://gretel.ccl.kuleuven.be>