

New resources for the study of Southern Dutch dialect syntax

The present paper reports (i) on the ongoing creation of a new resource to study the Southern Dutch dialects (East and West Flemish, French Flemish, Zeeland Flemish, Brabantian, and Limburgish), the “Parsed Corpus of the Southern Dutch Dialects” (project FWO18/KAN/009), and (ii) on the first results from using it to study two constructions in spontaneous speech.

Compared to other Germanic languages, the Southern Dutch dialects (as spoken in Belgium and bordering areas in France and the Netherlands) have been shown to have a number of striking typological characteristics some of which have been described and analysed in the literature, for instance, the occurrence of fronting without inversion as in (1) (e.g. Vanacker 1967, Saelens et al. 2016; Haegeman & Greco to 2018), or different kinds of discourse markers (e.g. Haegeman & Hill 2013), among which the former negative particle *en* seen in (2) (e.g. Breitbarth & Haegeman 2014, 2015).

Much of the more recent research on Southern Dutch dialects is either based on the big dialect atlases of Dutch, the FAND/MAND (phonological/morphological atlases; Goeman & Taeldeman 1996) and the SAND (syntactic atlas; Barbiers et al. 2005), or on introspective data (native speaker judgments). There are a number of problems with these methods regarding the study of structures such as (1) or (2). SAND, like FAND and MAND, is based on elicited data (questionnaires), which are ill-suited to reveal patterns requiring certain discourse contexts occurring in spontaneous speech, but are not easily elicited in constructed experimental settings. Similarly, the use of introspective data offers only a partial solution, in particular in light of the increasing dialect loss in Flanders (Ghyselen & Van Keymeulen 2014).

In fact, data such as (1), which appear to violate the V2 constraint, are underreported in the SAND. However, V2 violations with initial adjunct are relatively well attested in West Flemish (Saelens et al. 2016, Haegeman & Greco 2016, 2018). (1), taken from French Flemish, is special in that the initial constituent in the V2 violation is an argument, not an adjunct. The exact geographic spread of V2-violations with different kinds of constituents has so far not been studied in detail, and for the whole area where they may occur. (2) features the particle *en*, historically the preverbal (prefinite) negation particle, and a remnant of Jespersen’s Cycle, the development that established *niet* as the expression of sentential negation in all Dutch varieties. That *en* is retained in a number of Southern Dutch dialects is well-established (e.g. Haegeman 1995, Neuckermans 2008). Breitbarth & Haegeman (2014, 2015) have argued that the particle is used as a discourse marker with procedural meaning, expressing that a proposition is unexpected given the discourse background. Crucially, in (2), the particle appears in a non-negative context, and even more surprisingly, it precedes one of the non-finite verbs, not the finite verb as usual. In such cases, which have not been systematically studied so far, *en* only retains the procedural meaning signaling that the proposition is unexpected in the context. The exact spread of such constructions throughout the Southern Dutch dialects is as yet unknown, and has not yet been studied in spontaneous speech. The present paper presents our first findings regarding the spread of these constructions in our corpus.

The new Parsed Corpus of Southern Dutch Dialects is based on new transcriptions of a strategic selection of 35 tapes out of the 783 tape recordings of speakers born around 1900, from 550 places, made at Ghent University between 1963 and 1976 (cf. Vanacker & De Schutter 1967) and digitised in 2014 (www.dialectloket.be/geluid/stemmen-uit-het-verleden/). We report on the establishment of a two-tier transcription protocol, the POS-tagging, and the parsing pipeline. Compared to other dialect corpora, such as SAND, SyHD (Fleischer et al. 2015) or SADS (Glaser & Bart 2015), this resource is unique in being based exclusively on spontaneous speech, not questionnaire-based elicitations. Also, being based on recordings of speakers born around 1900 (the oldest in 1871), the corpus represents a historical stage of the dialects in an even stronger sense than the old adage that dialects are frozen language history. In case of French Flemish, it even represents the last witness of a now all but extinct language (Ryckeboer 2013).

Examples

- (1) *De nieuwe wagens me makten hé*
the new cars we made PRT
'The new cars, we made them.'
(recording of a speaker (°1901) from Bavinchove (French Flanders), 1965)
- (2) (about someone who fell into cow dung)
me zijn beste kleren aan ... g=ad dien e keer moeten en zien.
with his best clothes on you=had DEM a time must EN see
'With his best clothes on ... you should have seen this one!'
(recording of a speaker (°1903) from Pittem (West Flanders), 1970)

References

- Barbiers, S., H. Bennis, G. De Vogelaer, M. Devos & M. van der Ham. 2005. *Syntactic Atlas of the Dutch dialects*. Volume 1. Amsterdam: Amsterdam University Press.
- Breitbarth, A. & L. Haegeman. 2014. The distribution of preverbal *en* in (West) Flemish: syntactic and interpretive properties. *Lingua* 147: 69–86.
- Breitbarth, A. & L. Haegeman. 2015. 'En' *en is niet wat we dachten*: a Flemish discourse particle. *MIT Working Papers in Linguistics* 75: 85–102.
- Fleischer, J., A.N. Lenz & H. Weiß. 2015. Syntax hessischer Dialekte (SyHD). In R.Kehrein, A.Lameli & S.Rabanus (eds.), *Regionale Variation des Deutschen. Projekte und Perspektiven*, 261–287. Berlin: De Gruyter.
- Ghyselen, A.S. & J. Van Keymeulen. 2014. Dialectcompetentie en functionaliteit van het dialect in Vlaanderen anno 2013. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 130, 117–139.
- Glaser, E. & G. Bart 2015. Dialektsyntax des Schweizerdeutschen. In R.Kehrein, A.Lameli & S.Rabanus (eds.), *Regionale Variation des Deutschen. Projekte und Perspektiven*, 81–107. Berlin: De Gruyter.
- Goeman, A. & J. Taeldeman. 1996. Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal & Tongval* 48: 38–59.
- Haegeman, L. 1995. *The Syntax of Negation*. Cambridge: CUP.
- Haegeman, L. and C. Greco. 2016. V>2 in West Flemish. *Rethinking verb second: assessing the theory and data*. St John's college. University of Cambridge.
- Haegeman, L. & C. Greco. 2018. West Flemish V3 and the interaction of syntax and discourse. *Journal of Comparative Germanic Linguistics*.
- Haegeman, L. & V. Hill. 2013. The syntactization of discourse. In R.Folli, C.Sevdali & R.Truswell (eds.), *Syntax and its limits*, 370–390. Oxford: OUP.
- Neuckermans, A. 2008. *Negatie in de Vlaamse dialecten volgens de gegevens van de Syntactische Atlas van de Nederlandse Dialecten (SAND)*. PhD thesis, Ghent University.
- Ryckeboer, H. 2013. A West Flemish dialect as a minority language in the North of France. In F. Hinskens & J. Taeldeman (eds.), *Language and Space: An International Handbook of Linguistic Variation*. Vol. 3, Dutch: 782–800. Berlin: De Gruyter.
- Saelens, J. B. De Clerck, C. Lybaert & A. Willems. 2016. A corpus-based analysis of V2 variation in West Flemish and French Flemish dialects. Ms. Ghent University.
- Vanacker, V.F. 1967. Syntaktische Daten aus französisch-flämischen Tonbandaufnahmen. In *Verhandlungen des 2. Internationalen Dialektologenkongresses. Band II*, 844-855. (Zeitschrift für Mundartforschung. Beihefte, Neue Folge, Heft 4)
- Vanacker, V.F. & G. De Schutter. 1967. Zuidnederlandse dialecten op de band. *Taal en Tongval* 19: 35–51.