New ways of analyzing sociosyntactic variation

Background
Although sociosyntactic research is booming, it has repeatedly been observed that the Labovian method – which was originally conceived for the analysis of sociophonetic variation – cannot straightforwardly be extended to syntax (see a.o. Lavandera 1978, Romaine 1984, Cheshire 1987 and Winford 1996). The main objections against the extrapolation of the sociolinguistic variable-approach to syntax are fleshed out most explicitly in Romaine (1984):

- **the semantic equivalence condition** which is at the heart of the Labovian approach is easy to guarantee for phonetic variables, but nearly impossible to guarantee for syntactic variants, whose use is typically semantically or pragmatically conditioned.

- There is no **integrative (socio)linguistic theory** which can coherently deal with the semantic and the pragmatic factors which condition the choice for a specific syntactic variant. The reason for that is that most sociosyntactic research is Chomskyan in inspiration, as a result of which semantic and pragmatic factors are typically secondary considerations.

Workshop aim & structure
This workshop presents two case studies in support of a model of sociosyntactic analysis (Grondelaers & Speelman 2007; Grondelaers et al. 2009; Grondelaers 2009) whose central tenet is that pragmatic functions/processing strategies determine the choice between syntactic variants and motivate internal constraints which may, in turn, correlate with external constraints. We will focus on the theoretical underpinnings of this model – which cancels the traditional distinctions between sociolinguistics and theoretical syntax, and between sociolinguistics and psycholinguistics – but we will primarily zoom in on and demonstrate some of the methodological correlates of the approach:

- Corpus-based techniques to determine semantic equivalence, in particular *collostructional analysis*, a family of (easily applicable) techniques aimed at measuring the degree of attraction or repulsion that words exhibit to grammatical constructions (Stefanowitsch & Gries 2003);

- The operationalization of pragmatic/processing hypotheses into easily applicable linguistic constraints;

- The confirmation of the hypothesized pragmatic/processing functions with online experimental techniques (such as self-paced reading and eyetracking) and offline intuition gathering techniques;

- The use of logistic regression analysis to determine the impact of the internal and external constraints, as well as the quality of the model (in order to determine whether additional constraints have to be added to the model).

Following the presentations, the workshop also features a plenary discussion between the audience, the workshop conveners and invited discussant Sali Tagliamonte.
References


Papers

Dutch causatives and the semantic equivalence criterion

There is evidence that the choice of causative constructions with *doen* "make" and *laten* "let" in Dutch is governed by subtle semantic mechanisms as well as by sociolinguistic factors (in the broad sense). With respect to the semantic mechanisms that are involved, Kemmer and Verhagen (1994) propose the hypothesis that *doen* is the preferred form to express direct causation whereas *laten* is the form that is preferred when indirect causation is expressed. Follow-up research has provided additional empirical backup for this *(in)direct causation hypothesis* (a.o. Stukker 2005) even though some recent results suggest there is need for further refinements of the hypothesis (Speelman and Geeraerts 2009).

With respect to the sociolinguistic factors that are involved, Speelman and Geeraerts (2009) demonstrate that on top of the aforementioned semantic mechanisms, regional and register variation is at play: the *doen* variant has a stronger position in Belgian Dutch than in
Netherlandic Dutch and also has a stronger position in prepared speech than in spontaneous speech.

The absence of strict semantic equivalence can be argued to prevent us from treating the *doen-laten* alternation as a sociolinguistic variable in the traditional sense. So how then do we describe sociolinguistic variation in such a situation? Geeraerts (2010) conceptually distinguishes two approaches. The one is a *restrictive approach* in which you restrict your empirical dataset to a subset for which semantic equivalence does apply. The other is an *encompassing approach* in which you analyse the whole dataset and chart the interplay between both levels of variation. In this presentation we will discuss ways to put both approaches into practice, exploring the potential of, on the one hand, collostructional analysis and on the other hand mixed models to tease apart semantic and sociolinguistics variation in the use of causative verbs in Dutch, both in the context of the restrictive approach and in the context of the encompassing approach.

**References**


**Pragmatic motivation and constraint selection in a sociosyntactic analysis of existential construction choice**

This paper focuses on two unobserved but grave concerns in sociosyntax. By way of case study, we focus on the distribution of existential *er* “there” in Dutch locative inversion constructions such as *In de asbak lag (er) een hagelkorrel* “In the ashtray there was a hailstone”. Building on corpus and laboratory evidence, we have demonstrated that *er* is not an optional “dummy subject”, but an expectancy monitor inserted to facilitate the processing of unpredictable subjects such as *hailstone* (Grondelaers et al. 2009).

Observe to begin with that there is **no explicit constraint selection procedure** in variationist sociosyntax: constraints are typically added to a model on the basis of their significance in prior models, or on the basis of plausibility. We will argue that constraint selection should be theory-based and, whenever possible, function-based (see Grondelaers & Speelman 2007 and Grondelaers 2009). If *er*’s function is to facilitate the processing of unpredictable materials, then any context which determines the subject’s predictability is an internal constraint which
will improve our regression model of *er’s* distribution. In this sense, the internal constraints can be argued to be *motivated by* *er’s* function. Since functional motivation is the strongest constraint selection criterium, any sociosyntactician should investigate whether the variable under study is functionally motivated.

In much the same way, we propose that the syntactic variable under investigation should be modelled as exhaustively as possible. Sociosyntacticians seem to be satisfied when their variable can be shown to pattern with a number of internal and external main effects, but they rarely bother whether and to what extent these constraints interact, and *how much* variation they account for. A good model which explains most of the variation to be accounted for, however, not only returns more stable impact estimates (weights) for the constraints it contains. When syntactic variation can be successfully modelled, we have strong evidence that the function which is argued to motivate the constraints is the correct one. In the latter sense, variationist analysis is also theoretically relevant.

References

