

# Syntax of Dutch<sup>\*</sup>

## The data set

Hans Broekhuis

**Abstract:** This article is a rejoinder to the various reviews of the *Syntax of Dutch* (2012-2016) that have appeared in this and earlier volumes of *Nederlandse Taalkunde*. It focuses especially on one recurring theme in these reviews: the use of introspection for collecting data. Although many reviewers are of the opinion that data extracted from corpora are to be preferred, I will argue that such data are of limited use for the *Syntax of Dutch*, given that it involves competence rather than performance research: it aims at describing the internal structure of phrases and sentences and not the actual use of these structures.

**Key words:** reference grammar, competence, performance, introspection, corpus research.

## 1 Introduction

This article responds to several reviews of the *Syntax of Dutch* (henceforth: SoD) that have appeared in this and earlier volumes of *Nederlandse Taalkunde*. One recurring theme in these reviews is the data set offered in SoD, which has been received with mixed feelings by some reviewers. Although the general opinion seems to be that the empirical coverage of SoD is unparalleled by syntactic descriptions normally found in reference grammars, some reviewers nevertheless maintain that the data set does not fully come up to their expectations; especially the use of introspection in collecting data has met with objections. Since work on SoD will continue, it raises the question as to whether a revised version of SoD should be expanded by including data obtained by methods other than introspection, or whether the reviewers in question should look elsewhere in order to satisfy their specific needs. The conscious choice of introspection instead of corpus data is explicitly motivated in the preface of SoD (§4), where we discuss the delimitation of the object of description:

Our goal of describing the internal structure of phrases and sentences means that we focus on competence (the internalized grammar of native speakers), and not on performance (the actual use of language). This implies that we will make extensive use of constructed examples that are geared to the syntactic problem at hand, and that we will not systematically incorporate the findings of currently flourishing corpus/usage-based approaches to language: this will be done only insofar as this may shed light on matters concerning the internal structure of phrases.

---

<sup>\*</sup> I like to thank the editors of *Nederlandse Taalkunde* for their extensive comments on an earlier version of this article, as well as Frits Beukema for his willingness to correct my English. Obviously I am also greatly indebted to the authors of the reviews mentioned in this article.

Not surprisingly, the appreciation of this self-imposed restriction correlates with the reviewer's research interest and theoretical embedding: while it is considered "deplorable" by the corpus linguist and statistician Natalia Levshina (2016: §5), it is highly praised by the formal semanticist Hans Smessaert (2014: §3), to mention just two radically opposite positions.<sup>1</sup> This article will motivate the restriction in more detail by arguing that corpus data are of limited use for providing data pertaining to competence, as reflected by the speaker's unconscious knowledge of the core properties of the language system. It is important for the reader to keep in mind that core grammar refers to those aspects of the language system that arise spontaneously in the language-learning child by exposure to actual utterances, and that it stands in opposition to the so-called periphery, which refers to properties of the language that are often consciously learned by the speaker at some later age and that may be alien to the core system. Proverbs belong to this periphery because their meaning must be consciously learned, and it therefore need not surprise us that the verb *menen* in (1) occurs in a syntactic frame in which it cannot normally be used; the syntactic properties of this proverb are thus irrelevant for the syntactic description provided by SoD. The periphery furthermore includes specific properties of written and formal language, jargon, frozen expressions, historical relics, etc.

- (1)       Ieder       meent               zijn uil een valk te zijn.  
           Everyone is.of.the.opinion his owl a falcon to be  
           'Everyone believes his own to be the best.'

The discussion of the data set in SoD and the relevance of corpus research will address various issues raised in the reviews along the way. Some issues have been brought up in more than one review, more specifically those by Timothy Coleman, Helen de Hoop, Ernst Kotzé, Natalia Levshina, and Annelore Willems collected in this issue of *Nederlandse Taalkunde*, as well those by Maaïke Beliën and Freek Van de Velde in volume 19. Because of space limitations I cannot go into the details of each review and I therefore selected a number of representative cases from the contributions by De Hoop and Coleman in order to illustrate my position. Henk Verkuyl focuses on a purely theoretical issue concerning binary tense theory not directly related to the data collection in SoD, which I will therefore briefly address in a separate section.

## 2 The genesis of *Syntax of Dutch* (1992-2016)

The production of the present version of SoD has needed nearly 25 years. The idea for the SoD project was initiated in 1992 by Henk van Riemsdijk. In 1994 a pilot study was conducted at *Tilburg University*, and a steering committee was installed after a meeting with interested parties from Dutch and Belgian institutions. Unfortunately, the bilateral collaboration did not work out, as a result of which it took four more years before the

---

<sup>1</sup> Most of the reviewers objecting to the use of introspection for collecting data are "usage-based" linguists, who tend to deny the relevance of the distinction between competence and performance which is at the heart of competence linguistics; cf. Van de Velde (2014:89). This is not the place to discuss this chasm in detail but Section 5 will discuss various cases where, in my view, failing to make this distinction obscures the relevant syntactic generalizations.

project could actually start thanks to a substantial grant from the *Netherlands Organisation for Scientific Research* (NWO) obtained in 1998 and matching financing by *Tilburg University*. A writing group was formed consisting of Hans Broekhuis, Riet Vos and Marcel den Dikken. Den Dikken soon left the project for a position at the *City University of New York* and his work was continued by Evelien Keizer, who also left the project prematurely at the end of 2000 in order to take up a position at *University College London*.

The original plan was to write the full SoD in the period 1998-2001 but this turned out to be overly optimistic, although we managed to produce more or less final drafts of the AP-part (1999), the PP-part (2002) and the NP-part (2003), which were circulated on a small scale and then shelved for several years. Additional funding from the *Truus and Gerrit van Riemsdijk Stifting* enabled me to prepare the manuscripts for publication during 2008-2009 and to have them copy-edited in 2010-2012. The SoD project was later incorporated into the wider project *Language Portal Dutch/Frisian*, initiated by Hans Bennis and Geert Booij and funded by NWO in 2010-2015; during this period, Norbert Corver and I were able to write the missing part on verbs and verb phrases. In addition, the full SoD was converted into XML in order to make it available via the internet at *taalportaal.org*.

The survey above shows that, although the seven SoD volumes currently available were published by *Amsterdam University Press* between 2012 and 2016, the reader should be aware that the actual writing of these volumes took place in two phases: the precise periods for the four main parts are given below.

Production dates:

1. Nouns and noun phrases: 1998-2003
2. Adjectives and adjective phrases: 1998-1999
3. Adpositions and adpositional phrases: 1999-2002
4. Verbs and verb phrases: 2010-2015

Although the four parts together provide a full description of the main body of SoD, there are still various topics lacking which might be expected to be included in a comprehensive syntax of Dutch. In the main, these involve issues that could not easily be discussed within the overall macro-organization of the work; they will be discussed in a separate volume currently in preparation. Prominent examples are coordination and coordination reduction, which goes beyond sentence grammar in the strictest sense, because it may involve phrases of all types including sentences.

The genesis of SoD makes it clear that the present version of SoD should not be considered to be fully up-to-date, which was of course also indicated in the prefaces to the respective volumes. However, since prefaces tend to be skipped by readers, it is not surprising that some reviewers criticized SoD on account of ignoring specific, admittedly important, more recent publications: Van de Velde (2014) criticized the discussion of predeterminers in SoD-N9<sup>2</sup> because it did not address some of the

---

<sup>2</sup> This article will use the format *SoD-Xn* for references to SoD, where *X* can be V(erb), N(oun), A(djective) or P(reposition) and *n* refers to the relevant chapter of section in the relevant part: SoD-P3.1, for example, refers to section 3.1 of the SoD-part on adpositions.

diachronic issues discussed in Van de Velde (2009), while not taking into account that the published version of this chapter was more or less identical to the one he consulted in preparing his own work. Likewise, many of the references reported missing in Beliën's (2014) review of the PP-volume were simply not available at the time that this volume was written. Although one can regret the incompleteness of SoD on this score, it is simply unavoidable in a large-scale work of this type, which was written with limited resources, especially since it relates to a vital and highly productive field such as present-day formal syntax: it underlines the need of substantial structural funds for keeping reference works such as SoD up-to-date.

Having said this, it should be noted that the incompleteness of SoD is often less severe than suggested by the reviewers. For instance, Beliën (2014:§3-4) observes that SoD-P1.1.2.2 on the complementive use of PPs ignores certain intricate questions concerning the selection of temporal auxiliaries. However, this is not an accidental omission: because auxiliary selection is determined by the verb associated with it (and not by the syntactic function of PPs), the “wanting” information is given in SoD-V2.1.2. The reader should be aware that finding data in a sizeable grammar such as SoD is not always a trivial matter and requires some understanding of its overall organization as discussed in the preface of SoD (§5).

### **3 The object of description of *Syntax of Dutch***

The central concern of SoD is syntax in the strict sense: the study of how words are combined into larger phrases and, ultimately, sentences. The main body of SoD consists of four parts that focus on the four lexical categories (verbs, nouns, adjectives and adpositions) and their projections. Lexical categories have denotations and normally take arguments: nouns denote sets of entities, verbs denote states-of-affairs (activities, processes, etc.) that these entities may be involved in, adjectives denote properties of entities, and adpositions typically denote (temporal and spatial) relations between entities. The four lexical categories, of course, do not exhaust the set of word classes; there are also functional categories like complementizers, articles, numerals, and quantifiers. Such elements play a role in phrases headed by the lexical categories: articles, numerals and quantifiers are thus part of noun phrases and complementizers are part of clauses (that is, verb phrases). For this reason, functional elements are discussed in relation to the lexical categories.

As the reader of SoD will quickly notice, the focus on the internal structure of phrases does not preclude attention to other issues. Beliën (2014), for instance, expresses her surprise that SoD-P1 pays ample attention to the semantics of prepositions and prepositional phrases. This is warranted, however, by the fact that formal grammar takes syntactic structures to provide information about the relationship between forms and meanings: contrary to popular belief, semantics has played an important role in generative grammar at least since Fodor & Katz (1964). This does of course not imply that SoD should include semantics in its entirety, but at least some basic insights concerning the meaning of lexical items and phrases should be included. The same holds for certain notions concerning information structure: for example, there is reason to assume that the marked word order in (2) is not ungrammatical but unacceptable because it violates the tendency for phrases expressing discourse-old information to

precede modal adverbials. Information such as this is needed in order to appreciate the status of examples such as (2) in full; we ignore the fact here that the marked order becomes acceptable if the pronoun is assigned contrastive accent.

- (2) Jan heeft <hem> waarschijnlijk <\*hem> gezien.  
 Jan has him probably seen  
 ‘Jan has probably seen him.’

For similar reasons, information about language variation may be included in SoD. One phenomenon that has played an important role in the syntactic discussion on the northern and southern varieties of standard Dutch is the variation in relative word order within the verbal cluster and in the option of interspersing non-verbal material in these clusters. So, it is simply not true, as stated by Levshina (2016: §5), that SoD lacks “references to relevant works outside formal grammar” concerning variation, as evidenced by the extensive review of De Sutter (2005/2007) in SoD-V6/7. In my view, there is no principled reason for not including findings from diachronic, dialectal, typological and other types of linguistic research if they are relevant for issues discussed in SoD: for instance, Van de Velde’s (2009/2014) plea for including a discussion of the fact that predeterminers such as *al* and *heel* are on the decline in Dutch will certainly be acknowledged in a future version of SoD; similarly, incorporation of the results of the literature on extraposition mentioned in the final section of Willems (2016) is certainly an option. However, inclusion of information of this kind should be instrumental in the sense that it must contribute to a better understanding of the core issue addressed in SoD, that is, the syntactic description of the internal structure of phrases and sentences.

#### 4 The main goal of the *Syntax of Dutch*

The preface of SoD states that “the main objective of SoD is to present a synthesis of currently available syntactic knowledge of Dutch”, where syntax should be understood as indicated in the previous section. It further states that SoD aims at reviewing “the results of the formal linguistic research carried out over the last four or five decades that often cannot be found in the existing reference books” and emphasizes “that SoD is primarily concerned with language description and not with linguistic theory”. SoD aims at producing “a work of reference that is accessible to a large audience that has some training in linguistics and/or neighboring disciplines and that provides support to all researchers interested in matters relating to the syntax of Dutch”. I am happy to be able to say that most reviewers seem to agree that we did meet our main goal: the general feeling is that the empirical coverage of SoD is unparalleled by syntactic descriptions normally found in reference grammars, and that in general the discussions are accessible to linguists not specifically trained in formal linguistics.

It should be noted, however, that in my own view the present version of SoD does not fully succeed in presenting “a synthesis of *currently available* syntactic knowledge of Dutch”, because the formal linguistic literature simply turns out to be too extensive to be fully investigated with the limited means we have had to our disposal so far (about 16 man-years): there is still older material that we were not able to explore and (as was already indicated earlier) in the meantime a great deal of new material has become

available. This means that, although we were able to collect much material that “cannot be found in the existing reference books”, there is still a large amount of material waiting to be incorporated in an updated version of SoD.

Since the start of the SoD-project in the early 1990’s, the theoretical landscape in linguistics has changed considerably: attention has gradually shifted to performance which led to the current flourishing of corpus and usage-based grammars. This research has resulted in new data and insights that are sometimes also potentially relevant to syntax in the restricted sense intended here, and future versions of SoD may therefore profit from including such results. Section 5 will argue, however, that our hopes should not be too high given that competence and performance constitute two complementary linguistic research domains with different needs when data collection is at stake.

## 5 Introspection and corpus research

SoD is a competence grammar in the sense that it aims at describing the tacit knowledge a speaker of Dutch has of the syntactic structures in his language, and, in line with the generative tradition, the description is based on a data collection largely obtained by introspection. Various researchers have criticized SoD for using this method: this criticism is of course not exclusively directed at SoD as such but at formal linguistics more generally, as is clear from the fact that a substantial part of the data in SoD are taken over from the existing literature. Kotzé (2016) claims that the introspection method results “in what may be artificial or debatable exemplary material”. This section provides a reply to the implicit claim that corpus data are to be preferred across-the-board because they are not artificial or debatable, and argues that this position is rather naïve in that it reveals an unjustified trust in raw data, which, incidentally, can be observed more commonly in the literature based on corpus research.

### 5.1 Introspection data: artificial examples

Introspection research is done on the basis of constructed examples. Kotzé objects to this method of data collection because it may lead to artificial examples. There is no reason to deny this if the notion “artificial” is used to express that the examples in question are not spontaneously produced in context, but the question is whether this is objectionable. In my view this is not the case: SoD consciously aims at providing brief and maximally simple examples to illustrate the issues under discussion, in order to avoid interference of irrelevant factors. As Coleman (2016) correctly notes, competence research differs from other linguistic research in that there is only one thing that really counts, namely, whether or not a certain form is possible (with a certain meaning). The acceptability of transitive sentences such as *Jan kust Marie* ‘Jan is kissing Marie’, for instance, seems beyond dispute and performing corpus research in order to establish that such and similar transitive sentences can really be found would simply be a waste of valuable resources. Furthermore, corpus research is unable in principle to establish that a certain structure is impossible; see also the discussion in Section 5.3 below. It may of course be possible to establish unacceptability in an experimental setting but again in many cases this would be a waste of resources: that articles must precede nouns in Dutch (*de auto* ‘the car’ versus *\*auto de*) is again beyond

doubt. It should be stressed that in principle there is no objection against using data obtained by methods other than introspection for competence research, but this should be restricted to cases where using such methods has an added value.

## 5.2 Introspection data: debatable examples I (acceptability judgments)

It is unclear what Kotzé's notion of debatable example refers to. One potential interpretation of it may be related to the fact that researchers occasionally may have different acceptability judgments; this is exemplified by De Hoop's (2016) review of SoD-V13.2. De Hoop argues there that introspection is not a useful tool because intuitive judgments are partly theoretically biased. This claim is not supported by recent research: Sprout & Almeida (2010) formally tested a more or less random collection of judgment data on English (those found in Adger's text book *Core Syntax*) and their "results suggest that the maximum discrepancy between traditional methods and formal experimental methods is 2%". No doubt the discrepancy will be slightly higher in the case of an extensive reference grammar such as SoD because it discusses more complex and occasionally less-well studied examples, but I would be very surprised if it was much higher. I would like to add that although there are some exceptional cases where I suspect that there may be a theoretical bias in judgments, this is certainly not something that is common in the literature that I am familiar with: my estimate is that I agree with *at least* 98% of the Dutch data that I have seen so far in the syntactic literature that relies on introspection data *regardless of the theoretical orientation of the author*. This is completely in line with the conclusion in Sprout & Almeida (2010), but would be quite surprising if De Hoop's suggestion were correct.

It should further be mentioned that De Hoop misrepresents the discussion of the data in SoD. Since I cannot discuss all data, I will confine myself to the examples in (3) below, but the reader can verify himself that similar remarks can be made about the other examples cited by De Hoop. De Hoop claims that example (3b') is marked as ungrammatical in SoD. This is not true: the list of abbreviations in SoD states that asterisks mark examples as unacceptable.

- (3) a. Ik heb het aan Peter verteld. [speaker A]  
       I have it to Peter told  
       'I have told it to Peter.'
- b. Dan heb je waarschijnlijk de verkeerde ingelicht. [speaker B]  
           then have you probably the wrong.one prt.-informed  
           'Then you have probably informed the wrong person.'
- b'. \*Dan heb je de verkeerde waarschijnlijk ingelicht. [speaker B]  
           then have you the wrong.one probably prt.-informed

The distinction between grammaticality and acceptability is not a trivial one: *grammaticality* is a technical term that pertains to the question as to whether a certain example can or cannot be generated by the internalized grammar of the speaker, while *acceptability* is the term used for the speaker's judgments on linguistic objects, which may be prompted by his internalized grammar but may also be due to other (e.g. pragmatic) factors; see Newmeyer (1983: §2.2.1) for detailed discussion. Because SoD

aims at presenting the data made available by competence research but does not provide a formal model of the internalized grammar explaining these data for the reasons indicated in Section 4, grammaticality statements simply cannot be given, for which reason the notions *grammatical* and *ungrammatical* are rarely used in SoD. For instance, SoD-V13 mentions both notions only once; the notion *grammatical* is used on p.1601 in the sense that *any* grammar should be able to generate the example under discussion, and the notion *ungrammatical* is used on p.1612 in a discussion of predictions made by the flexible modification approach to A-scrambling.

De Hoop's misinterpretation is understandable in view of the fact that the asterisk is also used in the theoretical literature for indicating ungrammaticality, but the discussion of the examples in question leaves no doubt that acceptability judgments are intended: it is claimed that (3b) is "the neutral continuation of the discourse" started by speaker A, and that (3b') is possible with a contrastive accent on the noun phrase. This shows that the disagreement is less black and white than suggested by De Hoop in that the issue is not whether (3b') is grammatical or not, but whether it can be used as a neutral (non-contrastive) response to (3a): I claim that it cannot be used in this way, due to fact that A-scrambling affects the information structure of the clause, while De Hoop claims that it can, due to the fact that A-scrambling is essentially optional.

De Hoop correctly notes that it is not inconceivable that there is variation in speaker's judgments on A-Scrambling constructions. Consider the examples in (4) taken from Vikner (1994). These examples show that while speakers of Dutch normally reject word orders in which a direct object has scrambled across a nominal indirect object, this is fully acceptable for speakers of German; it would therefore not be surprising if speakers of the eastern varieties of Dutch more readily allow the German orders. Investigating this would indeed call for corpus research because it involves variational linguistics instead of competence research.

- (4) a. \*dat Peter het boek echt Marie <sub>t<sub>DO</sub></sub> getoond heeft. (Dutch)  
 b. dass Peter das Buch wirklich Maria <sub>t<sub>DO</sub></sub> gezeigt hat. (German)  
     that Peter the book really Marie shown has  
     'that Peter really showed Marie the book.'

De Hoop's suggestion that the judgments on the examples in question given in SoD are theoretically biased is clearly incorrect, given that these are mostly not of my own making but based on the existing literature; this also holds for the judgments on the (b)-examples in (3). The claim that A-scrambling affects the information structure of the clause, for example, is firmly rooted in the Dutch tradition that started in the late 1970's and culminated in Verhagen (1986), a work I value greatly but which is certainly not representative of my own theoretical orientation, and can in fact be extended to other West-Germanic languages like German, Afrikaans as well as Yiddish, as is clear from the review in Putnam (2007) and the references cited there. It is actually De Hoop's acceptability judgments that diverge from those found in the traditional literature without properly acknowledging this. This also holds for De Hoop's (2016, §2) judgments on the placement of neutral sentence accent ( $\approx$  the final non-contrastive main accent in the clause) in her examples (6) to (10). For instance, that neutral sentence accent may provide an important clue for determining the syntactic function of



constituents was already observed by, e.g., De Groot (1959:144) for predicative complements and adverbial phrases and by Gussenhoven (1992:87) for PP-complements and adverbial PPs. That A-scrambling affects the location of the sentence accent was furthermore observed in, e.g., Van den Berg (1978) and Verhagen (1986:§4.1.3.1) and, for German and Afrikaans, in Putnam (2007) and the references cited there.<sup>3</sup>

De Hoop's (2000/2003) claim that Dutch has truly optional A-scrambling of definite object-NPs is "empirically based" on a limited written corpus: the children's book *Otje* by Annie M.G. Schmidt. Given the target group (children from 5 to 12) and given that we are dealing with a quite specific written genre that may impose additional restrictions on language use, this corpus can hardly be considered representative of adult speech (also because it does not provide prosodic information); these corpus data are thus highly unreliable due to the interference of various unknown variables that may bias the results. This casts serious doubt on the validity of De Hoop's (2000/2003) conclusion that A-scrambling of definite object-NPs is truly optional, which is in fact highlighted by the fact that it diverges from the conclusion found in Van Bergen & De Swart (2010) that such NPs hardly ever scramble in adult speech, which is cited with apparent approval by De Hoop (2016, §6). The conclusion cannot but be that using "real language" data in linguistic research may likewise give rise to the problem of debatable examples.

Having said this, I do believe that corpus or experimental research may be of great help in the case of conflicting judgment data *provided it is sufficiently linguistically informed*. For example, there is good reason to believe that A-scrambling targets a specific well-defined position in the clause, namely the specifier of the functional head responsible for accusative case assignment; see Broekhuis (2008) and references cited there. This movement can be detected by various independently established facts based on introspection research, which are all neatly summarized in Verhagen (1986): A-scrambling involves movement across a relatively well-defined, semantically restricted set of adverbials and may have other side-effects such as the placement of sentence accent. By taking these effects into account, corpus research should *in principle* be able to establish whether or not A-scrambling of definite object-NPs affects the information structure of the clause but, unfortunately, the following section will show that such linguistically informed corpus research on A-scrambling simply does not yet exist; we can only hope that the review of A-scrambling found in SoD may help corpus linguists to set up their research in such a way that it also provides useful results for syntactic competence research.

---

<sup>3</sup> For more general theories on the relation between neutral sentence accent and syntactic structure, consistent with the Dutch, German and Afrikaans facts, we refer the reader to Gussenhoven (1992) and Cinque (1993). De Hoop's examples in (11) and (12) are not relevant for the present discussion because these involve monadic predicates, which were explicitly excluded from the discussion in SoD-VP13.1, sub III; I refer the reader to Baart (1987) and Gussenhoven (1992) for a discussion of the role of "new information" focus in determining the placement of sentence accent in such examples.

### 5.3 Introspection data: debatable examples II (frequency)

A second potential interpretation of the notion of debatable example would be that an example is debatable if it does not occur (frequently) in speech. Consider the two word orders in example (5); contrary to what is the case for definite object pronouns such as *hem* ‘him’ in (2), the syntactic literature reports that definite object-NPs either precede or follow modal adverbs. Assuming that the order in which the pronoun/NP precedes the modal adverb is derived by A-scrambling, this leads to the conclusion that this type of scrambling is obligatory with definite pronouns but not with definite NPs.

- (5) Jan heeft <de man> waarschijnlijk <de man> gezien.  
 Jan has the man probably seen  
 ‘Jan has probably seen the man.’

The previous section already referred to Van Bergen & De Swart’s (2010) claim, based on a sample extracted from the *Corpus Gesproken Nederlands* (Spoken Dutch Corpus), that in actual speech definite object-NPs such as *de man* in (5) scramble hardly at all. Does this mean that examples with a definite object-NP preceding a modal adverb are debatable? In my view, this would be an undesirable conclusion because all speakers of Dutch accept the scrambled order in (5), so something else must be going on. One reason for the discrepancy between the introspection data and the results reported by Van Bergen & De Swart may be that there are imperfections in their research design. There is reason indeed for assuming this: while the traditional literature claims that A-scrambling moves the object across adverbials of *certain semantic types* classified as comment modifiers by Verhagen (1986), Van Bergen & De Swart take *any* case in which an object follows an adverbial phrase to involve non-A-scrambling. This is clear from Section 3.1 of their article, where they state that they only exclude adverbial prepositional phrases and their pronominalized counterparts (e.g., *er ... in* ‘in it’) from their sample.<sup>4</sup> Now consider the examples in (6a&b), which are in fact the sole concrete (constructed) examples given in their article.

- (6) a. Sonja heeft gisteren de kaas opgegeten.  
 Sonja has yesterday the cheese prt.-eaten  
 ‘Sonja ate the cheese yesterday.’  
 b. Sonja heeft de kaas gisteren opgegeten.  
 Sonja has the cheese yesterday prt.-eaten  
 ‘Sonja ate the cheese yesterday.’

Contrary to what is suggested by Van Bergen & De Swart, the relative order of the temporal adverbial *gisteren* ‘yesterday’ and the object-NP in (6a) is not sufficient to show that the object did *not* A-scramble. This is clear from (7), where the object is

---

<sup>4</sup> The reason for this deviation may be that the corpus used for their study simply does not provide the information needed for identifying the relevant set of comment modifiers: I will return to this at the end of this section. Note that De Hoop (2016, §6) wrongly states that Van Bergen & De Swart investigate the relative order of direct objects and clause adverbs, and that she herself also fails to make the proper traditional delimitation of relevant adverbs in her articles discussed in section 5.2, as was already noted in SoD-N8 (p.1079).

scrambled into a position in between the temporal and a modal adverbial: this shows that A-scrambling does not necessarily lead to inverting the order of a temporal adverb such as *gisteren* and the object, and, consequently, (6a) may be of case of “invisible” A-scrambling as it does not cross any overtly realized material.<sup>5</sup>

- (7)      Sonja heeft gisteren    de kaas    waarschijnlijk opgegeten.  
           Sonja has yesterday the cheese probably prt.-eaten  
           ‘Sonja probably ate the cheese yesterday.’

It should be noted that concluding on the basis of (7) that (6a) may or may not involve A-scrambling is in fact the inverse of what Van Bergen & De Swart (2010:§3.1) do in excluding examples such as (7) from their sample because such sentences “could not be uniquely classified as scrambled or unscrambled”. It is rather remarkable for an article that aims at evaluating claims from the existing syntactic literature to investigate a sample that is not in line with the definition used in that literature, especially because this may considerably bias the statistical results: adopting the more traditional position would considerably reduce the number of non-A-scrambling cases by excluding examples such as (6a) as inconclusive and increase the number of A-scrambling cases by including examples such as (7).

For the sake of the argument, let us assume that a sample based on the more traditional definition of A-scrambling would also show that examples such as *Jan heeft de man waarschijnlijk gezien* ‘Jan has probably seen the man’, in which a definite object-NP (*de man*) precedes a comment modifier (*waarschijnlijk*), hardly ever occur in speech. This would still not justify the claim that A-scrambling of such object-NPs results in “debatable” examples, as there may be many plausible reasons for the lack of such cases in speech. We have already seen that the traditional literature suggests that scrambled objects refer to (non-contrastive) discourse-old information and it might be the case that in actual speech such information is preferably expressed by definite pronouns, while the use of definite noun phrases is reserved for new or otherwise salient information; cf. Du Bois (1987:816).

The discussion up to this point is intended to show that introspection provides a *different kind* of data than corpus research. While the former provide information about the acceptability of specific constructions, the latter provide information about their frequency in actual language use: because acceptable examples may fail to occur in actual speech for various reasons, corpus research is simply unable to provide proof that a specific example is not acceptable for speakers of the language. One easy way of determining this is simply by asking speakers for their acceptability judgments on the examples in question, that is, by appealing to their tacit knowledge of the language.

For competence research, Van Bergen & De Swart (2010) in fact reveals a much more serious problem with corpus research. That this study did not specifically address

---

<sup>5</sup> The results in Van Bergen & De Swart show that referential pronouns such as *hem* ‘him’ virtually obligatorily precede the adverbs in the larger set. This is in keeping with the fact that such pronouns are normally phonetically weak and that weak proforms are arguably moved into a structurally higher ( $\approx$  more leftward) position than A-scrambled object-NPs; see SoD-V13.4 for more detailed discussion.

the traditional claim that A-scrambling affects the information structure of the clause but instead performed a multifactorial analysis of their sample is related to the fact that the *Corpus Gesproken Nederlands* is enriched with tags mainly pertaining to syntactic category and does not contain tags relating to information structure.<sup>6</sup> This shows that the existing tag sets impose practical restrictions on what can and cannot be fruitfully investigated, which may make corpus linguistics of limited use for the evaluation of established linguistic insights based on introspection data. For example, Van Bergen & De Swart's research has shown that "the lower an object ranks in the definiteness hierarchy, the smaller its probability of occurring in scrambled position"; this simply confirms what we already knew on the basis of competence research and thus does not shed any new light on the traditional claim that A-scrambling of object-NPs depends on the information structure of the clause. The best we can say is that this conclusion is not incompatible with this claim. Another example: testing the traditional claim that A-scrambling crosses comment modifiers requires that the corpus provides semantic information about adverbials. Since this information is lacking in the *Corpus Gesproken Nederlands*, this corpus is unsuitable for the task at hand: Van Bergen & De Swart's "solution" of casting the net wider by including all (non-PP) adverbials simply introduces unwanted noise that makes the results unreliable. This shows again that relying exclusively on corpus data for competence research imposes undesirable -because scientifically irrelevant- restrictions on what can or cannot be successfully investigated: introspection data, on the other hand, do not impose such restrictions and therefore enable the competence researcher to activate his full potential of linguistic skills and creative power, and thus enhance scientific progress when it comes to competence research. In my view, this counterbalances Kotzé's objection that introspection research may lead to "debatable" examples.

#### 5.4 Incorporating corpus data can be harmful for competence research

Another drawback of appealing to introspection mentioned by Kotzé is that the "inherent variability of language, which lies at the foundation of diachronic change, is not considered or presented". This is clearly true but the question arises whether this is a valid argument against the use of introspection data, since using such data does not preclude the use of corpus data whenever that seems appropriate. There need not really be a debate on this issue, as is clear from the fact that corpus research is also at the heart of generative approaches to historical, typological and variational linguistics.<sup>7</sup> It will be

---

<sup>6</sup> See Van Eynde (2004) for a complete list of the tags used in the corpus; this tag set is intended to "connect with grammars for general use" such as Haeseryn et al. (1997). Some of the reviewers of this article object that this is not a principled objection to corpus linguistics because it is possible to extend the tag set. This is of course true but the crucial point to be made in the main text is that it is practically impossible to anticipate the actual needs of the researcher.

<sup>7</sup> That generative grammar has given rise to fruitful research programs in these fields shows that Van de Velde's (2014:89) argument based on language variation/change against the validity of the distinction between competence and performance (and in favor of the more holistic approach adopted in usage-based grammar) is a straw-man argument that does not do justice to the generative views on these issues; see Barbiers (2013) and Roberts (2007) for reviews of the generative literature on, respectively, language variation and language change.

clear by now, however, that I disagree with the view expressed by some reviewers that corpus data are to be preferred across-the-board and that I take the view that such alternative research is often of limited use when it comes to competence research; in fact I strongly believe that using findings of corpus research can be quite harmful. Consider Verhagen's (2005:124) finding that cases of so-called long *wh*-movement mainly occur "in the wild" when the subject of the matrix-clause is second person; examples such as (8b), for instance, require quite special contexts in order to be felicitous (indicated by the dollar sign).

- (8) a. Wat<sub>i</sub> denk je [dat hij *t<sub>i</sub>* moet lezen]?  
       'What do you think that he should read?'  
     b. \$Wat<sub>i</sub> denk ik [dat hij *t<sub>i</sub>* moet lezen]?  
       'What do I think that he should read?'

This observation is irrelevant for competence research because even if examples such as (8b) were entirely unusable (which they are not), this would not be a reason to exclude them from core grammar. Such exclusion would in fact be harmful because it requires the introduction of various *ad hoc* assumptions that would hamper establishing the correct syntactic mechanisms underlying the formation of sentences. The fact that (8a) is more common than (8b) is not syntactic but pragmatic in nature, and thus should receive an account in pragmatic terms. Claiming that (8b) is fully *grammatical* thus acknowledges the earlier-mentioned fact that acceptability judgments are not a matter of syntax only.

### 5.5 Corpus data: problems with raw data

It seems that corpus research is mainly useful for competence research in the case of unclear cases; this does not only hold for cases in which researchers disagree on acceptability judgments but also when the researcher himself is in doubt whether a certain construction is possible or not. This motivates the occasional use of Google searches in SoD, which normally serve the limited goal of showing that a certain "suspect" construction is or is not commonly used; this may help the researcher decide whether a certain structure should or should not be considered acceptable.<sup>8</sup> It does not work the other way, though, in that the fact that a certain construction can be found in a corpus does not provide foolproof evidence that the construction at hand should be part of the speakers' competence. A simple but telling example (involving internet data) is the following. SoD-A7 claims that partitive genitive constructions are unacceptable with the human pronoun *iemand* 'someone': cf. *iets leuks* 'something nice' versus *\*iemand leuks*. However, Eric Hoekstra pointed out to me that *iemand leuks* occurs quite frequently on the internet: a recent Google search [20/5/2016] on this string resulted in no less than 297 hits (after omission of the duplicates) and similar results arise with

---

<sup>8</sup> I agree with Van de Velde (2014:96) and Levshina (2016, §4) that Google counts are highly unreliable and should be used with care. I was unpleasantly surprised by the example Levshina gave from SoD because I had the impression that I had double-checked all Google counts with elimination of "double counting", which can normally be obtained automatically by browsing through the search results until Google notes that it has "omitted some entries very similar to the [ones] already displayed"; this often results in an astounding drop in the number of results. It now turns out that I have overlooked a limited number of cases, for which I apologize.

other adjectives. However, up to the present day I have not found a single speaker of Dutch who accepts such forms, and for this reason SoD-A7.1 (p.425) does not give *iemand leuks* as a possible option in standard Dutch (although it leaves open the possibility that we are dealing with an innovation). This shows that the raw data made available by corpus research are simply not well adapted to the needs of competence researchers and do not make introspection superfluous; it remains necessary to rid extracted samples of unwanted noise, regardless the degree of sophistication of the search method.

The need of cleansing samples collected by corpus research of unwanted noise can also be demonstrated by evaluating the corpus examples given by Coleman (2016) against the claim in SoD that *krijgen*-passives systematically involve ditransitive verbs.<sup>9</sup> First, consider his example (1), repeated here in an abbreviated form as (9a), which means something like “the CML was made responsible for certain funds earmarked for Leiden Centraal”. Apart from the fact that (9a) is taken from a highly formal text (minutes of an advisory body), which might in fact already be a good reason to exclude it from core grammar, the inclusion of this example in the set of counterexamples is based on the naïve presupposition that any case of *krijgen* + *participle* constitutes an instantiation of the *krijgen*-passive. It seems quite likely, however, that (9a) is not a *krijgen*-passive but a so-called semi-copular construction of the type illustrated in (9b), which was extensively discussed in SoD-A6.2.1: the participle *ondergebracht* is not verbal but adjectival in nature.<sup>10</sup>

- (9) a. Het CML heeft het potje voor Leiden Centraal ondergebracht gekregen.  
       The CML has the pot for Leiden Centraal under-brought got  
       b. Jan heeft het raam open gekregen.  
       Jan has the window open got  
       ‘Jan managed to get the window open.’

That Coleman’s example (2), repeated in a shorter form as (10a), is not suited for refuting the claim that *krijgen*-passives systematically involve ditransitive verbs is clear from the fact that the verb *toevoeren* ‘to supply’ does occur as a ditransitive verb in older stages of Dutch; see the citations in WNT, *toevoeren*<sup>1</sup>. Example (10b) shows that it is in fact still possible to find active ditransitive constructions on the internet today. This shows that Coleman incorrectly presumes that *toevoeren* is not a ditransitive verb.

<sup>9</sup> Coleman wrongly interprets the SoD claim such that *krijgen*-passivization is a transformational rule changing an active ditransitive sentence into a *krijgen*-passive sentence; such rules relating sentences were indeed part of early generative grammar but they were abandoned in the 1970’s. Fortunately, there is no reason for assuming that this misconception seriously affected Coleman’s argument, since it is an extended version of his remark on Van Oostendorp (2014), where he did formulate the generalization as intended.

<sup>10</sup> The use of adjectival participles in this construction is marked in standard Dutch but quite common in the southern varieties of Dutch which seemingly use “double time auxiliaries” in perfect tenses; see Koenen et al. (2011), who correctly analyze these “perfect doubling constructions” as perfect semi-copular constructions. An attested example of the semi-copular construction with *krijgen* from Brabantish spoken in the area of Tilburg is: *Ze hebben<sub>aux</sub> de weg niet gevonden<sub>A</sub> gekregen<sub>past participle</sub>* ‘They have not been able to find the way’.

- (10) a. De hartspier krijgt niet voldoende zuurstof toegevoerd.  
 the heart.muscle gets not sufficient oxygen prt.-supply  
 ‘The heart muscle is not supplied with sufficient oxygen.’  
 b. Glutamine voert het haar zwavel toe, [...] <sup>11</sup>  
 Glutamine supplies the hair sulfur prt.  
 ‘Glutamine supplies the hair with sulfur.’

Colleman’s example (3) is unacceptable to me, as indicated by my judgment on the shorter form in (11a). This example may be acceptable in the southern varieties of Dutch (the example is taken from a Belgian newspaper), but then it should also be mentioned that it is possible to find examples such as (11b), which show that at least some speakers can use *aanspannen* as a ditransitive verb. Example (11a) can therefore not be used to argue against the claim that *krijgen*-passives systematically involve ditransitive verbs.

- (11) a. ??Ze kreeg een rechtzaak aangespannen.  
 she got a legal.procedure prt.-start  
 ‘A legal procedure was started against her.’  
 b. ??[...] de ander spant je een rechtzaak aan. <sup>12</sup>  
 the other start you a legal procedure prt.  
 ‘[...] the other starts a legal procedure against you.’

Colleman’s example (4), repeated here as (12), is problematic because the verb in *tegenfluiten* in example (4) is a new coinage, which seems to be used in an attempt to translate the English collocation *to whistle a foul (against ...)*. It therefore seems that we are dealing with deliberate, contrived language use, which should not be included in a synchronous description of core grammar. Furthermore, the use of *tegenfluiten* seems restricted to sports commentaries and should thus (in as far as it is indeed well-established) be considered as technical jargon, which is also excluded from core grammar for the principled reason that it is normally not learned spontaneously in infancy but learned consciously at some later age.

- (12) Hij kreeg een fout tegengefloten.  
 he got a foul against-whistled  
 ‘A foul was whistled against him.’

Despite the fact that Colleman’s final example, repeated here in a shorter form as (13a), is taken from a Flemish newspaper, I have difficulties in assigning it a proper interpretation (although its context makes more or less clear what is intended). But even if we accepted that (13a) is part of core grammar, we should note that it is again easy to find constructions like (13b&c) on the internet in which *adviseren* is used as a ditransitive verb; I will not digress on the fact that *adviseren* ‘to advise’ is normally ditransitive if it takes a direct object clause, as in *Hij adviseerde mij dat boek te lezen* ‘He advised me to read that book’.

<sup>11</sup> cf. [juvel-5.nl/haar-direct.htm?websale8=juvel-5.nl&ci=haardirect](http://juvel-5.nl/haar-direct.htm?websale8=juvel-5.nl&ci=haardirect)

<sup>12</sup> cf. [mirosjabin.wordpress.com/page/45/?archives-list=1](http://mirosjabin.wordpress.com/page/45/?archives-list=1).

- (13) a. NTGent kreeg één van de hoogste bedragen geadviseerd.  
 NTGent got one of the highest sums advised.
- b. Meijers adviseert hem een ding: [...] <sup>13</sup>  
 Meijers advises him one thing
- c. [De opticien] adviseerde mij een bril. <sup>14</sup>  
 the optician advised me glasses

The discussion above is meant to show that it is quite hazardous to use raw corpus data for settling linguistic disputes. It is clear that there may be a wide variety of reasons why specific examples should be excluded from the sample. There may be cases that should be dismissed as irrelevant because they are incorrectly included in the sample, such as (9a) which in all likelihood is not a *krijgen*-passive, or because they do not show what they purport to show, such as (10a), (11a) and perhaps (13a), which all contain a verb that can also be used as a ditransitive verb by at least some speakers. Furthermore, the sample may include cases that are not part of core grammar but of the periphery: this holds for cases that are part of a specific restricted (written or formal) register, such as (9a) and perhaps (13), or that belong to jargon, such as (12) and perhaps also (9a) and (10a), or that are alien to the language in question due to language contact or borrowing, such as (12). Furthermore, the sample may include problematic cases that may receive an alternative (e.g., diachronic) explanation, such as (10a), or are restricted to a subset of the speakers investigated or even due to idiosyncrasies in individual speakers, such as (10a&b), (11a&b) and (13a). And of course, there may be cases that involve speech/writing/printing errors, jokes or swagger, and there is certainly a long list of other problems that can be added. This shows that samples collected by corpus research meet their own difficulties and shortcomings if used for competence research.

The general neglect of weeding out the noise from the raw data reveals that corpus researchers tend to put too much trust in their samples: the discussion of Coleman's data again shows that corpus research does not necessarily result in less "debatable" data than introspection research. Enhancing the quality of corpus data for the research task at hand is of course possible, but it seems that eliminating distortions of the sort mentioned above cannot be done properly without making an appeal to introspection, and will therefore be subject to similar objections as traditional introspection research. What is perhaps more harmful is that Coleman's discussion is based on a tacit (perhaps even *unconscious*) appeal to introspection, as is clear from the fact that Coleman's examples in (9) to (13) can only be used for arguing against the claim that *krijgen*-passives systematically involve ditransitive verbs on the assumption that the main verbs in these examples cannot be used as ditransitive verbs. That this claim is based on introspection is clear from the fact that it was not further examined by Coleman, although it is not difficult to collect internet data that suggest that this assumption may be false for no less than three out of the five verbs mentioned.

Note in this connection that Coleman's (2010/2016) claim that the verb *kopen* 'to buy' does not allow *krijgen*-passivization in those varieties of Dutch that allow it with a

<sup>13</sup> cf. [managementscope.nl/magazine/artikel/178-deloitte-roger-dassen](http://managementscope.nl/magazine/artikel/178-deloitte-roger-dassen)

<sup>14</sup> cf. [meningeoom.wordpress.com/2012/04/13/verhaal-van-een-meningeoomgenote/](http://meningeoom.wordpress.com/2012/04/13/verhaal-van-een-meningeoomgenote/)



benefactive object seems likewise based on a tacit appeal to introspection, as no source for the acceptability judgments motivating this claim is given. The observation is of course interesting since it suggests that *krijgen*-passivization of ditransitive verbs with a benefactive may be restricted after all (which cannot be established on the basis of standard Dutch). However, more research is needed to establish this conclusion, as it is easy to find such examples from German on the internet: *Kinder sehen auch nicht, dass andere von ihren Eltern ein- zweimal die Woche ein Eis gekauft kriegen, was bei ihnen vielleicht zweimal im ganzen Sommer vorkommt* ‘... that other (children) are bought an ice-cream by their parents once or twice a week ...’.<sup>15</sup> Whatever the outcome of this research, one thing is for sure: including a discussion of this kind in SoD-V3.2.4.1 would be undesirable given its limited goal of showing “that, contrary to what is sometimes assumed in the literature, the *krijgen*-passive is fairly productive” (p.444).

It seems fair to conclude on the basis of the discussion in this section that we simply have to live with the fact that raw data collected by corpus research cannot be used to refute claims made by competence research because the relevance of the data should first be evaluated.

## 5.6 Core grammar and periphery

The discussion above has made clear that SoD focuses on the description of core grammar. A well-known problem discussed in Los (2016) is that the distinction between the core (unconsciously learned part) and the periphery (consciously learned part) of grammar is not always as obvious as we would like it to be. There are of course many cases that are clear-cut: that the rule that articles precede nouns is part of the core grammar of Dutch, while the various morphological case residues are part of the periphery seems uncontroversial. However, there are also many cases for which it is not evident whether or not they should count as part of core grammar. This is related to the fact mentioned by Los that, although the periphery may include all kind of material (historic relics, loan forms from other languages, forced language use as found in jargon, etc.), there is no *a priori* reason to assume that the periphery is a domain, where irregularity is the rule rather than the exception. Los correctly points out that the periphery may contain more or less coherent subsystems. One unclear case mentioned earlier concerns the predeterminers *al* ‘all’ and *heel* ‘whole’: SoD-N7 allocated these predeterminers to core grammar because Den Dikken and I turned out to have quite clear acceptability judgments on their meaning, distribution and syntactic behavior, while Van de Velde (2014) concluded on the basis of his diachronic investigation that they are in fact “living fossils”, which should therefore be relegated to the periphery. It seems that there are no hard and fast criteria that can be used to decide who is right. The issue is perhaps less important for synchronic language descriptions of the type provided by SoD, as exclusion of peripheral matters is specifically important for the formalization of grammars in order to avoid inclusion of postulates/rules that are alien to the language in question. For this reason, it is probably best to include the borderline cases in SoD and leave the question as to whether they are part of core grammar to formal linguistics by “letting the theory decide” whether the construction should be considered grammatical or not. If the borderline cases follow automatically from the

<sup>15</sup> cf. [spiegel.de/forum/politik/kinderarmut-sind-staatliche-hilfen-der-richtige-weg-thread-4381-67.html#js-article-comments-box-form](http://spiegel.de/forum/politik/kinderarmut-sind-staatliche-hilfen-der-richtige-weg-thread-4381-67.html#js-article-comments-box-form)

proposed grammar, this may be reason to ascribe them to core grammar but if they require the introduction of special machinery that is not independently motivated, this may be reason to ascribe them to the periphery. For this reason, SoD does include data that some researcher may put aside as peripheral in their syntactic analysis. I refer the reader to Newmeyer (1983, §2.2) for a more detailed discussion of the way in which unclear/borderline cases are treated in formal linguistics.

### 5.7 Some concluding remarks

SoD reports the results of more than 40 years of generative competence research on standard Dutch. The fact that a coherent and comprehensive work such as SoD could be written on the basis of this research is a clear illustration of the maturity of the generative program and the fertility of the method of collecting data by means of introspection. Introspection data have much to recommend themselves: they can be obtained fairly easily, they can be manipulated such that they are optimized for use as illustrations for specific linguistic problems by eliminating interference of irrelevant factors, and there are few practical limitations on the type of data that can be obtained. In these respects they differ markedly from corpus samples, which are difficult to obtain, contain a great deal of noise and are highly dependent on the queries that the available corpora can be asked successfully. Although the use of introspection is not ideal and may give debatable results in specific cases, there is no clear evidence that the overall results differ markedly from other (less economical) methods of collecting data when it comes to competence research.

Competence research imposes different conditions on data collection than, e.g., performance research; often it suffices to establish whether a certain construction is possible or not. It is not *a priori* clear that corpus research improves the data set for competence research, especially because at present it is suitable only for investigating relatively shallow linguistic phenomena that are readily observable or, more precisely, searchable in the existing corpora: co-occurrence of certain forms (such as *krijgen* + participle), word order (such as A-scrambling and word order variation in verbal clusters), the optionality of specific elements (such as *om* in certain infinitival clauses mentioned by Levshina 2016), and so on. It is not clear, however, how corpus research could be of help when it comes to more complicated syntactic issues or more profound linguistic questions as a result of the limitations imposed by the available tag sets in the existing corpora. My impression is that at least 90% of SoD could not have been written if we had had to rely only on data obtained by means other than introspection.

Corpus data include noise that is irrelevant for competence research and this may hamper the detection of the syntactic mechanisms underlying the formation of sentences. The samples extracted by corpus research have no immediate meaning for competence research as such, apart from the fact that they may reveal accidental omissions in the data set obtained by introspection. Just like introspection data, corpus data are in need of further analysis before they can be used for competence research: are the data indeed relevant for the problem at hand, do they indeed show what they purport to show, etc.? Making extracted samples useful for specific research tasks may trigger similar objections as introspection research: there is in fact no evidence for assuming

that corpus data are less “debatable” than introspection data when it comes to their relevance for a specific linguistic problem.

Those reviewers that suggest that competence research should impose the same methodological restrictions on data collections as performance research simply illustrate their failure to appreciate the difference between the two types of research. There is of course no principled reason for excluding corpus data from SoD because any type of linguistic data can in principle be included in SoD. However, since it normally suffices for syntactic theorizing to know whether a certain structure is possible or not, data obtained by introspection will normally be preferred for reasons of economy; as long as sufficient progress in competence research can be made by using introspection, corpus research should be set aside for types of research that crucially depend on it, such as diachronic or variational linguistics.

The quotes from the preface of SoD given at the start of this article make explicit that the rationale for writing SoD has never been to consider what other linguistic subdisciplines have to offer to competence research. For this reason, it is extremely difficult, well-nigh impossible to field the repeated complaints that SoD does not pose the “right” questions, namely those pertaining to language use. A clear example is Levshina’s (2016) remark that “the use and omission of the optional complementizer *om* [...] is not discussed in the chapter in sufficient detail” because we do not consider the restrictions on actual use. SoD-V5.2 purports to show that from a syntactic point of view “infinitival argument clauses can be divided into three main types: *om* + *te*-, *te*- and bare infinitivals” (p.765). We can only hope that the discussion of this matter may be of help in answering the type of questions that, for instance, corpus and usage-based researchers are interested in but we cannot be expected to answer these questions ourselves. The starting point of the SoD-project is that competence research has made available a wealth of new information (both data and linguistic insights), which is accessible to a limited group of linguists only, as it is mostly buried in highly technical discussions. We believe that this information may also be relevant to other types of linguistic research, and SoD should be seen as an attempt at making this information available to a larger group of linguists. It is now up to this group to investigate whether SoD provides material that can be used fruitfully in their own research. Of course, we hope that this will be the case, not only because we want to be of service, but also for the more selfish reason that it will increase the chance that the output of that research will connect more easily with the specific needs of competence researchers than is the case at this moment.

The self-imposed restrictions on SoD are of course not imperative for grammar writing, and can in principle be relaxed or changed. I personally believe that corpus and usage-based linguistics has insufficiently matured to be able to produce a comprehensive reference work comparable to SoD: it is still fragmented and anecdotal in nature, due to the fact mentioned earlier that it mostly covers easily observable/searchable phenomena. However, in the event that I am wrong, I would certainly welcome a work of this type. In this connection, I would like to point to the fact that the *Virtuele instuut vir Afrikaans* has taken the initiative to produce a grammar of Afrikaans based on the material on Dutch and Frisian available on [taalportaal.org](http://taalportaal.org),

and that the resulting *Syntax of Afrikaans* will more or less follow the overall structure of SoD but make more use of corpus-based research. Hopefully, this will give us the opportunity in the near future to compare the strictly competence-based SoD with a more performance-based *Syntax of Afrikaans*, and no doubt this will also teach us more about the possible added values that corpus research might have for SoD.

## 6 Conclusion

The reviews of SoD make clear that the synchronic description provided in SoD does not have the last say; there is no reason to deny that it can be highly profitable if we connect it with information about other or older varieties of Dutch, other languages, language acquisition and deficiencies, etc. For example, the claim made in SoD-V11 that the first position of the sentence in subject-verb inversion constructions is normally filled by a constituent with a special information-structural status (interrogative phrase, topic or focus) is indirectly supported by Los' (2016) conclusion on the basis of Old English that the left periphery of the sentence is used to "satisfy various communicative requirements". Since the seminal article by Rizzi (1997) the description of these information-structural aspects have in fact become part of the generative program, that is, the relation between word order and information structure is now considered part of core grammar. There is no *a priori* reason to exclude the possibility that other "functional" aspects of language use may become relevant in future version of this program and for this reason the current flourishing of usage-based research is to be applauded: for example, Willems' (2016) discussion of the factors affecting extraposition is also relevant from the perspective of competence research, and this holds for more issues raised in the reviews. The reviews of SoD contain a lot of information that may find their way in future versions of SoD in the form of corrections and additions; some revisions made on the basis of the three reviews published in *Nederlandse Taalkunde* 19 can already be found in the internet version of SoD found at [taalportaal.org](http://taalportaal.org). The observation that connecting the results of the various types of linguistic research is likely to deepen our insights in fact motivated our attempt to collect the data and insights made available by competence research, and to present these in such a way that they can be used by researchers who would normally not be willing or able to consult the formal linguistic literature. In short, we would like to encourage all linguists to make use of SoD in any way they feel proper.

## 7 Binary tense theory: a reply to Verkuyl (2016)

Verkuyl (2008) argues convincingly that the mental representation of tense involves the three binary features in (14). Following Te Winkel (1866), he further claims that Dutch expresses all three oppositions within the verbal system: inflection expresses [ $\pm$ PAST], the verb *zullen* 'will' expresses future, and the temporal auxiliaries *hebben* 'to have' and *zijn* 'to be' express perfectivity.

- (14) a. [ $\pm$ PAST]: present versus past  
       b. [ $\pm$ POSTERIOR]: non-future versus future  
       c. [ $\pm$ PERFECT]: imperfect versus perfect

Broekhuis & Verkuyl (2014) argue against the claim that *zullen* ‘will’ expresses future: it is an epistemic modal verb that locates the eventuality in the realm of possible worlds and its future interpretation results from the pragmatic fact that possible worlds are located after speech time by default. This claim is supported by at least two facts: (i) similar future readings are also found with other epistemic verbs such as *moeten* ‘must’ and *kunnen* ‘may’; (ii) the future reading of modal verbs (including *zullen*) can be overridden in examples such as (15), which can be used if the speaker is underinformed about the actual situation at speech time, that is, if the split-off point of the possible worlds precedes speech time: cf. SoD-V1.5.2.

- (15) Jan zal/moet/kan gisteren al vertrokken zijn.  
 Jan will/must/may yesterday already left be  
 ‘Jan will//must/may already have left (namely yesterday).’

Verkuyl’s review of SoD-V1.5.4 is based on the presumed fact that SoD eliminates the feature  $[\pm\text{POSTERIOR}]$  in (14b). This is a misunderstanding, which is probably due to the fact that Verkuyl has misinterpreted the phrase *the Dutch verbal system* as “the Dutch tense system” in quotations like the following: if the future reading of *zullen* is only due to pragmatic considerations, “the Dutch verbal system is based on just the binary features  $[\pm\text{PAST}]$  and  $[\pm\text{PERFECT}]$ , and therefore does not make an eight-way, but only a four-way tense distinction” (p.157). That the two phrases have different meanings is clear from the clarification of these terms in the first paragraph of V1.5.4 (p.156): the tense oppositions in (14) can be expressed “within the verbal system by means of inflection and/or auxiliaries, but may also involve the use of [other means]” (p.156).

Misunderstandings of this sort sometimes have fortunate consequences, and this may happen to be the case here. Verkuyl (2016) argues that dropping the opposition in (14b) is impossible within the binary tense system because the notion present  $j$  of eventuality  $k$  used in the formal definition of  $[\pm\text{POSTERIOR}]$  in (14b&b’) functions as a bridge between the notion present/past tense interval  $i$  in the (slightly simplified) definition of  $[\pm\text{PAST}]$  in (14a&a’) and the notion eventuality  $k$  used in the definition of  $[\pm\text{PERFECT}]$  in (14c&c’); cf. SoD-V1.5.1 (p.117).

- (16) a. Present:  $i \circ n$  [i includes speech time  $n$ ]  
 a’. Past:  $i \circ n'$  [i includes virtual speech-time-in-the-past  $n'$ ]  
 b. Non-future:  $i \approx j$  [i and j synchronize]  
 b’. Future:  $i_a < j$  [ $i_a$  precedes  $j$ ]  
 c. Imperfect:  $k \leq j$  [ $k$  need not be completed within  $j$ ]  
 c’. Perfect:  $k < j$  [ $k$  is completed within  $j$ ]

Verkuyl (2016) now tries to accommodate the presumed claim in SoD by redefining the notion future (posteriority) as  $j \leq i_\diamond$  (cf. his example (7)), where  $i_\diamond$  may include speech time  $n$ . If I understand Verkuyl correctly, this definition expresses that present  $j$  of eventuality  $k$  can but need not be fully encompassed by  $i_\diamond$  and thus allows that  $j$  is also partly situated in  $i_a$ . The clarification of this definition below his example (7) further suggests that posteriority and epistemic modality can/should be equated, and I believe that this is indeed a promising step. A more direct way of expressing this, however,

would be by saying that  $j$  is located in the temporal interval following the so-called split-off point of the possible worlds, which already played a prominent role in Broekhuis & Verkuyl (2014), Verkuyl & Broekhuis (2014), and SoD-V1.5.2/4. This would enable us to reinterpret Verkuyl's binary tense system as a modular system that arises from the interaction of the temporal, modal and aspectual distinctions in (17), where  $i_{pw}$  refers to the temporal interval starting at the split-off point of the possible worlds (which equals speech time  $n$  in the default case but may also precede it when the speaker is underinformed).

- (17) a. Tense: [-PAST]:  $i \circ n$  — [+PAST]:  $i \circ n'$   
 b. Modal: [-IRREALIS]:  $j \approx i$  — [+IRREALIS]:  $j < i_{pw}$   
 c. Aspect: [-PERFECT]:  $k \leq j$  — [+PERFECT]:  $k < j$

The binary TMA theory in (17) does not change anything in the default cases where the split-off point of the possible worlds coincides with the (virtual) speech time (in the past), but simplifies the system for the marked cases in which this split-off point precedes it by allowing us to locate  $j$  (and  $k$ ) directly in the restricted temporal domain  $i_{pw}$ , which can be determined on the basis of contextual information. Perhaps, the more modular view may enable us to expand binary TMA theory so that it provides us with a more comprehensive model for spatio-temporal representations; I refer to SoD-V8.2.3 for evidence that spatial and temporal adverbial phrases may play a similar role in restricting the location of  $j$  and  $k$  on the temporal axis.

## References

- Baart, Joan (1987). Focus en accent in het Nederlands. *GLoT* 2, 293-312.  
 Barbiers, Sjeff (2013). Microsyntactic variation. In: Marcel Den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*. Cambridge University Press, 899-926.  
 Beliën, Maaïke (2014). Complementing *Adpositions and adpositional phrases*: a review of Broekhuis (2013). *Nederlandse Taalkunde* 19, 291-304.  
 Broekhuis, Hans (2008). *Derivations and evaluations: object shift in the Germanic languages*. Berlin/New York: Mouton de Gruyter.  
 Broekhuis, Hans, Norbert Corver, Marcel Den Dikken, Evelien Keizer, and Riet Vos (2012-2016). *Syntax of Dutch (7 volumes)*. Amsterdam: Amsterdam University Press.  
 Broekhuis, Hans, and Henk Verkuyl (2014). Binary tense and modality. *Natural Language & Linguistic Theory* 32, 973-1009.  
 Cinque, Guglielmo (1993). A null theory of phrase and compound stress. *Linguistic Inquiry* 24, 239-297.  
 Coleman, Timothy (2010). Lectal variation in constructional semantics: benefactive" ditransitives in Dutch. In: Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. De Gruyter Mouton, 191-221.  
 Coleman, Timothy (2016). Over werkwoordalternanties in de *Syntax of Dutch*. *Nederlandse Taalkunde* 21 (this issue).

- De Groot, A.W. (1959). Een nieuwe Nederlandse Grammatica. *De Nieuwe Taalgids* 52, 141-144.
- De Hoop, Helen (2000). *Otje en scrambling in het Nederlands*. *Tabu* 30, 97-112.
- De Hoop, Helen (2003). Scrambling in Dutch: Optionality and Optimality. In: Simin Karimi (ed.), *Word Order and Scrambling*. Malden, Ma (USA)/Oxford (UK), Blackwell Publishing.
- De Hoop, Helen (2016). Wordvolgordevariatie: theorie versus empirie? *Nederlandse Taalkunde* 21 (this issue).
- De Sutter, Gert (2005). Rood, groen, corpus! Een taalgebruikersgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. University of Leuven: PhD thesis.
- De Sutter, Gert (2007). Naar een corpusgebaseerde, cognitief-functionele verklaring van de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. *Nederlandse Taalkunde* 12, 302-330.
- Du Bois, John B. (1987). The discourse basis of ergativity. *Language* 63, 805-855.
- Fodor, Jerry A., and Jerrold J. Katz (1964). *The structure of language; readings in the philosophy of language*. Englewood Cliffs, N.J.: Prentice-Hall.
- Gussenhoven, Carlos (1992). Sentence accents and argument structure. In: I.M. Roca (ed.), *Thematic structure. Its role in grammar*. Berlin/New York, Foris Publications, 79-106.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten C. van den Toorn (1997). *Algemene Nederlandse spraakkunst*. Groningen: Nijhoff.
- Koenenman, Olaf, Lekakou. Marika, and Sjef Barbiers (2011). Perfect doubling. *Linguistic Variation* 11, 35-75.
- Kotzé, Ernst (2016). Aspects of congruence and divergence relating to adjectives in Dutch and Afrikaans - A discussion of certain descriptive categories on the basis of Broekhuis' *Syntax of Dutch*. *Nederlandse Taalkunde* 21 (this issue).
- Levshina, Natalia (2016). Control, causation and Google counts. *Nederlandse Taalkunde* 21 (this issue).
- Los, Bettelou (2016). The rise of verb-second. *Nederlandse Taalkunde* 21 (this issue).
- Newmeyer, Frederick J. (1983). *Grammatical Theory. Its limits and possibilities*. Chicago/London: University of Chicago Press.
- Putnam, Michael T. (2007). *Scrambling and the Survive Principle*. Amsterdam / Philadelphia: John Benjamins.
- Rizzi, Luigi (1997). The fine structure of the left periphery. In: Liliane Haegeman (ed.), *Elements of grammar. Handbook of generative grammar*. Dordrecht/Boston/Londen, Kluwer Academic Publishers, 281-337.
- Roberts, Ian (2007). *Diachronic syntax*. Oxford/New York: Oxford University Press.
- Smessaert, Hans (2014). Telwoorden en kwantoren: syntaxis versus semantiek. *Nederlandse Taalkunde* 19, 77-86.
- Sprouse, Jon, and Diogo Almeida (2010). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics* 48, 609-652.
- Te Winkel, L.A. (1866). Over de wijzen en tijden der werkwoorden. *De Taalgids* 8, 66-75.
- Van Bergen, Geertje, and Peter de Swart (2010). Scrambling in spoken Dutch: Definiteness versus weight as determinants of word order variation. *Corpus Linguistics and Linguistic Theory* 6, 267-295.

- Van de Velde, Freek (2009). *De nominale constituent. Structuur en geschiedenis*. Leuven: Universitaire Pers Leuven.
- Van de Velde, Freek (2014). Nederlandse predeterminatoren als levend fossiel. *Nederlandse Taalkunde* 19, 87-103.
- Van den Berg, Evert (1978). Fokus presuppositie en NP-preposing. *De Nieuwe Taalgids* 71, 212-222.
- Van Eynde, Frank (2004). Part of speech tagging en lemmatisering van het Corpus Gesproken Nederlands. Leuven, University of Leuven, Center of computer Linguistics.
- Van Oostendorp, Marc (2014). *Het meisje kreeg de sleutels overhandigd*. Wat we nog niet weten over het werkwoord (3). [neerlandistiek.nl/2014/12/het-meisje-kreeg-de-sleutels-overhandigd](http://neerlandistiek.nl/2014/12/het-meisje-kreeg-de-sleutels-overhandigd).
- Verhagen, Arie (1986). *Linguistic theory and the function of word order in Dutch. A study on interpretive aspects of the order of adverbials and noun phrases*. Dordrecht: Foris Publications.
- Verhagen, Arie (2005). *Constructions of intersubjectivity: discourse, syntax, and cognition*. Oxford/New York: Oxford University Press.
- Verkuyl, Henk (2008). *Binary tense*. Stanford: CSLI Publications.
- Verkuyl, Henk (2016). In the Mood for tense? *Nederlandse Taalkunde* 21 (this issue).
- Verkuyl, Henk, and Hans Broekhuis (2014). Temporaliteit en modaliteit. *Nederlandse Taalkunde*, 306-323.
- Vikner, Sten (1994). Scandinavian object shift and West Germanic scrambling. In: Norbert Corver and Henk van Riemsdijk (eds.), *Studies on Scrambling. Movement and non-movement approaches to free word-Order phenomena*. Berlin/New York, Mouton de Gruyter, 487-517.
- Willems, Annelore (2016). Een extra position? *Nederlandse Taalkunde* 21 (this issue).

### About the author

Hans Broekhuis  
 Meertens Institute, Amsterdam  
[hans.broekhuis@meertens.knaw.nl](mailto:hans.broekhuis@meertens.knaw.nl)