

Meaning in-between Structure and the Lexicon. Project description.

1. State-of-the-art

After having been temporarily relegated to the periphery of linguistic enquiry in the heyday of "autonomous syntax", the relationship between morphosyntax and semantics has in recent years regained its position on linguistic centre stage. According to Langacker (1999), for instance, the interaction between the form of sentences and word groups and their meaning is the most crucial issue in contemporary linguistics. The formal side of utterances can be divided in two components: the individual words or morphemes on the one hand (the lexical component), and the way these individual lexical elements are combined into grammatical utterances on the other hand (the syntactic, structural component). Until fairly recently, the predominant view in linguistic theory (in the generative as well as in the (structural)-functional tradition) was that the lexicon is the sole determining factor in the form/meaning-relationship. That is to say, the overall meaning of a sentence was taken to be projected from the lexical-semantic properties of the individual words (with the main verb providing the relational meaning and determining the structure of the sentence). The rise of paradigms like Cognitive Grammar (Langacker 1987, 1991) and (especially) Construction Grammar (Goldberg 1995, Kay & Fillmore 1999), however, has led to the (re?)establishment of the grammatical construction as a meaningful element in its own right. In sentence (2), for example, the ditransitive syntactic structure definitely contributes to the overall meaning of the sentence. Just like the sentences in (1), sentence (2) describes a transfer of possession: Sue's action brings about that Bill receives a cake. In (1), this transfer element can be attributed to the lexical semantics of the respective verbs; that is to say, *give*, *sell*, etc. lexicalise a transfer of possession. In (2) however, the transfer element can hardly be attributed to the lexical semantics of the verb, since *throw* does not normally denote a transfer of possession. Here, it is the ditransitive syntax itself which triggers the transfer interpretation, while the verb provides additional information about the means or manner of the transfer.

- (1) Sue gave / sold / served / handed / ... Bill a cake.
- (2) Sue threw Bill a cake.
- (3) Sue * wanted / * moved / ?? ate... Bill a cake.

In this way, both the lexical and the structural component contribute to the overall meaning of clauses. It is obvious, however, that lexicon and structure are interdependent factors: the syntactic structure of utterances is to a certain extent determined by the valency requirements and restrictions of their lexical elements. The ungrammaticality of the sentences in (3), for example, can be ascribed to the incompatibility of their respective verbs with the ditransitive syntactic structure.

Although the Construction Grammar approach has undoubtedly generated a number of new insights into the relationship between form and meaning, it has not as yet succeeded in completely elucidating the complex interaction between lexicon, structure and meaning. What is more, it is by no means sure whether this new approach will ever be able to provide a realistic picture of this complex interaction. The main flaws in the Construction Grammar methodology are the following (it should be stressed that these are definitely not typical of Construction Grammar but are characteristic of other recent work on the syntax:semantics-interface as well):

1. All too often, the semantic contributions of the structural and lexical components are studied separately, their manifest interdependency being downplayed. That is to say, particular semantic elements are often attributed to either the lexical or the structural component. This 'splitting' tendency is counterintuitive, since it is obvious that in natural

language use, speakers do not produce lexical elements in isolation, nor 'bare' structural patterns without lexical filling (as pointed out by Croft 2003 among others).

2. Scholars often choose their lexical examples from a number of well-studied and (at least in theory) clearly demarcated semantic fields, such as the 'verbs of movement'. This leads to a narrow and possibly biased view of the facts.

3. Thorough contrastive investigations of data from several languages are scarce. This is unfortunate, because it would definitely enhance understanding to consider the facts from a contrastive-linguistic angle. The theoretical apparatus of Construction Grammar and other modern approaches to the syntax/semantics-interface is largely based on English data, but the relationship between lexicon, structure and semantics is not necessarily exactly the same in other languages.

4. All too often, central hypotheses about the syntax/semantics-interface rest on introspective data. Crucial claims about the meanings of syntactic constructions are supported with invented example sentences. These are provided to illustrate that certain verb-construction combinations are allowed while others are excluded. Such grammaticality judgements, however, are usually based exclusively on the linguist's (fallible) intuitions, and are not (or at least, not systematically) tested against 'analyser-independent' sources of linguistic data (large corpora of real language, experimental psycholinguistic research). There is a real danger that both the selection and the evaluation of these example sentences are strongly influenced by the hypotheses they are meant to verify. Also, when exclusively relying on invented example sentences, the syntactic intricacies of naturally occurring language may easily be overlooked. In this regard, a lot of contemporary research hardly differs from the generative research of the past decades.

The research group submitting this proposal (named *Contragram* in 1995) possesses the necessary expertise to make a contribution to the study of the syntax/semantics-interface that does not suffer from the above-mentioned drawbacks. This expertise was partly obtained through the previous projects (GOA 12052095 and GOA 12050399), which have provided us with the opportunity to carry out thorough scholarly research into the relationship between lexicon and structure. The outcome of these projects, the *Contrastive Verb Valency Dictionary* (CVVD, cf. <http://bank.rug.ac.be/contragram/>), contains a large number of detailed descriptions of the valency structures of verbs in three languages (Dutch, French and English). This large database offers reliable information about the question which verbs are compatible with which syntactic structures and hence will allow us to get a clear view of the mutual restrictions of lexicon and structure (at the level of the verb and clause-level constructions, that is). In addition to listing valency patterns, the CVVD also contains detailed information on the (often highly polysemous) semantics of the verbs included, hence it also provides a good starting point for the study of the relationship between the semantic pole and the two formal components. The current project (GOA 12050399) has already initiated some fundamental research into the domain under discussion: the PhD research of each of the three researchers attached to the project focusses on the semantics of one or more (related) morphosyntactic structures. Coleman (in preparation), Defrancq (2002) and Noël (2003) provide empirically justified descriptions of the double-object and prepositional-dative constructions, the embedded interrogative clause, and the alternation between a finite complement clause and an infinitival clause with *verba dicendi* and *verba cogitandi*, respectively. In each case, the semantics of the construction(s) under investigation is analysed in relation to the semantics of the lexical elements entering into those constructions.

The lexical-semantic scope of the CVVD is broad: it contains information about verbs from a wide range of semantic fields. Also, the three languages involved are treated on a par, that is, the analysis is equally thorough and exhaustive for all three languages. In other words,

the empirical basis provided by the CVVD database is far wider than that of the average study (three languages, a variety of semantic fields). In addition, the key methodological principle of the CVVD is the belief that reliable linguistic facts should first and foremost be drawn from large text corpora, and not exclusively from the linguist's own intuitions. The above-mentioned PhD dissertations employ a corpus-based methodology as well.

The project promoters have conducted fundamental research into the relationship between morphosyntactic structure and the lexicon, and into the relative value of several sources of linguistic evidence. As an extension of her earlier work on verbal constructions in French, D. Willems' recent research has focussed on the establishment of a typology of verbs and verbal constructions, which integrates lexical, syntactic and semantic elements and offers a suitable framework for the study of meaning extensions and the organization of semantic fields (Willems 2002, 2004). M. Devos devoted some of her research attention to the semantics of fixed verb-preposition patterns (Devos and Vandeweghe 2004). Both D. Willems and A.-M. Vandenberghe have looked into the methodological issue of how to collect reliable linguistic data: the former reflected on the complementarity of intuition and empirical observation and on the role of text corpora in linguistic description, the latter on the importance of translation corpora for contrastive linguistics (Willems 2000a,b, Aijmer and Simon-Vandenberghe 2003, 2004).

2. Aim

The main objective of the submitted project is to provide an empirically grounded account of the respective contributions of the lexical and structural components to the overall meaning of Dutch, French and English clauses, thus contributing to the development of a typological theory of the relationship between meaning, structure and the lexicon in the world's languages. The submitted project will complement existing typological research in two major respects. First, while it is common practice to offer an (often rather shallow) overview of a small number of formal patterns and/or semantic elements in a large number of languages (e.g. Talmy 2003), the Contragram team aims to provide an in-depth analysis of a large number of formal means for the expression of semantic content in a small number of languages. We believe that this approach will result in a more representative picture of the relationship between lexical and constructional meanings. Second, the synchronic analysis will be supplemented with an important diachronic component: data from earlier language phases will be taken into account in order to track the evolutions which have led to the present-day situation, and to determine whether the languages under investigation have evolved towards greater typological consistency.

Synchronic component

For obvious reasons, we shall focus on the verbal lexicon and verb patterns (i.e. clause-level constructions like monotransitive, ditransitive, etc). The meanings of both the main verb and the clause structure are not (strictly) referential, but relational in nature: both express a relationship between the referential elements of the clause. Hence, they may be regarded as "rival" formal means of expression par excellence and so constitute a natural area of investigation for the submitted project.

The competition and interaction between verbs and structural patterns becomes evident in structure alternations and lexicalisation patterns. Structure alternation is a general term for all kinds of phenomena where a single lexical item can be combined with several morphosyntactic structures, as in the following examples with the English verb *believe*:

- (4) a. They believed that he was ill.
 b. They believed him to be ill.
 c. They believed him ill.

The standard view of several modern approaches to the syntax/semantics-interface - including Cognitive Grammar and Construction Grammar - is that the meaning of the verb *believe* remains the same in (4a-c) and that the subtle differences in meaning between the three sentences should be attributed to the semantic import of the alternate syntactic constructions. Elucidating those subtle meaning distinctions, however, is an altogether different matter: though this particular alternation has definitely not been understudied, there is still no consensus about the exact nature of the semantic differences between the sentences in (4) (cf. Noël 2003). A thorough analysis of a large number of contextualized corpus examples according to the CVVD methodology - which implies a more refined syntactic analysis than the rather shallow analyses typical of, for instance, some versions of Construction Grammar - will enable us to formulate an empirically grounded picture of those meaning distinctions, and to compare them to the situation in the other languages. In this particular case, the possibilities in Dutch and French seem to be more restricted than in English:

- (5) a. Ils croyaient qu'il était malade.
 b. * Ils le croyaient être malade. (but: Jean, qu'ils croyaient être malade.)
 c. Ils le croyaient malade

- (6) a. Ze geloofden dat hij ziek was.
 b. * Ze geloofden hem ziek te zijn.
 c. * Ze geloofden hem ziek.

At first sight, and in this particular case, these restrictions would seem to suggest a less diverse semantic import of the structural component in Dutch and French than in English. But obviously, contrastive claims of this kind are in need of much stronger argumentative support than what is offered here. In any event, languages differ in their strategies for expressing semantic content, and an in-depth comparison of those different strategies would be an essential contribution to existing typological research. Needless to say, languages do not only differ in their inventory of verb patterns, but also in the organisation of their verbal lexicon. For instance, a well-known typological claim is that verbs which lexicalize a manner component are used much more often in the Germanic languages than in the Romance languages (Slobin 2000). Obviously, these different lexicalisation tendencies interact with the division of labour between the lexical and structural components: if a particular meaning element (e.g. manner) is lexicalised in the verb, there is no need for other formal means of expression. Consider the contrast between (7a-b) on the one hand and (7c) on the other, each of which can be considered the unmarked expression for the meaning 'They caused him to die by means of a gunshot'. In both English and Dutch, the means (i.e., the gunshot) is lexicalised by the verb, while the change of state of the patient (i.e., his ending up dead) is signalled by a resultative phrase (which is optional in English, and to a certain extent incorporated into the verb in Dutch). In French, by contrast, the verb lexicalises the change of state and the means is expressed in a prepositional phrase.

- (7) a. They shot him (dead/to death).
 b. Ze hebben hem dood-geschoten.
 c. Ils l'ont tué par balle.

This brief discussion of these two selected areas of investigation (i.e., structure alternations and lexicalisation patterns) should suffice to give an impression of the specific topics that will

be investigated, but other phenomena which illustrate the interplay between the lexical and structural components, may be included in the investigation.

Diachronic component

The synchronic analysis will be supplemented with an extensive diachronic component. It is an interesting question whether or not languages evolve towards greater typological consistency, with similar meanings being expressed by similar formal means of expression. In the course of their evolution, each of the three languages under investigation has been subject to shifts in the relationship between the lexical and structural components. For instance, while the combination of the verb *geloven* ('believe') with an accusative-plus-infinitive complement is impossible in contemporary Dutch (witness the ungrammaticality of example sentence 6b above), such sentences do occur in 16th and 17th Century Dutch (Duihoven 1991). In present-day English, accusative and infinitives are relatively infrequent as compared to so-called nominative and infinitives, with a passive matrix verb (as in 4b', the passive pendant of 4b). As argued by Noël (2001), this passive is developing into an auxiliary-like element with evidential meaning (in other words, from a lexical into a grammatical element).

(4) b'. He was believed to be ill.

In investigating transitions of this kind, the project will link up with existing research on grammaticalization (i.e., the process of language change in which lexical elements are gradually transformed into elements of structure), which is undoubtedly one of the major issues in contemporary historical linguistics.

Output

For each of the three languages under investigation, the synchronic and diachronic study of the relationship between meaning, structure and the lexicon, will be the subject of a Ph.D dissertation by one of the junior researchers. The senior researchers will be responsible for the contrastive-typological aspects of the investigation, and will assist the promoters in the supervision of the language-particular (corpus) research. The ongoing research will be regularly reported upon: we aim to submit at least two articles for publication in internationally-refereed journals per year per person involved in the project. In addition to its theoretical relevance, the submitted project may have a number of practical applications as well. We expect that the project will bring forth a number of potentially relevant insights for the field of automatic language processing (concordancing, alignment software, analysing software, translation software). Language software in general attributes a dominant role to the lexicon in the relationship between meaning and form, thus underestimating the semantic import of the structural component. This may well be one of the reasons for the limited success rate of many applications. A more realistic approach to the relationship between lexicon and structure will probably enhance the performance of language software applications.

3. Methodology and timing

The empirical basis of the submitted project will be an extensive corpus-based database of formal patterns and their meanings. The first phase of the project will consist in the selection of a varied set of relational meanings to be investigated. This selection process will be carried out on the basis of the existing CVVD database, containing hundreds of thousands of both formally and semantically labelled corpus examples from the 27- and 38-million-word-corpora of the Institute for Dutch Lexicology for Dutch, the *Le Monde sur cd-rom* corpus for French and the Lancaster-Oslo/Bergen Corpus and British National Corpus for English. By focussing on elements that recur significantly often in the semantic descriptions of the CVVD lemmas, a varied but manageable set of potentially relevant elements of relational meaning will be compiled (such as possession, transfer, change of state,... cf. Dixon 1991; Willems 2002).

The synchronic study of the formal means used for expressing these selected elements of meaning will make up the second phase of the project. On the basis of large-scale corpus research, a detailed inventory will be made of the verb/structure-combinations displaying the selected relational meanings. This inventory will allow us to investigate which elements of meaning remain the same across all possible combinations of a given verb with different structures, and which elements of meaning remain the same across all possible combinations of a given structure with different verbs. The corpus data will also allow us to analyse the frequencies of structures and lexical items in order to discover relevant correlations (do verbs with vaguer lexical semantics require more structural input, for instance). Only extensive corpus research can provide us with the necessary data for statistically relevant conclusions. We are confident that this will prove to be a fruitful approach to determining the respective contributions of the lexical and structural components to the overall meaning of sentences. The CVVD material will provide a useful starting point, but the increasing amount and diversity of available corpus resources calls for additional corpus research. For instance, while the British National Corpus contains some 10 million words of spoken English, the corpora that have been used for Dutch and French in the present project do not contain any spoken language data. Therefore, we envisage to bring in data from the CGN-corpus ('Corpus spoken Dutch') and the *Corpus de français parlé*. Also, for French, the findings from the *Le Monde* newspaper corpus will be complemented with data from the on-line Frantext corpus of mainly literary texts. Finally, a license for the American National Corpus will be purchased (when this becomes fully available), so as to be able to better represent the internal variation of English.

The findings of these monolingual corpus investigations will have to be systematically compared. This comparison, which is at the heart of Contragram's contrastive concerns, involves analysing the extent to which structures and lexical items are semantically equivalent across languages, putting side by side the frequency data of the three languages, and actively prospecting correlations and possible typological generalisations. We envisage a crucial role for translation corpora in the developing and testing of contrastive hypotheses: the Oslo Multilingual corpus and the Namur corpus seem good candidates for this task.

In the third phase of the project, the focus will shift to the diachronic study of expression means. The data for this diachronic component of the project will be drawn from existing corpora, viz. the Helsinki Corpus of English texts, the *Base de français médiéval* and the Integrated Language Database of 8th-21st century Dutch (when this becomes available), as well as from other historical text resources. Basically, the diachronic research will be carried out with the same aims and methodology as the contrastive part of the project: the comparison of different language stages is generally considered to be part and parcel of contrastive research.

Finally, in a fourth and concluding phase, the findings of the synchronic and diachronic components will be translated into an empirically-justified theory of the relationship between lexicon, structure and meaning in the three languages under investigation.

In the final two phases of the project, the junior researchers will be able to devote a sufficient amount of their research time to the finalisation of their PhD dissertations.

4. References

- Aijmer, K. & A.-M. Simon-Vandenberghe (2003) The discourse particle *well* and its equivalents in Swedish and Dutch. *Linguistics* 41, 6: 1123-1161.
- Aijmer, K. & A.-M. Simon-Vandenberghe (2004) A model and a methodology for the study of pragmatic markers: the semantic field of expectation. *Journal of Pragmatics*. (to appear)
- Colleman, T. (in preparation) De datiefalternantie in het Nederlands: Een corpusgebaseerd onderzoek naar de Nederlandse dubbelobject- en oblieke constructies. Ph.D. dissertation, Ghent University.
- Croft, W. (2003) Lexical rules vs. constructions: A false dichotomy. In Hubert Cuyckens *et al.* (eds.) *Motivation in Language*. Amsterdam: Benjamins. 49-68.
- Defrancq, B. (2002) L'interrogative enchâssée : Sa nature sémantique et ses rapports avec l'élément recteur. Ph.D. dissertation, Ghent University.
- Devos, M. & W. Vandeweghe (2004) Relationele subtypes bij voorzetselobjecten. *Leuvense Bijdragen* (to appear)
- Dixon, R.M.W. (1991) A new approach to English grammar, on semantic principles. Oxford: Clarendon Press.
- Duinhoven, A.M. (1991) Dat siet men wit ende reine wesen. A.c.i.-constructies in het Nederlands. *Nieuwe Taalgids*, 84: 409-30.
- Goldberg, A. (1995) *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Kay, P. & C. J. Fillmore (1999) Grammatical constructions and linguistic generalizations: the *What's X doing Y?* construction. *Language* 75: 1-33.
- Langacker, R. W. (1987/1991) *The Foundations of Cognitive Grammar. Volumes 1 & 2*. Stanford: Stanford University Press.
- Langacker, R. W. (1999), *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.
- Noël, D. (2001) The passive matrices of English infinitival complement clauses: Evidentials on the road to auxiliarihood? *Studies in Language* 25, 2: 255-296.
- Noël, D. (2003) Believe-type matrix verbs and their complements: Corpus-based investigations of their functions in discourse. PhD dissertation, Ghent University.
- Slobin, D. (2000). Verbalised events. In S. Niemeier & R. Dirven (eds.), *Evidence for linguistic relativity*. Amsterdam: John Benjamins. 107-138.
- Talmy, L. (2003) *Toward a Cognitive Semantics*. Cambridge, Mass.: MIT Press
- Willems, D. (2000a) Le rapport entre théorie et données. Le cas du passif en français. In H.-L. Andersen en A.-B Hansen (eds), *Le français parlé* (= *Etudes Romanes* 47), Copenhagen. 13-29.
- Willems, D. (2000b) Objet d'étude, théories et données. Sur la place des corpus dans la recherche linguistique contemporaine. In M. Bilger (ed.), *Corpus, méthodologie et applications linguistiques*. Paris : Champion. 149-155.
- Willems, D. (2002) Classes et relations verbales. A la recherche d'une typologie intégrée. In J. Chuquet en M. Paillard, *Morphosyntaxe du lexique 1*, Travaux linguistiques du Cerlico 15: 285-301.
- Willems, D. (2004), *Typologie des procès et régularités polysémiques*. In D.Bouchard en I. Evrard, *Représentation du sens II* (to appear)