A radically data-driven Construction Grammar: Experiments with Dutch causative constructions[1]

Natalia Levshina[†] and Kris Heylen[‡]
[†]University of Jena/University of Marburg
[‡]University of Leuven,
RU Quantitative Lexicology and Variational Linguistics

Abstract

In this paper we propose a novel, radically data-driven approach to constructional semantics. It is based on Semantic Vector Space models, which are commonly used in computational linguistics to model the semantic relationships between words on the basis of their distribution in a large corpus. In a case study of the near-synonymous Dutch causative constructions with *doen* 'do' and *laten* 'let' we show this method in action by testing a variety of distributional models and clustering options applied to the constructional collexemes. The method opens new perspectives for generating hypotheses about constructional semantics, providing a quick estimation of large amounts of data. The paper also contributes to bridging the gap between the neostructuralist distributional approaches still predominant in computational linguistics, on the one hand, and the non-reductionist constructionist approaches to grammar, on the other hand.

1. The need for objective data-driven semantic classes

Constructions are commonly defined as pairings of form and function (Goldberg 1995; Goldberg 2006). The meaning, understood here as the concept or conceptual structure associated with a construction, is a crucial aspect of the latter's function. Although the meaning of a construction cannot be always reduced to the meaning of its components (e.g. Goldberg 1995), the semantic properties of its slot fillers can be used as a convenient heuristic to access the conventional uses of the construction in question. For instance, the central sense

[1] This research project was partly funded by a grant from the Research Foundation of Flanders (FWO) (G.0330.08) awarded to Dirk Geeraerts and Dirk Speelman, the Quantitative Lexicology and Variational Linguistics Research Unit at the University of Leuven.

of the caused-motion construction (X CAUSES Y to MOVE Z, e.g. *She threw a coin in the Trevi fountain*) commonly involves a verb of directed physical action (*threw*), a movable physical object (*a coin*) and another physical object that can serve as a location (*the Trevi fountain*). Semantic classes of the slot fillers can be helpful in two ways. First, they can indicate the differences between the senses of one construction; second, they may reflect the division of 'semantic labour' between two or more near-synonymous constructions.

Quantitative corpus-based studies of constructional semantics (including lexical semantics) frequently employ semantic classes. Table 1 lists some of the existing quantitative methods and approaches. They differ with regard to the research perspective: the researcher can either focus on the internal semantic structure of a construction, most commonly on its polysemy, or compare the distinctive features of functionally related constructions, for instance, near-synonyms, or 'alternations'. This distinction corresponds to the semasiological and onomasiological perspective, in more traditional semantic terms (Geeraerts et al. 1994). The other distinction is whether the semantic classes are determined *a priori* and form the basis of the subsequent analyses, or they are inferred *a posteriori* in order to interpret the results (e.g. lists of distinctive collexemes). Yet, all those methods involve semantic classes of the slot fillers – exclusively or alongside other semantic features.

	***a priori* semantic classes**	***a posteriori* semantic classes**
semasiological perspective (polysemy)	- Behavioural Profiles of a word's senses (Gries 2006) - Multidimensional Scaling-based semantic maps of linguistic categories (Levshina 2011)	(standard) Collostructional Analysis (Stefanowitsch & Gries 2003)
onomasiological perspective (near-synonymy)	- regressing on functionally similar constructions (e.g. Heylen 2005; Bresnan et al. 2007) - Behavioural Profiles of lexemes (Gries & Divjak 2006) - Correspondence Analysis maps of constructional spaces (Levshina et al. In press)	Distinctive Collexeme Analysis (Gries & Stefanowitsch 2004)

Table 1. Quantitative corpus-based methods in usage-based approaches to constructions (a selection).

However, the use of classifications is often problematic. If a researcher applies an *ad hoc* intuitive classification, (s)he runs the risk of missing some important distinctions or imposing irrelevant ones. Trying to avoid this caveat, many linguists apply ready-made classifications, such as the ones available in Levin (1993) or WordNet (Fellbaum 1998), which are based on more or less definite criteria or conventions. Still, this practice involves several conceptual and practical difficulties. First of all, ready-made conventional classifications are not available for many languages besides English. Second, the existing classifications tend to be incomplete, so that the researcher has to decide what to do with a large chunk of data that fall outside the classifications. In addition, many classifications, such as the WordNet, are tree-like and contain several levels. In this situation, choosing the level of classification granularity (i.e. how deep one should prune the classification tree) becomes an empirical problem. One of the goals of the present paper is to develop a strategy of finding the optimal level of granularity on the basis of objective quantitative criteria.

The greatest problem, however, is that even the conventional classifications are largely introspective. More recently, there have been attempts to classify constructional slot fillers on the basis of large-scale corpus evidence. For instance, Gries and Stefanowitsch (2010) have attempted to classify constructional collexemes with the help of a set of contextual features found in the corpus. The classes were evaluated qualitatively, the main criterion being semantic interpretability of the classes. Yet, if one adheres to the principles of empirical semantics (e.g. Geeraerts 2010a), it is at least just as important to present objective quantitative evidence that the choice of classification is justified by the facts of usage.

In this paper, we propose a novel objective distributional approach based on large-scale corpus data and rich contextual information. The core of the approach is Semantic Vector Spaces (Lin 1998), a method widely used in computational models of language. We demonstrate how the method can be used to choose between hundreds of possible classifications, arriving at the optimal one in terms of parsimony and predictive power for every particular set of near-synonymous constructions. We illustrate how the method works on the 'alternation' of Dutch causative constructions with the auxiliaries *doen* "do" and *laten* "let".

The structure of the article is as follows. In the following section, we introduce the object of the case study, the causative constructions in Dutch. Section 3 presents the general principles of the distributional models of Semantic Vector Spaces, followed by a description of the data and specific models in Section 4. Section 5 reports the results of our classification

experiments. In Section 6, we discuss the results from the constructionist perspective and suggest some steps for future research.

2. Dutch causative constructions

Dutch periphrastic causatives consist of an auxiliary predicate (*doen* or *laten*), an effected predicate and several nominal slots, as shown in the example below:[2]

(1)	*De*	*politie*	*deed/liet*	*de auto*	*stoppen*.
	the	police	did/let.Past	the car	stop
	Causer		Auxiliary	Causee	Effected
			Predicate		Predicate

"The police stopped the car (let the car stop)."

Most of the corpus-based studies of these constructions (Kemmer and Verhagen 1994; Verhagen and Kemmer 1997; Stukker 2005) suggest that *doen* is an auxiliary that expresses direct. It is used to categorise causative situations in which the Causer uses its own energy to produce the caused event encoded by the Effected Predicate. On the other hand, the auxiliary *laten* refers to indirect causation, when "some other force besides the initiator is the most immediate source of energy of the effected event" (Verhagen and Kemmer 1997: 67). The semantics of *laten* also covers situations of letting. In fact, it represents a continuum from coercion to enablement and permission (Verhagen and Kemmer 1997; Speelman and Geeraerts 2009) with some ambiguous cases in between. For example, the construction in (2) suggests two interpretations:

(2)	*Hij*	*liet*	*iedereen*	*zijn*	*romaan*	*lezen.*
	he	let	everyone	his	novel	read

"He had/let everyone read his novel."

[2] Some examples also contain an Affectee, which is the object of the Effected Predicate and the end point of the causation chain, e.g. *the window* in *The strong wind caused the tree to break the window*. Since Affectees expressed by NPs are infrequent in the corpus, the Affectee slot will not be considered in our experiments. See also Stukker (2005), who shows that the semantic classes of Affectees are not relevant for the choice between *doen* and *laten*.

The most typical uses of *doen* are described as physical and affective causation. The former usually involves an inanimate Causer and Causee and a non-mental observable caused event:

(3) *De aardbeving* *deed* *de muren* *trillen.*
the earthquake did the walls shake
"The earthquake made the walls shake."

Affective causation typically involves an inanimate stimulus (Causer), a human cognizer (Causee) and a mental caused event:

(4) *Je kapsel* *doet* *me* *denken aan* *een vogelnest*.
your hairstyle does me think to a bird-nest
"Your hairstyle reminds me of a bird's nest."

As far as *laten* is concerned, its prototype is considered to be inducive causation, with a human Causer affecting a human Causee, normally intentionally and by means of communication (Stukker 2005). An example of inducive causation is given below:

(5) *De trainer* *liet* *de spelers* *loopoefeningen* *doen*.
the coach let the players run-exercises do
"The coach had the players do running exercises."

Therefore, one can expect human Causers to be typical of *laten*, and non-human ones to favour *doen*. This is also what was found in the previous quantitative multivariate studies (Speelman and Geeraerts 2009; Levshina 2011). The inherent semantic classes of the Causee (human being, abstract entity, artifact, etc.) have never shown strong effects in the previous analyses (Levshina 2011). However, the thematic roles of the Causee (quasi-patient or agent) have shown significant effects: *laten* is favoured by relatively agentive Causees, whereas *doen* is associated with patient-like affected Causees. The low relevance of the inherent semantic class of the Causee can be explained by the diverse roles of human Causees, who can be both relatively passive experiencers, as in (4), and active agents, as in (5).

As far as the Effected Predicate is concerned, our previous research (Levshina 2011) revealed a few distinctive verb classes at different levels of semantic specificity, from specific verbs to medium-grained semantic classes *à la* Levin and to highly abstract distinctions. This

is in line with the constructionist approach, which claims that both exemplars and generalisations are stored in the speaker's memory (Langacker 1987; Goldberg 2006: Ch. 3). On the most lexically specific level, some verbs, such as *denken aan* "think of" are used exclusively with *doen*, and some others, such as *weten* "know" and *wachten* "wait", occur predominantly in the combination with *laten*. These expressions form low-level constructional pairings with specific meaning. Some exemplars of this type form clusters. For instance, the verbs of perception (*zien* "see" and *horen* "hear") are normally used with *laten*, whereas most predicates that designate internal mental processes – for instance, belief (*geloven* "believe", *vermoeden* "suppose"), emotion (*vrezen* "fear") and decision (*besluiten* "decide") – tend to occur with *doen*. In addition, verbs of quantitative change along a scale (*stijgen* "rise", *toenemen* "increase") also prefer *doen*. These clusters form the middle level of generalisation. Finally, on the most abstract level, *laten* is in general preferred by semantically and syntactically transitive verbs (*maken* "make", *doden* "kill"), whereas *doen* usually occurs with intransitive verbs with a patient-like first argument (*verdwijnen* "disappear", *smelten* "melt").

To summarise, one can expect a high effect of the semantic classes of the Causer and the Effected Predicate slots, and a weak effect of the Causee slot in predicting the choice between the two constructions. As for the Effected Predicate, it will also be interesting to see which level of granularity will be the optimal one in distinguishing between the constructions.

3. Semantic vector spaces

3.1. Origin

Semantic Vector Spaces (SVS's) have become the mainstay of modeling lexical semantics in Computational Linguistics over the last 20 years. Based on the hypothesis that semantically similar words tend to be used in similar contexts, these corpus-based approaches model the meaning of a word in terms of the contexts in which it appears. They have been applied to a wide variety of computational tasks – from Question Answering and Information Retrieval to automated essay scoring (Landauer and Dumais 1997) or the modeling of human behavior in psycholinguistic experiments (Lowe and McDonald 2000). SVS's were first developed during the so-called statistical turn in Natural Language Processing (NLP) in the 1990's, when NLP moved away from the then prevalent rule-based approach. They addressed the need to model semantics in a bottom-up, automated fashion from large amounts of corpus data, rather than having to rely on the time-consuming manual construction of lexical resources. As such, this

data-oriented development in Computational Linguistics was not unlike the empirical and statistical turn observable today in Theoretical Linguistics, and in Cognitive Linguistics and Construction Grammar in particular. We will argue that SVS's also can be useful in more theoretically-oriented linguistic research in Construction Grammar. Thanks to their fully automatic, bottom-up analysis of the distribution of a word, SVS models are not only able to deal with enormous quantities of data; they also bypass the need for subjective human judgments and may bring to light patterns that escape the human eye.

The origin of SVS's can be traced back to a fundamental linguistic insight already expressed in the 1950's. Back then, a number of linguists and philosophers stressed the dependency, or even the identity, between the meaning of a word and its use. This view inspired John Rupert Firth's quote that "you shall know a word by the company it keeps" (Firth 1957), Ludwig Wittgenstein's "the meaning of a word is its use in the language" (1953), and Zelig Harris' (1954) insight that semantically similar words are used in similar contexts – a view which is now often referred to as the distributional hypothesis. In the (mainly) British tradition of Corpus Linguistics this hypothesis was put into practice by investigating the collocational behavior of words and identifying their idiomatic usage. SVS's can be seen as an extension and generalisation of collocational analysis. Instead of identifying only a restricted number of significant collocations as input for further qualitative analysis, SVS's track a word's co-occurrences with all other words in the corpus, resulting in a sort of over-all collocational profile that is the input for further quantitative analysis. More specifically, the similarity of collocational profiles is measured mathematically. The hypothesis is that words with a similar collocational profile will be semantically related and can thus be grouped into semantic classes.

3.2. Practical implementation

In practice,[3] SVS's record the co-occurrence frequencies of a set of target words with a large set of context words in a given window around the target words. The choice of target words depends on the task and can range from all words in the corpus (e.g. to automatically identify taxonomic relations in the whole of the vocabulary), or it can be limited to a set of words, like in our case, where we want to group only the nouns and verbs occurring in the Dutch causative constructions into semantic classes. The choice of context words can be said to be dependent on how well they are able to represent the semantics of the target words. A stop-list

[3] See Turney and Pantel (2010) for on overview of implementations and applications.

of highly frequent (function) words is usually excluded because they occur with almost all words and cannot therefore discriminate one set of semantically related words from another. Context words with very low frequencies simply do not occur with enough target words to be a basis of comparison. Most SVS's therefore use a few thousand highly frequent context words minus a stop list of function words. The co-occurrence frequencies between target and context words are stored in vectors and collected in a large matrix. Table 2 illustrates such a co-occurrence matrix with a handful of context words. In reality, the matrix is high-dimensional with thousands or target and context words: The length of the vectors, i.e. the number of columns, is equal to the number of context words, and the number of vectors (the rows) to the number of target words. As shown, many co-occurrence counts are zero, making matrices usually quite sparse. This is because words tend to co-occur with a limited set of context words, which is exactly the property that allows the technique to capture word semantics through context. In the toy matrix in Table 2, it is clear that *kiss* and *hug* must be semantically related because they have high co-occurrence frequencies with the same context words (*lovingly*, *mother*, *lovers*). The same holds for *kill* and *murder* that share high co-occurrence frequencies for *gun*, *psychopath*, *knife* and *cruelly*. *Soap*, on the other hand, has high co-occurrence frequencies with very different context words and thus is not related.

	gun	*psychopath*	*knife*	*cruelly*	*lovingly*	*mother*	*lovers*	...	*detergent*
kiss	2	2	0	0	89	56	98	...	0
hug	3	1	2	5	77	49	88	...	0
kill	10	59	67	69	0	8	12	...	1
murder	97	65	58	81	0	9	9	...	0
...	...	...				...	...	...	...
soap	0	0	0	0	1	0	1	...	67

Table 2. A matrix with imaginary co-occurrence frequencies of target words (rows) and contextual features (columns).

To capture these collocational properties even better, the raw co-occurrence frequencies are usually weighted to represent collocational strength (e.g. Pointwise Mutual Information or Log Likelihood Ratio). This has the effect of giving a higher weight to very informative context words, i.e. those that co-occur only with a limited set of semantically related target words. For the vector comparison, Semantic Vector Spaces use a geometrical approach (hence Vector *Space*): the weighted co-occurrence frequencies can be seen as co-

ordinates defining a point in a high-dimensional context feature space. Points closer together in the space are then semantically more related. Figure 1 shows a 2D subspace of the high-dimensional space where *kiss* and *hug* are close together based on their shared relatively high co-occurrence frequency with *lovingly* and relatively low co-occurrence frequency with *cruelly*, and vice versa for *kill* and *murder*.

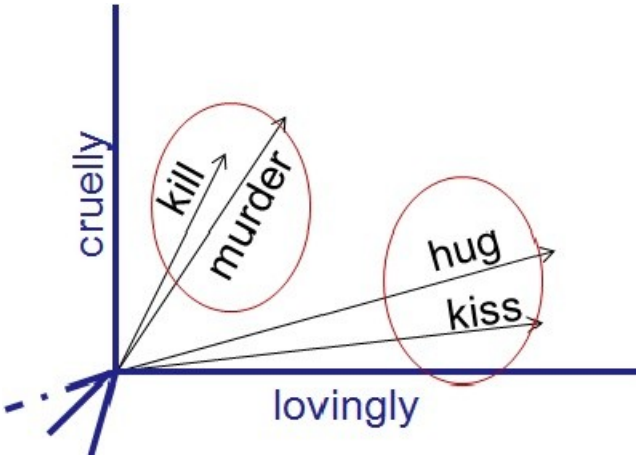

Figure 1. An imaginary 2D subspace of semantic vectors.

As a proximity measure, most implementations use the cosine of the angle between the word vectors. If the angle is small, like between *kill* and *murder,* the cosine will be close to 1, indicating high similarity. If the angle is large, like between *kill* and *kiss*, the cosine is close to 0, indicating low similarity[4]. The computation of all pairwise cosine similarities between word vectors results in a word-by-word similarity matrix, shown in Table 3. Since the cosine is a symmetric similarity measure (the cosine of vector A with vector B is the same as the cosine of vector B with vector A), also the matrix as a whole is symmetric with 1's on the diagonal (a target word is always completely similar with itself). Based on the similarity matrix, we can now derive for each target word a similarity ranking of all other target words. Depending on the application, the top most similar words can then be synonym candidates for inclusion in a thesaurus, or possible replacements in a search query to a Question Answering system. Important for this paper is that the similarity matrix can also be the input for a clustering algorithm that will try to identify groups of similar target words and hence semantic classes.

	kiss	*hug*	*kill*	*murder*	...	*soap*
kiss	1	0.88	0.24	0.19	...	0.08

[4] See Weeds, Weir & McCarthy 2004 for an overview and precise mathematical characterisation of different similarity measures, including the cosine.

hug	0.88	1	0.18	0.26	...	0.11
kill	0.24	0.18	1	0.91	...	0.14
murder	0.19	0.26	0.91	1	...	0
...	...	...	...	...	1	...
soap	0.08	0.11	0.14	0	...	1

Table 3. Pairwise cosine similarities between word vectors (imaginary data).

3.3. Different definitions of Context

Although all Semantic Vector Spaces share the basic ingredients described above, they come in many different flavors and implementations. To start with, there is a wide variety of possible settings for the more technical parameters like the frequency weighting function, the feature cut-off function or the similarity measure. However, the most important difference between the various implementations is how they define the notion of *context*. Recall that the basic principle behind SVS's is that a word's meaning can be modeled through the context it appears in. How context is defined, will necessarily influence the kind of semantics that is captured. In the description above, context was simply defined as the words co-occurring in a window around the target word, e.g. 5 words to the left and 5 words to the right. However, many other context definitions exist. A first obvious extension is varying the size of the context window. The extremes on this spectrum are, on the one hand, models that simply take the whole document as context, in which case words are similar if they tend to occur in the same documents. On the other hand, models can use simple bigrams, where only one word on the left or right is taken into account. Previous research (Peirsman, Heylen & Geeraerts 2008) has shown that larger context windows lead to SVS's that capture looser, more associative semantic relations like doctor-hospital or bird-sky. Smaller context windows tend to find tighter, taxonomic relations like doctor-nurse (co-hyponym) or bird-robin (hyponym).

A second important aspect in the definition of context is the type of relation that holds between target and context words. In the models discussed so far, the relation is simply one of proximity without any further distinction between the context words within the chosen window. These models are therefore also called bag-of-words models. However, these models also record co-occurrences that are not very relevant or informative for the semantics of a given target word.

(6) *The teacher was startled by a barking dog on her way to work.*

In a sentence like (6), a bag-of-words model would include for the target word *dog* context words like *barking* – which is very informative – but also *teacher*, *startle*, *way* and *work* – which are much less informative. Therefore so-called dependency-based SVS models restrict the possible relation between target and context words by their syntactic or dependency relation. For example, an SVS can be construed to only include modifiers of the target noun, so that in sentence (6) only the highly informative *barking* is counted as context word. This results in fewer, but more semantically relevant context words. Of course, different dependency relations are relevant for different parts of speech: modifying adjectives are relevant for nouns but not for verbs. Therefore, these dependency-based SVS's are constructed for each part of speech separately. However, multiple dependency relations that are relevant for the same part of speech can be combined in a single SVS; A model for verbs can include adverbs, subject nouns, direct object nouns, prepositional complement nouns etc. In this case, most models define a context feature as the tuple [dependency relation, context word], so that [subject, dog] will be a different context feature from [object, dog]. In other words *dog* will be treated as a different context feature for *bite* in the sentences *A dog bites a man* and *A man bites a dog*. Additionally, the dependency relation between context and target word need not be binary: It can in principle consist of a dependency path of arbitrary length. In practice, only paths of three are commonly used to capture dependency relations involving prepositions, e.g. prepositional complements to verbs or modifying prepositional phrases to nouns.

(7) *We listen to the radio*.

In a sentence like (7), the target verb *listen* then has the context feature [object, to, object, radio], indicating that *listen* has a prepositional complement *radio* introduced by the preposition *to*[5]. Previous research (Heylen et al. 2008, 2009) has shown that dependency-based models tend to find even tighter semantic relations than the small-window bag of words models. They are especially good at identifying near-synonyms like hospital–clinic or monkey–ape.

Finally, a third way of defining context is specifically relevant for verbs and relies purely on a verb's syntactic behavior. Whereas the previous approaches all looked at the

[5] See Padó & Lapata (2007) for an overview of possible applications.

lexical context of a target word (be it syntactically restricted or not), these models only use the schematic syntactic slots or *subcategorisation frames* that a verb occurs with.

(8) *Ellen bought her girlfriend a book for her birthday.*

In (8), the context feature extracted for *give* would be the tuple [subject, indirect object, direct object, prepositional adjunct], indicating only the attested syntactic arguments without any lexical information. Subcategorisation frame SVS's have been developed in a separate research tradition (see Schulte im Walde 2009 for an overview) based on the insights of (neo)structuralists like Lucien Tesnière, Juri Apresjan and Beth Levin that a verb's meaning is closely connected to its syntactic behavior and the arguments it governs. More specifically, semantically similar verbs are said to show the same alternation patterns in the syntactic arguments they take, which allows to derive semantic verb classes similar to Levin's (1993) classification. The implementations vary further in two respects. Firstly, they differ in the number of possible argument types taken into consideration. Some only look at the bare NP arguments (subject, direct object, indirect object), others include all arguments (prepositional complements, obligatory adverbs etc.), and the most inclusive models also take into account adjuncts (e.g. temporal, locational or benefactive adverbial expressions). Secondly, subcategorisation-frame models also differ with respect to the amount of lexico-semantic information they take into account. As discussed above, the basic approach is to include only schematic syntactic valency information, but some models add some high-level semantic information: this might be semantic classes of the NP arguments, e.g. distinguishing between animate and non-animate arguments, or information about the specific preposition introducing a prepositional complement. The extracted context feature for *buy* in (8) then becomes [animate subject, animate indirect object, inanimate direct object, inanimate prepositional adjunct introduced by *for*]. Previous research indicates that subcategorisation frame models are indeed able to capture verb classes based on event-type, like manner-of-motion verbs, position verbs or perception verbs (Schulte im Walde 2006).

On a more theoretical note, it is necessary to make a qualification regarding the status of the dependencies. SVS's and most other computational distributional models originate from the structuralist reductionist tradition of relational semantics (see Geeraerts 2010b: 165–170), which assumes the existence of the categories like 'subject' and 'direct object' and the corresponding dependency relations. However, these categories have a controversial status in the contemporary usage-based constructionist approaches. Whereas Langacker's Cognitive

Grammar interprets such categories in an essentialist way, as primary and secondary focal points (e.g. Langacker 2005), the non-essentialist Radical and Cognitive Construction Grammars question this assumption (e.g. Croft 2001; Goldberg 2006: Ch. 10). However, they do not claim that there can be no generalisations over similar arguments, which share the same form, in different constructions. It is necessary to emphasise, though, that the cognitive reality of such categories and the criteria for their demarcation need empirical support (Goldberg 2006: 223). Our use of the pre-determined categories and corresponding dependencies in this study is explained by solely historical and practical reasons: most computational models, including syntactic parsers, use this kind of neostructuralist approach. Still, we hope that these categories and dependencies provide an acceptable approximation of the possible generalisations over the similar constituents and relationships across different constructions.

Different definitions of context will result in Semantic Vector Spaces that capture different types of semantic relations. A clustering based on these SVS's will therefore also lead to different groupings of nouns and verbs into semantic classes. It's not clear *a priori* which semantic classification is the most relevant for modeling the choice between the causative constructions in Dutch. Additionally, it is also unclear how many semantic classes we need to distinguish to model the constructional variation. In this paper, we will therefore construct a range of SVS's with different context definitions, both for the nouns that fill the Causer and Causee slot and for the verbs that fill the Effected Predicate slot. Based on these SVS's, we will cluster these nouns and verbs at different levels of granularity and test which classification predicts the use of *laten* and *doen* the best. The following section describes the procedure in detail.

4. Data and design

4.1. Data

For this case study, we took 6863 occurrences of the constructions with *doen* and *laten* from two large corpora of Dutch: the Twente News Corpus (Ordelman et al. 2007) and Leuven News Corpus.[6] The newspaper data consisted of equal samples of articles about politics, economy, music and football. Because the corpora were syntactically parsed, the information

[6] Leuven News Corpus is a large corpus of contemporary newspaper Dutch in Flanders. The corpus was compiled by the Quantitative Lexicology and Variational Linguistics Research Unit at the University of Leuven.

about the lexical fillers of the three main slots – the Causer, the Causee and the Effected Predicate – was extracted automatically and later checked manually. Some of the nominal slots were empty, especially the Causee in transitive constructions, as in (9):

(9) *Ik* *liet* *het huis* *bouwen.*
I let the house build
"I had the house built."

The objects of the classifications were all explicit non-pronominal slot fillers treated as types, or lemmata. In total, we obtained 2700 common and proper nouns in the function of the Causer, 1810 nouns filling the Causee slot, and 1155 verbs in the function of the Effected Predicates.

4.2. Distributional classes

The Semantic Vector Spaces were all constructed from the Twente Nieuws Corpus (380 million words, see Ordelman et al. 2007)[7]. For the SVS's using syntactic dependency information, we used the version parsed with the Alpino dependency parser (van Noord 2006) whose dependency triples have been shown to be 90% accurate in the Twente Nieuws Corpus (Plank & van Noord 2010). Following our discussion in 3.3, we built SVS models with 3 different types of context definition and with the slot fillers of the Dutch causative construction as target words:

1) bag-of-words models for nouns and verbs;
2) dependency-based models for nouns and verbs;
3) subcategorisation frame models for verbs only.

For the nouns filling the Causer and Causee slots in the *doen* en *laten* constructions, we only constructed one bag-of-words SVS and one dependency-based SVS (subcategorisation frame models are not relevant for nouns). This choice was based on previous research (Heylen et al. 2008, Peirsman et al. 2009) where these two models gave the best performance in finding tight semantic relations. The bag-of-words SVS had a relatively small context window of 5 words left and right of the target noun. The dependency-based SVS used 8 dependency relations distinguished by the Alpino parser that a noun can be

[7] No separate SVS's were constructed for Belgian Dutch (Leuven News Corpus). As a consequence, some typically Belgian Dutch verbs or verb meanings might have been disregarded.

involved in. These relations are listed in Table 4 with examples (the target noun is in italic and the context feature resulting from the dependency relation is underlined).[8] In the first four relations (su, obj1, pc, advPP), the target noun is regarded as the dependent element and the governing verb is counted as a context feature. In the next three relations (pmPP, adj, app) the noun is the head and the dependent adjective and/or noun is counted as a context feature. Finally, the co-ordination relation is a symmetric one and always generates two target noun/context noun pairs.

Abbr.	Dependency relation	example
su	subject	De *baby* slaapt
obj1	direct object	Hij eet een *appel*
pc	prepositional complement	Ze luistert naar de *radio*
advPP	adverbial prepositional phrase	Hij woont in een *dorp*
pmPP	post-modifying prepositional phrase	het *meisje* met de jurk
adj	adjective	de gelaarsde *kat*
app	apposition	de *koningin*, een wijze vrouw
cnj	co-ordination	de *krekel* en de mier de krekel en de *mier*

Table 4. Syntactic dependency relations of nouns. The target noun is in italic and the context feature resulting from the dependency relation is underlined.

Because previous research has shown several possible classification criteria of the Effected Predicates (see Section 2), we focused on exploring different Semantic Vector Spaces for verbs. Within each of the 3 context definition types, we therefore varied the specific number and type of context features. Within the bag-of-word models, we varied the size of the window around the target verb and constructed a first model with a relatively small window size of 4 context words to the left and right, and a second model with a relatively large window of 15 words on either side. Within the dependency-based models we varied the number of different dependency relations. Based on the Alpino dependency parser, we can distinguish 23 different dependency relations that a verb can engage in. They are listed in Table 5. A first model only takes context features into account that are based on the 3 bare NP arguments (su, obj1, obj2). A second model uses all 7 arguments to extract context features (su, obj1, obj2, pc, ld, ldprep, predc) and, finally, the third model took all 13 dependency relations with a lexically full element into account (su, obj1, obj2, pc, ld, me, ldPP, adv, advPP, predm, predc, invomte, invte)[9].

[8] For a full description of the Alpino parsing scheme, see van Noord (2006).

Abbr.	Relation	Example
su	subject	Het meisje *slaapt*
sup	cataphoric subject	Het *blijkt* dat...
obj1	direct object	Hij *eet* een appel
pobj1	cataphoric object	...of ik het *betreur* dat...
obj2	indirect object	Ze *geeft* papa een kus
se	reflexive	Hij *schaamt* zich
svp	separable affix	Je *lacht* me uit
pc	prepositional complement	Ze *luistert* naar de radio
me	measure complement	Het *kost* 20 euro
ld	locative complement	Ze *werkt* thuis
ldPP	locative prepositional phrase	Ze *rijdt* naar huis
adv	adverbial complement	Je *zingt* goed
advPP	adverbial prepositional phrase	Hij *komt* over 2 weken
predm	predicative modifier	Hij *kwam* dronken thuis
predc	predicative complement	Dat *smaakt* lekker
ccl	complement clause	Hij *zegt* dat hij komt
cclof	complement clause (choice)	Ze *vraagt* of je komt
cvte	complement verb	Z *staat* te praten
cvom	complement verb (goal)	we *reizen* om te leren
invaux	auxiliary verb	ik kan *lezen*
invte	semi-auxiliary verb	hij ligt te *slapen*
invomte	semi-auxiliary (goal)	ik probeer om te *slapen*
invaanhet	progressive marker	Ik ben aan het *lezen*

Table 5. An overview of syntactic dependency relations of verbs. The target verb is in italic and the context feature resulting from the dependency relation is underlined.

Within the subcategorisation frame models, we varied both the syntactic positions that could be included in a subcategorisation frame and the amount of lexico-semantic information about the elements filling the syntactic positions. The syntactic positions are based on the same 23 dependency relations in Table 5. Again, we made 3 subdivisions for the types of relations: the 5 bare NP arguments (su, sup, se, obj1, obj2); all 9 arguments (su, sup, se, obj1, obj2, pc, ld, ldprep, predc); all 23 dependency relations. The lexico-semantic information on the syntactic position fillers could be of 4 types: 1) no lexico-semantic information; 2) the specific preposition introducing the dependent in pc, ldPP and advPP; 3) the semantic class of a dependent noun; 4) both the specific preposition and the semantic noun class. For the semantic noun classes, we used the second highest ancestor of the noun in the Dutch WordNet (Vossen 1998). This resulted in 11 semantic noun classes: animate being, object, situation, action, utterance, property, thought, part, group, place, time.

[9] Since dependency-based models, like bag-of-words models only select lexically full words as context words (excluding function words), we only took those dependency relations into account where there is a lexically full dependant of the verb (a noun, adjective or other verb). This excludes relations like cataphoric object (pobj1) or reflexive (se), but also the clausal arguments where the dependent is not a single lexically full word.

If the noun was not present in the Dutch WordNet, we reverted to a syntax-only subcategorisation frame feature. If the noun belonged to more than one semantic class (because of polysemy), the most frequent overall class was used.

An overview of the models can be found in Table 6, together with the abbreviations used in the rest of the article. For all models, the maximum number of context features was restricted to the 4000 most frequent context features (excluding 122 function words). Both target and context words were processed on the lemma level (i.e. generalising over word-forms). In all models, the co-occurrence frequencies were weighted with the Pointwise Mutual Information index. For all the SVS models, the similarity between target word vectors was measured with the cosine. This resulted in 16 similarity matrices for verbs, and 2 similarity matrices for nouns.

	Causer and Causee (nouns)		**Effected Predicate (verbs)**	
Context definition	**feature selection**	**no. clusters**	**feature selection**	**no. clusters**
Bag of words	5 words left and right: *BOW5*	2–100	4 words left and right: *BOW4*	5–100
			15 words left and right: *BOW15*	5–100
Dependency-based	8 dependencies: *DEPREL8*	2–100	3 dependencies: *Vbarel*	5–100
			7 dependencies: *Varel*	5–100
			13 dependencies: *rVrel*	5–100
Subcat. frame	-	-	Syntax only	
			5 dependencies: *5syn*	5–100
			9 dependencies:: *9syn*	5–100
			23 dependencies: *23syn*	5–100
			Preposition information	
			9 dependencies: *9relprep*	5–100
			23 dependencies: *23relprep*	5–100
			Sem.Class information	
			5 dependencies: *5sclass*	5–100
			9 dependencies: *9sclass*	5–100
			23 dependencies: *23sclass*	5–100
			Prep. + Sem.Class information	
			5 dependencies: *5richsubcat*	5–100
			9 dependencies: *9richsubcat*	5–100
			23 dependencies: *23richsubcat*	5–100

Table 6. An overview of the models and classifications.

The similarity matrices were the input for a hierarchical cluster analysis (Everitt et al. 2001) that groups the noun and verb lemmata into semantic classes. We experimented with different numbers of classes, ranging from 2 to 100 for the Causers and Causees (all numbers from 2 to 10 and then intervals of 5, totalling 27 different clusterings), and ranging from 5 to 100 for the Effected Predicates (intervals of 5, totalling 20 different clusterings). Together with the different context definitions and feature selection criteria, this gives 54 possible semantic classifications of the Causer and Causee nouns and 240 possible semantic classifications of the Effected Predicate verbs.

4.3. The objective criterion for evaluation of classifications

The next question is, how to choose the optimal classification from this richness? In previous research, the classifications were evaluated against an *a priori* manually created 'gold standard' (Schulte im Walde 2006) or *a posteriori* intuitively (Stefanowitsch and Gries 2010). In this study, we propose an entirely objective and data-driven criterion. The evaluation of the models is carried out according to the power of the classes in predicting the use of *doen* or *laten* in the above-mentioned sample of 6863 observations. Obviously, the predictive power of a classification will be greater if the lexemes that belong to one and the same class will also tend to be used in the contexts with only one of the two auxiliaries. In other words, the predictive power is actually an operationalisation of the success of discrimination between the lexemes that tend to be used with *doen* and the ones that are attracted to *laten* (cf. distinctive collexemes in the Distinctive Collexeme Analysis developed by Gries and Stefanowitsch 2004).

In this case study, we used several well-known statistical measurements of the predictive power: C, Somer's D_{xy}, Nagelkerke's R^2, Gamma and AIC (Hosmer & Lemeshow 2000; Baayen 2008). All statistical analyses were performed in R (R Development Core Team 2010). Since most of the parameters displayed very similar behaviour, in the discussion of the results we will limit ourselves to the concordance index C, which is believed to be one of the most objective estimators (Hosmer and Lemeshow 2000: 160–164). This statistic usually falls in the range between 0.5 (random prediction) and 1 (perfect prediction). If $C < 0.7$, this suggests no discrimination; if $0.7 \leq C < 0.8$, the prediction is acceptable; if $0.8 \leq C < 0.9$, the model has excellent discrimination; and if $C \geq 0.9$, the prediction is outstanding. A good model combines a low number of classes with high predictive power. The results of our experiments are presented in the next section.

5. Results of classification experiments

In this section, we discuss the results of the classification experiments for every slot individually. In the last subsection, we also present the results for all three slots taken together.

5.1. Classification of the Causers

As mentioned in the previous section, we had two SVS models of the Causer nouns: the one with the lexical information only (*BOW*), and the one where the lexical information was enriched with the syntactic information about the eight dependency relations (*DepRel8*). For both models, we also tested different clustering solutions with the number of classes from 2 to 100. Figure 2 shows how the *C* index rapidly goes up from the very beginning, which indicates that the relevant semantic distinctions are captured by a relatively small number of classes. The syntactically enriched model performs much better than the simple bag-of-words model. This finding corroborates the results in Gries & Stefanowitsch 2010 (Section 2.2), which also compared a bag-of-words model with a syntactically more precise one. The starting value for the syntactically enriched model with two classes only is already 0.69, and for six classes it is 0.80, which is considered to be good. The 100-cluster syntactically enriched solution has the highest value ($C = 0.89$), but its predictive power is not dramatically different from the more parsimonious classifications with a smaller number of classes.

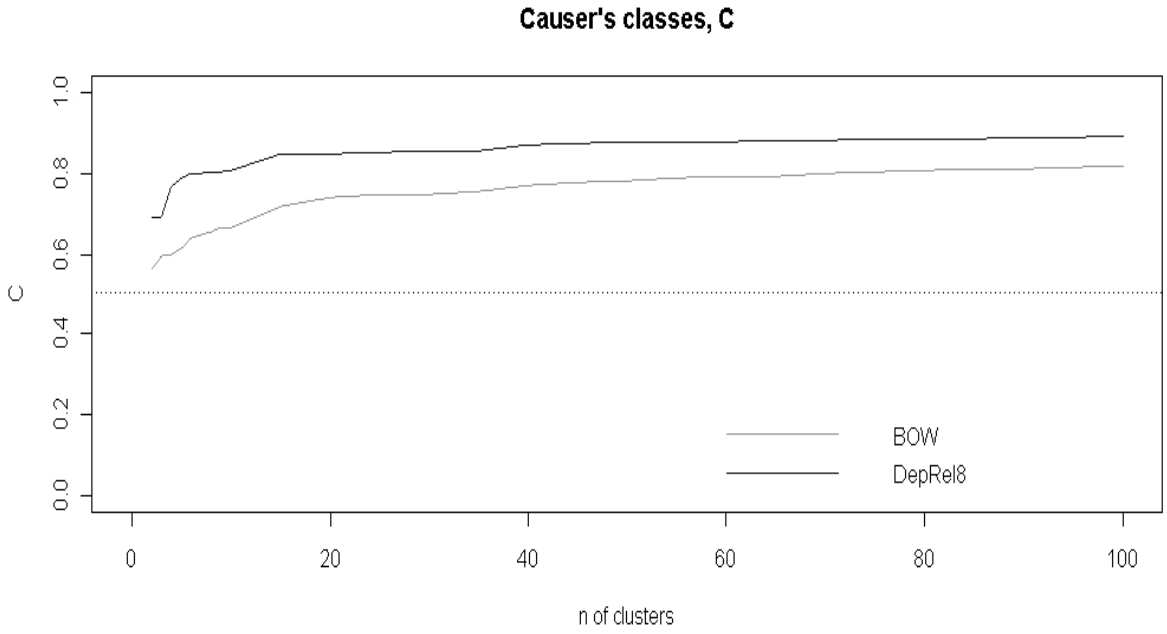

Figure 2. The Causer: predictive power of two models, for different number of classes.

Let us consider the classification with 6 clusters, which is quite successful in discriminating between the *doen*- and *laten*-observations. The classification includes a cluster with predominantly inanimate concrete and abstract nouns: *cd* "cd", *cijfer* "digit", *herstel* "recovery", *stem* "voice", *aanslag* "attack", *afwezigheid* "absense", *resultaat* "result", etc. Cluster 2 (the numbering is arbitrary) contained mostly football- and music-related nouns denoting people and organisations: *Feyenoord* (a football club in the Netherlands), *dirigent* "conductor", *speler* "player", *orkest* "orchestra" etc., although there were a few exceptions, such as *beurs* "stock exchange", which comes from economy-related articles. Cluster 3 included some proper names of conductors and common and proper nouns denoting political and other agents: *Gergiev* (a Russian conductor), *Van Hecke* (a Belgian politician), *secretaris-generaal* "General Secretary", *Harnoncourt* (an Austrian conductor) etc. Cluster 4 contained many geographical names, which are frequently used in newspaper articles to refer to the government metonymically: *Verenigde Staten* "the US", *Amerika*, *Washington*, etc. Cluster 5 included mostly common nouns, which denote people in charge and organisations: *regering* "government", *minister* "minister", *bedrijf* "company", *trainer* "trainer". The sixth cluster contained only 7 nouns with very low collocability due to an extremely low frequency.

The majority of the observations that contain the Causers from Cluster 1 (inanimate entities) are instances of the construction with *doen*, whereas the nouns from Clusters 2–5 (people and organisations) occur more frequently with *laten*. Cluster 6 was too small for

evaluation. The findings therefore support the results of the previous studies. The distinction between animate and inanimate Causers had very high predictive power in all previous multivariate analyses.

5.2. Classification of the Causees

The same models were tested on the Causee nouns. Neither of them showed much predictive power, as displayed in Figure 3, although the predictive power slowly grows with the number of clusters. This is not surprising: the higher the granularity, the better the individual observations are fitted, but at the cost of parsimony. This poor predictive power (maximum C = 0.74) corroborates the previous studies, according to which the inherent semantic classes of the Causee are largely irrelevant, in contrast with the thematic role of the participant in bringing about the effected event.

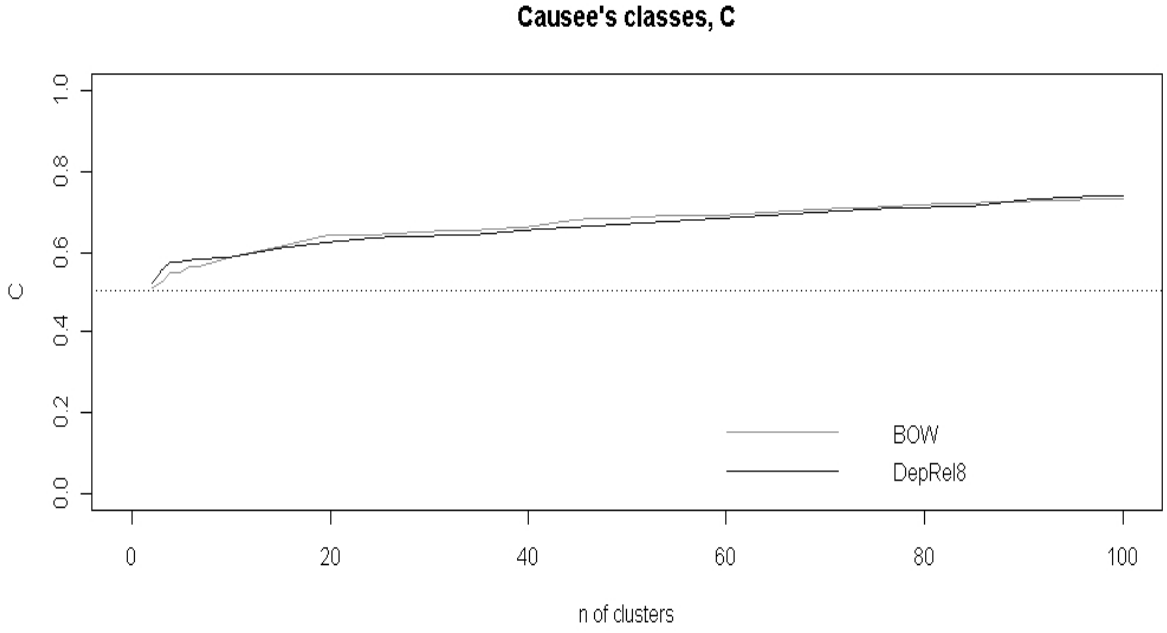

Figure 3. The Causee: predictive power of two models, for different number of classes.

5.3. Classification of the Effected Predicates

Finally, let us consider the Effected Predicate. Figure 4 shows the predictive power of 16 models. According to the analysis, the best-performing model was *23syn* (the upper line), the model with information about the subcategorisation frames based on 23 syntactic relations without any additional information, although some other models were more successful for a

very small number of classes, as one can see from Figure 4. It is interesting that the model *9richsubcat*, which was the leader when the number of clusters was very small ($C = 0.68$ with 5 classes), contained information about the subcategorisation frames enriched with the information about the prepositions and semantic noun classes. As the number of clusters grew, the leadership was taken over by other models. Namely, the next two leaders (for 10 and 15 clusters) were based on the subcategorisation frames enriched with the information about the prepositions (*9relprep* with 10 clusters, $C = 0.73$) and semantic noun classes (*23sclass* with 15 clusters, $C = 0.79$). From 20 verb classes on, the simple subcategorisation frames based on 23 dependencies without additional information yielded the best results (*23syn* with 20 clusters, $C = 0.83$), although the model was closely followed by *9relprep*, which involved subcategorisation frames based on 9 syntactic relationships and prepositions, especially for large numbers of clusters. The bag-of-words models performed on average worse than most other models, although the large window model (*BOW15*) was slightly more successful than the small window one (*BOW4*). Another poorly performing model was *9sclass* (subcategorisation frames based on only 9 syntactic relations, enriched with the information about noun classes). All this suggests that the abstract constructional information is vital for a successful verb classification, and that this information should be very detailed and go beyond the main arguments. We can also conclude that the semantic and prepositional information does not add much, and even makes the classification less successful, when we have 20 and more classes.

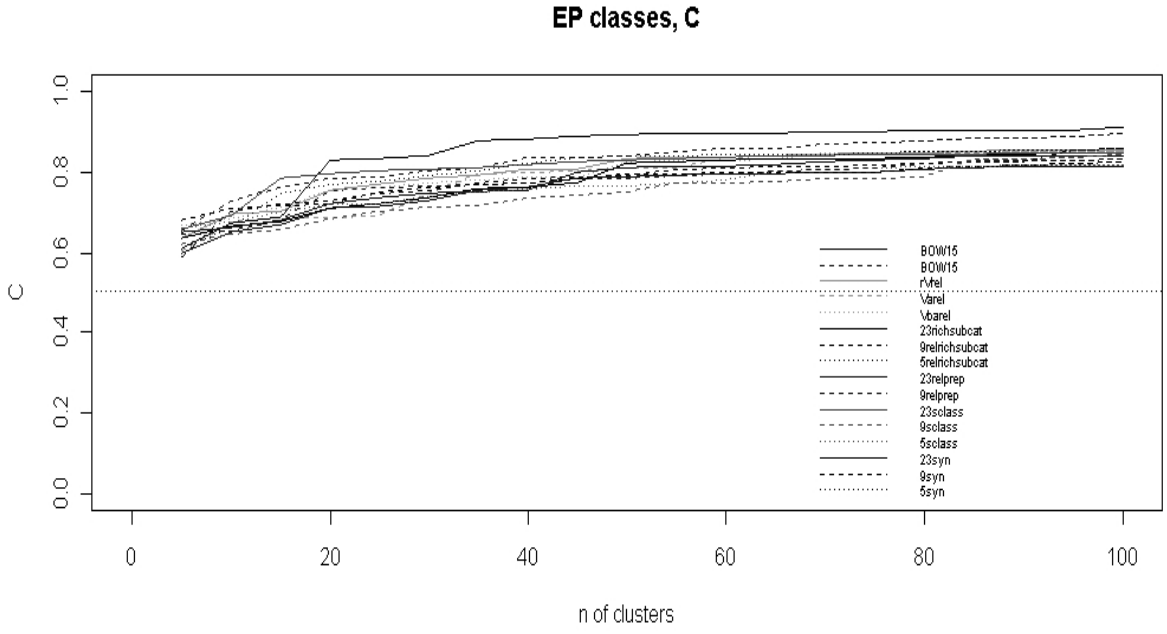

Figure 4. The Effected Predicate: predictive power of sixteen models, for different number of classes.

The increase in the predictive power was gradual, so that the optimal classification into 35 clusters ($C = 0.88$) – the number after which the predictive power does not increase substantially – was medium-grained. Only three of these clusters contain verbs that predominantly co-occurred with *doen* in the test sample. These classes are interpretable as verbs of qualitative and configurational change of state (e.g. *herleven* "come to life again", *kantelen* "tip over", *smelten* "melt", *verslappen* "weaken", *vervagen* "fade"), verbs of quantitative change along a scale (e.g. *stijgen* "go up", *dalen* "go down", *groeien* "grow", *zakken* "fall") and verbs of various mental processes, emotions and beliefs (e.g. *denken* "think", *vermoeden* "suppose", geloven "believe", *besluiten* "decide, conclude", *vrezen* "fear", *hopen* "hope").

Most of the clusters showed a higher proportion of *laten*, as was the case with the Causer (see 5.1). One of them was the cluster containing predominantly verbs of communication: *schrijven* "write", *uitleggen* "explain", *adviseren* "advise", *vertellen* "tell". Another cluster contained many verbs of change of possession and possessional deprivation: *geven* "give", *ontnemen* "take, rob", *leveren* "deliver", *verkopen* "sell". Yet another one mainly consisted of verbs of searching, active perception and testing, e.g. *onderzoeken* "explore", *bekijken* "have a look (at)", *toetsen* "test". Most of these imply a volitional human Causee and a human Causer, who represents an authority and gives orders (10). The presence of such semantic frames and scenarios of causation implies that the combinations of the three slot fillers are not arbitrary. From this follows that the combined effect of the semantic classes of the Causer, Causee and Effected Predicate on the choice of the auxiliary is probably not additive. We will come back to this observation later.

(10) *Obama laat alle kerncentrales VS onderzoeken*.
Obama lets all nuclear-stations US check
"Obama orders to check all nuclear power stations in the US."

Another cluster contains verbs related to putting and bringing (*leggen* "lay", *zetten* "set", *stellen* "put" , *brengen* "bring"), which imply an active Causee that brings about a change in the location or position of another entitity. Yet another one had verbs of motion, such as *draaien* "turn round", *glijden* "slide", *rijden* "ride", *vliegen* "fly", which also involve a certain degree of autonomy on the part of the Causee.

However, many other classes were more difficult to interpret. For instance, one of the clusters with a very high proportion of *laten* contained verbs of perception (*zien* "see", *voelen* "feel"), the verbs *weten* "know", *kennen* "know, be acquainted", *maken* "make", *doen* "do", *blijken* "appear", *schijnen* "shine, appear", *zijn* "be", *worden* "become" and *hebben* "have". It is difficult to say why these verbs go together. Probably this is an effect of the overall high frequency and broad constructional repertoire of these verbs. In this respect, this cluster was similar to another one, which contained the semi-auxiliary or light verbs prototypically related to motion or position: *gaan* "go", *zitten* "sit", *staan* "stand", *vallen* "fall", *komen* "come". Note that many of these verbs in the vague clusters are strongly associated with *laten* (see Section 2).

In general, many of these classes strikingly resemble Levin's (1993) classification. This similarity can be explained by the fact that both Levin's 'alternations' and subcategorisation frames reflect the distribution of verbs in constructions. However, our approach is finer-grained (more possible constructions are examined), probabilistic and treats every construction on its own, not as a part of an alternation pair.

5.4. Combined classes of three slots

In addition, we compared the models with the three slots taken together. The 5 to 100 classes of the Effected Predicates were combined with 2, 5, 10, 15, 20 and 30 classes of the Causer and the Causee. The best-performing models only were tested: *depRel8* for the Causer and *23syn* for the Effected Predicate. For the Causee, also *depRel8* was chosen, although it did not perform better than the bag-of-words model. The *C* index for these different classifications is displayed in Figure 5.

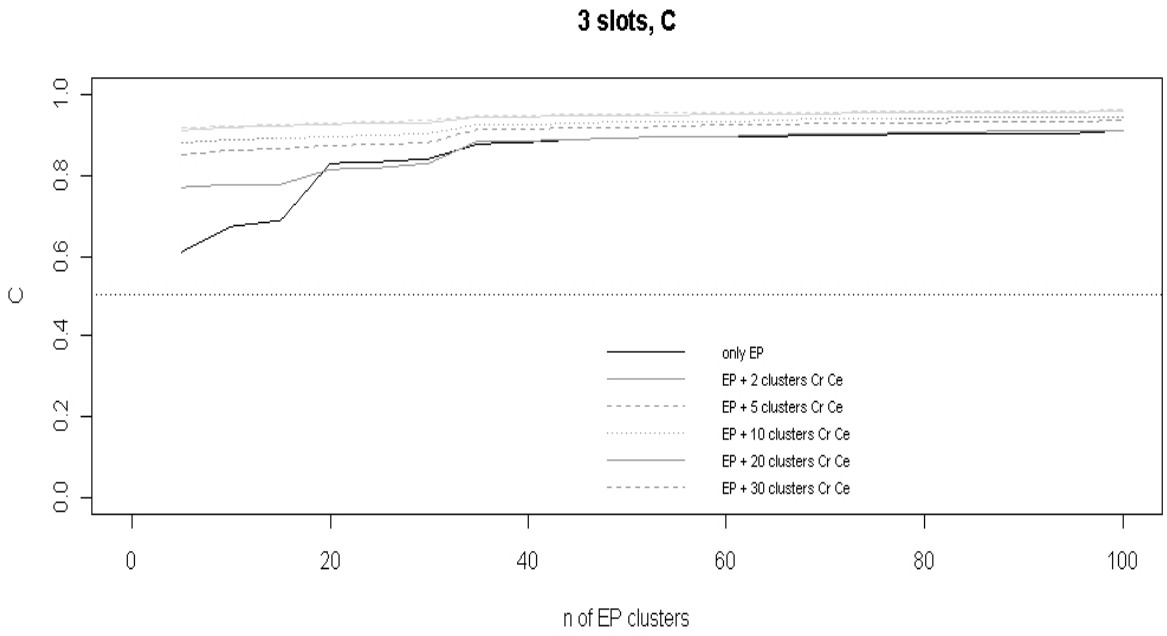

Figure 5. Thee slots together: predictive power of the best-performing individual classifications.

The results of the combined classifications show that adding the semantic information about the Causer and Causee to the information about the Effected Predicate does increase the predictive power, but the effect is non-additive. The effect is significant for the small number of the Effected Predicates, but as the verb classes become finer-grained and more successful in the prediction, the impact of the nominal classes becomes smaller. Also the difference in the granularity of the nominal slot classifications becomes less evident as the number of the verb classes and the predictive power grow. [10]

6. General discussion and conclusions

In this paper, we have demonstrated how Semantic Vector Spaces, an approach in computational linguistics, can be transferred to Construction Grammar and used to model constructional semantics. Our approach offers the following epistemological advantages in comparison with introspective manual semantic analysis (in the extreme case, mere eye-balling of the contexts):

[10] Unfortunately, we were unable to test the statistical interactions of the slot fillers due to data sparseness.

- it allows one to find the optimal semantic classification of constructional slot fillers, based on their usage;
- it takes into account the frequencies of semantic classes in near-synonymous constructions, which has a cognitive interpretation in terms of the cue validity of collexeme classes;
- it provides objective evidence of the optimal level of semantic granularity of semantic descriptions;
- it allows one to detect semantic regularities, which otherwise might be unnoticed;
- it allows one to explore larger amounts of data than it would be possible manually, and, in principle, can provide distribution-based classifications for all collexemes that occur in the corpus. As a result, the conclusions are robust from both the quantitative and qualitative perspectives.

In addition, the co-occurrence matrices that were used for this case study can be easily 'recycled' for other construction-oriented case studies in Dutch.

Apart from the demonstration of a new method for studying constructional semantics, the results of the experiments have theoretical consequences. For instance, in 5.4 we found that the effects of the semantic classes of the different slots are non-additive. In other words, the predictive power of the three slots combined together is smaller than the sum predictive power of the slots taken separately. Why should this be the case? A possible explanation might be as follows. If a word is compatible with the meaning of the construction, as assumed in Construction Grammar, then there should be coherence among all slots. This statement is called the Principle of Semantic Coherence by Stefanowitsch and Gries (2005:11). The coherence can be based on different types of knowledge, for instance, frame-semantic relationships, as in example (10). The information stored in one slot therefore interacts with the information from the other slots. This non-additivity can serve as evidence that the semantics of constructions is not reducible to the semantics of the constructional components.

Another interesting finding is that the most parsimonious classification of the Causers has fewer classes than that of the Effected Predicates. In general, noun classifications tend to be organised taxonomically as trees with long branches, for instance, 'entity' – 'concrete object' – 'living being' – 'animal' – 'mammal' - 'carnivore' - 'canine' – 'dog' – 'bulldog', whereas verbs constitute far less hierarchically structured 'bushes' (the hyper- and hyponymy chains in the WordNet provide a nice illustration). This difference can be explained by the complex semantic structure of verbs, which normally involves a configuration of participants, spatiotemporal characteristics of the event, image schemata, social frames and scenarios, and

other information. To organise these heterogeneous abstract semantic structures in a consistent tree-like hierarchy with a high level of generalisation is problematic. It seems plausible that verbs are organised in a large number of small local clusters. We need very detailed contextual information (in our approach, many possible subcategorisation frames) to capture these small classes.

Perhaps the most intriguing finding concerns the type of distributional context. Both in the case of the Causers and the Effected Predicates, the maximally syntactic (constructional) models perform the best (cf. Gries and Stefanowitsch 2010). One may wonder why this should be the case. We would like to propose the following explanation. The abstract syntactic context features highlight the syntactically (constructionally) relevant semantic properties of lexemes (e.g. animacy – inanimacy). These features may also be relevant for the prediction of other constructions beside the causatives with *doen* and *laten*. According to this hypothesis, one can expect the more syntactically enriched models to perform better than more lexically specific models in all sets of near-synonymous syntactic constructions. Future research will show if this hypothesis is correct.

The approach also has to face a few challenges in the future. Fist of all, a more realistic approach would require word sense disambiguation. For instance, it would be necessary to treat *denken* in the sense "think (of)" separately from *denken* as "think (that)" and even "think (about)". Another important problem is pronominal reference resolution, which would allow us to use all occurrences of the constructions in the testing set and get a more complete picture.

The results also suggest that a medium level of granularity is optimal for classification of the Effected Predicates. However, as we pointed out in Section 2, the speaker's knowledge about the constructions may have several interacting levels of generalisation. So far, we have pruned the clustering tree only at one level. In future experiments, we are planning to test different levels of granularity simultaneously, which may allow us to integrate both higher-level generalisations and specific exemplars of the constructions.

From a broader interdisciplinary perspective, this study has demonstrated that the neostructuralist distributional models perform quite well in finding relevant semantic similarities between constructional slot fillers. We expect that fully construction-based bottom-up models, free from any aprioristic assumptions about the syntactic categories and relations, will further improve the performance and make the approach even more radically data-driven. One of the possibilities, for instance, would be to use unsupervised stochastic grammar induction from raw data on the basis of existing Data Oriented Processing models

(e.g. Beekhuizen and Bod, this volume). A practical implementation of this approach will bridge the gap between the computational models of semantics and usage-based Construction Grammar.

References

Baayen, Harald, 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press, Cambridge.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Kramer, & Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Everitt, Brian S., Sabine Landau & Morven Leese. 2001. *Cluster analysis*. London: Arnold.

Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In John R. Firth (ed.), *Studies in Linguistic Analysis*, 1–32. Oxford: Blackwell.

Geeraerts, Dirk. 2010a. The doctor and the semantician. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*, 63-78. Berlin/New York: De Gruyter Mouton.

Geeraerts, Dirk. 2010b. *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The structure of lexical variation. Meaning, naming, and context*. Berlin/New York: Mouton de Gruyter.

Goldberg, Adele E. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Gries, Stefan Th. 2006. Corpus-based methods and Cognitive Semantics: The many senses of *to run*. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, 57–99. Berlin/New York: Mouton de Gruyter.

Gries, Stefan Th. & Dagmar Divjak. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory 2*, 23–60.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9 (1), 97–129.

Gries, Stefan Th. & Anatol Stefanowitsch. 2010. Cluster analysis and the identification of collexeme classes. In John Newman & Sally Rice (eds.), *Empirical and experimental methods in cognitive/functional research*, 73–90. Stanford, CA: CSLI.

Harris, Zellig. 1954. Distributional structure. *Word* 10(2/3), 146–162.

Heylen, Kris. 2005. A quantitative corpus study of German word order variation. In Stephan Kepser & Magda Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives*, 241–264. Berlin: Mouton de Gruyter.

Heylen, Kris, Yves Peirsman, Dirk Geeraerts, and Dirk Speelman. 2008. Modelling word similarity: An evaluation of automatic synonymy extraction algorithms, *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008), Marrakesh, Morocco, 23–30 May 2008, 3243–3249. Marrakesh: European Language Resources Association.

Heylen, Kris, Yves Peirsman and Dirk Geeraerts. 2009. Automatic Synonymy Extraction. A Comparison of Syntactic Context Models. In *Proceedings of Computational Linguistics in the Netherlands* (CLIN) 2008, Nijmegen, Holland.

Hosmer, David W., Lemeshow, Stanley, 2000. *Applied logistic regression.* Wiley, New York.

Kemmer, Suzanne & Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5, 115-156.

Landauer, Thomas K. and Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge, *Psychological Review* 104(2), 211–240.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Vol. 1. *Theoretical prerequisites*. Stanford: Stanford Univeristy Press.

Langacker, Ronald W. 2005. Construction Grammars: Cognitive, Radical, and Less So. F. J. Ruiz de Mendoza Ibáñez and M. Sandra Peña Cervel (eds.) *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, 101–159. Berlin/New York: Mouton de Gruyter.

Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.

Levshina, Natalia. 2011. *Doe wat je niet laten kan. A usage-based analysis of Dutch causative constructions.* PhD diss. Katholieke Universiteit Leuven.

Levshina, Natalia, Dirk Geeraerts & Dirk Speelman. In press. Mapping constructional spaces: A contrastive study of English and Dutch analytic causatives. To appear in *Linguistics*.

Lin, Dekang. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics*, Montreal, Canada, August 1998, 768–774.

Lowe, Will. & Scott McDonald. 2000. The direct route: mediated priming in semantic space. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (CogSci 2000), 675–680. Wheat Ridge, CO: Cognitive Science Society.

Ordelman, Roeland, Franciska de Jong, Arjan van Hessen & Henri Hondorp. 2007. TwNC: a Multifaceted Dutch News Corpus. *ELRA Newsletter* 12 (3–4). Available at: http://doc.utwente.nl/68090/ (last accessed 20 October 2011).

Padó, Sebastian & Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics* 33(2): 161–199.

Peirsman, Yves, Kris Heylen & Dirk Geeraerts. 2008. Size Matters. Tight and Loose Context Definitions in English Word Space Models. In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, Hamburg, Germany.

Plank, Barbara & Gertjan van Noord. 2010. Dutch Dependency Parser Performance Across Domains. In *Proceedings of the 20th Meeting of Computational Linguistics in the Netherlands,* 123-138.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org/.

Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–188.

Schulte im Walde, Sabine. 2006. Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics* 32(2), 159–194.

Schulte im Walde, Sabine. 2009. The Induction of Verb Frames and Verb Classes from Corpora. In Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*, 952–972. Berlin: Mouton de Gruyter.

Speelman, Dirk & Dirk Geeraerts. 2009. Causes for causatives: the case of Dutch *doen* and *laten*. In: Ted Sanders & Eve Sweetser (eds.), *Linguistics of Causality*, 173–204. Berlin: Mouton de Gruyter.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–243.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1), 1–43.

Stukker, Ninke, 2005. *Causality marking across levels of language structure*. PhD diss., University of Utrecht.

van Noord, Gertjan. 2006. At last parsing is now operational. In Piet Mertens et al. (eds.), *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*, 10–13 April 2006, Leuven, 20–42.

Verhagen, Arie & Suzanne Kemmer. 1997. Interaction and Causation: Causative Constructions in Modern Standard Dutch. *Journal of Pragmatics* 27, 61–82.

Vossen, Piek (ed.). 1998. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*, Kluwer, Dordrecht.

Weeds, Julie, David Weir & Diana McCarthy. 2004. Characterising measures of lexical distributional similarity. *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*1015).

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.