

An integrative exemplar-based model of semantic structure: The Dutch causative construction with *laten*¹

Natalia Levshina
University of Marburg
Research Group “Language Typology and Quantitative Linguistics”
natalevs@gmail.com

Abstract

The paper addresses an under-investigated issue of the structure of constructional meaning. While it is commonly assumed that a construction should have 'senses', one of which should be the 'prototypical' one, the present study questions this assumption and presents an innovative corpus-based bottom-up approach that can be used to model semantic structures. The method is based on the visualization of the semantic similarities between the exemplars of a construction in a semantic map created with the help of Multidimensional Scaling. The study explores the main semantic dimensions and 'senses' of the Dutch causative construction with the auxiliary *laten* “let”. The quantitative analyses of 731 exemplars from corpora of spoken, computer-mediated and newspaper Dutch, which were coded for 35 various contextual variables, show that the constructional semantics is organized as a doughnut, with an empty centre and extensive periphery. The main senses (clusters of similar exemplars) are not discrete, but represent a continuum. These findings are contrasted with the previous accounts, which assumed a discrete prototype-based semantic structure of the category.

1. Introduction: different models of semantic structure

Construction Grammar treats constructions as pairings of form and meaning. In order to describe the semantics of a construction, or to explain and predict its variation and change, it is important to know how the meaning is organized. While it is commonly accepted both in psychology and linguistics that natural languages categories have fuzzy boundaries from the intercategory point of view, and degrees of category membership from the intracategory point of view (see Geeraerts 2010: 183–192), the more specific structure has not been investigated very thoroughly. Many semanticists rely on traditional century-old lexicographic practices without questioning them. Today, with the advances of empirical quantitative approaches to semantics and constructions, we have sufficient data and a broad range of tools to test our hypotheses about semantic structures.

In Cognitive Semantics and Construction Grammar, there have been a number of suggestions about how meaning can be organized. These suggestions differ with regard to two major distinctions, which are described below.

1) Distinct senses vs. semantic dimensions. Most semantic studies follow the traditional lexicographic practice and describe related but distinct senses. Some of the best-known examples are studies of the semantics of the preposition *over*, e.g. Brugman (1983), Lakoff (1987), Tyler and Evans (2001). These studies differ in many respects, which are beyond the scope of this article, but most of them propose a

¹ A significant part of this research was implemented at the University of Leuven as a part of my PhD project under the supervision of Dirk Geeraerts and Dirk Speelman. The project was financially supported by a grant from the Research Foundation – Flanders (FWO). I also thank the anonymous reviewer for valuable comments and suggestions. All remaining mistakes and imperfections, of course, are solely mine.

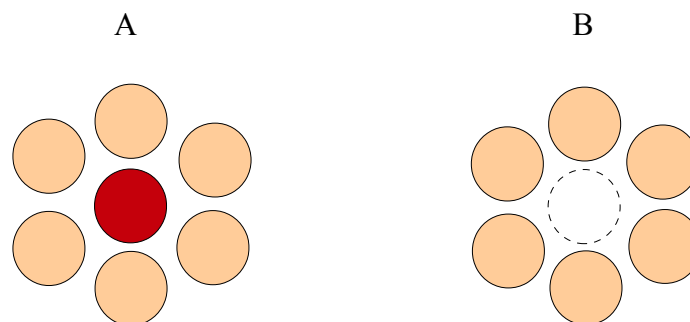
number of distinct senses, which are frequently represented as nodes in a radial polysemy network. For instance, the sense formulated as “on the other-side-of”, e.g. *The village is just over the river*, and the “above-and-beyond” sense, e.g. *The arrow flew over the target* (Tyler and Evans 2001).

An alternative to this approach is Geeraerts' (1998) theoretical justification and Coleman's (e.g. 2009) corpus-based implementation of a multidimensional approach to semantics. Extensions from the basic sense are organized along several dimensions of variation. For instance, there is variation in the Dutch ditransitive construction with regard to the direction of transfer (causing to receive vs. causing to lose), as exemplified by the ditransitive predicates *geven* “give” and *ontnemen* “take away from”, respectively (Coleman 2009). Another dimension is the polarity of transfer, i.e. whether or not the transfer takes place. The examples are, again, the predicate *geven* “give” (successful transfer) and the verb *weigeren* “refuse”, which denotes unsuccessful transfer (*ibid.*).

In fact, these two approaches are closer than it might seem. On the one hand, some authors who describe distinct senses do not always exclude a continuum between these senses, at least theoretically (e.g. Brugman 1983). This continuity can be regarded as a manifestation of dimensionality. On the other hand, the multidimensional approach can lead to discrete senses if the dimensions are of a categorical nature, e.g. the above-mentioned direction of transfer has only two possible values. A list of all possible combinations of such binary values would be a list of discrete senses. To the best of my knowledge, continuous dimensions of constructional semantics have not been modelled yet, with the exception of an attempt in Levshina *et al.* (In press).

2) Presence vs. absence of the central sense or exemplar. The overwhelming majority of studies in Cognitive Linguistics and Construction Grammar assume the existence of a basic sense (primary sense, central sense, prototype, protoscene, etc.), the other meanings being extensions from the prototype. For instance, Goldberg (1995) in her analysis of the English ditransitive construction treats the actual physical transfer (e.g. *I gave him the book*) as the basic sense, and transfer of information (e.g. *She told me the news*) as a metaphorical extension from the basic sense. Yet, in some cases it is difficult to find the semantic centre. An example is the subject-auxiliary inversion (SAI) construction in Goldberg (2006: Ch. 8). She shows that the category is organized in a family resemblance fashion, with different 'senses' sharing some of the functional features of the construction, but there is no central element that would share all of these features (Goldberg 2006: 176). However, such an analysis is more of an exception than a rule.

If we combine these distinctions, we can distinguish four extreme types of semantic models. These four types are represented schematically in Figure 1. Type A (discrete senses plus the 'core' sense) is the most popular in Cognitive Semantics, whereas Type D (no discrete senses, no outspoken semantic core) has not been reported yet. Type B with the empty centre is probably the most similar to Goldberg's (2006) analysis of the SAI construction. Type C is the model with continuous dimensions and the core/periphery distinction. It is represented as a fully coloured area because in its purest form any value on one dimension can be combined with any value on the other dimension. To the best of my knowledge, this type has not been described yet, either.



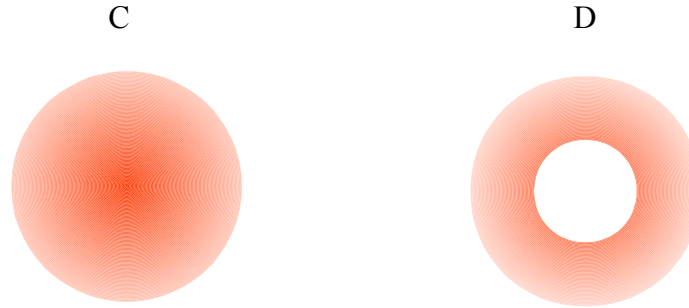


Figure 1. Different types of semantic structures: A – distinct senses with the central sense; B – distinct senses without a directly expressed central sense; C – a continuous structure with the central sense; D – a continuous structure without a central sense. Colour intensity corresponds to semantic centrality.

In fact, this predominance of the model with distinct senses and the single 'prototype' should be taken critically. In the Prototype Theory of categorization (e.g. Rosch 1975; Rosch and Mervis 1975), prototypes are commonly understood as abstract combinations of the most typical features shared by all members of the category. These combinations may be represented by any specific category member, which is normally regarded as the most representative member of the category (e.g. a robin is a good instance of the category BIRD). It has been shown that the more prototypical members are more easily learnt, identified and reproduced in various kinds of experimental tasks than the less prototypical ones. Still, the one and only most prototypical member that has all characteristic properties of the category does not necessarily exist; it can be only an abstract representation at the intensional level without any extensional counterpart.²

The idea of distinct senses is equally problematic. In the studies of concrete categories (e.g. BIRD or FURNITURE), it is relatively easy to come up with a list of subcategories, such as sparrows, swallows, swans, or chairs, tables, sofas. As we move to more abstract words and constructions, a classification of the instances into usages or senses becomes increasingly difficult. One of the main problems is the level of granularity and detail (Tyler and Evans 2001). In principle, there exist an unlimited number of strategies, from the radical 'splitter' (every exemplar is treated separately) to the ultra 'lumper' (the category is described as a whole), although most researchers find themselves at different points in between.³ For instance, Brugman (1983) suggests several image schemata for the 'covering sense', including the full coverage sense (*She spread the cloth over the table*) and the multiplex trajector and incomplete coverage sense (*The bushes are scattered over the field*), whereas Tyler and Evans (2001) mention only 'covering' as one sense. Another obvious problem is the boundary between closely related senses. To summarize, a description of discrete senses involve quite a few difficult decisions.

However, the dimensional model is not always an easy solution, either. There may arise

² In her later works (e.g. Rosch 1978), Rosch is even more cautious about the use of the word 'prototype'. It is only a convenient shortcut to refer to the prototypicality effects, and it should not be interpreted as the mental representation of the category.

³ The influential Exemplar Theory (e.g. Medin and Schaffer 1978) assumes that the general representation of a category as a prototype is not relevant. The radically exemplar-based view has been criticized by Goldberg (2006), who dedicated the entire volume to prove that speakers can learn and store generalizations above the level of the specific occurrences, although there is evidence of the relevance of the exemplars understood as low-level schemata (e.g. Bybee and Eddington 2006, Zeschel 2010). These findings do not contradict each other from the non-reductionist usage-based perspective, which assumes the co-existence of different levels of abstraction in the speaker's knowledge of constructions.

problems with establishing whether any two dimensions are distinct or they represent one underlying conceptual dimension. For instance, the enabling use of the ditransitive construction (*Mary offered John a glass of wine*) and the benefactive meaning (*John cooked Mary a risotto*) can be seen as two separate dimensions, or as one superordinate dimension “X causes Y to receive Z”.

In the present study, all these issues are regarded as empirical questions, which have to be answered for every particular construction in question. I propose a quantitative corpus-based methodology that can be applied to model semantic structures. The method can be used to establish the dimensions and senses, as well as the centre and periphery of a category in a bottom-up fashion. I demonstrate how the method works in a case study of the Dutch causative construction with *laten* “let”. In previous research the category was treated as the one with the prototype and distinct extensions (Type A). This study tests these assumptions empirically.

The paper is organized as follows. Section 2 introduces the main results of previous research in the semantics of *laten*. Next, I present the data and an innovative multivariate method of representing the semantic structure. Section 4 reports the results of the statistical experiments. Section 5 concludes with a general discussion.

2. The Dutch causative construction with *laten*

The Dutch construction with *laten* is a periphrastic causative. In many languages, this is a special type of constructions that combines the causing and caused events in one causative chain. The causing event is expressed by a Causative Auxiliary (such as *make, cause, have, get* in English) and is highly schematic. Consider (1), an example with the auxiliary *laten* in the past form (*liet*). The causing event (what the general did) is unspecified, although one can guess that he probably gave his army orders. The caused event (the effect of the general's actions) is that the army destroyed the city. This action is expressed by the Effected Predicate *vernielen* “destroy”, a bare infinitive.

(1)	<i>De generaal</i>	<i>liet</i>	<i>het leger</i>	<i>de stad</i>	<i>vernielen.</i>
	the general	let	the army	the city	destroy
	CAUSER	AUX.	CAUSEE	AFFECTEE	EFFECTED
		PRED.			PRED.

“The general ordered the army to destroy the city.”

In addition, the construction includes several nominal slots:

- the Causer, who performs the roles of the initiator and/or the responsible entity (the general in the above-mentioned example);
- the Causee, who carries out the caused event expressed by the Effected Predicate (the army);
- the Affectee, the end point in the causation chain (the city), available only in the case of transitive Effected Predicates.⁴

The periphrastic causative with *laten* is highly frequent and semantically broad. Although etymologically related to the English *let*, which expresses only enablement and permission, the Dutch *laten* was used to express both the semantics of letting and coercion already in the earliest attested examples (e.g. van der Horst 1998). Note that in Talmy's force dynamics (Talmy 2000: Ch. 7, the terminology adjusted to the purposes of the present paper), letting involves a Causer who fails, deliberately or not, to override the Causee's intrinsic tendency towards rest or motion. As a result, the

⁴ I follow Verhagen and Kemmer (1997) in naming the main constructional slots.

Causee is maximally autonomous in bringing about the effected event. On the other hand, causation *per se* (coercion, impingement) means that the stronger Causer overrides the intrinsic tendency of the Causee towards rest or motion. The contemporary *laten*, like its German cognate *lassen*, expresses both force-dynamic situations. In fact, there are many cases when the difference between the two is neutralized. Compare the letting context in (2a) with an ambiguous sentence in (2b) and a coercive one in (2c). The letting/coercive distinction is not encoded, but inferred with a higher or lower probability.

- (2) a *De politie liet de dader ontsnappen.*
 the police let the criminal escape
 “The police let the criminal escape.”
- b *Hij liet iedereen zijn roman lezen.*
 he let everybody his novel read
 “He had/let everyone read his novel.”
- c *De trainer liet de spelers loopoefeningen doen.*
 the coach let the players run-exercises do
 “The coach had the players do running exercises.”

This broad and abstract meaning of the construction with *laten* can be interpreted as that of indirect causation (Kemmer and Verhagen 1994). Indirect causation is observed when “it is recognized that some other force besides the initiator is the most immediate source of energy in the effected event” (Verhagen and Kemmer 1997: 67). Its opposite is the so called direct causation, i.e. when “there is no intervening energy source ‘downstream’ from the initiator: if the energy is put in, the effect is the inevitable result” (Verhagen and Kemmer 1997: 70). The (more) direct causation is covered by the construction with *doen* “do”, and the very infrequent and lexically restricted *aan het V maken* (*brenge*) “bring to V”.

As Verhagen and Kemmer (1997) demonstrate, indirectness of causation is closely associated with the particular configurations of the semantic classes of the Causer and the Causee. Thus, if both the Causer and the Causee are animate (human), one can expect the causation to be indirect because a human being cannot affect another mind directly, telepathy disregarded (Verhagen and Kemmer 1997: 71). This type of causation is often labelled as inductive causation, e.g. (3a). In the case of volitional causation, with an animate causer and an inanimate causee, a human being can also influence the world indirectly, e.g. with the help of automation as in (3b), or by using natural forces, such as gravity (3c).

- (3) a *De trainer liet de spelers loopoefeningen doen.*
 the coach let the players run-exercises do
 “The coach had the players do running exercises.”
- b *De machinist liet de motoren draaien.*
 the engine-driver let the engines run
 “The engine driver had/let/left the engines run(running).”;
- c *Hij liet de water weglopen.*
 he let the water away-run
 “He let the water drain out.”

In addition to inductive and volitional causation, Verhagen and Kemmer (1997) and Stukker (2005) also speak about so called physical and affective causation. The former involves an inanimate Causer and

Causee, as in (4), and the latter contains an inanimate Causer but an animate sentient Causee, as in (5). However, these two uses are less frequent than inductive and volitional causation.

(4) *De bommenwerpers laten hun dodelijke lading vallen.*
The bombers let their deadly cargo fall
“The bombers drop their deadly cargo.”

(5) *Het laat ons het beste verhopen.*
it lets us the best hope-for
“It makes/allows us (to) hope for the best.”

Note that Talmy's use of the term 'inductive causation', or caused agency (e.g. 2000: Section 5.6) is somewhat different from the operationalization in Verhagen and Kemmer (1997) and Stukker (2005). Talmy's inductive causation involves intentionality, or volitionality of the Causee's actions. For Verhagen *et al.*, inductive causation is defined with the help of the semantic classes of the Causer and the Causee (both are animate entities). It is thus a more coarse-grained operationalization, although in most cases, e.g. (3a), the two interpretations overlap. The difference between the definitions can be seen in the following examples:

(6) *Smoke getting in its eyes made the squirrel leave its tree.*

(7) *He let me know that he is leaving.*

The example (6) would be a case of inductive causation according to Talmy because the Causee (the squirrel) left the tree intentionally. In Verhagen – Kemmer – Stukker's operationalization, this context could be interpreted as affective causation because of the inanimate Causer (*the fact of smoke getting in the squirrel eyes*). In contrast, in (7) the Causee (*me*) is a non-volitional experiencer, so the causation would not be inductive in Talmy's interpretation. For Verhagen *et al.*, the sentence would be an example of inductive causation because both the Causer and the Causee are animate sentient beings. In this paper I will try to take into account both Talmy's more direct conceptual interpretation and Verhagen – Kemmer – Stukker's approximation.

According to Stukker (2005), inductive causation (as a combination of an animate Causer and a sentient Causee) is the semasiological prototype of the *laten*-construction. The other senses (volitional, physical and affective causation) are extensions from this sense. Thus she assumes the semantic structure of Type A with the inductive causation in the centre and the other causation types on the periphery. In what follows I test and refine Stukker's hypothesis, both at the level of discreteness and centrality of the senses with the help of an innovative technique, which allows for the representation of exemplars (occurrences) of constructions in a low-dimensional space. The data and approach are represented in the following section.

3. The exemplar space of *laten*: data and method

The approach presented here was developed in Levshina (2011). It allows for the representation of the exemplar space of constructions as a low-dimensional semantic map, which can display both the dimensions of semantic variation, and the senses as clusters of similar exemplars. It is important to mention that the exemplars are most commonly understood in psychology as unique instances (see the discussion in Murphy 2002: 58–60), e.g. every fact of presentation of a stimulus to the subjects in an experiment. In this study, exemplars are regarded as occurrences of the construction in a corpus.

Exemplars in the other use of the word (e.g. Bybee and Eddington 2006) – fully or partly lexically specified constructions – are treated here as low-level abstractions. The following subsections describe the steps of the analysis.

3.1. Selection and preparation of the data

I used the materials from three popular registers:

- the fundamental register of spontaneous face-to-face conversations: data from the Corpus of Spoken Dutch (Oostdijk 2002);
- the newspaper register: data from Twente News Corpus (Ordelman *et al.* 2007) and Leuven News Corpus, constructed at the University of Leuven (the Quantitative Lexicology and Variational Linguistics research unit);
- the online communication register: postings from several Belgian and Dutch online discussion groups (the Usenet) collected by Tom Ruetten (University of Leuven).

All three subcorpora contained samples of Dutch spoken in the Netherlands and Flanders (the Dutch-speaking part of Belgium). Since previous studies (Levshina 2011) have shown that there is no substantial conceptual variation in the semantics of *laten* in the three registers and the two countries, the variational aspect is ignored here. For this study I created a sample of 731 occurrences of the causative *laten* randomly selected from the above-mentioned subcorpora. The exemplars were then coded manually for 35 contextual variables. These variables can be subdivided into several groups:

- the variables related to the nominal slot fillers (the Causer, the Causee and the Affectee, if available): the semantic class, syntactic expression, part of speech, grammatical person, number and definiteness;
- specific features of the Causee: volitionality and whether the participant undergoes or causes a change;
- the variables describing the relationships between the main participants: relationships of coreferentiality and possession;
- the features of the Effected Predicate: transitivity (in a broad sense, including ditransitivity, copula functions, etc.), and the type of prepositional complements. Also, the semantics of the caused event was considered, both in the literal and metaphorical sense (if applicable). The specific lemmata of the predicates were considered a separate variable, as well, because many of them occurred many times;
- the variables related to different modifiers: polarity (negation), adverbial modifiers, modal verbs modifying the auxiliary;
- the syntactic function of the construction;
- the more general properties of the clauses and sentences where the construction was found, such as the mood and tense of the clause, the syntactic type of the clause (main, relative, adverbial, etc.), and the communicative type of the sentence.

This list of variables represents all possible contextual variables that could be described at an acceptable level of objectivity, with the help of linguistic markers or simple tests. This comprehensive approach is similar to the one in Gries' (2006) corpus-based analysis of the verb *run*, which also involved a large number of heterogeneous variables. Although many of these variables are associated, this redundancy of linguistic cues is natural when we learn a new word or construction.

In some observations the relevant information was missing. In this case, I had either to depend

on the context (e.g. the semantic class of implicit Causees), or, if the contextual clues were insufficient, to code the feature as 'Not Applicable' (e.g. the grammatical properties of the missing Affectee).

3.2. Multivariate analyses of the sample

The matrix with the individual exemplars (731 rows) coded for the categorical variables (35 columns) was used as the input for a series of statistical analyses in R (R Development Core Team 2011). With the help of Gower's distance (Gower 1971) – a universal distance metric for numeric and categorical variables – I created a matrix of distances between the exemplars. A distance matrix looks like a chart of distances between cities. The distances are defined on the basis of the shared semantic features: the more features two observations share, the smaller the distance between them. The exemplars with the same features have the distance of 0. If one or both exemplars in a pair contained a missing value, the corresponding features was disregarded in the calculation of the distance between the exemplars.

Next, this distance matrix was represented spatially with the help of Multidimensional Scaling (MDS). In this study I used the highly successful SMACOF algorithm developed by de Leeuw and Mair (2009), see also Borg and Groenen (1997). Both the metric and non-metric solutions were tested, and the representations were nearly identical. In what follows, the metric solution is reported. This MDS map serves as a visualization of the differences and similarities between the exemplars of the construction. It can be treated as a semantic, or conceptual map of semantic categories. One can evaluate the general structure of the category, interpret the main dimensions, and explore the different clusters, or 'senses'. Since the exemplars with the same values will have the same coordinates on the map, it may also be useful to represent the semantic structure in a density map, which shows how densely different semantic regions on the map are 'populated' by the exemplars. I will use a 2D kernel density estimator in the package MASS (Venables & Ripley 2002) to create such a map.

4. The results of the quantitative analyses

The map in Figure 2 shows the two-dimensional solution for the data discussed in the previous section. The stress of the solution was only 9%, which means that more than 90% of the variation is captured by the map. The subsequent dimensions (3, 4 and 5) do not add more than 5% and are difficult to interpret, so they will not be discussed here.

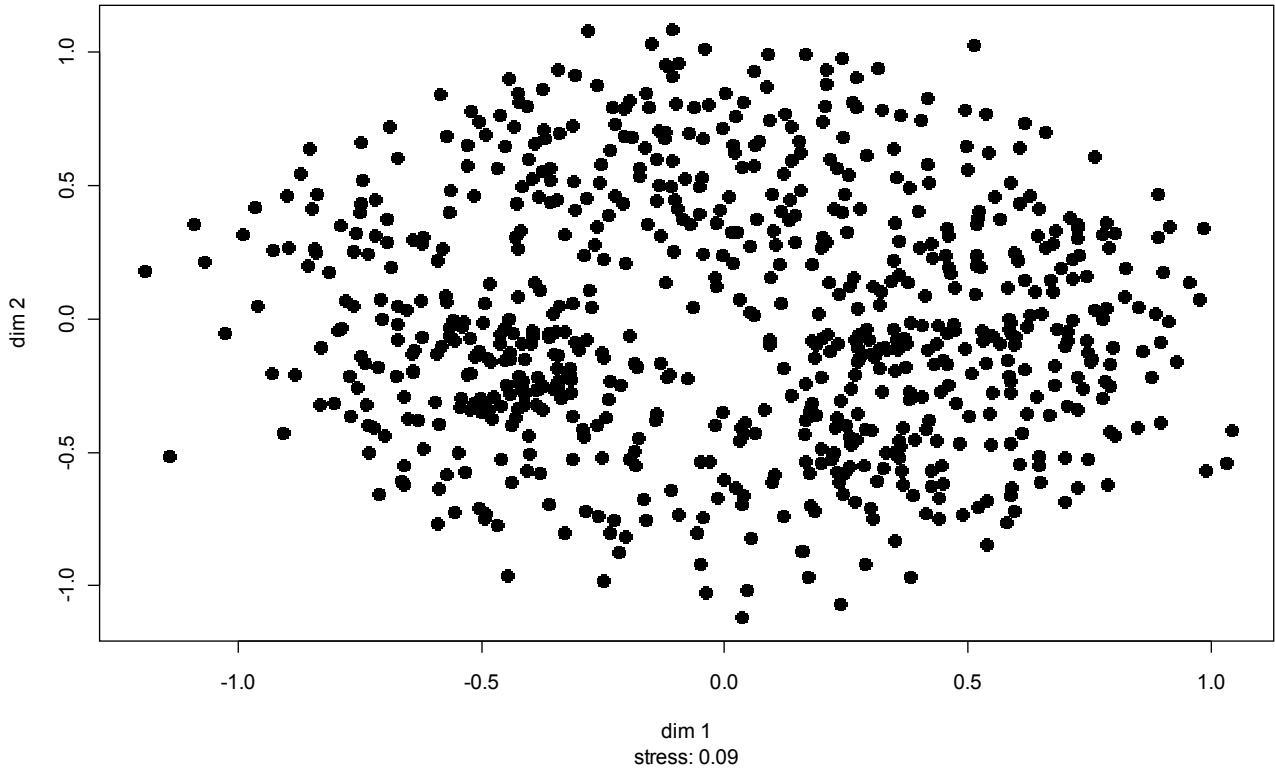


Figure 2. The exemplar space of *laten*.

As one can see, the semantics of the *laten*-construction has a somewhat irregular doughnut-like structure, with a relatively empty centre and broad periphery. This indicates that there is no central sense, which would be equidistant from all others. There is no evidence of discrete senses, although one can see the presence of a few very fuzzy clusters. Before describing them, it is necessary to check if the solution reveals any conceptual dimensions. For this goal, I used two complementary methods. The first was intuitive and visual. For each variable, I created a map with all exemplars represented as different symbols corresponding to the values of the variable. An example is shown in Figure 3. The variable describes the semantics of the caused event, with the values 'Mental' (as in *Ik liet hem weten dat...* “I let him know that...”), 'Physical' (*Ik liet hem mijn huis schilderen* “I had him paint my house”), and 'Social' (*Ik liet hem zijn verhaal doen* “I let/had him tell his story”). One can see the mental caused events mostly in the bottom left part of the map, and the social events predominantly in the top right part, intermingled with the physical caused events.

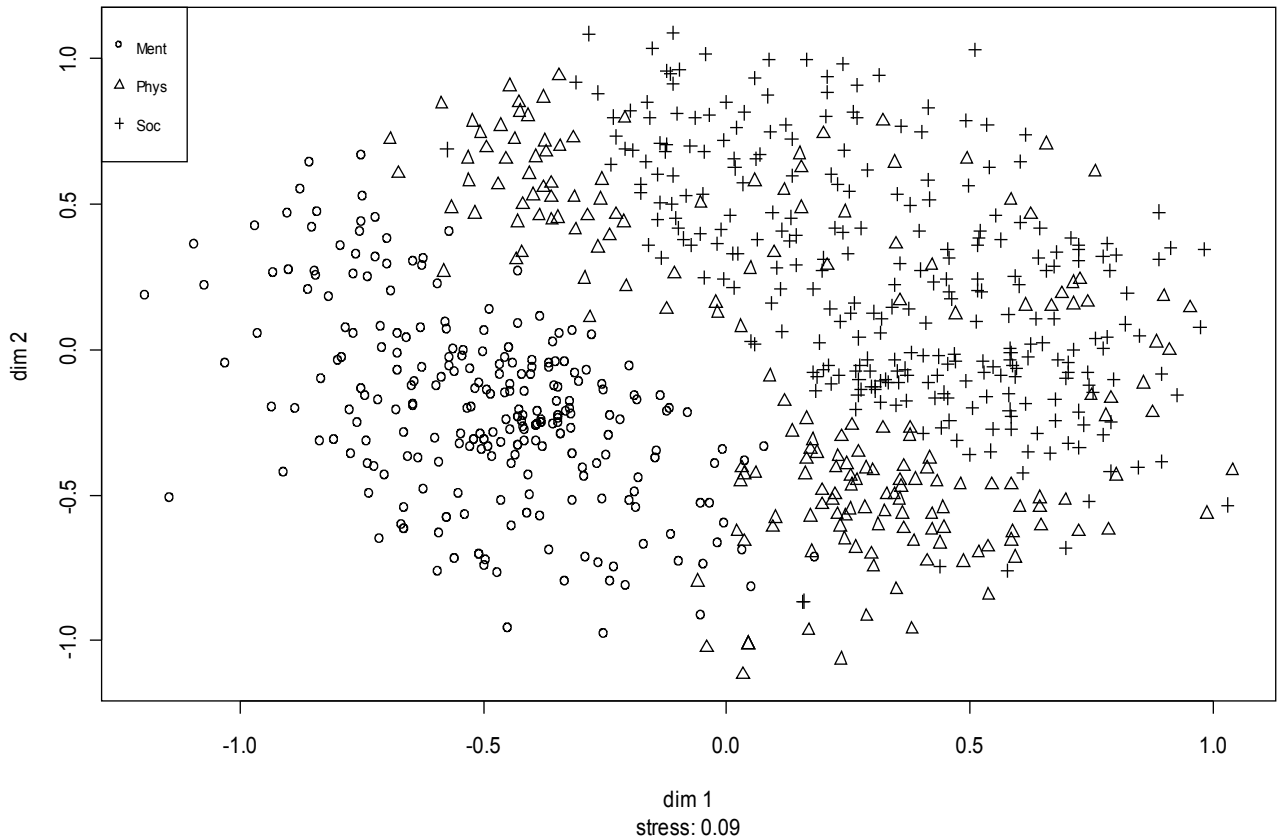


Figure 3. Distribution of the semantic domain of the caused events in the exemplar space of *laten*.

The second approach was quantitative and was applied only to the horizontal and vertical dimensions established by the MDS algorithm. For every variable,⁵ I carried out two simple linear regression analyses and ANOVAs with the coordinates of the points on the horizontal and vertical dimensions as the response and the variable as the predictor. Next, I compared the explained variance statistics (R^2 and F -test) of the variables to see which ones explained the position of the exemplars the best. The results can be found in the Appendix (Table 1).

Both approaches showed very similar results. The horizontal dimension mainly corresponds to the distinction between mental and non-mental caused events: ANOVA F -scores for the semantics of the caused event (in the case of figurative expressions, the figurative meaning was coded) was 476.1 on 2 and 716 d.f., $p < 0.001$; the explained variance R^2 is 0.57. These also are the highest observed scores for Dimension 1 (see the numbers in bold in the Appendix). Compare the example from the extreme left of the map with a mental caused event (*ontmoedigen* “discourage”) in (8), and the one from the extreme right with a physical caused event (*rijden* “ride, drive”) in (9):

- (8) *Laat u niet ontmoedigen en zet uw bijdragen voor deze*
 let you not discourage and set your contributions for this
nieuwsgroep gewoon voort!
 newsgroup as usual forward
 “Don’t let yourself be discouraged and continue contributing to this newsgroup as usual!”

⁵ Except for the Effected Predicate lemmata, which have too many hapax legomena for a meaningful analysis.

- (9) *'s avonds willen ze geen bussen meer laten rijden.*
 in the evening want they no buses more let ride
 “They want to cancel buses in the evening.”

The vertical dimension is associated the most strongly with intentionality, or volitionality of the Causee's actions (unintentional at the bottom, intentional at the top, with a few ambiguous cases in the middle). The F -statistic is 325.1 on 2 and 728 d.f. with $p < 0.001$, and R^2 is 0.47 (again, these are the largest values for Dimension 2, as can be seen from the table in the Appendix). Intentionality is followed by the role of the Causee (the Causee undergoes a change – there is no change – the Causee causes a change): $F=191.9$ on 2 and 727 d.f. with $p < 0.001$, $R^2 = 0.34$. Both features imply the distinction between the patient-like and relatively agentive autonomous Causees. Compare (10), where the implicit Causee is the food that is being prepared, and (11), an exemplar with an agentive implicit Causee (someone who has power to exclude a redundant holder). The exemplars are located in the extreme bottom and top areas of the map, respectively.

- (10) *heel e*a heel even heel even laten roerbakken heel even en klaar.*
 very e*a just a little bit just a little bit let stir-fry just a little bit and ready
 “Just a little bit, let (it) stir-fry just a little bit, and it's ready.”

- (11) *Is er een reden voor om 4 volmachthouders te hebben?*
 is there a reason for in-order 4 authorized holders to have?
Waarom die andere drie niet laten schrappen?
 Why those other three not let drop?
 “Is there any reason for having 4 authorized holders? Why not have the other three dropped?”

These two variables - intentionality and semantic role of the Causee – correspond closely to Verhagen and Kemmer's (1997) distinction between direct and indirect causation, discussed in Section 2. The map shows that the distinction is a matter of continuum, and that the exemplars of *laten* have different values in this continuum.

One can also see in Figure 3 that the mental caused events are located on average a little lower than the physical and social ones. The mental events represented by *laten* on average involve less control by the Causee because most of them refer to perception and knowing, as in the collocations *laten weten* “let know, inform” and *laten zien* “let see, show”, which will be discussed below.

However, if one examines the distribution of the variables more closely, one can find additional semantic dimensions. The most important one is transitivity of the Effected Predicate, which cuts the map diagonally, as shown in Figure 4. Most intransitives are located in the bottom right part of the map, whereas the transitives populate the top left part. Transitivity of the Effected Predicate and the number of participants were discussed in detail in Kemmer and Verhagen (1994). They write that the transitive (three-participant) and intransitive (two-participant) constructions are related to the ditransitive (three-participant) and transitive (two-participant) simple clause constructions, respectively. In the two-participant constructions, causative or not, the second entity (the Causee or the direct object) is the affected one, whereas in the three-participant constructions, the second participant (the Causee or the indirect object of a ditransitive predicate) is less affected and more peripheral. Thus, the degree of affectedness ('patientivity') of the Causee is greater in intransitive causative constructions than in transitive causatives, and the degree of the Causee's autonomy (agentivity) is smaller. For illustration, compare two above-mentioned examples: (10), which contains a patient-like Causee, has an intransitive Effected Predicate, whereas (11), which has a relatively agentive implicit Causee, is an example of the

transitive *laten*-construction.

This is exactly the picture that we see in the maps. Most of the intransitive Effected Predicates (the bottom right part) co-occur with the patient-like Causees (the bottom), whereas the majority of the transitive Effected Predicates (the top left part) are combined with the agentive autonomous Causees (the top). This is in line with Kemmer and Verhagen's interpretation. Yet, these two dimensions – transitivity of the Effected Predicate and affectedness/autonomy of the Causee – do not fully coincide. There are quite a few transitive Effected Predicates in the bottom left part of the map, which involve less agentive Causees, as in (12):

- (12) *Ik liet een vriendin een song horen.*
 I let a friend-FEM a song hear
 “I let a friend hear a song.”

On the other hand, some intransitive Effected Predicates in the top right part of the map are quite autonomous entities. Consider (13):

- (13) ... *daarom lieten we rechtsback Bryssinck almaar mee oprukken.*
 that's why let we right-back Bryssinck continuously with advance
 “That's why we had the right-back Bryssinck push up all the time, too.”

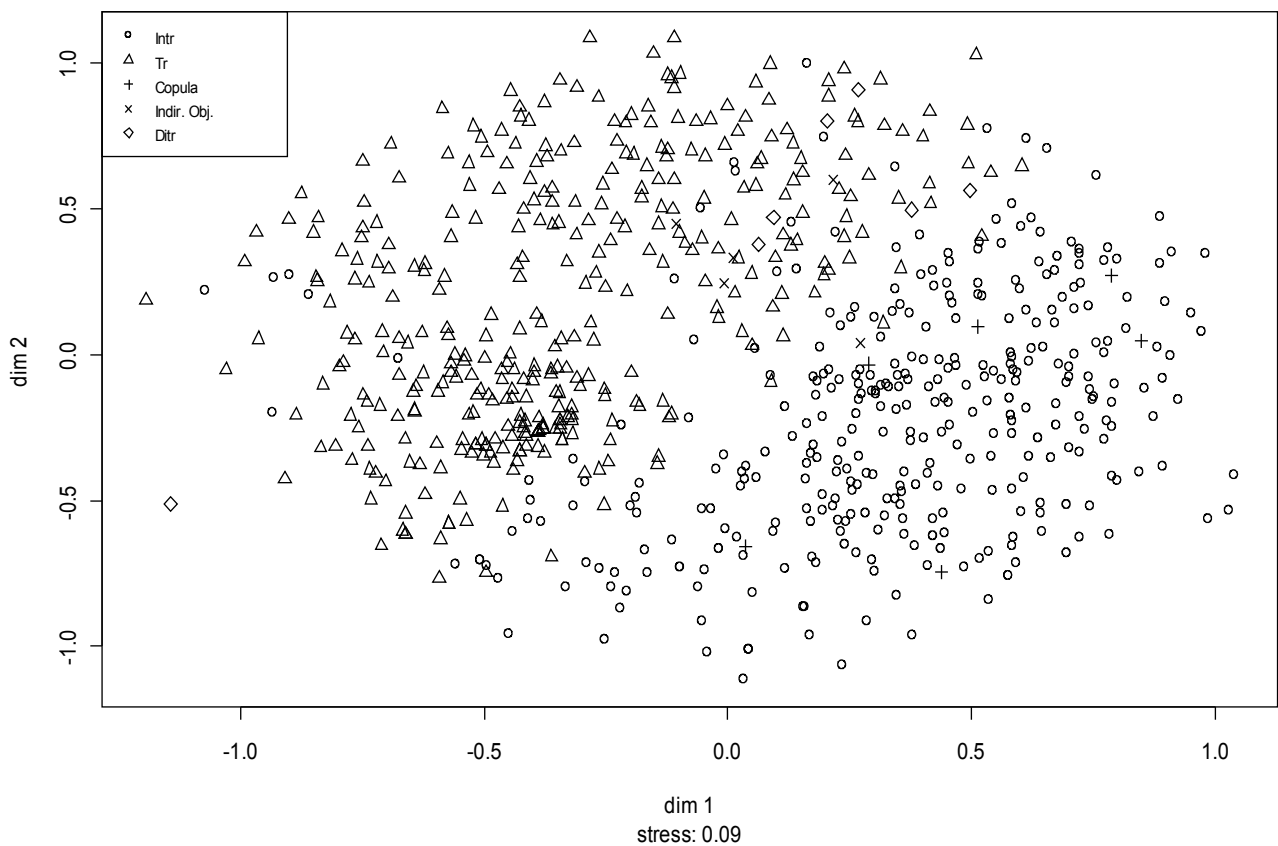


Figure 4. Distribution of the Effected Predicate transitivity patterns in the exemplar space of *laten*.

Next, we can examine the clusters of the exemplars, which may represent the constructional

'senses'. Looking at the map, we can see three main regions, although these are not very distinct. The density plot in Figure 5 gives a clearer picture of the structure. The contour lines delineate the regions with different density of the exemplars. One can see two regions with very high density. A closer look reveals that the one on the left is populated mostly by the instances of *laten weten* “let know, inform”, as in (14), and the exemplars of *laten zien* “let see, show” and some other mental caused event constructions, as in (15), repeated here for the sake of convenience:

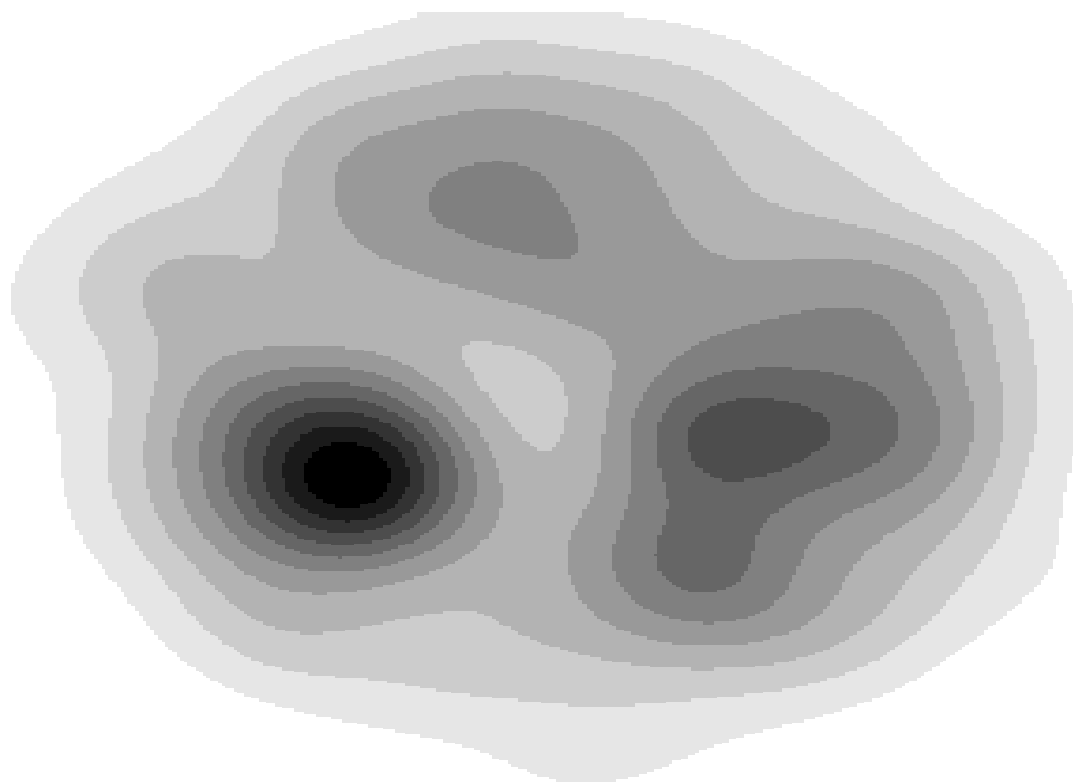


Figure 5. A density map of the causative construction with *laten*, based on the coordinates of the individual exemplars (see the previous maps).

- (14) *Berlusconi liet gisteren weten de functie van Ruggiero voor zeker*
 Berlusconi let yesterday know the function of Ruggiero for sure
zes maanden waar te zullen nemen.
 six months true to shall take
 “Berlusconi said yesterday that he will fill in the function of Ruggiero for at least six months.”

- (15) *Ik liet een vriendin een song horen.*
 I let a friend a song hear
 “I let a friend hear a song.”

The densely populated area in the non-mental part also contains some frequent fixed expressions, e.g. (*links*) *laten liggen* “ignore” and *laten vallen* “drop, abandon”, although this region is less lexically homogeneous. Most of these exemplars, however, are conceptually similar. They represent metaphorical expressions with the sense of leaving, abandoning or missing an opportunity. For instance:

- (16) *AS Roma liet gisteren de kans liggen om naast Inter aan de leiding te komen.*
 AS Roma let yesterday the chance lie in-order near Inter to the leadership to come.
 “A.S. Roma missed the chance to become a leader next to Inter yesterday.”

The area on the top with medium density corresponds to transitive Effected Predicates and maximally autonomous, usually implicit Causees. The examples commonly convey the sense of delegated causation, e.g. the service frame (17) or administrative interaction (18):

- (17) *De makkelijkste manier van beleggen is om het iemand anders te laten doen.*
 the easiest way of investing is in-order it someone else to let do
 “The easiest way to invest is to have someone else do it.”

- (18) *Daarom vroeg het aan het parket om het complex opnieuw te laten onderzoeken.*
 that's why asked it to the public prosecutor in-order the complex again to let search
 “That's why they asked the public prosecutor to have the complex searched again.”

On the periphery of this region, mostly in the top left part of the area, one can find a few examples of middle voice events (e.g. Davidse and Heyvaert 2003) with the coreferential Causers and Affectees, as in (19):

- (19) *Het cultuur laat zich niet makkelijk exporteren.*
 the culture lets itself not easily export
 “The culture cannot be exported easily.”

Interestingly, the Causers in the contexts like (17) and (19) are to a certain degree affected by the effected event: they are either the beneficiaries of the delegated causation, as in (17), or the semantic objects of the action denoted by the Effected Predicate in the reflexive constructions, as in (19). Recall that this area of the map involves agentive Causees. Thus, it also contains non-agentive Causers. This inverse correlation of the agentivity of the Causer and the Causee is logical: as the causation becomes more indirect, the impact of the Causee increases, and the role of the Causer as the main driving force becomes less prominent.

Because these semantic regions are not distinct, one should also explore the transitional zones between them. First, the transitional zone between the 'abandonment' sense with *laten vallen*, etc. and

delegated causation is exemplified by (20) about a football club selling a player. This is an intransitive construction with a relatively autonomous Causee (see the discussion above):

- (20) *FC Utrecht weigerde evenwel de door de eigen fans geadoreerde*
 FC Utrecht refused as-well the by the own fans adored
cultheld te laten gaan.
 cult-hero to let go
 “FC Utrecht also refused to sell the cult hero, adored by his own fans.”

Between the 'abandonment' cluster and the cluster with *laten weten* and similar expressions one can also find a few hybrid metaphorical examples with the abandonment of mental states, conceptualized as physical objects.

- (21) *De directeuren hebben hun vooroordelen blijkbaar schielijk laten varen.*
 The directors have their prejudices apparently suddenly let sail
 “Apparently, the directors have suddenly abandoned their prejudices.”

Finally, between the giving-information cluster and the region with delegated causation, there are a few reflexive constructions that resemble the above-mentioned middle voice contexts. The Causer yields to or resists a negative mental influence, as in (22), repeated here for the sake of convenience:

- (22) *Laat u niet ontmoedigen en zet uw bijdragen voor deze*
 let you not discourage and set your contributions for this
niewsgroep gewoon voort!
 newsgroup as usual forward
 “Don't let yourself be discouraged and continue contributing to this newsgroup as usual!”

This brief examination shows that the continuum between the main 'senses' that appears on the map can be also interpreted in semantic terms. The relationships between the semantic regions are those of family resemblance.

After the analysis of the exemplar space of *laten*, we can address the main question of this study: what is the semantic structure of the construction really like? The first important issue is whether the category is organized along dimensions or represents a set of distinct senses. The map suggests that both types are involved. On the one hand, there is a clear continuum along the two main dimensions – the semantic domain of the caused event and the indirectness of causation. However, the exemplars are not distributed homogeneously along the dimensions. They form clusters, albeit very fuzzy ones.

Interestingly, the distinctiveness of a cluster correlates positively with the frequency of the lexicalized expressions in it. Recall that the most autonomous cluster is the one with the collocations *laten weten* and *laten zien*. According to Bybee (2010: Ch. 3), highly entrenched lexically specific constructions are reproduced by speakers without invoking the general schema of *laten*. This results in a 'shortcut' routine, which is an important source of linguistic change. As the routine becomes more entrenched, the link of the cluster with the general schema may be lost, and an independent construction with its own peculiar functions will be formed. The method presented in this article can be useful in capturing such frequency effects and predicting the future development of a construction.

The other question is the existence of any central sense, such as inductive causation. The map displays a lack of any sufficiently frequent sense, which could be the origin of most other senses. In fact, inductive causation in Talmy's terms corresponds to intentionality of the Causee, which constitutes a dimension, and not a regular cluster. As far as Verhagen – Kemmer – Stukker's approximation is concerned, most exemplars with the animate Causers and Causees are located in two quite distinct

regions: the delegated causation region at the top of the map and the cluster with mental Effected Predicates (*zien* “see”, *weten* “know”, etc.). These two rather different senses are difficult to interpret as a single starting point for the other extensions.

5. Conclusions

To conclude, the analysis to a large extent supports Geeraerts – Coleman's multidimensional model of semantics, according to which the dimensions of semantic variation are more important than specific discrete senses, although the map also displays some discontinuities, which form fuzzy 'senses' for the Dutch causative construction with *laten*. On the other hand, the study shows that there is no outspoken conceptual centre (a prototype, protoscene, etc.). The exemplars of *laten* are connected in a family-resemblance fashion without an explicit central subschema. Needless to say, the results of the corpus-based analysis need support from other types of evidence. For instance, it is necessary to know to what extent the distances between the exemplars reflect their perceived similarity in speakers' minds. This information can be collected from similarity judgments or sorting tasks.

From a more general perspective, the study demonstrates that some assumptions about the semantic structure (such as the existence of distinct senses and prototypes) should not be taken for granted. In every specific case, we should rely on empirical evidence, and not on intuitions about the units of semantic analysis and the relationships between them. Although the notion of prototype may be useful for teaching and learning purposes (e.g. Lindstromberg 2010), a prototypical structure should not be assumed *a priori*. This awareness is especially important for the emergent area of empirical constructional semantics.

Appendix

Variable	Dimension 1 (horizontal)		Dimension 2 (vertical)	
	<i>F</i> -score	<i>R</i> ² (adjusted)	<i>F</i> -score	<i>R</i> ² (adjusted)
Causer's Semantics	11.96***	0.09	1.43	0.004
Causer's Syntactic Expression	0.69	0	13.27***	0.03
Causer's POS	0.03	0	1.29	0
Causer's Definiteness	1.87	0.001	1.3	0
Causer's Person	7.73***	0.02	16.8***	0.04
Causer's Number	12.25***	0.02	12.27***	0.02
Causee's Semantics	32.2***	0.26	61.41***	0.41
Causee's Syntactic Expression	156.5***	0.52	30.77***	0.17
Causee's POS	3.02*	0.01	5.28**	0.03
Causee's Definiteness	1.56	0.001	0.02	0
Causee's Person	23.9***	0.1	11.35***	0.05
Causee's Number	0.01	0	2.55	0.004
Affectee's Semantics	2.02.	0.02	18.9***	0.25
Affectee's Syntactic Expression	9.04**	0.02	123.3***	0.23

Affectee's POS	10.13***	0.05	5.94**	0.03
Affectee's Definiteness	3.94*	0.01	10.16**	0.03
Affectee's Person	5.93**	0.03	0.08	0
Affectee's Number	8.19**	0.02	7.06**	0.02
Causee's Volitionality	10.32***	0.02	325.1***	0.47
Causee's Role (affecting or affected)	112.1***	0.23	191.9***	0.34
Coreferentiality of main participants	48.07***	0.11	63.75***	0.15
Possession relationships between main participants	13.51***	0.05	5.45**	0.02
Effected Predicate Transitivity	136.2***	0.48	38.93***	0.21
Prepositional Complements of the Effected Predicate	1.45	0.01	1.83*	0.01
Caused Event Semantics (if metaphor, literal interpretation)	467.5***	0.57	77.19***	0.18
Caused Event Semantics (if metaphor, figurative interpretation)	476.1***	0.57	69.8***	0.16
Negation	1.03	0	1.34	0.003
Adverbial Modifiers	1.83.	0.01	0.55	0
Modals	2.73*	0.01	0.82	0
Syntactic Function of the construction	8.36***	0.02	0.48	0
Clause Mood	14.79***	0.04	26.56***	0.07
Clause Tense	2.54*	0.01	1.59	0.004
Clause Type	2.29.	0.01	5.8***	0.03
Sentence Type	10.59***	0.03	24.78***	0.06

Table 1. ANOVA F -scores and linear regression R^2 (the adjusted version) for 34 variables (the lemmata of effected predicates were excluded). The stars indicate the level of significance:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1.

References

- Borg, Ingwer & Patrick Groenen. 1997. *Modern multidimensional scaling: theory and applications*. New York: Springer.
- Brugman, Claudia M. 1983. *The story of over*. Bloomington: Indiana University. Linguistics club.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan L. & David Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2): 323–355.
- Colleman, Timothy. 2009. The semantic range of the Dutch Double Object Construction. A collostructional perspective. *Constructions and Frames* 1: 190–221.
- Davidse, Kristin & Liesbet Heyvaert. 2003. On the so-called 'middle' construction in English and Dutch. In Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds.), *Corpus-based approaches to contrastive linguistics and translation studies*, 57–73. Amsterdam/New York: Rodopi.

- Geeraerts, Dirk. 1998. The semantic structure of the indirect object in Dutch. In Willy van Langendonck & William Van Belle (Eds.), *The Dative. Vol. 2. Theoretical and contrastive studies*, 185–210. Amsterdam/Philadelphia: John Benjamins
- Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- 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.
- Gower, J. C. 1971. A general coefficient of similarity and some of its properties, *Biometrics* 27: 857–874.
- 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.
- van der Horst, Joop M. 1998. *Doen* in Old and Early Middle Dutch: A comparative approach. In Ingrid Tieken-Boon van Ostade, Marijke van der Wal and Arjan van Leuvensteijn (eds.), *'Do' in English, Dutch and German. History and present-day variation*, 53–64. Munster: Nodus Publicationen.
- Kemmer, Suzanne & Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5: 115–156.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
- de Leeuw, Jan & Patrick Mair. 2009. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software* 31(3): 1–30. <http://www.jstatsoft.org/v31/i03/>. (Last access 17.11.2011)
- Levshina, Natalia. 2011. *Doe wat je niet laten kan: A usage-based analysis of the Dutch causative constructions*. PhD dissertation, 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*.
- Lindstromberg, Seth. 2010. *English Prepositions Explained*. Revised edn. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Medin, Douglas L. & Schaffer, Marguerite M. 1978. Context theory of classification learning. *Psychological Review* 85: 207–238.
- Murphy, Gregory L. 2002. *The big book of concepts*. Cambridge, MA: MIT Press.
- Oostdijk, Nelleke H.J. 2002. The design of the Spoken Dutch Corpus. In Pam Peters, Peter Collins and Adam Smith (eds.), *New frontiers of corpus research*, 105–112. Amsterdam: Rodopi.
- Ordelman, Roeland, Franciska de Jong, Arjan van Hessen & Henri Hondorp. 2007. TwNC: a multifaceted Dutch News Corpus. *ELRA Newsletter* 12 (3–4), <http://doc.utwente.nl/68090/> (last access 17.11.2011)
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org> (last access 17.11.2011)
- Rosch, Eleanor. 1975. Cognitive representation of semantic categories. *Journal of Experimental Psychology* 104 (3): 192–233.
- Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (Eds.), *Cognition and Categorization*, 27–48. Hillsdale: Lawrence Erlbaum Associates.
- Rosch, Eleanor & Carolyn B. Mervis. 1975. Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology* 7: 573–605.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.
- Tyler, Andrea & Vyvyan Evans. 2001. Reconsidering Prepositional Polysemy Networks: The Case of *Over*. *Language* 77 (4): 724–765

- Stukker, Ninke. 2005. *Causality marking across levels of language structure*. University of Utrecht dissertation.
- Venables, William N. and Brian D. Ripley. 2002. *Modern Applied Statistics with S*. New York: Springer.
- Verhagen, Arie & Suzanne Kemmer. 1997. Interaction and Causation: Causative Constructions in Modern Standard Dutch. *Journal of Pragmatics* 27: 61–82.
- Zeschel, Arne. 2010. Exemplars and analogy: Semantic extension in constructional networks. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in Cognitive Semantics: Corpus-driven approaches*, 201–219. Berlin/New York: De Gruyter Mouton.