Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Pragmatics 52 (2013) 34–48

journal of PRAGMATICS

www.elsevier.com/locate/pragma

Towards a 3D-grammar: Interaction of linguistic and extralinguistic factors in the use of Dutch causative constructions

Natalia Levshina [a,b,*], Dirk Geeraerts [a], Dirk Speelman [a]

[a] Research Unit Quantitative Lexicology and Variational Linguistics, Katholieke Universiteit Leuven, Belgium
[b] Philipps-University Marburg, Germany

Received 4 November 2011; received in revised form 31 May 2012; accepted 28 December 2012

Abstract

The integration of three main dimensions of linguistic usage and variation – formal, social and conceptual – can be seen as a major ambition of the Cognitive Sociolinguistics enterprise. The paper illustrates this theoretical approach with a corpus-based study of near-synonymous causative constructions with *doen* and *laten* in the Belgian and Netherlandic varieties of Dutch. A series of quantitative analyses show a complex interplay of the dimensions at different levels of constructional schematicity. At the more schematic level, the results indicate that the effects of transitivity and coreferentiality on the probability of the two constructions are slightly different in the two varieties. However, incorporating the effected predicate slot fillers in a mixed-effect model reveals that these differences can be explained to a large extent by the country-specific lexical patterns. These findings suggest that the interplay of the lectal and conceptual factors in constructional variation should be studied at varying degrees of constructional schematicity.

Keywords: Causative constructions; Dutch; Logistic regression; Mixed models

1. Introduction

Cognitive Linguistics, with its emphasis on the semiological function of language, has been traditionally concerned with variation in meaning (polysemy, prototypicality, metaphor and other phenomena). Sociolinguistics, on the other hand, has been studying how linguistic variables are distributed across social groups and situations. Cognitive Sociolinguistics is a project that integrates the two approaches, aiming at a more realistic model of variation (Kristiansen and Dirven, 2008). This approach is a vivid example of the ongoing recontextualization of language in modern linguistics at large (Geeraerts, 2009), which is based on the conception of language as a heterogeneous dynamic entity that exists in the form of lects. A linguistic analysis then should take into account three dimensions of variation: formal (what linguistic units are used?),[1] conceptual (what is said?) and social, or lectal (by whom, for whom, when, where, why, etc. is it said?). This integration is achieved in onomasiological lectally enriched grammar and lexicology, which explore how conceptual and lectal factors co-influence the speaker's choice from available categorization options. This approach has been implemented in a number of usage-based studies in Cognitive Linguistics and related fields, such as lectally enriched Cognitive/Construction Grammar, variationist psycholinguistics and probabilistic linguistics (e.g. Geeraerts et al., 1994; Rosenbach,

* Corresponding author. Tel.: +32 16 324 807; fax: +32 16 324 767.
E-mail address: natalevs@gmail.com (N. Levshina).

[1] In this paper the term 'formal dimension' describes a range of linguistic alternatives potentially available to the speaker in the given context, regardless of what underlies the existence of this choice: lectal or semantic factors. This interpretation should not be confused with 'formal onomasiological variation' in studies by Geeraerts et al. (1994) and Geeraerts (2010), where it is applied to a set of naming options for the same concept across or within lects.

0378-2166/$ – see front matter
http://dx.doi.org/10.1016/j.pragma.2012.12.013

2002; Heylen, 2005; Tummers et al., 2005; Bresnan et al., 2007; Glynn, 2007; Grondelaers et al., 2002, 2007; Bresnan and Hay, 2008; De Sutter et al., 2008; Speelman and Geeraerts, 2009; Bresnan and Ford, 2010; Colleman, 2010; Szmrecsanyi, 2010). These studies focus on near-synonymous constructions, i.e. constructions with similar but not identical semantics. The differences between the near-synonyms are normally captured with the help of linguistic contextual features. The typical lectal factors studied by Cognitive Linguists include national varieties of language, communication channels and degrees of formality. Most studies that address lectal and conceptual variation of near-synonymous constructions usually find significant effects of both dimensions, although the language-internal factors tend to be more important. A possible explanation might be the principles of optimal organization of information in language: if two variants are highly distinct from the semantic point of view, their chances of conveying some significant social and stylistic distinctions are small (cf. Goldberg, 1995:67). Another factor is schematicity of constructions and their relatively low frequencies (cf. Cheshire, 2003), which may prevent variants from becoming consciously identified with certain social groups.

Another important strand of research is semasiological grammaticalization studies and contextualized Construction Grammar (e.g. Fried and Östman, 2005; Fried, 2009; Traugott, 2011), which reveal a complex interplay of formal, encyclopaedic, historical, interactional and other factors in semantic variation and change, including the ones that exist in the ''grey zone'' between cognition and communication *per se*: information structure, maxims of communication, pragmatic presuppositions, face-saving strategies and so on. Although this direction of research is very illuminating, we shall focus on the above-mentioned onomasiological corpus-driven studies, which seek to disentangle the different sources of variation in a methodologically robust way. On the basis of these quantitative studies we suggest three possible ways of interplay between semantic and lectal factors in co-determining constructional choices:

(a) *Independent lectal and conceptual variables.* This type of variation can be observed when lectal and conceptual factors influence the probabilities of near-synonyms independently. This requires some amount of free, semantically unconditioned variation, where the lectal factors can manifest themselves. However, there is very little evidence of this classical sociolinguistic type of variation in the usage-based research of near-synonymous constructions. One of the few examples is Bresnan et al. (2007), who find an independent effect of the spoken/written modality on the choice between two dative constructions in American Standard English.[2] Probably, this paucity of evidence in usage-based linguistics is due to the problematic nature of free constructional variation as a theoretical concept, which contradicts the Principle of No Synonymy (Goldberg, 1995).

(b) *Lectally moderated semantic variation.* This type occurs when the impact of the conceptual factors varies across different lects. The social dimension acts as a *moderator* of the conceptual factors (Baron and Kenny, 1986; Aguinis, 2004). The numerous examples include slightly different constraints on the use of presentative *er* in Belgian and Netherlandic Dutch (Grondelaers et al., 2002), different effects of animacy in the dative constructions in American and New Zealand English (Bresnan and Hay, 2008), difference in the use of the Norman vs. Saxon genitive in several varieties and registers of English (Szmrecsanyi, 2010).

(c) *Socioconceptual variation*. In some disciplines researchers deal with situations when one variable accounts for the causal relation between two other variables. This effect is called mediation. For example, there is a causal relationship between work stressors (e.g. a poor relationship with the boss) and alcohol use, but it is not direct, being mediated by anxiety and other psychological factors (Aguinis, 2004:5–6). Similar effects can be observed in linguistics, for example, when samples from different genres or varieties have different frequencies of the conceptual features that trigger the use of the one or the other near-synonym. This may result in a higher ratio of one variant for the given sample, and a stronger effect of the lectal factors when the relevant semantic features are not taken into account. The relationship between the social and the formal dimensions in this case is indirect: the social dimension influences the formal dimension via the conceptual dimension. An example is the study of the prepositional vs. double object dative in Bresnan et al. (2007). The proportion of the double object constructions was higher in their spoken data than in the written texts because the ratio of pronominal recipients, which increase the chances of the double object construction, was higher in the spoken subcorpus. Therefore, there was an indirect effect of the channel of communication on the formal variation via the language-factors (cf. note 2). The higher frequency of pronominal recipients in the spoken data might be explained by a more personal type of communication in the spoken corpus (the Switchboard collection of telephone conversations) in comparison with the written data (the Wall Street Journal corpus). We believe that such variables, which can be called socioconceptual, are an integral part of a fully recontextualized three-dimensional model of grammar, and should be taken into account in linguistic descriptions.

[2] However, according to our own logistic regression analysis of the data from Bresnan et al. (2007), this effect is very weak when the immediate contextual features (pronominality, givenness of the slot fillers, etc.) described in their paper are controlled for. This suggests that lectal effects may mask socioconceptual variation (see below).

Disentangling these sources of variation is a vital task for any usage-based study of constructions (the latter are understood here in a maximally broad sense, as symbolic units at any level of compositionality and schematicity). In the following sections we demonstrate how this can be done in a case study of two near-synonymous causative constructions in Belgian and Netherlandic Dutch. The rest of the article is organized as follows. The next section introduces the Dutch causative constructions with *doen* ''do'' or *laten* ''let'' and discusses the previous studies of the constructions. In section 3 we present the data and the statistical methods that can help us disentangle the interacting dimensions. We report the results of our analyses in section 4, followed by a discussion in section 5.

2. Dutch causative constructions: evidence from previous usage-based studies

Dutch causative constructions consist of the auxiliary predicate (*doen* or *laten*), the effected predicate and several nominal slots, as shown in the example below:

(1)

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

The difference between *doen* and *laten* as causative auxiliaries (like the English causative *make* and *let*, they are commonly used as lexical verbs, especially the multifunctional *doen*) has been explored in a number of corpus-based studies. Some of them (Kemmer and Verhagen, 1994; Verhagen and Kemmer, 1997; Stukker, 2005) argue that *doen* is a direct causation auxiliary, which is used in situations when the causer uses its own energy to produce a certain (inevitable) result. 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). It is important to note that the difference lies in the construal of events as direct or indirect causation, not in the events themselves, although some situations more naturally invoke the direct construal than the others, and vice versa. In (1), *deed* (the past form of *doen*) might be used in a situation when the police made efforts to bring about the effected event, for example by blocking the road. If *laten* is used, it is more probable that the police only signalled the car to stop, and the caused event took place because of the driver's conscious effort. However, this information is not encoded in the auxiliaries themselves.

The sentence with *laten* can also mean that the police allowed the car to stop. In fact, the semantics of *laten* is more schematic than that of *doen.* In the context of this paper, as in recent theorizing in Cognitive Linguistics in general, we will say that an expression is more 'schematic' to the extent that it is less specified, either in a strictly semantic sense, or in its combinatorial possibilities, than another one. Like the German *lassen*, *laten* represents a continuum from coercion to enablement and permission (Verhagen and Kemmer, 1997; Speelman and Geeraerts, 2009). Although it originates from the Gothic *lētan* ''let'', the coercive meaning of *laten* was attested already in the earliest Dutch texts available (van der Horst, 1998). Often it is impossible to tell whether the construal involves proper causation or letting. For example, the construction in (2) is difficult to interpret as strictly permissive or coercive:

(2)

Zij	*liet*	*de agent*	*haar*	*rijbewijs*	*zien*
she	let	the officer	her	driver's-license	see

''She showed the officer her driver's license.''
(from Verhagen and Kemmer, 1997)

Capturing such subtle differences in construal is by no means an easy task. A possible solution could be an analysis based on a number of lexical, syntactic and semantic features that could serve as some kind of circumstantial evidence of directness or indirectness. Some of such contextual features have been discussed previously. In a study based on a corpus of Netherlandic Dutch, Verhagen and Kemmer (1997) consider animacy of the causer and the causee, expression and marking of the causee, and transitivity of the effected predicate. They argue that these features are associated with the construal of causation as direct and indirect. For instance, in the case of transitives, the causation chain is longer, and the causer does not affect the end point directly, whereas intransitives imply a shorter causation chain and therefore more direct influence. Compare (3a) and (3b):

(3) a

Hij	*liet*	*zijn*	*huis*	*ontwerpen*	*door*	*de*	*architect.*
he	let	his	house	design	by	the	architect

''He had his house designed by the architect.''

b *De crisis deed de prijzen dalen.*
the crisis did the prices descend
''The crisis made the prices go down.''

In (3a), the causer affects the end point (the house) with the help of the causee (the architect), whereas in (3b), the crisis is construed as directly influencing the prices.

In Dutch the causee can be marked with the prepositions *door* ''by'' and *aan* ''to'', which indicate its roles of an agent (instrument) and recipient, respectively. Instrumental and dative marking shows that the causee is less affected than in the case of zero marking, which means that the causation is less direct (Kemmer and Verhagen, 1994; Verhagen and Kemmer, 1997). For instance, the causer's impact on the causee (the architect) is more direct in (4a) than in (4b).

(4) a *Hij liet de architect het huis ontwerpen.*
he let the architect the house design
''He had (let) the architect design the house.''

b *Hij liet het huis ontwerpen door de architect.*
he let the house design by the architect
''He had the house designed by the architect.''

c *Hij liet het huis ontwerpen.*
he let the house design
''He had the house designed.''

In addition, the causee can be left implicit, as in (4c). In such contexts the causee is also considered to be less affected than in cases like (4a). However, the realization of this rule is not straightforward. Most importantly, it applies only in the (predominant) cases of coercion or mild causation, but not in the cases of letting, where the causee is both explicit and minimally affected (cf. Loewenthal, 2003).

As for the animacy of the main participants is concerned, the configurations of animate causers and causees refer to inducive causation, which is normally indirect, since a human being cannot influence another human mind directly (Verhagen and Kemmer, 1997). An example of inducive causation is (5a). If both main participants are inanimate, one speaks of physical causation, as in (5b), which is usually construed as direct. The same construal is typical of affective causation, which involves an inanimate causer (stimulus) affecting a cognizer, as in (5c). Finally, an animate causer in a combination with an inanimate causee, as in (5d), often indicates so-called volitional causation, which does not show any clear 'preference' for the causation construal and, therefore, for an auxiliary.

(5) a *De trainer liet de spelers loopoefeningen doen.*
the coach let the players run-exercises do
''The coach had the players do running exercises.'' [inducive]

b *De aardbeving deed de muren trillen.*
the earthquake did the walls shake
'The earthquake made the walls shake.'' [physical]

c *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.'' [affective]

d *De machinist liet de motoren draaien.*
the engine-driver let the engines run
''The engine driver had/let/left the engines run(running).'' [volitional]

Some other features, such as animacy of the affectee and volitionality of causation, were explored by Stukker (2005) and Degand (2001), respectively, but were found to be less relevant for the division of labour between *doen* and *laten*. Speelman and Geeraerts (2009), in addition to transitivity of the effected predicate and animacy of the causer, also operationalized lexical and conceptual attraction between the auxiliary and the effected predicate, using an approach similar to collostructional analysis (Stefanowitsch and Gries, 2003). They also explored coreferentiality of the causer and other participants. An example is (6), where the first-person causer is coreferential with the affectee. Such contexts are semantically similar to the middle voice.

(6) *Ik liet me inspirieren door haar boek.*
I let me inspire by her book
''I let myself be inspired by her book.''

Speelman and Geeraerts (2009) found that coreferential contexts favour *laten* and argued that this fact contradicts the (in)directness hypothesis. According to them, the identity of the causer with the coreferential argument implies the minimal distance between them, and thus the maximal directness. That is why one should expect *doen* in coreferential contexts. However, if one interprets directness and indirectness in terms of the division of labour between the participants in bringing about the effected event, rather than in terms of their contiguity, it is in fact possible to explain the preference of *laten* in coreferential contexts. In (6), the coreferential causer is not the main source of energy, but the entity responsible for the effected event taking place (Loewenthal, 2003). The main cause of the event is the causee (the book). Therefore, this and other coreferential contexts do fit the indirectness construal.

The present study continues and elaborates the previous quantitative analyses of *doen* and *laten*. Of course, such an approach does not, and cannot provide an exhaustive interpretation of the construal for every particular context found in a corpus (cf. Verhagen and Kemmer [1997], who examine a few cases where *doen* is used for ironical and other stylistic purposes in contexts where one would expect to see *laten*). Such interpretation would involve a close examination of the relevant encyclopaedic frames, communication settings, and other background information, which, unfortunately, is extremely difficult to capture with the help of observable linguistic features. What our approach yields, is the probabilities of use of *doen* or *laten* in a combination with a given contextual feature, or a combination of features. These probabilities can interpreted in terms of directness/indirectness or, possibly, another semantic distinction. However, this seeming indirectness of evidence and coarse granularity of the semantic description are compensated by the objectivity of such evidence and quantification of the results – the advantages that by far outweigh the shortcomings.

The lectal dimension of the *doen* and *laten* use has been studied, as well. Speelman and Geeraerts (2009) show that *doen* is more frequently used in Belgian Dutch than in the Netherlandic variety, and more commonly in formal registers of spoken communication than in informal ones. They interpret these findings as evidence of a gradual decay of the causative *doen.* The Netherlandic variety is traditionally considered to be the leader in various linguistic changes, whereas the Belgian variety tends to 'lag behind' due to a number of historical and cultural reasons, such as the belated standardization and persistence of dialects in Belgian Dutch (see Geeraerts et al., 1999:13–18; Auer, 2005:25). On the other hand, formal language tends to retain more archaic features than informal one. The multivariate analyses in Levshina (2011) based on both spoken and written data corroborate Speelman and Geeraerts' conclusions. The present study examines one register of written communication – newspaper Dutch in the Netherlands and Belgium (more precisely, its Dutch-speaking part Flanders), and focuses only on the geographic variation. The goal is to examine as many potential interactions between the dimensions of variation as possible.

As for the methodology, the present study employs cutting-edge statistical techniques that allow for disentangling the different dimension of variation. The analyses in Verhagen and Kemmer (1997) were based, as many studies in Cognitive Linguistics, on several tables with cross-tabulated frequencies of one or more features, on the one hand, and the two auxiliaries, on the other hand. The differences between *doen* and *laten* in the proportions of the features were interpreted as evidence of semantic differences. In her bivariate analysis, Stukker (2005) used a more robust statistical hypothesis-testing technique, which shows whether one can reject the null hypothesis that there is no difference in the use of *doen* and *laten* with regard to one or more specific features. This approach allows one to extrapolate the results from a sample to the entire population and thus overcome the limitations of the available data. Speelman and Geeraerts (2009) combined the multiple features with the statistical power in their multivariate logistic regression model of *doen* and *laten*. The present study enhances this multivariate approach by using mixed-effect regression, which enables us to take into account the lexical attraction of effected predicates to *doen* and *laten*. We also add several new local contextual variables to the model. The variables and the method are presented in the next section.

3. Data and method

From Twente News Corpus (Ordelman et al., 2007) and Leuven News Corpus,[3] two corpora of Dutch and Flemish newspapers, we created a regionally and thematically balanced corpus of 8 million words, selecting the articles that represent four subject domains: economy, politics, music and football. The corpus was syntactically parsed with the Dutch Alpino parser (Bouma et al., 2001), which allowed us to extract all contexts with the causative auxiliaries *doen* and *laten* automatically (the lexical uses of the verbs were filtered out). After some spurious hits had been deleted manually, we had

[3] Leuven News Corpus is a large (over 1 billion words) syntactically parsed corpus of Belgian newspapers in Dutch compiled by the Quantitative Lexicology and Variational Linguistics research unit at the University of Leuven.

a total of 6855 observations. 83% of them contained *laten*, and only 17% were with *doen*. This asymmetry supports the previous results (see above), which suggest that *doen* is a marginal form.

The formal variation of the two constructions was coded as the *Auxiliary* variable with the values ''doen'' and ''laten''. The lectal dimension was represented in this study by the Country variable (Belgium or the Netherlands). Below is the list of the conceptual features that were considered relevant for the *doen* and *laten* distinction, according to the previous research (see section 3) and our own pilot studies. Although some of them are 'semantic' in the classical sense, and the others are 'syntactic', the symbolic nature of grammar in Cognitive Linguistics (e.g. Langacker, 1987) allows us to treat both as conceptual in a broad sense. More importantly, for our present purposes, we refer to these features as 'conceptual' because they involve distributional differences in the linguistic context of the auxiliaries that could be diagnostic for semantic differences between the verbs. In other words, any semantic differences will be reflected in the way in which the auxiliaries combine with certain classes of causers, causees, etc. In addition, both the syntactic and semantic features, in combination with related encycopaedic and situational information, serve as interpretive clues (cf. Fried, 2010) that help the hearer to grasp the meaning of the construction, and thus can be regarded as signs (cf. Verhagen, 2009). Most original values of the variables were conflated to avoid data sparseness, so we present them here as they appear in the final analyses:

- *CrSem*: semantic class of the causer (animate or inanimate);
- *CeSem*: semantic class of the causee (animate or inanimate);
- *CeSynt*: syntactic expression of the causee (''central'', when the causee is expressed by a nominal phrase or a clause; or ''peripheral'', when the causee is implicit or expressed by a nominal phrase with the prepositions *aan* or *door*);
- *EPTrans*: transitive or intransitive effected predicate, transitivity defined as the presence of the direct object;
- *CdEventSem*: whether the caused event is mental or non-mental (physical or social);
- *Coref*: if the causer is coreferential with the other participants or not;
- *Possess*: whether or not there are relationships of possession between the causer and the other participants (understood broadly)[4];
- *Polarity*: whether or not the clause with the causative construction contains a negation.

The main tool of our quantitative analysis is logistic regression, a technique that allows the researcher to explore the role of one or more factors in predicting a binary outcome (the response variable), e.g. whether the speaker will use *doen* or *laten* in a given context. This method allows for measuring the association of each feature (variable) with *doen* and *laten*, while controlling for all other variables. The technique has been used in many quantitative studies of near-synonyms, as well as in the traditional sociolinguistic research. The resulting model is a mathematical equation with numeric estimates for each factor. Special techniques are available to perform evaluation and diagnostics of the model (see, for example, Hosmer and Lemeshow, 2000). On the basis of the diagnostic tests we excluded two untypical influential observations with *doen* and improved the fit.

A useful addition to logistic and other regression techniques is so-called mixed-effect models, which combine fixed effects (such as the factors that we described above) with random effects. Usually, random effects are adjustments made to individual subjects or items to compensate for a potential correlation between their repeated measurements (see Baayen, 2008 for examples). In this study, the random effects were the infinitives in the effected predicate slot (cf. Bresnan et al., 2007).

Table 1 summarizes the operationalization of the three dimensions as variables.

4. Variation of Dutch causative constructions

4.1. 2D-models

Though our ultimate goal is a 3D-model of the *doen*- and *laten*-variation, it seems convenient to begin the discussion with the 2D-models, which describe the general semantic and geographic tendencies in our data.

4.1.1. Conceptual variables: a general picture

In order to obtain a general picture of the use of the constructions in both varieties, we fitted a logistic regression model with all possible semantic variables and their two-way interactions. All conceptual factors described above were found to be significant (here and in all subsequent analyses the significance level is set at $\alpha = 0.05$). The model also contained

[4] We used possessive pronouns and nouns in the possessive case as indicators of relationships of possession.

Table 1
Operationalization of dimensions as variables.

Formal dimension	Conceptual dimension		Lectal dimension
	Fixed effects	Random effects	
Auxiliary	*CrSem* *CeSem* *CeSynt* *EPTrans* *CdEventSem* *Coref* *Possess* *Polarity*	1166 effected predicates	*Country*

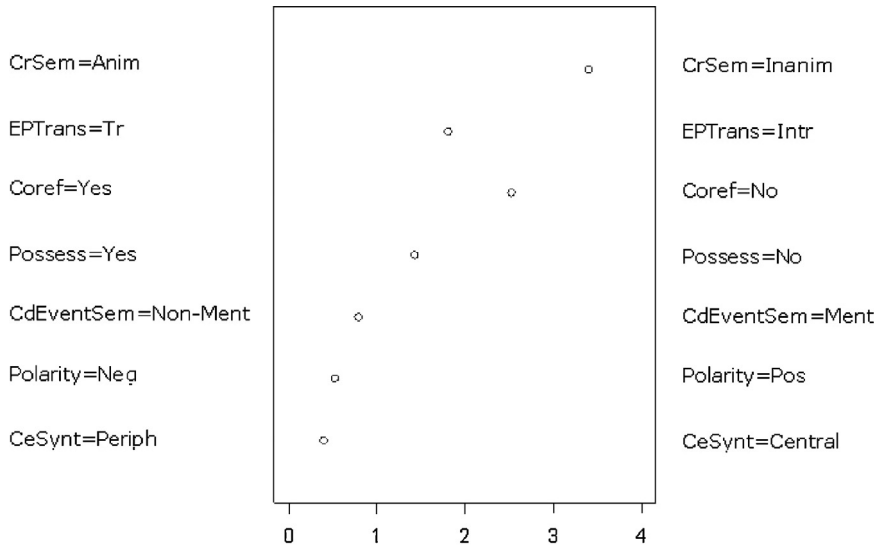

Fig. 1. Effect of conceptual factors on log-odds of *doen* (main effects only).

eight significant interactions, which we address later. The main effects are shown in Fig. 1. On the left are the values of the variables that increase the probability of *laten*: animate causers, transitive effected predicates, coreferentiality, possession relations, non-mental caused events, negative polarity, implicit or prepositionally marked causees. On the right side are those which boost the chances of *doen*: inanimate causers, intransitive effected predicates, lack of coreferentiality and possession relations, mental caused events, lack of negation, explicit causees without prepositional marking. The scale measures the increase in the probability (more precisely, log odds ratio) of *doen* for the given factor. The greater the value on the scale, the greater role the corresponding variable plays in discriminating between the auxiliaries.

The most prominent role in the model is played by the causer and its relationships with the other participants. *CrSem* is the most powerful predictor and the most distinctive feature. Inanimate causers increase the chances of *doen* with log odds 3.39, which can be transformed into a simple odds ratio of 29.67. This means that the *doen/laten* ratio is almost 30 times higher for inanimate causers than for animate causers when the other factors are controlled for. In other words, animate causers strongly increase the probability of *laten*, and inanimate ones boost the chances of *doen*. Note that *CeSem* (the semantic class of the causee) is not shown because it is significant only in the presence of the interactions. In that case, inanimate causees slightly increase the probability of *doen* (with the log odds of 0.27 and $p = 0.034$). This fits the previous observations about the direct and indirect causation construals (see section 2): whereas animate causers are distinctive of inducive causation, which is associated with *laten*, inanimate ones are typical of physical and affective causation, which are typical of *doen*. As for the causee, physical and affective causation involve inanimate and animate causees, respectively, which makes the semantics of the causee a poor criterion for the prediction. The variables *EPTrans*, *CeSynt* and *Coref* support the indirectness/directness distinction suggested by Kemmer and Verhagen (1994), as well. The same seems to hold for the predominance of possessing causers in the contexts with *laten*. Compare (7a) and (7b), which are identical with regard to all variables except for *Possess*:

(7) a *De Post wil de bedrijven doen betalen om hun*
the Post wants the companies make pay in-order their
zendingen 's morgens voor 8 uur te ontvangen.
parcels in-the-morning before 8 o'clock to receive
''The Post wants to make the companies pay if they want to receive their parcels before 8 a.m.''

b *De trainer liet zijn ploeg veel te voorzichtig aan de wedstrijd beginnen.*
the trainer let his team much too cautiously at the match begin
''*The trainer let his team begin the match far too cautiously*.''

In (7a), the Post forces the other companies to pay more, so that the causation is inevitable and direct. The context in (7b) is the one where the contrast between coercion and letting is neutralized. On the one hand, the players might have followed the trainer's directions to be careful. On the other hand, the trainer might have failed to correct the performance of his team. In either case, the players have some freedom to carry out the effected event. The trainer is not the driving force in bringing about the event. Instead, he is the one responsible for it. This responsibility is encoded in the possessive pronoun (*his* team). As in the case of coreferential contexts like (6), possession (with its generalized extensions) indicates the domain of the causer's responsibility, which enables the indirect causation construal.

As far as *Polarity* is concerned, a closer examination of the data shows that negative contexts usually involve the semantics of not letting, where the actual source of energy is most often the causee and the causer resists causation instead of initiating it:

(8) *Ik liet me niet intimideren door hem.*
I let me not intimidate by him
''I didn't let him intimidate me.''

Moreover, the contexts with negation are often coreferential, as in (8). In such cases the causer is not only the entity responsible for his or her own mental state, but also the participant potentially or actually affected by the causee.

Finally, we have to account for the association between mental caused events and *doen*. Consider (9), repeated here for the sake of convenience:

(9) *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.''

This finding does not seem directly related to the causation construal. Recall that the main senses of the causative *doen* are affective (changing a mind), as in (9), and physical (changing the world). In the broader context of the ongoing qualitative and quantitative decrease in the use of *doen* (Speelman and Geeraerts, 2009; Levshina, 2011), we can hypothesize that the use of *doen* is becoming restricted to affective causation.

One of the advantages of the method is that it allows for evaluation how successfully all factors together account for the use of the near-synonyms, and whether the researcher misses any important predictors. This information is provided by the so-called goodness-of-fit statistics. The first row of Table 2 lists some of them for the complete model with interactions. The proportion of correct predictions shows how many uses of *doen* and *laten* are predicted by the model correctly if the probability of *doen* greater than 0.5 is counted as a prediction of *doen*, and the probability of *doen* less than 0.5 is interpreted as a prediction of *laten*. The *C* measure is the index of concordance between the predicted probability and the observed response with the range of values from 0.5 (random predictions) to 1 (a perfect match).

Table 2
Some statistics of two models with conceptual features only.

Model	Number of observations	Proportion of correct predictions	*C*	Nagelkerke's R^2
Conceptual variables (fixed effects only)	6853 (1168 with *doen*, 5685 with *laten*)	91%	0.92	0.60
Conceptual variables as fixed effects and EPs as random effects		99.96%	0.99	NA

Table 3
Top five verbs attracted to *doen* and *laten*, added as random effects.

Pro-*laten* verb	Random effect	Pro-*doen* verb	Random effect
horen ''hear''	−4.59	*respecteren* ''respect''	6.10
zien ''see''	−4.21	*fronsen* ''frown''	5.41
weten ''know''	−4.10	*kantelen* ''topple''	5.00
liggen ''lie''	−3.92	*opmerken* ''notice''	4.77
wachten ''wait''	−3.49	*vergeten* ''forget''	4.67

Nagelkerke's R^2 is a coefficient analogous to R^2 in linear models, where it stands for a proportion of variation explained by the model.

The statistics show that our first model fits reasonably well. However, as the previous studies (Speelman and Geeraerts, 2009; Levshina et al., 2011) demonstrated, there is another important factor that should be taken into account: lexical attraction between the effected predicate slot fillers and *doen* or *laten*. Some of the verbs prefer *doen* (e.g. *denken* ''think'' in *doen denken aan* X ''make think, remind of X''), whereas the others are used predominantly with *laten* (e.g. *zien* ''see'' in *laten zien* ''show''). This was the reason why we added the effected predicates as random effects to the analysis. The resulting model had an outstanding predictive power (see the statistics in the second row of Table 2). Table 3 displays five verbs that were the most strongly attracted to *laten*, five verbs with the highest preference for *doen*, and their individual effects (adjustments to the intercept). Interestingly, most of the verbs strongly attracted to *doen* and *laten* are mental, but the three top *laten*-verbs *horen* ''hear'', *zien* ''see'' and *weten* ''know'' refer to communication and sharing information, as in *he let me know that*, whereas the top *doen*-predicates describe other types of intellectual and emotional impact. Communication presupposes an addressee, who often benefits from the information and even actively looks for it (e.g. *let me know what you think!*). Moreover, the role of the causee in such constructions is similar to the role of the recipient, a marginally affected participant in the ditransitive construction. Therefore, such situations can be construed as indirect causation.

In the new model the semantic class of the causee ceased to be significant, even marginally. The other main fixed effects in the new model did not change dramatically, so the general tendencies that we described above were still valid. However, only three out of the eight interactions found in the previous model retained their significance: *CrSem*Coref*, *Coref*Possess* and *EPTrans*CeSynt*. Interactions are observed when the effects of two or more variables taken together are non-additive, i.e. they cannot be predicted by the sum of the individual effects of the variables. For example, the effect of age on the probability of a disease can be different for males and females. Sometimes the effects can be opposite. This is what we find when we examine the interaction *EPTrans*CeSynt*.

This interaction, visualized in Fig. 2, can be interpreted as follows. For transitive effected predicates, leaving out the causee or marking it with a preposition leads to a lower probability of *doen* (and, inversely, higher chances of *laten*). If the

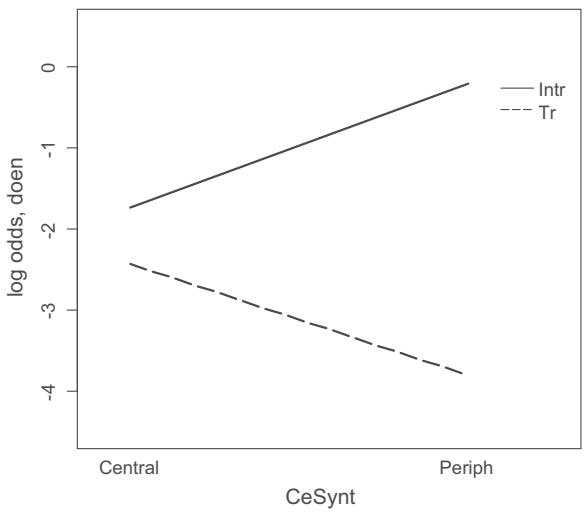

Fig. 2. Interaction *EPTrans*CeSynt* in the mixed-effect model with conceptual factors only.

effected predicate is intransitive, however, syntactic peripherality of the causee leads to a higher probability of *doen*. Compare (10a) and (10b):

(10) a *Ik liet mijn huis ontwerpen.*
I let my house design
''I had my house designed.''

b *De eerste repetitie doet verlangen naar meer.*
the first rehearsal does long to more
''The first rehearsal makes you long for more.''

Both contexts have an implicit causee (which is by far the most frequent type of the peripheral causees). In (10a), where the effected predicate is transitive, the causee's peripherality means that he or she is only an intermediary between the focal participants, which is typical of indirect causation. In (10b), an intransitive construction, the implicit causee is the final affected entity, the end point of the energy flow. It is focal and should normally be explicit. However, its discourse prominence in the contexts like (10b) is low due to the generic nature of the causee: the music is so good that it can make anyone want to hear more. This pragmatic shift from the process of causation to the evaluation of the stimulus is common in newspaper music reviews, where this example was found. Culturally triggered avoidance of the first person reference (cf. Goldberg, 2005) could also be an explanation because the context represents the journalist's personal evaluation of the music.

The other significant interactions are *CrSem*Coref* and *Coref*Possess*. The first one is quite mild. It shows that the effect of animacy or inanimacy of the causer is weaker in coreferential contexts with a high probability of *laten* than in non-coreferential contexts. We suppose that the lack of variation in the coreferential contexts is due to the high proportion of the cases of (not) letting, as in (8), which is conveyed exclusively by *laten*. Ideally, one should exclude the letting contexts from the study, making the approach more restrictive (cf. Geeraerts, 2010). However, as was shown in (2), the schematic semantics of *laten* makes this problematic. Nevertheless, even if we ignore numerous ambiguous cases when it is not clear whether the construction conveys letting or coercion, 53% of all coreferential contexts with *laten* can be interpreted as clear cases of letting, as opposed to only 9% for the non-coreferential contexts.

The presence of the *Coref*Possess* interaction is not surprising because the two variables reflect similar conceptual contents: what is going on is within the causer's domain of control and responsibility. A closer look at the interaction pattern reveals that coreferentiality does not boost the probability of *laten* when possession is already present, and possession relations do not increase the probability of *laten* in coreferential contexts.

We would like to emphasize that the statistical insignificance of the other interactions that were significant in the fixed-effect model – for example, the interaction *CrSem*EPTrans*, which showed a smaller effect of the semantic class of the causer for transitives than for intransitives – does not mean that they cannot exist. This loss of significance means only that these interactions can be fully or to a large extent predicted on the basis of lexical information. However, this is not a reason for them to be absent from a constructional description, as was argued in section 2. According to the usage-based non-reductionist constructionist approach (e.g. Goldberg, 2006), to ignore such generalizations would be an example of the exclusionary fallacy, which separates lexicon ('lists') and grammar ('rules') (Langacker, 1987). Even if constructions are fully predictable from their components, they do exist if they occur with a sufficient frequency (Langacker, 2005; Goldberg, 2006). The same should be true about onomasiological constructional constraints at different levels of generalization (see a unified account of constructions and constraints in Cappelle, 2009). The mixed-effect models are thus an important tool in detecting lexically predictable patterns, but they should be used in a combination with the fixed-effect ones.

To summarize, our 2D-analysis of the semantics of *doen* and *laten* supports the direct/indirect causation distinction, which was proposed in previous research. The behaviour of the previously described and novel contextual variables with regard to the choice between the auxiliaries can be interpreted within the existing theory. The only exception is the preference of mental caused events for *doen*, which probably indicates a semantic specialization of *doen* in affective causation. This explanation fits the existing accounts of the ongoing linguistic change in Dutch. The statistical models predict the choice between the auxiliaries well, which suggests that we have managed to take into account the most influential conceptual factors of variation.

4.1.2. Geographic variation: a general picture

This time, we fitted a simple regression model with *Country* as the only predictor. According to this model, the *doen*/*laten* ratio in Belgium is 2.15 times higher than in the Netherlands, as shown in Fig. 3. This finding fits the idea of *doen* as an obsolecent form, which persists in the more archaic variety. However, this is only a general tendency. In what follows we explore the causes of this variation and whether it is stable across different semantic contexts.

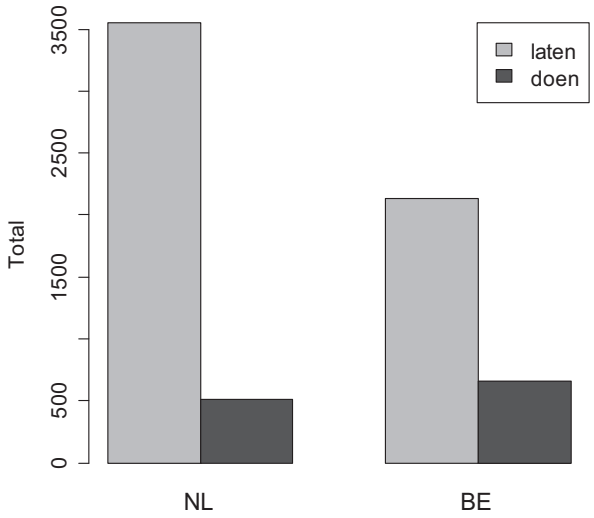

Fig. 3. The national distribution of *doen* and *laten* in the corpus.

4.2. Interaction of the conceptual and lectal dimensions

Our next step is to test whether *Country* and the semantic factors co-influence the choice of *doen* and *laten* independently, or they interact. We fit two models, which allowed all possible two-way interactions between *Country* and the conceptual factors.[5] The first one was with fixed effects only; the second one also contained effected predicates as random effects. In the first model we found two interactions between the conceptual and lectal factors: *EPTrans*Country* ($p < 0.001$) and *Coref*Country* (with a borderline significance: $p = 0.06$), shown in the 3D-plots in Fig. 4.[6]

Both interactions are mild and non-crossover. To examine the difference in the semantic constraints in the two varieties, one has to look at the planes from the right-hand side and compare the slopes corresponding to the Belgian and the Netherlandic samples. The left plot demonstrates that the slope in the Netherlandic data is steeper than that in the Belgian sample. This means that (in)transitivity of the effected predicate has a more dramatic effect in the Netherlands, where the probability of *doen* for transitives is extremely low. The right plot shows that *Coref*, on the contrary, allows more variation in Belgium than in the Netherlands. Literally changing the perspective and looking at the plots from the left-hand side, one can see that *doen* and *laten* are better sociolinguistic variables in transitive than in intransitive contexts, and in non-coreferential than in coreferential contexts, according to the steepness of the corresponding semantic slopes. In fact, the national difference is almost neutralized in coreferential contexts, which have a relatively low probability of *doen*. The reader may recall the interaction pattern *CrSem*Coref* discussed in section 4.1.1, where the effect of *CrSem* was blocked in the coreferential contexts due to the prominence of the semantics of letting, which is expressed exclusively by *laten*. In the case of *Coref*Country* the explanation may be similar. It is natural that the coreferential semantic contexts, which exhibit less formal variation, are less influenced by lectal factors than the non-coreferential contexts with a stronger competition between two forms.

The second model included the effected predicates as random effects of two types: (a) the general attraction of the verb towards *doen* or *laten*, as in the previous mixed model (adjustments to the intercept); (b) the country-specific nuances in the direction and degree of this attraction (adjustments to the slope, or estimate, of the *Country* variable). Fig. 5 shows all effected predicates that occurred both in Belgium and the Netherlands, plotted according to their adjustment values.

The numeric values are the adjustments that should be added to obtain the chances of *doen* and *laten* (expressed as logarithmically transformed odds ratios) with the given verb. The horizontal dimension corresponds to the log odds ratio of *doen* vs. *laten* in the Netherlands; the vertical dimension shows the probabilities in Belgium. The dotted diagonal line above the grey line corresponds to the fixed effect of the country (Belgium) with the log odds ratio of 1.137 and $p < 0.001$. One can see that there is a high correlation of the values in the two varieties. This suggests that in most cases the caused

[5] We have also tested all possible three-way interactions, but adding them did not result in a substantial increase in the predictive power of the model (based on Akaike's Information Criterion), so we left them out.

[6] Because the interacting variables also took part in a few other interactions not discussed here, the estimates for each plot were taken from the model without the other interactions that contained the same variable.

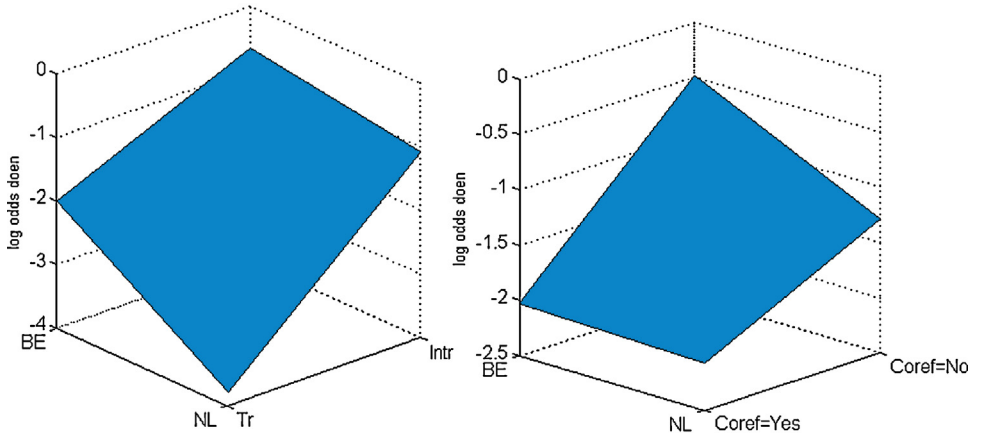

Fig. 4. Interactions *EPTrans*Country* and *Coref*Country* in the fixed-effect model.

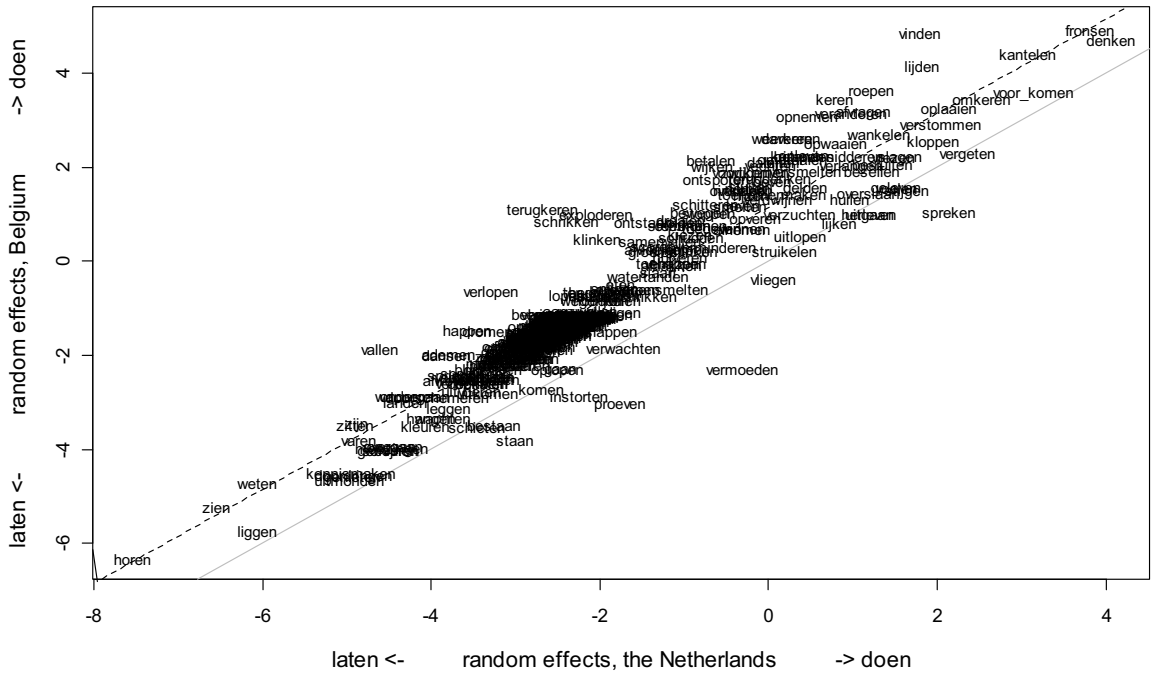

Fig. 5. Random effects of effected predicates in the mixed model.

events associated with specific verbs trigger similar construals in both countries. At the same time, there are a few national peculiarities. If the verb is above the solid grey diagonal line, the probability of *doen* in the reference context is higher in Belgium than in the Netherlands. For example, the verb *terugkeren* ''return'' increases the chances of *doen* in Belgium, but it is boosts the chances of *laten* in the Netherlands. On the contrary, the verbs below the line (e.g. *vermoeden* ''suppose'') are more strongly attracted to *doen* in the Netherlands than in Belgium. It is very difficult to say what causes these individual preferences. To some extent, they are due to the presense of idiomatic expressions: for example, in the Netherlandic sample *spreken* ''speak'' is very often used in the idiom *van zich doen spreken* ''lit. make (people) talk about oneself, get noticed''. This can be treated as evidence of a mediation-like effect of the fine-grained meanings expressed by the collocations that are more frequent in one sample than in the other one.

A zoom-in on the verbs in the big black blurb shows that most of them are used only once with *laten*, and therefore no significant preferences for *doen* or *laten* can be established for them. The high type frequency of such verbs indicates that *laten* as a constructional schema is stronger than *doen.* Most of the low-frequency verbs with *laten* from the black blurb have slightly higher chances to be used with *laten* in the Netherlandic variety (they are above the grey line) because most

of them occur in the Netherlandic sample. This indicates that that the overall preference for *laten* in the Netherlandic variety is to a large extent due to these low-frequency verbs, whereas *doen* persists in frequent fixed collocations.

The general semantic patterns expressed by the conceptual features were similar to the ones discussed earlier, but the two interactions of the conceptual variables with *Country* that we observed in the previous model lost in effect sizes and significance.[7] For the *EPTrans*Country* interaction, the *p*-value changed dramatically from $p < 0.001$ to $p = 0.96$, which means that this interaction emerged due to the lexical effects. Note that the three highly frequent verbs *horen* ''hear'', *zien* ''see'' and *weten* ''know'' (see the verbs in the bottom left corner in Fig. 5) are nearly exclusively *laten*-verbs, appearing predominantly in transitive constructions. At the same time, they are the most frequent effected predicates in the Netherlandic sample, accounting for 28.3% of all Dutch observations versus only 3.4% of all Belgian cases. The interaction *EPTrans*Country* can thus be explained to a large extent by these lexical patterns, and the cause of the interaction is different frequencies of specific collocations in the two samples.

The significance of the interaction *Coref*Country*, however, changed less (from $p = 0.06$ to $p = 0.12$). A closer look at the interaction reveals that the new pattern was very similar to the one shown in Fig. 4, but the difference between the varieties became less outspoken in non-coreferential contexts, probably due to the fact that the lexical effects are incorporated in the model.

5. Conclusions

The 2D and 3D analyses support the previous findings and add new observations. From the conceptual perspective, there is evidence that the key semantic difference between *doen* and *laten* is related to the direct and indirect causation construals. This conclusion is supported by our multivariate analyses with different semantic and syntactic contextual features, which serve as 'circumstantial evidence' of the construals. We have also found that *doen* is preferred by mental caused events, which indicates a specialization of the auxiliary towards affective causation in contemporary Dutch, with the exception of the verbs of a few verbs related to obtaining information, which are attracted to *laten*. This fact should be interpreted within a broader perspective – a gradual decline of the causative *doen* from the quantitative and qualitative points of view.[8] The geographic variation in the use of the auxiliaries supports this conclusion because the proportion of *doen* is lower in the Netherlandic subcorpus than in the Belgian one, and the Netherlandic variety is considered to be the leading one with regard to language change.

As for the interaction between the conceptual and lectal factors, our results demonstrate a complex mixture of different variation types at more schematic and more lexically specific levels. Like in most previous studies of constructional variation, the influence of the social dimension on the probabilities of the two variants is much weaker than the integral impact of conceptual factors. This influence – again, similarly to most other studies – manifests itself indirectly, through moderation- and mediation-like effects. Moreover, these effects can be explained to a large extent by the country-specific lexical patterns.

More specifically, we found evidence of lectally moderated semantic variation at different levels of schematicity, although the national difference in the semantic constraints is a matter of degree, not the substance of these constraints. The difference between coreferential and non-coreferential contexts is more outspoken in the Belgian sample, whereas (in)transitivity of the effected predicate seems to play a more important role in the Netherlandic data. Our mixed-effect modelling revealed that the transitivity effect emerged due to the difference in the distribution of effected predicates in the corpus, such as the predominance of transitive verbs *zien* ''see'', *horen* ''hear'' and *weten* ''know'', which form fixed expressions with *laten*, in the Netherlandic sample. In other words, the moderating effect of the *Country* variable at the more schematic level of constructional variation can be explained by the mediating effect of the collocations at the more specific constructional level. The interaction with coreferentiality, however, is less dependent on the specific verbs. It seems to be based on the restricted formal variation in the coreferential contexts due to the predominance of the meaning of letting, which is expressed exclusively by *laten*.

An important question is whether we can extrapolate these findings from the newspaper data set to the entire language. The tendencies discussed here may be caused by country-specific peculiarities of the newspaper genres. A close examination of the contexts suggests that this might well be the case. For example, the majority of the uses of *laten weten* ''let know, inform'' refer to official communication reported in the news, as in (11):

[7] We have tested all other interactions between the region and the conceptual variables, together and one by one, but none of them was significant at $\alpha = 0.05$, or improved the predictive power of the model measured as the correlation of the predicted values with the observed values.

[8] Although Verhagen (2000) does not regard the causative *doen* as obsolescent, his diachronic analysis suggests that the use of the auxilary has narrowed since the 18th century, at least with regard to inducive causation.

(11) *Minister Zalm heeft dat dinsdag aan de Amerikaanse autoriteiten laten weten.*
Minister Zalm has that Tuesday to the American authorities let know
''Minister Zalm informed the American authorities about that on Tuesday.''

In addition, an examination of spontaneous face-to-face conversations from the Corpus of Spoken Dutch by Levshina (2011) shows no evidence of a higher frequency of these specific *laten*-constructions in the Netherlandic subcorpus in comparison with the Belgian data. Thus, these national differences are characteristic of the newspaper discourse. Still, this strong entrenchment of the marginally causative collocations with the mental verbs in the Netherlandic register is surprising. Interestingly, the greater prominence of some mental effected predicates in the construction with *doen* is also observed in a Netherlandic Dutch corpus (Levshina et al., 2011).

The role that the effected predicate slot fillers played in the interaction with transitivity highlights the importance of incorporating patterns at different levels of lexical specificity in a study of constructional variation. Ideally, we should shift from the analysis of individual constructions to entire networks of constructions. On the other hand, more social variables could be added: dialects, registers, genres and styles. A challenge for empirical semantics is an operationalization of relevant encyclopaedic, discursive and situational parameters, which could be added as variables to the model (e.g. Bergs and Diewald, 2009). A fully statistical comprehensive analysis of that kind will require huge samples and very sophisticated methods, which are yet to be developed. The present study is only one of the first steps in that direction.

Acknowledgements

This research was supported with a grant from FWO – Research Foundation of Flanders. We are also grateful to Joan Bresnan for helping to access the data from Bresnan et al. (2007), Kris Heylen for his help in extracting the data used in this study, and Benedikt Szmrecsanyi for his critical comments on the previous version of the paper. The remaining errors are, of course, ours. In our analyses we used R (R Development Core Team, 2010), an open source environment for statistical computing and graphics, with the special packages *Design* (now *rms*) for regression analysis and *lme4* for mixed models. The graphics were produced with the help of R and MATLAB.

References

Aguinis, Herman, 2004. Regression Analysis for Categorical Moderators. Guilford, New York.
Auer, Peter, 2005. Europe's sociolinguistic unity or: a typology of European dialect/standard constellations. In: Delbecque, N., van der Auwera, J., Geeraerts, D. (Eds.), Perspectives on Variation: Sociolinguistic, Historical, Comparative. Mouton de Gruyter, Berlin, pp. 8–42.
Baayen, Harald, 2008. Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge University Press, Cambridge.
Baron, Reuben M., Kenny, David A., 1986. The moderator–mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology 51, 1173–1182.
Bergs, Alexander, Diewald, Gabriele (Eds.), 2009. Contexts and Constructions. John Benjamins, Amsterdam.
Bouma, Gosse, van Noord, Gertjan, Malouf, Robert, 2001. Alpino: wide-coverage computational analysis of dutch. In: Dalemans, W, et al. (Eds.), Computational Linguistics in the Netherlands 2000: Selected Papers from the Eleventh CLIN Meeting. Rodopi, Amsterdam, pp. 45–59.
Bresnan, Joan, Ford, Marilyn, 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. Language 86 (1), 186–213.
Bresnan, Joan, Hay, Jennifer, 2008. Gradient grammar: an effect of animacy on the syntax of give in New Zealand and American English. Lingua 118 (2), 245–259.
Bresnan, Joan, Cueni, Anna, Nikitina, Tatiana, Baayen, Harald, 2007. Predicting the dative alternation. In: Bouma, G., Kraemer, I., Zwarts, J. (Eds.), Cognitive Foundations of Interpretation. Royal Netherlands Academy of Science, Amsterdam, pp. 69–94.
Cappelle, Bert, 2009. Contextual cues for particle placement: multiplicity, motivation, modeling. In: Bergs, A., Diewald, G. (Eds.), Context in Construction Grammar. John Benjamins, Amsterdam, pp. 145–192.
Cheshire, Jenny, 2003. Social dimensions of syntactic variation: the case of *when* clauses. In: Britain, D., Cheshire, J. (Eds.), Social Dialectology. John Benjamins, Amsterdam, pp. 245–261.
Colleman, Timothy, 2010. Lectal variation in constructional semantics: benefactive ditransitives in Dutch. In: Geeraerts, D, Kristiansen, G, Peirsman, Y (Eds.), Advances in Cognitive Sociolinguistics. Mouton de Gruyter, Berlin/New York, pp. 191–221.
De Sutter, Gert, Speelman, D., Geeraerts, D., 2008. Prosodic and syntactic–pragmatic mechanisms of grammatical variation: the impact of a postverbal constituent on the word order in Dutch clause final verb clusters. International Journal of Corpus Linguistics 13, 194–224.
Degand, Liesbeth, 2001. Form and function of causation. In: A Theoretical and Empirical Investigation of Causal Constructions in Dutch Peeters, Leuven.
Fried, Mirjam, 2009. Representing contextual factors in language change: between frames and constructions. In: Bergs, A., Diewald, G. (Eds.), Context in Construction Grammar. John Benjamins, Amsterdam, pp. 63–94.
Fried, Mirjam, 2010. Constructions and frames as interpretive clues. Belgian Journal of Linguistics 24, 83–102.
Fried, Mirjam, Östman, Jan-Ola, 2005. Construction grammar and spoken language: the case of pragmatic particles. Journal of Pragmatics 37, 1752–1778.
Geeraerts, Dirk, 2009. Recontextualizing grammar: underlying trends in thirty years of cognitive linguistics. In: Tabakowska, E. (Ed.), Cognitive Linguistics in Action: From Theory to Application and Back. Mouton de Gruyter, Berlin, pp. 71–102.

Geeraerts, Dirk, 2010. Lexical variation in space. In: Auer, P., Schmidt, J. (Eds.), Language in Space. An International Handbook of Linguistic Variation. Theories and Methods, vol. 1. De Gruyter Mouton, Berlin/New York, pp. 821–837.

Geeraerts, Dirk, Grondelaers, Stefan, Bakema, Peter, 1994. The Structure of Lexical Variation: Meaning, Naming, and Context. Mouton de Gruyter, Berlin.

Geeraerts, Dirk, Grondelaers, Stefan, Speelman, Dirk, 1999. Convergentie en Divergentie in de Nederlandse Woordenschat. Een Onderzoek Naar Kleding- en Voetbaltermen. (Convergence and Divergence in the Dutch Lexicon. A Study of Clothing and Football Terms). Meertens Instituut, Amsterdam.

Glynn, Dylan, 2007. Mapping meaning. Towards a usage-based methodology in cognitive semantics. PhD diss., Katholieke Universiteit Leuven.

Goldberg, Adele E., 1995. Constructions: A Construction Grammar Approach to Argument Structure. University of Chicago Press, Chicago.

Goldberg, Adele E., 2005. Argument realization: the role of constructions, lexical semantics and discourse factors. In: Östman, J.-O., Fried, M. (Eds.), Construction Grammars: Cognitive Grounding and Theoretical Extensions. John Benjamins, Amsterdam, pp. 17–43.

Goldberg, Adele E., 2006. Constructions at Work. The Nature of Generalization in Language. Oxford University Press, Oxford.

Grondelaers, Stefan, Speelman, Dirk, Geeraerts, Dirk, 2002. Regressing on *er*. Statistical analysis of texts and language variation. In: Morin, A., Sébillot, P. (Eds.), 6th International Conference on the Statistical Analysis of Textual Data. Institut National de Recherche en Informatique et en Automatique, Rennes, pp. 335–346.

Grondelaers, Stefan, Geeraerts, D., Speelman, D., 2007. A case for cognitive corpus linguistics. In: Gonzalez-Marquez, M., Mittelberg, I., Coulson, S., Spivey, M.J. (Eds.), Methods in Cognitive Linguistics. John Benjamins, Amsterdam/Philadelphia, pp. 49–169.

Heylen, Kris., 2005. A quantitative study of german word order variation. In: Reis, M., Kepser, S. (Eds.), Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives. Mouton de Gruyter, Berlin/New York, pp. 241–263.

Hosmer, David W., Lemeshow, Stanley, 2000. Applied Logistic Regression. Wiley, New York.

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

Kristiansen, Gitte, Dirven, René, 2008. Introduction: cognitive sociolinguistics: rationale, methods and scope. In: Kristiansen, G, Dirven, R (Eds.), Language Variation, Cultural Models, Social Systems. Mouton de Gruyter, Berlin, pp. 1–17.

Langacker, Ronald W., 1987. Foundations of Cognitive Grammar. Theoretical Prerequisites, vol. 1. Stanford University Press, Stanford.

Langacker, Ronald W., 2005. Construction grammars: cognitive, radical and less so. In: Ruiz de Mendoza Ibáñez, F.J., Peña Cervel, M.S. (Eds.), Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction. Mouton de Gruyter, Berlin, pp. 101–159.

Levshina, Natalia, 2011. Doe wat je niet laten kan (Do what you cannot let): a usage-based analysis of Dutch causative constructions. PhD diss., University of Leuven.

Levshina, Natalia, Geeraerts, Dirk, Speelman, Dirk, 2011. Changing the world vs. changing the mind: dictinctive collexeme analysis of the causative construction with doen in Belgian and Netherlandic Dutch. In: Gregersen, F., Parrott, J., Quist, P. (Eds.), Language Variation – European Perspectives III. Selected Papers from the 5th International Conference on Language Variation in Europe (ICLaVE5), Copenhagen, June 2009. John Benjamins, Amsterdam, pp. 111–123.

Loewenthal, Judith, 2003. Meaning and use of causeeless causative constructions with *laten* in Dutch. In: Verhagen, A., van de Weijer, J. (Eds.), Usage-Based Approaches to Dutch. LOT, Utrecht, pp. 97–130.

Ordelman, Roeland, de Jong, Franciska, van Hessen, Arjan, Hondorp, Henri, 2007. TwNC: a multifaceted Dutch news corpus. ELRA Newsletter 12 (3–4), Available at:. http://doc.utwente.nl/68090/, (last accessed 07.08.10).

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/.

Rosenbach, Anette, 2002. Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies. Mouton de Gruyter, Berlin.

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

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

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

Szmrecsanyi, Benedikt, 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In: Geeraerts, D., Kristiansen, G., Peirsman, Y. (Eds.), Advances in Cognitive Sociolinguistics. Mouton de Gruyter, Berlin/New York, pp. 141–166.

Traugott, Elizabeth C., 2011. Pragmatics and language change. In: Allan, K., Jasczcolt, K. (Eds.), The Cambridge Handbook of Pragmatics. Cambridge University Press, Cambridge, pp. 549–565.

Tummers, José, Speelman, Dirk, Geeraerts, Dirk, 2005. Inflectional variation in Belgian and Netherlandic Dutch: a usage-based account of the adjectival inflection. In: Delbecque, N., van der Auwera, J., Geeraerts, D. (Eds.), Perspectives on Variation. Sociolinguistic, Historical, Comparative. Mouton de Gruyter, Berlin/New York, pp. 93–110.

van der Horst, Joop M., 1998. Doen in old and early middle Dutch: a comparative approach. In: Tieken-Boon van Ostade, I., van der Wal, M., van Leuvensteijn, A. (Eds.), 'Do' in English, Dutch and German. History and Present-day Variation. Nodus Publicationen, Münster, pp. 53–64.

Verhagen, Arie, 2000. Interpreting usage: construing the history of Dutch causal verbs. In: Barlow, M., Kemmer, S. (Eds.), Usage Based Models of Language. CSLI Publications, Stanford, pp. 261–286.

Verhagen, Arie, 2009. The conception of constructions as complex signs. Emergence of structure and reduction to usage. Constructions and Frames 1, 119–152.

Verhagen, Arie, Kemmer, Suzanne, 1997. Interaction and causation: causative constructions in modern standard Dutch. Journal of Pragmatics 27, 61–82.