Abstract

The paper examines three dimensions of linguistic variation: formal (what linguistic units are used), social (in what circumstances they are used) and conceptual (what those units mean). The integration of these in a coherent theoretical framework can be seen as a major ambition of the Cognitive Sociolinguistics enterprise. The authors argue that a grammar should take into account all three dimensions by systematically studying their interactions. The paper illustrates this theoretical approach through a usage-based study of Dutch causative constructions with *doen* and *laten*, in which the authors explore the co-influence and interaction of regional (Belgian vs. Netherlandic variety of Dutch) and various semantic factors in predicting the Dutch speaker’s preference for *doen* or *laten* in a given context. The paper examines whether there is difference in the effect of the conceptual factors on the choice of *doen* or *laten* depending on the region and vice versa, thus combining two perspectives inherent in Cognitive Sociolinguistics: conceptual and sociolinguistic. The first analyses suggest that the regional factor indeed moderates the predictive power of the conceptual dimension on different levels. However, a closer look at the data reveals that this moderation is to a large extent due to lexical idiosyncrasies of the regional samples. This also implies, from the variationist perspective, that the most outspoken divergence between the regions with regard to *doen* and *laten* is to be found at the level of lexical collocations.

1. Three dimensions of language variation.

Cognitive Linguistics has always been concerned with variation in meaning (polysemy, prototypicality and other phenomena), as well as different ways of expressing similar concepts (e.g. synonymy). Sociolinguistics, on the other hand, has been studying how linguistic variables are distributed along social and geographic dimensions. Cognitive Sociolinguistics is a project that integrates the two approaches to the mutual benefit of the parties (Kristiansen & Dirven 2008), aiming at a more nuanced, and also more realistic theoretical model of variation. An ideal contemporary grammar then should take into account three dimensions of variation: formal (what linguistic units are used?)¹, conceptual (what is said?) and social (by whom, for whom, when, where, why, etc. is it said?) (cf. Geeraerts 2010). Every dimension allows for a wide range of different phenomena that can be covered by these notions. The formal dimension may refer to any linguistic alternatives that coexist synchronically or diachronically. The conceptual dimension includes all kinds of meaning, from truth-conditional ones to communicative intentions and social, or stylistic, meanings (e.g. Eckert 2008). Finally, the social dimension potentially covers any aspect of communicative acts, from idiosyncratic characteristics of individual speakers and situations to cross-country and cross-cultural differences.
Approaches (examples)

<table>
<thead>
<tr>
<th>Approaches (examples)</th>
<th>Formal dimension</th>
<th>Conceptual dimension</th>
<th>Social dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>A naïve idealized view of language</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Traditional studies of polysemy</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Linguistic demography and language geography</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Semasiological lectally enriched grammar/lexicology</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Generative Grammar (Standard Theory)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Traditional studies of synonymy</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>First-wave sociolinguistics</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Onomasiological lectally enriched grammar/lexicology</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1. Three dimensions of language variation exemplified by different approaches and views.

Table 1 presents examples of approaches that consider different dimensions of linguistic variation, or their combinations. Although all professional linguistic approaches mentioned here may be useful, nevertheless it is the onomasiological lectally enriched grammar and lexicology that model the speaker's actual behaviour best by taking into account the situation, the variety of possible meanings and the range of linguistic alternatives available. This approach has been implemented in quite a number of cognitively oriented studies, such as Grondelaers et al. 2002, Heylen 2005, Glynn 2007, Speelman and Geeraerts 2009. We should admit that there have been attempts to incorporate all three dimensions in one model from the variationist side, too, especially actively in the studies of social meanings as indexical fields (see Eckert 2008). However, only first attempts are being made to systematically study conceptual variation within a range of possible formal alternatives. In addition, there remains a major methodological problem of identification and interpretation of fluid stylistic meanings.

A study of the three dimensions of variation can be carried out in two complementary ways. The first method aims at disentangling the dimensions by studying the relationships within a pair of dimensions (formal alternatives and conceptual factors, formal alternatives and social factors, and social factors and conceptual factors). When doing so, it is crucial to control for the third remaining dimension, which is especially difficult in the case of the conceptual one; cf. the debate between Lavandera and Labov in 1978 on the meaning of syntactic variables (see Romaine 1984 for an overview), although
quite a number of strategies have been successfully implemented to overcome this obstacle (see Geeraerts 2010). For example, in the study of regional variation of clothing terms in Belgium and the Netherlands (Geeraerts et al. 1994) the researchers controlled for the conceptual factors by providing maximal referential synonymy. More recently, multivariate statistical techniques such as multiple regression have been employed, which enable the researcher to control for conceptual factors while studying the influence of social variables, and vice versa (e.g. Speelman and Geeraerts 2009).

However, it is not enough to determine, for example, how the choice of a specific linguistic unit is influenced by social factors and by conceptual factors separately. The next step is a systematic study of interactions between the latter two. For instance, one can ask a question whether the role that certain conceptual factors play in making the speaker prefer one linguistic unit to another depends on the social or geographical factors (e.g. Grondelaers et al. 2002), or if certain conceptual features boost the impact of social and geographic factors on formal variation (e.g. Geeraerts & Speelman, In press).

In this paper, we attempt to place Dutch analytic causatives with *doen* and *laten*, two near-synonymous constructions, in the framework that has been outlined. In a study based on a newspaper corpus of Netherlandic and Belgian Dutch we address the following questions:

- a. if there is semantic difference between the constructions both in Belgium and the Netherlands;
- b. if there are any general regional preferences in the use of the constructions;
- c. whether there are different semantic constraints on the use of the constructions in the two regions;
- d. whether there are any conceptual features that trigger divergence between the varieties with regard to the preference for *doen* and *laten*.

Questions a and b represent the first step of our study, exploration of two-dimensional relations, while questions c and d refer to interaction of the dimensions. At the same time, questions a and c relate to the conceptual perspective, traditionally adopted by Cognitive Linguistics, with the meaning as the primary focus of attention, whereas the variationist perspective is more prominent in questions b and d.

The rest of the article is organized as follows. Section 2 briefly introduces the object of our case study, causative constructions with *doen* and *laten*. In Section 3, the data and methods are described. Section 4 shows how the regional and conceptual dimensions affect the formal dimension separately (see questions a and b). In Section 5, we examine the interplay of these factors from the conceptual and variationist perspectives (cf. questions c and d, respectively). We conclude with a summary of our findings and suggestions for future research.

2. Dutch causative constructions.

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

(1) De tovenaar deed/liet zijn dienaars een kasteel bouwen.
The magician made (let) his servants a castle build
The difference between *doen* and *laten* has been examined in a number of corpus-based studies (Kemmer and Verhagen 1994, Verhagen and Kemmer 1997, Stukker 2005). They argue that *doen* is a direct causation verb, which is used in situations when the causer is the main source of energy required for the caused event, and *laten* refers to indirect causation (which includes permission), when another entity (usually the causee) is the primary source of energy. In (1), *liet* (the past form of *laten*) means most likely that the magician gave an order to his servants and they proceeded on their own (there is also a possibility of a permissive reading), whereas *deed* (the past form of *doen*) would activate a situation when the magician used his magical powers to enforce his will upon his servants (who might be unwilling to work) and thus would be more directly involved in the process of causation.

Needless to say, operationalization of the highly abstract distinction between directness/indirectness, which manifests itself in such subtle context-specific nuances as in (1), is a challenge for an empirical researcher. In addition, in comparison with the studies of concrete lexemes (e.g. clothing terms in Geeraerts et al. 1994), studying the referents of syntactic constructions is more difficult because we deal with situations – complex, dynamic and abstract entities. A possible solution could be an analysis based on a number of lexical, syntactic and semantic features of the linguistic context that could serve as indirect indicators, or ‘circumstantial evidence’, of directness or indirectness of causation. A few of them were suggested in the previous studies, such as transitivity of the effected predicate. However, interpretation of such indicators in terms of directness/indirectness is sometimes problematic. For example, Speelman and Geeraerts (2009) argue that constructions with the coreferential causer and the causee should refer to the most direct type of causation possible; yet, contrary to the expectations, those contexts favour *laten*. Therefore, we should try to select as many contextual features as possible (the complete list can be found in the next section) and only then try to suggest an interpretation, directness/indirectness being only one of potential ways of explaining our findings. The results of such a bottom-up analysis are presented in Section 4.

3. Data and design on the study.

The study is based on a large sample of real language data (see argumentation in favour of the usage-based approach in Geeraerts 2005). From TwNC and LeNC, two corpora of Dutch and Flemish newspapers, we built up a regionally and thematically balanced subcorpus of 8 million words in total, 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 occurrences of the constructions with *doen* and *laten* and their contexts automatically. After some spurious hits had been deleted manually, we had a total of 6782 observations. The formal variation of the two constructions was coded as the *Auxiliary* variable.

The social dimension was represented in this study by the *Country* variable (Belgium or the Netherlands). The conceptual dimension was split into two
subdimensions: the macro-conceptual subdimension, which was based on the subject domain of the article where an observation was found (the SubjectDomain variable), and various micro-conceptual factors, which represent an attempt to describe the causative situation via indirect indicators. Below is the list of the semantic and syntactic features (micro-conceptual variables) that were considered relevant for the causative event:

- **CrSem**: semantic class of the causer (human or non-human)
- **CeSem**: semantic class of the causee (human or non-human)
- **CeSyntPresence**: explicitness or implicitness of the causee
- **CeSyntCase**: oblique or non-oblique case marking of the causee
- **EPTransitivity**: transitive or intransitive effected predicate
- **EPAspect**: whether the effected predicate describes a dynamic process, punctual event or state
- **CausedEventSem**: whether the caused event is mental or non-mental
- **Coreferentiality**: if the causer is coreferential with the other participants or not
- **Possession**: whether or not there are relationships of possession between the causer and the other participants
- **AffirmOrNeg**: whether or not the clause with the CC contains a negation

There were two more formal variables included, which checked for the degree of idiomatic lexical attraction between the effected predicate and *doen* or *laten*. The reason for that was obvious idiomaticity involved in some frequent collocations, for example, *laten zien* ‘show’ or *doen denken* ‘remind’. Apparently, the speaker does not have much choice between *doen* or *laten* when conveying these meanings, and the influence of the other factors should be minimal. We operationalized this factor in two variables: **LexAttraction**, which tells whether the effected predicate is significantly attracted to the given auxiliary, and **CausAttraction**, which indicates whether the effected predicate is significantly attracted to the causative construction CAUS + INF in general. For every effected predicate, both of the values were defined with the help of Fisher's exact test, which estimated whether the observed frequency of the effected predicate with *doen* and/or *laten* was significantly higher than expected (see Speelman and Geeraerts 2009 for a detailed description of the method of measurement, see also Stefanowitisch and Gries 2003 and later works on collostructional analysis for a justification of the approach).

Table 2 summarizes the operationalization of the three dimensions as variables in our study.
Table 2. Operationalization of dimensions as variables.

After coding, we carried out a series of logistic regression analyses. Logistic regression is a technique that allows the researcher to explore the role of one or more factors (predictors) in predicting a binary outcome treated as the response variable, e.g. whether the journalist will use *doen* or *laten* in a given context. The resulting model is represented as a mathematical equation with numeric estimates for each factor. Special techniques are also available to perform evaluation and diagnostics of possible models and select the one that provides the best fit to the data. The results of our regression analyses and their interpretation follow in the next sections.

### 4. Disentangling the dimensions.

Table 3 shows the results of the main-effect logistic regression analysis in a very simplified form. All predictors were tested, but three of them, those related to the causee (*CeSem, CeSyntCase* and *CeSyntPresence*), do not seem to play a role when the other variables are taken into account, and therefore they are not retained in the final model. The estimates, expressed in logits, reveal whether the specified value (e.g. *CrSem* = NonHuman) boosts *doen* (a positive estimate) or *laten* (a negative estimate) in comparison with the reference level (e.g. *CrSem* = Human). The p-values show if there is a significant relation between the predictor and the response variable. The global model statistics allowed us to conclude that the model has some real predictive capacity.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate (in logits)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>-8.282</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>CrSem</em> = NonHuman (vs. Human)</td>
<td>3.335</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>EPTransitivity</em> = Intransitive (vs. Transitive)</td>
<td>1.904</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>Coreferentiality</em> = No (vs. Yes)</td>
<td>2.497</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>Country</em> = Belgium (vs. Netherlands)</td>
<td>0.733</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>CausedEventSem</em> = Mental (vs. Non-mental)</td>
<td>0.874</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>EPAspect</em> = State (vs. Process)</td>
<td>-1.202</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>EPAspect</em> = PunctEvent (vs. Process)</td>
<td>-1.127</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>Possession</em> = No (vs. Yes)</td>
<td>1.172</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>SubjectDomain</em> = Music (vs. Economy)</td>
<td>0.708</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>SubjectDomain</em> = Football (vs. Economy)</td>
<td>0.635</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><em>SubjectDomain</em> = Politics (vs. Economy)</td>
<td>0.635</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Table 3. Results of the main-effect logistic regression.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LexAttraction</strong></td>
<td>Significant (vs. Insignificant)</td>
<td>1.794</td>
</tr>
<tr>
<td><strong>CausAttraction</strong></td>
<td>Significant (vs. Insignificant)</td>
<td>-1.551</td>
</tr>
<tr>
<td><strong>AffirmOrNeg</strong></td>
<td>Affirm (vs. Neg)</td>
<td>0.634</td>
</tr>
</tbody>
</table>

The order of the predictors in the model reflects the order in which they were added to the model in the forward selection procedure, and therefore correlates with the strength of the factors in predicting the speaker's choice. We see that the three micro-conceptual variables at the top of the list (*CrSem, EPTtransitivity, Coreferentiality*) are the most powerful. They are followed by the regional preferences for *doen* in Belgium. The controlling collocation-related variables (*LexAttraction* and *CausAttraction*) are significant, too, but their low position in the list suggests that there is still some place for conceptual factors in explaining the variation. Interestingly, the macro-conceptual subdimension (*SubjectDomain*) seems to be less important than most of the micro-conceptual factors.

Although the estimates are quite revealing on their own, there is an important nuance. We should ask ourselves, whether, for example, the estimate for *Country* means (a) *doen* is preferred in Belgium, (b) *doen* is disfavoured in the Netherlands, or both (a) and (b)? From the additional bivariate analyses that we performed it seems that *laten* is the default verb for any value of almost every predictor; i.e. only (b) is true because both in Belgium and in the Netherlands *laten* is more popular than *doen*. The only exception is the semantic class of the causer (which is by the way the strongest predictor in the model): it seems that human causers prefer *laten* and disfavour *doen*, and for non-human causers the tendency is reversed.

From our results and also following the previous study by Speelman and Geeraerts (2009), we suggest that *laten* is the default causative auxiliary while *doen* is a verb with a specialized meaning, restricted to short and simple causation schemes initiated by non-human entities and involving processual effected predicates and mental caused events. The semantic restrictions tie up with the collocational restrictions of the effected predicates of *doen*. The *laten* auxiliary covers a variety of mental, interpersonal and physical caused events and states, causation chains of different length; *laten* is also the auxiliary for situations of self-causation and non-occurrence of impingement. The causation is usually initiated by human causers, who may exert control over the other participants (in the case of ownership). To conclude, it is necessary to note that all these differences can hardly be explained in terms of directness/indirectness of causation, at least, not in the way the distinction was formulated in the previous research. Especially interesting is the fact that the causee, whose properties should be crucial for distinguishing between direct and indirect causation, does not play any role in our analysis.

This model helped us to disentangle the regional and conceptual causes of the formal variation and answer the questions that were asked in Section 1: (a) there is a conceptual difference between *doen* and *laten*, which we have just briefly outlined, and (b) we observe overall preference for *doen* in the Belgian sample. The next step is to investigate if (c) different conceptual factors play different roles in the two regions, and (d) whether the regions diverge in the use of *doen* and *laten* from each other in some
circumstances more than in the others. To answer these questions, we have to look at the interactions of the dimensions.

5. Studying interactions of the conceptual and social dimensions.

5.1. The method.

Grondelaers et al. 2002 proved the existence of different sets of constraints underlying the use of presentative constructions with and without er in Belgian and Netherlandic Dutch. Having compared two regression models for two regional samples, the researchers found out that the conceptual factors were overall stronger in the Netherlands, where er is less frequent. Another, more direct, way to answer this question is to study the interactions of the conceptual factors with the regional ones in a global two-regional model. Interaction in the statistical sense means that there is a non-additive effect between two explanatory variables. We fit a model with all possible two-way interaction terms selected stepwise with the help of Akaike's Information Criterion (cf. Note 5). In this paper, we focus only on those involving the Country variable. There are only two conceptual variables that interact significantly \((p < 0.05)\) with Country: EPTransitivity and SubjectDomain. There are a number of marginally significant interactions, which are nevertheless retained in the model to improve its power: Country*LexAttraction, Country*CrSem and Country*AffirmOrNeg. We will zoom in only on the significant ones.

5.2. Interaction Country*EPTransitivity

The interaction is illustrated by a three-dimensional interaction plot in Fig. 1. The vertical axis represents the formal dimension; its direction corresponds to increase in the probability of doen. The horizontal axes refer to the regional (Country) and the conceptual (EPTransitivity) dimensions. The plot is slightly artificial because it shows the categorial variables as continuous ones, but it still serves the purpose of illustration.
The plots shows the same tendency that we observed earlier. Belgian Dutch favours *doen* (the Belgian-side slope is higher than the Netherlandic-side one), both for intransitives and transitives. The semantic factor also behaves as predicted in the main-effect model: the transitives favour *laten* more than the intransitives do. The interaction is ordinal (noncrossover) and quite moderate.

Before interpreting the interaction itself, it is necessary to note that this three-dimensional plot allows the viewer to literally distinguish between the two perspectives that Cognitive Sociolinguistic is meant to integrate: variationist and conceptual. If one's original focus is regional variation and one is interested how it behaves across different conceptual levels, one can face the regional axis in the plot and look at the slopes of the inner surface that correspond to the conceptual levels. Then he or she can compare the steepness of the slopes and establish which conceptual level triggers more regional variation than the other, leading to more semantic convergence or divergence between the varieties. For example, in Fig. 1, the steepness of the slope is more outspoken in the case of transitive effected predicates, therefore, transitive causatives are more responsible for divergence between the lects with regard to the preference of *doen* or *laten* than intransitives. We also suggest that the choice between *doen* and *laten* can be potentially used as a more salient sociolinguistic marker of Dutchness of Flemishness in transitive CCs than in intransitive ones. If this can actually be the case depends on whether the dialects belong to one language system or constitute two separate systems for the speakers that we select (cf. argumentation in Geeraerts 2005: 173), as well as on the question if and to what extent conceptually different syntactic phenomena may also be different pragmatic (stylistic) markers (cf. Goldberg 1995: 67). These issues, however, remain beyond the scope of this paper.
To study the interaction from the other perspective, that of the conceptual factors moderated by the region, the viewer should face the conceptual axis on the plot and watch the two regional slopes of the inner plane. The steepness of the slopes suggests that the importance of the conceptual factor in predicting the choice of *doen* or *laten* is higher in the Netherlands than in Belgium.

We should add perhaps that the two perspectives do not imply any causal relationship between the dimensions. They simply describe the same tendencies in the data in two different ways. It is the task of the researcher to interpret the observed phenomena and find the causes. We intend to do so after examining the second interaction.

5.3. Interaction *Country: SubjectDomain.*

As *SubjectDomain* is a factor with 4 levels, it would be more convenient to build traditional two-dimensional interaction plots (Fig. 2). The vertical axis in both plots refers to the probability of *doen* predicted by the model. The upper plot profiles the divergence between the regions across the four domains, and the lower one displays the divergence between the domains depending on the region.
Fig. 2. 2-D plots of interaction between Country and SubjectDomain.

The upper plot in Fig. 2 shows that the largest regionally diverging domain is Economy, with a strong avoidance of doen in the Netherlands and a high acceptance of this auxiliary in Belgium; the most converging one seems to be Music. From the macro-conceptual perspective shown in the lower plot, one can see that the distance between the different domains is slightly larger in Belgium, especially between Economy and Football, although Music and Politics slightly converge in comparison with the Netherlands. One can also conclude that the line of the economy-related articles, which crosses the other more or less co-directed domains, must be the main cause of this interaction being significant.

Evidently, thematic difference should be tied up with the use of different lexemes. We zoomed in on the effected predicates in the subcorpora from the four domains and the two regions.

<table>
<thead>
<tr>
<th>Belgium</th>
<th>The Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>liggen</em> 'lie' 9.17%</td>
<td><em>zien</em> 'see' 21.58%</td>
</tr>
<tr>
<td><em>stijgen</em> 'go up' 4.38%</td>
<td><em>weten</em> 'know' 16.54%</td>
</tr>
<tr>
<td><em>vallen</em> 'fall' 4.17%</td>
<td><em>vallen</em> 'fall' 2.84%</td>
</tr>
<tr>
<td><em>vermoeden</em> 'suppose' 3.33%</td>
<td><em>wachten</em> 'wait' 2.2%</td>
</tr>
<tr>
<td><em>verstaan</em> 'understand' 3.13%</td>
<td><em>liggen</em> 'lie' 1.98%</td>
</tr>
<tr>
<td><em>wachten</em> 'wait' 2.5%</td>
<td><em>horen</em> 'hear' 1.88%</td>
</tr>
<tr>
<td><em>optekenen</em> 'record' 2.5%</td>
<td><em>stijgen</em> 'go up' 1.68%</td>
</tr>
</tbody>
</table>
Table 4 shows the relative frequencies (in %) of ten verbs in the economy-related texts. From this table, one can observe that the verbs zien 'see' and weten 'know' account together for 38.12% of all observations in the Dutch sample (as opposed to only 1.88% in its Belgian counterpart). This tendency is also found in the other domains, although to a lesser extent, with the smallest frequencies of these two verbs in the music-related texts. It is an important fact that the differences between the percentages of these two verbs in the two varieties across the domains correlate with the distances between the regions across the domains that we observed in the upper plot in Fig. 2. Table 5 demonstrates this correlation.

<table>
<thead>
<tr>
<th>Rank of divergence according to Fig. 2</th>
<th>Difference in relative frequencies (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.36</td>
</tr>
<tr>
<td>3</td>
<td>20.31</td>
</tr>
<tr>
<td>4</td>
<td>8.79</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 5. Difference between the regions in the frequencies of zien and weten and their divergence with regard to preference for doen or laten across the domains.

The specificity of these verbs is that they form highly idiomatic collocations with laten (laten zien 'show' and laten weten 'let know, inform'), which is the almost exclusive auxiliary for these verbs, with only 3 exceptions from the total of 935 in both regions. Interestingly, the Dutch economy-related articles use the construction laten zien predominantly in the metaphorical sense as in (2) with abstract causers, and not in the more common sense of interpersonal interaction.

(2) Volgend jaar zal de economie een groei laten zien van 1,25 procent.
“Next year the economy will report a growth of 1.25 per cent”

From this we can assume that the interaction between Country and SubjectDomain is significant due to the lexical specificity of the affected predicates in the two varieties of Dutch. Knowing that both zien and weten are transitive verbs, can we assume that the first interaction, Country*EPTransitivity, is caused by the same factor? The following section puts these hypotheses to test in a small experiment.

5.4. The experiment.
To test our hypotheses, we carried out a series of regression analyses, deleting from the data set the observations with the top ten most frequent affected predicates verb by verb, and refitting the regression models with main effects and interactions. By doing so, we wanted to check stability of our predictors and their interactions. The verbs were zien ‘see’, weten ‘know’, horen ‘hear’, liggen ‘lie’, denken ‘think’, vallen ‘fall’, gaan ‘go’, vermoeden ‘suppose’, wachten ‘wait’, spelen ‘play’. They accounted for 34% of all observations in our data set.

The experiment showed that the main-effect model that we discussed in Section 4 was quite stable. The exceptions were SubjectDomain, which ceased to play any role in the model, and EPTransitivity, which retained its sign, but in five iterations out of ten had only marginal significance. In addition, some causee-related predictors occasionally showed up and disappeared again, except for CeSyntCase, which steadily gained significance throughout the entire experiment, boosting laten for the causees marked with door and aan (cf. Kemmer and Verhagen 1994). All this does not undermine our interpretation of the conceptual difference between the two constructions, adding only that laten, as the default auxiliary, also allows for different ways of marking the causee. The order of the predictors slightly changed, too, with the Country variable going down from the fourth rank in the initial model to the ninth place in the last regression.

As expected, the interaction with SubjectDomain disappeared quite quickly, after excluding zien and weten from the data set. The Country*EPTransitivity interaction was more persistent, lurking in the analyses now and then, but after six iterations it disappeared for good, too. These facts corroborate our predictions made in the previous section. However, nearly at the onset of the experiment there appeared two new significant interactions (marginally significant in the first model) that exhibited stable performance. Those were Country*LexAttraction (Fig. 3) and Country*CrSem (Fig. 4).
Fig. 3. A 3-D plot of interaction between Country and LexAttraction.

Fig. 4. A 3-D plot of interaction between Country and CrSem.

These plots are based on the most ‘stripped-off’ data set without ten most frequent effected predicates (with only 66% of the observations left). The upper one shows that strong collocation ties boost doen only slightly less in the Netherlands than in Belgium, but there is more substantial regional divergence for ‘free’ effected predicates. In the Netherlands doen is restricted to idiomatic uses (cf. Levshina et al. (forthcoming)), and ‘free’ verbs tend to be accompanied by laten. In Belgium doen is used in stable lexical collocations, too, but it is more often tolerated with ‘free’ verbs.

The second plot is very similar to the first one, although the interaction is even milder. The semantic class of the causer seems to be slightly more influential for the Netherlandic sample, and there is more flexibility in the Belgian subcorpus. If the causer is human, then Belgian Dutch allows for more variation.

The experiment demonstrated that the lexical effects are indeed very powerful in the regional samples. Sometimes these effects can overshadow subtle conceptual differences between the varieties. Therefore, the importance of lexical differences should not be underestimated in a study of grammatical variation.

6. Conclusions.

In the beginning of the paper, we asked four questions about the variation of Dutch causatives. These questions, we believe, reflect the approaches and perspectives that are relevant to any Cognitive Sociolinguistic research that predicts or explains linguistic choices on the basis of social and conceptual factors. We would like to conclude this paper
with the answers to the questions, even though we already discussed them in different parts of the text.

a. There is conceptual variation between *doen* and *laten*. Our newspaper corpus showed that *laten* is the default auxiliary, covering a wide range of causative situations, whereas *doen* is specialized in short and simple causation initiated by non-human causers and involving a mental caused event.

b. There is also regional variation in the use of *doen* and *laten*. While both Belgium and the Netherlands prefer *laten* in most situations, the use of *doen* is even more restricted in the Netherlands than in Belgium.

c. There is no evidence of outspoken difference in semantic constraints on the use of *doen* and *laten* in the two regions, although there seems to be slightly more freedom in the Belgian variety than in the Netherlands with regard to the most important conceptual factor, the semantic class of the causer. In all other cases, the moderating effect of the region and transitivity upon the micro- and macro-conceptual features turns out to be caused by lexical specificity of the regional samples.

d. Therefore, the main locus of regional divergence in the use of the two constructions is to be found at the idiomatic level.

Apart from the objectives formulated in the first section of the paper, our analyses also demonstrated how important it is to take into account lexical effects in a study of syntactic variation. From the conceptual perspective, our study also raises questions about the boundary and interrelation between the abstract conceptual properties, such as the general schematic force-dynamic meaning of a construction, and idiomatic restrictions connected with the lexical slot fillers. In case of a conflict, does the schematic meaning override the lexical effect, or is it the other way round? From the variationist perspective, it would be interesting to include other social factors, such as the register, and investigate their interaction with the conceptual and regional factors. This could be an important step in finding out whether *doen* and *laten* can be treated as indexical sociolinguistic markers or mere indicators of some geographic and social categories (Labov 1971) and, hopefully, would shed new light on the potential of syntactic variables as socially meaningful phenomena. This line of research could form a basis for integration of social meaning with all other kinds of meaning in one framework and therefore bridge the remaining gap between the cognitive and variationist approaches.

Notes

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: regional, conceptual or any other factors. This interpretation should not be confused with 'formal onomasiological variation' in studies of Geeraerts et al. 1994 and later works, where it is usually applied to a set of naming options for the same concept across different lects. The same applies to the terms 'conceptual dimension' in this paper and 'conceptual semasiologic/onomasiologic variation' (ibid.).

2 The thematic subcorpora were created automatically with the help of keywords provided for every article in the two corpora.
The aspectual characteristics of the effected predicates were assigned automatically, on the basis of specific lexical markers of dynamicity and durativity co-occurring with the verbs in a larger corpus. We used possessive pronouns and nouns in the possessive case as indicators of relationships of possession. Due to space limitations, we discuss here only the main-effect model. All possible two-way interactions have been studied, too; most of them turned out to be ordinal and moderate, which implies that the main-effect model can be regarded as valid.

According to the following model statistics: Gamma-index = 0.824, $C = 0.91$, Somer’s $D_{xy} = 0.82$, generalized $R^2 = 0.58$.

References


