Article

Chiara Paolini*, Hubert Cuyckens, Dirk Speelman and Benedikt Szmrecsanyi

The English dative alternation in vector space: how much does meaning matter?

https://doi.org/10.1515/cllt-2024-0095
Received September 5, 2024; accepted April 3, 2025; published online August 25, 2025

Abstract: Variationist linguists study alternations between different ways of saying the same thing (Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press: 188). Some linguists tend to believe that this sort of optionality is not predicted by theory and dysfunctional. We empirically explore if the two positions can be reconciled by assuming interchangeability between grammatical variants but admitting semantic information about the material in argument slots of competing constructions into variationist modeling. As a test case, we marshal corpus-based methods to study the well-known dative alternation after the verb *to give* in English. We specifically reanalyze the dataset investigated in Bresnan et al.'s seminal (2007) study. To assess the lexical semantics of the dative constituents, we rely on Semantic Vector Space modeling (Lenci, Alessandro. 2018. Distributional models of word meaning. *Annual Review of Linguistics* 4(1). 151–171). The meaning distinctions that we thus establish are subsequently submitted to variationist modeling. The distributional-semantic dative alternation model is then benchmarked against a traditional dative alternation model considering mostly higher-level, i.e., fairly abstract formal predictors (e.g., end-weight, definiteness) à la Bresnan (2007). The analysis shows that traditional variationist modeling à la Bresnan (2007) outperforms distributional-semantic modeling.

Keywords: isomorphism; synonymy; variationist linguistics; vector space modeling; distributional semantics; dative alternation

***Corresponding author: Chiara Paolini**, Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium, E-mail: chiara.paolini@kuleuven.be. https://orcid.org/0000-0001-7961-3624
Hubert Cuyckens, Dirk Speelman and Benedikt Szmrecsanyi, Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium

1 Introduction

In this paper, we take a fresh look at an in principle extremely well-studied case of grammatical variation, or *optionality* (the term we will use in the remainder of this contribution, as it unambiguously refers to intraspeaker variation) in language – the dative alternation after the verb *to give* in English. In English, language users have the choice between two functionally broadly equivalent ways to express dative relations involving a recipient and a theme: the ditransitive dative variant, as in (1), and the prepositional dative variant, as in (2).

(1) **The ditransitive dative variant**

(a) And so, you know, [we]$_{\text{subject}}$ ['ll give]$_{\text{verb}}$ [him]$_{\text{recipient}}$ [fifteen]$_{\text{theme}}$ which will teach him a lesson but it's not just, you know, horrible (DAT-2644)[1]

(b) But [they]$_{\text{subject}}$ [give]$_{\text{verb}}$ [the guy]$_{\text{recipient}}$ [a job]$_{\text{theme}}$ in prison and make him pay his damn debt. (DAT-2772)

(2) **The prepositional dative variant**

(a) if [I]$_{\text{subject}}$ [gave]$_{\text{verb}}$ [it]$_{\text{theme}}$ [to the government]$_{\text{recipient}}$ they would just waste it. (DAT-4100)

(b) [The judge]$_{\text{subject}}$ [will usually, uh, give]$_{\text{verb}}$ [custody]$_{\text{theme}}$ [to the mother]$_{\text{recipient}}$ ninety-seven percent of the time. (DAT-4067)

There is, needless to say, a huge literature on the dative alternation in English (see Szmrecsanyi and Grafmiller 2023: chap. 2 for a review). The considerable interest that the dative alternation has sparked demonstrates how grammatical variation between "alternate ways of saying 'the same' thing" (Labov 1972: 188) is intriguing and, frankly, a bit puzzling to many analysts. To summarize the discussion to follow: foundational principles in Functional and Cognitive Linguistics are often interpreted as not predicting the existence of grammatical optionality (see, e.g., Haiman 1980; Goldberg 1995). Instead, the default assumption is that syntactic differences should correlate with meaning differences. In variationist linguistics, "alternate ways of saying 'the same' thing" (Labov 1972: 188) – i.e., linguistic variables, or alternations (in the realm of grammar) – in practice cover the spectrum from truth-conditional equivalence to discourse-functional equivalence, given that semantic differences

1 Here and in the following, labels are unique identifiers in the dataset under study (Bresnan et al. 2017).

(to the extent that they exist) are considered to be often subject to neutralization in discourse (see Sankoff 1988: 153 for discussion). Of course, choice between different (semantically/functionally broadly equivalent) ways of saying the same thing thus defined (the dependent variable) may be probabilistically predictable to some extent from various contextual, pragmatic, or social factors, which is why these predictors are typically included as independent variables in variationist modeling. It is important to note that variationists tend to be primarily interested in the predictors (contextual, pragmatic, or social factors), rather than in variables/alternations *per se*.

In this study, we investigate the extent to which we can remain true to the empirical fact that dative constructions after *to give* in English are broadly interchangeable while acknowledging that meaning differences may play a role in dative choice after all, via semantic properties of the materials in the theme and recipient slots of dative constructions. We note at the outset that the dataset under analysis here only covers dative constructions after the verb *to give*, so the verb slot does not come under the remit of the present study. Therefore, the empirical question that we are asking is the following:

> How adequately can we predict dative choices as a function of the lexical semantics of dative *constituents* (i.e., recipient and theme), as opposed to the semantics of dative *constructions* per se?

So, in example (1b), what is the extent to which the theme *job* triggers the ditransitive dative variant? In (2b), what is the extent to which the recipient *mother* triggers the prepositional dative variant? And so on. To address this question, we will marshal Semantic Vector Space modeling and submit the output to state-of-the-art variationist analysis.

Our theoretical point of departure is that skepticism about grammatical variation and optionality is fairly widespread outside of variationist linguistics (and perhaps to some extent outside of generative linguistics, where transformations, or their equivalents in nontransformational approaches, are thought to be able to trigger optionality). In variationist circles, this skepticism has been called the "Doctrine of Form-Function Symmetry" (Poplack 2018: 7). Prescriptivist grammarians and language mavens, for example, are typically committed to eradicating variation in language (see, e.g., Yáñez-Bouza 2006 for discussion). For them, "[l]anguage standardization involves minimizing variation" (Curzan et al. 2023: 18). But variation and optionality are also controversial in more theoretically oriented linguistics. Take the famous Principle of Isomorphism, formulated by Haiman citing authors such as Bloomfield and Bolinger, among others: "[…] the commonly accepted axiom that no true synonyms exist, i.e. that different forms must have different meaning […]" (Haiman 1980: 516). If different forms must have different

meaning, then true variation and optionality should not exist. A different incarnation of this axiom is the "Principle of No Synonymy" in Construction Grammar: "If two constructions are syntactically distinct, they must be semantically or pragmatically distinct. […]" (Goldberg 1995: 67). No Synonymy is very much an article of faith for many Construction Grammarians and Cognitive Linguists, some recent debate notwithstanding (see Leclercq and Morin 2023 for discussion). For reasons of space, we must eschew a more detailed discussion of the above Principles. That said, we wish to acknowledge that the reasoning underpinning the Principles is likely more nuanced than portrayed here, and it is also likely that the Principles can be interpreted responsibly, such that at the end of the day, there are no major disagreements between variationist and functional/cognitive linguists. But then again, Haiman and Goldberg use strong and non-nuanced language in their original formulations ("must have different meaning", "must be […] distinct"), and so we feel that reading these Principles at face value is intellectually not dishonest. What is more, the notion that optionality is somehow doubtful is *exactly* the message (intended or unintended) that many people take away from reading Haiman and Goldberg. It is this popular (mis) conception that we are discussing here – so if we are deconstructing a straw-man, then it's one that many colleagues believe in.

We also wish to add that faith in form-function symmetry has also informed thinking in historical linguistics, where when we find variation a common default assumption is that in the long run, one of the variants must die, or competing variants must find their unique niches in functional space (see De Smet et al. 2018 for critical discussion). We conclude that for many theorists, the existence of different ways of saying the same thing is not predicted because variation and optionality without meaning distinctions is allegedly dysfunctional and "unmotivated" (De Smet 2019).[2]

And this brings us to the dative alternation in English. According to Gerwin (2014: 19), the pertinent literature can be divided into two camps: studies that adopt a single-meaning approach (predominant in variationist/probabilistic linguistics) versus studies that adopt a multiple-meaning approach (popular in the generative community, but also in Construction Grammar). Take the dative verb *give* (as in (1) and (2)). The multiple-meaning approach essentially assumes that different ways of conceptualizing the *giving* event trigger different constructional variants: change-of-state or change-of-possession conceptualizations supposedly favor the ditransitive

2 Of course, one may argue that the type of contextual/pragmatic/social probabilistic conditioning that variationists study as independent variables via, e.g., regression modeling does come under the remit of "meaning differences" or "non-equivalence" (see Leclercq and Morin 2023). In this case, there is probably no disagreement. But the price is that the various Principles then simply boil down to stating that absolutely free and unconstrained optionality (i.e., synonymy under all circumstances) does not exist.

dative variant, while change-of-place or movement-to-a-goal conceptualizations are said to favor the prepositional dative variant (e.g., Green 1974; Oehrle 1976; see also Langacker 1990: 13–15). There are serious empirical problems with these largely intuited meaning distinctions and the multiple-meaning approach, however, as Bresnan et al. demonstrate in their seminal (2007) study. Consider *give*-idioms: *give somebody the creeps* does not involve movement toward a goal, so the ditransitive pattern should be mandatory. Yet Bresnan et al. (2007) show that the prepositional dative pattern (*give the creeps to somebody*) is actually attested.[3] The conclusion, now widely accepted in variationist circles and beyond, is that "English dative verbs have more syntactic flexibility than we thought, occurring more freely in alternative constructions," and that "we cannot predict the dative alternation from meaning alone" (Bresnan et al. 2007: 11). And this, in turn, leaves us with the position that dative variants fundamentally have the same meaning and/or the same function. That is the assumption that has informed variationist-probabilistic research on the dative alternation ever since the publication of Bresnan et al. (2007). We note in passing that further evidence that dative variants are different ways of saying the same thing comes from picture naming tasks as in Bock (1986): participants are shown the exact same picture, but some participants describe the picture using the ditransitive dative variant while others describe the picture using the prepositional dative variant. This variance seems to suggest that there is optionality indeed.

What really seems to fuel the dative alternation, according to recent variationist-probabilistic research, is higher-level formal predictors (a.k.a. constraints), such as length or weight of the constituents, their definiteness, or their pronominality. Furthermore, analysts have taken into account information status-related predictors, as well as – in a limited way – selected top-down semantic predictors relating to the materials in the constructional slots (i.e., constituent animacy, or semantics of the dative verb). In addition, state-of-the-art probabilistic modeling also takes lexical effects on board: what is the extent to which particular theme, recipient, and dative verb lemmas trigger particular dative variants? This may all sound a bit simplistic, but in fact recent variationist-probabilistic modeling of the dative alternation is spectacularly successful. For example, Szmrecsanyi et al. (2017) model the dative alternation in four regional varieties of English based on nine fixed-effect predictors (recipient/theme type, recipient/theme definiteness, recipient/theme animacy,

3 One could argue, of course, that *give the creeps to somebody* does not actually constitute an empirical problem to the multiple-meaning approach: *give the creeps to somebody* can be seen as an infrequent and/or marked variant of *give somebody the creeps*. In that sense, one could still maintain that a change-of-possession conceptualization *favors* the ditransitive dative variant. That is to say, the term "favors" does not imply that the occasional counterexample does not exist. That said, the cognitive/functional isomorphism-adagio is mostly construed as categorical, and in that sense, Bresnan's counterexamples do pose an empirical problem.

recipient/theme length, and semantics of the dative verb give; see Section 2.1 for more discussion) plus theme and recipient lemmas as random effects. Their models achieve Index of Concordance (*C*) scores between 0.97 and 0.99. Note that *C* scores can range between 0.5 and 1; values >0.9 are customarily interpreted as indicating "outstanding" (Levshina 2015: 259) goodness of fit. Of note, models such as those in Szmrecsanyi et al. (2017) are built on the assumption that dative constructions are functionally and semantically equivalent. Nonetheless, they account for the observable variation astonishingly well.

It follows that meaning is indeed irrelevant in the dative alternation. Or is it? Not quite. Models such as those reported in Szmrecsanyi et al. (2017) do include top-down (if rather coarse-grained) semantic predictors, such as constituent animacy or the meaning of the dative verb ("transfer" vs. "communication" vs. "abstract"). Additionally, these models include lexical lemma effects in their random effects structure, which treat lexico-semantic triggers as mere "nuisance factors," but which may actually hide away semantic generalizations to be potentially had. On top of this, traditional higher-order formal predictors such as constituent length may mask semantic information. For example, we know that Germanic words are shorter and tend to refer to more basic, Swadesh list-type concepts (consider *job* in (1b)), while Latin and French-derived words tend to be longer and refer to less basic concepts (consider *custody* in (2b)). The upshot is that while state-of-the-art modeling of the dative alternations assumes semantic/functional equivalence of the dative constructions themselves, they do rely, implicitly or explicitly, on some lexico-semantic information about the materials in the constructional slots. But the extent to which we would be able to account for grammatical variation if we restricted attention to properly operationalized lexico-semantic predictors remains largely unknown.

Our objective in this paper, therefore, is to investigate if variationist dative alternation modeling can be reconciled with form-function symmetry à la Haiman (1980: 516) and Goldberg (1995: 67) by focusing on meaning differences not of the constructions per se but of the lexical materials in the theme and recipient slots. To responsibly model the meaning of these materials in a replicable, nonintuition-based way, we will marshal a robust corpus-based technique from distributional semantics (Lenci 2018) called Semantic Vector Space modeling. We subsequently use the resulting meaning distinctions as semantic predictors in variationist modeling. We will refer to these bottom-up, cognitively solid (Lenci 2018: 152) predictors as Vector Space Modeling (VSM)-semantic predictors. This is our corpus-based methodology in short. Of course, the idea that materials in constructional slots can attract certain grammatical variants is not new – it is what fuels distinctive collexeme analysis (Gries and Stefanowitsch 2004) and related techniques. These techniques, however, cannot adequately distinguish between purely lexical effects (i.e., particular lexemes attracting certain variants, for whatever reason) and generalizable semantic effects

(i.e., particular *meanings* attracting certain variants), which is precisely why we bring in Semantic Vector Space modeling. Semantic Vector Space modeling, on the other hand, has sure enough been used before in research on grammatical constructions (for example, Perek and Hilpert 2017), albeit to the best of our knowledge not as a technique to systematically generate predictors for alternation research (but see Levshina and Heylen 2014 for earlier Construction Grammar work roughly along the same lines). We thus endeavor to break new ground by using Semantic Vector Space modeling as an auxiliary technique for variationist linguistics, for the sake of understanding better how grammatical variation works.

Against this backdrop, our case study reanalyzes the US-American section of the publicly available *give*-dative alternation dataset available at https://purl.stanford.edu/qj187zs3852 (Bresnan et al. 2017). This is one of the corpus-derived datasets analyzed in Szmrecsanyi et al. (2017), and it is largely identical to the dataset investigated in Bresnan et al.'s seminal (2007) study.

Key findings include the following. The vector space model (VSM) with the optimal parameter configuration generates 15 VSM-semantic predictors for the recipient slot in dative constructions and 15 VSM-semantic predictors for the theme slot. The clusters are surprisingly coherent. Feeding these predictors into variationist modeling via mixed-effects regression reveals that the resulting model has an outstanding fit. However, in the traditional benchmark model, the traditional fixed-effect predictors do considerably more work than the semantic fixed-effects predictors do in the VSM-semantic model (which relies more on lexical effects via the random-effects structure). This is another way of saying that traditional predictors outperform VSM-semantic predictors.

This paper is structured as follows. Section 2 details our methodology. Section 3 reports the results. Section 4 offers a discussion and some concluding remarks.

2 Methods and data

Our methodology can be summarized as follows. We reanalyze a publicly available and richly annotated dative alternation dataset, which is largely identical to the dataset investigated in Bresnan et al.'s seminal (2007) study, and which is based on the Switchboard corpus of American English (Godfrey et al. 1992). To this dataset, which covers dative constructions after the verb *to give*, we add a layer of VSM-semantic predictors that we generated utilizing the Corpus of Contemporary American English (COCA) (Davies 2019) as a training corpus. We are specifically interested in the semantics of the recipient and theme slots in dative constructions. Subsequently, we engage in a round of mixed-effects binomial logistic regression modeling of the dative alternation, a well-established statistical technique in

variationist linguistics to account for speakers' choices between variants. Our goal is to benchmark the explanatory power of distributional semantic information against traditional state-of-the-art dative alternation modeling.

What follows is a description of the methodological steps necessary to conduct the analysis.

2.1 Step 1: Prepare a dataset about the alternation under study, where observations are annotated for various contextual characteristics and what we will call "traditional" predictors, such as, e.g., constituent length

This paper reanalyzes the US-American section of the dative alternation dataset compiled by Bresnan and colleagues (2017). Dative observations were extracted from the Switchboard corpus of American English (Godfrey et al. 1992). The Switchboard corpus covers telephone conversations collected between 1990 and 1991. The variable context is defined as follows:

> Attention was restricted to the dative verb *give*. The definition of interchangeable ditransitive and prepositional dative variants broadly follows Bresnan et al. (2007), which essentially means that all instances of give with two argument NPs minus non-interchangeable constructions were considered (Szmrecsanyi et al. 2017: 6).

The dataset has annotation for the following language-internal traditional predictors:

- *Recipient/Theme.type:* pronominal (personal pronouns, demonstrative pronouns, impersonal pronouns) versus nonpronominal (noun phrase)
- *Recipient/Theme.definiteness:* definite (definite, definite proper noun) versus indefinite
- *Recipient/Theme.animacy:* animate (human and animal) versus inanimate (collective, temporal, locative, inanimate)
- *Semantics of the dative verb:* (i) transfer; (ii) communication; (iii) abstract
- *Length difference:* log10 difference of the length of the recipient and theme phrases in orthographically transcribed words (Bresnan and Ford 2010): *log*(-Recipient.length) – *log*(Theme.length)

The dataset covers a total of N = 1,190 observations. After exclusion of observations with missing data, we end up with N = 1,164 observations, of which 153 (13.1 %) are prepositional dative constructions and 1,011 (86.9 %) are ditransitive dative constructions.

2.2 Step 2: From the dataset, extract information about the lexical material in the argument slots of variant constructions, called target words in vector space modeling. For the dative variants, we extract the theme and recipient constituent head lemmas, i.e., the single most important content word of the constituent slot

For each individual dative observation in the dataset, we consider the theme and recipient constituent head lemmas, which are preannotated in the dataset under analysis (columns Recipient.head and Theme.head). Take, for instance, example (2b) (*the judge will usually give custody to the mother*): in this example, the theme head lemma is *custody*, and the recipient head lemma is *mother*. We pay attention only to the head lemmas, and not to the entire constituents, because these are going to be the so-called target words in semantic vector space modeling. Note that we also include pronominal head lemmas in our analysis, as in (1a) (*we'll give him fifteen*) – as Bresnan et al. (2007: 89) point out, pronominal recipients are extremely frequent in spoken English, and excluding those would be unrealistic.

2.3 Step 3: Choose a training corpus to train the distributional semantic models

Vector space models require balanced and large training corpora to generate solid and reliable distributional representations (Lenci and Sahlgren 2023). For this reason, we will use the Corpus of Contemporary American English (COCA; Davies 2019) as a training corpus. COCA is widely used, also in distributional semantics (Lenci and Sahlgren 2023:32). We specifically restrict attention to the spoken section of COCA (approx. 127 million words), given that our dative dataset likewise covers spoken English.

2.4 Step 4: Check if the target words also occur in the training corpus wordlist

Steps 4 and 5 identify the target words to be used in semantic vector space modeling. Importantly, each of the head lemmas in the dative should be attested at least once. If a head lemma does not appear in COCA, the source dative observation is excluded from the analysis. Accordingly, we had to exclude 23 dative observations, resulting in a new total of N = 1,164 valid and modellable dative observations

(down from 1,187), of which 153 (13.1 %) are prepositional dative constructions and 1,011 (86.9 %) are ditransitive dative constructions.[4] The target words thus obtained from the Bresnan et al. (2017) dataset (recipient and theme head lemmas combined) were tagged using the CLAWS7 PoS tag set (Garside and Smith 1997), the same PoS tag set used for COCA.

2.5 Step 5: Run a quality control check on the final target words list

To fix issues with PoS tagging errors, we hand-checked all the PoS-tagged lemmas. We removed from the wordlist the incorrectly tagged entries and corrected PoS tags where needed (for example, the lemma *red* can be both a proper noun (tagged as *pn*) and an adjective (tagged as *adj*)). The final target word list counts 590 PoS tagged head lemmas, of which 129 are recipient head lemmas, and 461 are theme head lemmas. It is important to note that these numbers reflect the high token frequency of pronouns and anaphorical markers in the recipient slot, in contrast to the more diversified lexical space in the theme slot.

2.6 Step 6: Define parameter settings for the distributional models

We are now getting ready for the actual distributional semantic vector space modeling. For the present analysis, we chose to employ a count-based type-level vector model, built with the python package Nephosem (QLVL 2021). This model allows us to build vectors as an aggregation over all the attestations of a given lemma in a given corpus, taking the form of an overall numerical profile of the lemma. The type-level vectors thus obtained abstract away from individual occurrences (i.e., tokens) that realize the (potentially) polysemous meanings of a lexeme, encoding instead "the patterns within the type-level matrices that are indicative of different senses" of the lemma (Heylen et al. 2015: 157).

We begin by fine-tuning the ability of the model to define and represent different configurations of the distributional context space by using different

4 These observations were excluded because the dataset source corpus (Switchboard) and the training corpus (COCA) are not the same. Distributional modeling requires reasonably large and balanced training corpora (see (Lenci and Sahlgren 2023), Section 2.1.2 for discussion). For this reason, we use COCA as a training corpus. A small number of lemmas in Switchboard do not occur in COCA. These lemmas typically involve proper nouns and/or hapax legomena – elements that are anyway challenging to model and not critical to the analysis.

implementations of parameter settings. As demonstrated by Montes (2021), this is a critical process that severely affects the final output, as parameters are heavily corpus- and task-dependent. After several rounds of assessment and testing, we selected a set of parameters as a function of the specificities of our spoken dataset (see Lenci 2018 for a review). The combination of the different parameters listed below yields a total of 432 distributional models to be considered:

- Window-size parameters, namely the selection of right and left context size: 3, 4, 7 tokens (applied in step 8)
- Collocation matrix parameters, that is, the dimensionality of the type-level vector: 5,000, 10,000, 50,000 context words for each target word (applied in step 8)
- Lexical context filters: no filtering, only nouns context (applied in step 8)
- Association strength measures: Positive Pointwise Mutual Information (PPMI), and log likelihood (applied in step 9)
- Similarity measures: cosine (applied in step 10)
- Amount of reduction: none (no dimensionality reduction), reduction to 2 dimensions, reduction to 10 dimensions (applied in step 10)
- Clustering: 3, 5, 8, 15 partitions (applied in step 11)

2.7 Step 7: Selecting the context words that will be used in subsequent steps, on the basis of a number of filtering criteria, most importantly their overall frequency in the training corpus

To start up the distributional semantics process, the first thing to be computed is a frequency list of *all* the lemmas in COCA, the training corpus. The lemmas contained in the list are the so-called context words, constituting the distributional lexical context from which our target lemma vectors will be modeled. Using the `vocab` class provided by the Nephosem package, the context words are extracted from COCA and stored in a python data structure called dictionary, where each lemma is paired with its frequency.[5] In this way, the distributional context can be filtered by PoS tags. Specifically, we employed both a full (excluding punctuation and numbers) and a noun-only context words list for the sake of assessing if a nominal, hence less sparse, distributional context could help building more representative target word vectors.

5 For more technical details, see https://qlvl.github.io/nephosem/tutorials/vocab.html.

2.8 Step 8: Using the target words list and the filtered context words lists, create co-occurrence matrices, in which rows correspond to the target words and the columns to the context words

At this point, we are ready to compile the large co-occurrence matrices at the core of our VSM-semantic predictors, as a function of the parameters defined in Step 6. These matrices are based on the target word lists and context word lists, which we previously computed. Each row represents a target word from the theme or recipient slot of each dative observation in the dataset, while each column represents a context word from the COCA corpus. Consider example (2b): this observation yields two rows with the target words *custody* and *mother*. The aggregation of the absolute co-occurrence frequencies between each target word and all the context words from COCA, a long array of numbers, constitutes a *word-type vector*, a geometrical representation of a lemma in the distributional semantic space. In order to obtain diversified and comparable distributional representations, we used a range of different co-occurrence configurations and so built a total of 18[6] raw co-occurrence matrices as a function of the different parameters defined in Step 6 (context and window-size filters and vector dimensionalities).

2.9 Step 9: Weighting the raw co-occurrence matrices using association strength measures

Absolute co-occurrence measures (or raw frequencies) are, however, not a solid ground on which to build meaningful semantic representations: following Zipf's law on skewed lexical frequency distributions, target words would only co-occur with a small number of context words, resulting in sparse, uninformative vectors. A customary practice in distributional semantics, therefore, is to turn raw co-occurrence values into significance weights using association strength measures. This lessens the effect of skewed frequency distributions and allows us to bring to the fore more informative semantic relationships between lemmas. Table 1 illustrates this based on a selection of recipient target lemmas from examples (1) and (2). The table shows the weighted co-occurrence values using the popular, reinforced version of Pointwise Mutual Information (PMI), called Positive PMI (PPMI; Bullinaria and Levy 2007): while maintaining the original goal of PMI, that is, to compare the

6 The 18 raw co-occurrence matrices are the result of the combination of 3 possible settings for the window-size filter parameter, 3 possible settings for the vector dimensionality parameter, and 2 possible settings for the context-filter parameter; hence, $3 \times 3 \times 2 = 18$.

Table 1: Small, constructed example of a co-occurrence matrix showing PPMI weighted type-level vectors of a selection of target lemmas from examples (1) and (2). The columns represent the PPMI association value between each target and context words.

Target	daughter	europe[a]	it	dad	troop
government	0	2.1	0.4	0	1.74
mother	3.5	0	0.9	3.2	0
him	0.2	0	4.1	0.6	0
guy	1.5	0	1.2	0.3	0

[a]In the co-occurrence matrix, "europe" comes without the capital letter because all the lemmas in the model are represented in their lowercase form.

probability of a target-context word pair to be distributionally closer with the expected probability if the two lemmas were independent, the PPMI relies on zero values instead of PMI negative values to better manage inaccuracies (Jurafsky and Martin 2023: 109).[7] From the previous 18 raw co-occurrence matrices, we obtained 36 matrices by applying the two parameter settings on each raw co-occurrence matrix.

2.10 Step 10: Compute similarity matrices to represent the distributional similarity between the target words

From the matrices generated in step 9, we next compute similarity matrices to represent the distributional similarity between target words, using the classic pairwise similarity measure cosine. The more the two target words are similar, the closer the value will be to 1; the more the two target words are different, the more the cosine value will be close to zero. In Table 2, we can observe how *guy* and *him* are closer in vector space because of some context words shared in the co-occurrence

Table 2: Cosine-similarity matrix generated from the co-occurrence matrix in Table 1.

Target	government	mother	him	guy
government	1	0	0.05	0
mother	0	1	0.6	0.3
him	0.05	0.6	1	0.8
guy	0	0.3	0.8	1

7 We also experimented with another classic association strength measure, log likelihood, but found that PPMI yielded more robust models.

matrix (Table 1), such as *dad* and *it*. We would expect, then, to find *government* in a different part of the vector space.

However, to assess if our data are meaningful and not skewed, we compared cosine-based similarity matrices to dimensionality-reduction-based similarity matrices. Nonlinear dimensionality reduction is a technique applied to reduce sparse, high-dimensional vector spaces into low-dimensional spaces that encapsulate the meaningful essence of the data. There are many ways to reduce dimensionality. We applied Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2020). UMAP essentially organizes high-dimensional data into a graph to find similarity patterns and then reorganizes these patterns in low-dimensional space using Riemannian geometry. Concretely, UMAP (a) first calculates how similar the target lemmas are in the original high-dimensional space using some distance measure (in this study, cosine) and then (b) reduces the data to a lower-dimensional space (e.g., 2 or 10 dimensions), where relationships between words are modeled using the UMAP-standard Euclidean distance measure.

In our case study, Step 10 yields 108 similarity matrices for each of the two dative construction slots, so $108 \times 2 = 216$ similarity matrices in total.[8]

2.11 Step 11: On the basis of the similarity matrices, cluster the target words with separate clusters for theme and recipient

On the basis of the recipient and themes similarity matrices generated in Step 10, theme and recipient heads are now – always separately – going to be clustered. This process yields groupings of semantically related types which will serve as semantic predictors of dative variant choice. Specifically, using the Partition Around Medoids (PAM) clustering algorithm (Kaufman and Rousseeuw 1990) in its Python implementation provided by the scikit-learn-extra module for machine learning from the package scikit-learn (Pedregosa et al. 2011), the central members of the clusters, called medoids, are identified in the data. Then, the rest of the type-word vectors, i.e., the distributional representations of our dative head lemmas, are grouped around these medoids based on the cosine similarity measure. The partition values (i.e., the number of clusters to create) are set to 3, 5, 8, 15: for each of the 108 similarity

8 Based on the previous number of 36 weighted co-occurrence matrices, we obtained $36 \times 3 = 108$ similarity matrices for each slot given the 3 possible settings for the parameter "Amount of reduction."

matrices, we obtain four clustering models, for a final total of 108 × 4 = 432 clustering models per theme and recipient slot. We can now let the distributional machine rest and focus on the evaluation of the models and their predictors.

2.12 Step 12: Evaluate the vector space models obtained and their clusters, and choose the optimal model according to principled criteria

After obtaining the clustering models, the next step consists of selecting the model that clusters best. In other word, we need to choose the VSM-semantic predictors that best capture generalizations about semantic similarities between the recipients/themes in our dataset. To this end, the clustering models are evaluated by computing (i) the Concordance-C index (C-value)[9] of their Conditional Random Forest (CRF), a multivariate statistical method that we utilize to predict dative variant choice based on only VSM-semantic predictors, and (ii) the negative token silhouette (Rousseeuw 1987) of the dative lemmas, i.e., the percentage of tokens of each dative lemma assigned to the "wrong" (or suboptimal) clusters, to gain a better understanding of data dispersion during clustering.

We considered only models with a negative token silhouette of less than 25 % in both recipient and theme clusters, and a CRF C-value higher than 0.75. After identifying the model with the lowest negative token silhouette and the highest CRF C-value,[10] we also inspected the clustering consistency of the next best fifteen models to rule out possible biases. As the clusters of these next-best models share similar semantic patterns, we proceeded to pick the best model. This was a model with a window size of 4, without lexical context filtering, with a vector dimensionality of 5,000, using PPMI as association strength measure, and with a dimensionality reduction of 10 and 15 clusters for both dative slots (see above for details). Everything is now ready for using these clusters as VSM-semantic predictors of the dative alternation.

9 The C-value is a nonparametric measure for determining goodness-of-fit, i.e., how well a statistical model fits a set of observations. C = 1 corresponds to optimal model prediction, and C <= 0.5 represents inability to predict (Levshina 2015: 259).

10 The best model was chosen using an R script: after assessing that the models using log-likelihood would perform worse than the model using PPMI, we selected all the models with PPMI. From here, we filtered all the models with recipient negative token silhouettes ≤ 0.25 and theme negative token silhouette ≤ 0.25. We then ranked the filtered models as a function of the C-values yielded by conditional random forest analysis. The best model is the top-ranked model in this list.

2.13 Step 13: Visualize the VSM-semantic predictors using t-SNE for a qualitative assessment

The last two steps in this workflow are dedicated to integrating VSM-semantic predictors into customary variationist alternation analysis. First, our methodology includes a module that qualitatively investigates VSM-semantic predictors using visualization tools. Inspired by the visualizations of the distributional clouds of tokens in Montes (2021), we map out VSM-semantic predictors using the stochastic algorithm t-SNE (Maaten and Geoffrey 2008) in its R implementation Rtsne (Krijthe 2015). From the same family of dimensionality reduction algorithms as UMAP, t-SNE is widely used to represent high-dimensional data in low-dimensional spaces (i.e., 2- or 3-dimensional spaces), maintaining and enhancing the meaningful features and semantic relations among the data. What we obtained are two plots, one for the VSM-semantic predictors relating to recipients, and the other relating to themes. The plots visually depict the distributional semantic geography of the predictors: how they are interconnected, which semantic patterns they show, and what their internal structure is. Additionally, the plots afford insights on misclassified data points (i.e., lemmas with negative token silhouette). In short, visualization is a tool that allows us to better understand the semantic structure and offers guidance for a responsible interpretation of the quantitative behavior of VSM-semantic predictors in regression analysis (see next step).

2.14 Step 14: Conduct two logistic regression modeling runs, creating models (a) with only traditional predictors as fixed effects and (b) with only VSM-semantic predictors as fixed effects. Determine the effect direction of VSM-semantic predictors and compare the goodness-of-fit of the two models

We utilize mixed effects logistic regression analysis, which is of course widely used in variationist/probabilistic linguistics (Tagliamonte and Baayen 2012). The model with only traditional predictors is structurally similar to the regression model in, e.g., Bresnan et al. (2007). The model with only VSM-semantic predictors uses binary categorical predictors in its fixed-effects structure. The binary structure of VSM-semantic predictors is simple: the recipient/theme lemma is a member of the cluster/predictor in question versus the recipient/theme lemma is *not* a member of the

cluster/predictor in question.[11] Hence, the regression model assesses whether a specific recipient or theme VSM-semantic predictor is a significant predictor of dative choices and evaluates effect directions (i.e., resulting preferences for the ditransitive or the prepositional variant).

Following best practice in variationist linguistics, we started out with a maximal model including all the predictors as fixed effects, and the language-external factor *speaker*[12] as well as the lexical effects Theme and Recipient heads as random effects (see, e.g., Röthlisberger et al. 2017). We include Theme and Recipient head lemmas as random effects even in the model with only VSM-semantic predictors because we want to tease apart the effect of purely lexical factors (via Theme and Recipient head lemma as random effects) from the effect of VSM-semantic predictors (via cluster memberships in the fixed-effects structure). After a round of model optimization to address multicollinearity and convergence issues, we pruned the models: first, we pooled infrequent random effect categories into "other" categories (thresholds: 5 for *speaker* and 2 for *Theme and Recipient heads*). Subsequently, those predictors that lacked explanatory power, namely the ones with the highest p-value, were removed one by one, until we ended up with minimal adequate models.

Data availability statement: the dataset(s) and analysis scripts are available at our OSF repository (https://osf.io/kfqsr/?view_only=71f776905c2645eb9dbb3b9f33aa42c5).

3 Results

In this section, we report the output of our methodology from Step 12 onward. We begin by presenting the distributional-semantic clusters of recipients and themes in the dative alternation dataset in Section 3.1. In Section 3.2, we assess the extent to which these clusters can help us predict dative choices.

3.1 Clustering theme and recipient lemmas as a function of semantic similarity

The evaluation process carried out in Step 12 (described in Section 2.12) identified the optimal semantic vector space model. As previously stated, an optimal model should

11 To obtain these binary VSM-semantic predictors, we converted the "Recipient.cluster" and "Theme.cluster" predictors, which map each recipient/theme head to its VSM-semantic predictor, into single-cluster VSM-semantic predictors. For more information about the coding, consult our OSF repository.

12 The variable *speaker* contains speakers IDs and represents idiolectal variation via the random effect structure of the regression model.

be associated with a very low negative token silhouette value for both recipients and themes (≤0.25), and with a conditional random forest *C*-value > 0.75. The optimal model is an excellent one, coming with a negative token silhouette of only 0.016 % for recipients and 0.06 % for themes, as well as with a *C*-value of 0.95 for a conditional random forest model predicting dative variant choices. We note that most of the parameter settings in the optimal model and its clustering are actually shared by the top fifteen models, indicating that the clustering output is fairly robust. The optimal model has the following parameters: 4-size context window, no vocabulary filter, dimensionality of 5,000, PPMI as weighting measure, dimensionality reduction to 10 dimensions, and 15 clustering partitions.[13] This means that the optimal model yields fifteen clusters for each dative slot, for a total of thirty clusters of dative recipients and themes.

Before proceeding to the variationist analysis, we visually depict the clusters in step 13 (described in Section 2.13). The interpretation of the resulting plots is fairly straightforward: individual dots represent distributionally modeled recipient/theme lemmas, with color coding indicating cluster membership. Each cluster is labeled by its central medoid. The extra labels in black with the arrows locate the lemmas in examples (1) and (2).

Figure 1 plots the recipient lemmas. PAM clustering yields all in all 15 internally reasonably homogeneous clusters, which are listed in Table 3.

Figure 2 plots the theme lemmas in our dataset. Again, PAM clustering yields in all 15 internally reasonably homogeneous clusters, as described in Table 4.

3.2 Variationist modeling: to what extent does distributional-semantic information predict dative choices?

In Step 14 (described in Section 2.14), we set a baseline by fitting a binary logistic regression model (Table 5) with only "traditional" predictors (i.e., those that old-school dative alternation research usually considers – recall that these come pre-annotated in the dataset). The response variable is dative variant choice (predicted odds are for the ditransitive dative). After model trimming (Zuu et al. 2009: 121–122), the minimal adequate model has a *C* score of 0.990; the condition number kappa[14] is 6.541. Suffice it to say that the traditional predictors all have the expected effect

13 See the supplementary materials in our OSF repository for more detailed information about the code and for the files related to the best model.

14 The condition number kappa is calculated using the *kappa.mer()* function from the R package *JGmermod* (https://github.com/jasongraf1/JGmermod). Values less than 6 indicate no collinearity to speak of, values less than 15 indicate medium collinearity (Baayen 2008: 200).

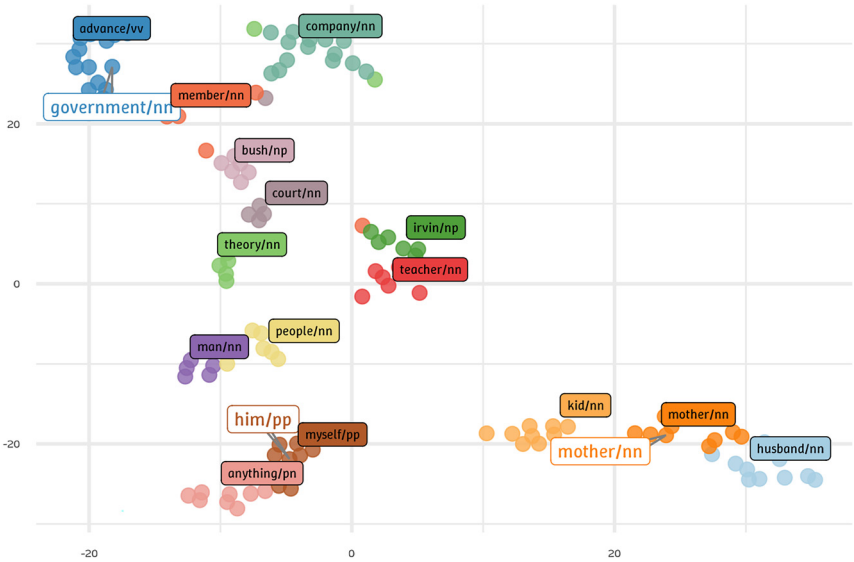

Figure 1: Scatterplot of recipient clusters. Color coding indicates cluster memberships. Clusters are labeled by their central medoid. Items in white rectangles with arrows refer to examples (1) (*we'll give him fifteen, but they give the guy a job*) and (2) (*I gave it to the government, the judge will usually give custody to the mother*).

direction, given the rich literature on the dative alternation (see, e.g., Bresnan et al. 2007; Szmrecsanyi et al. 2017 for models based on essentially the same dataset). We would like to emphasize that a C score of 0.990 indicates outstanding discrimination (Levshina 2015: 259) and underscores how well-understood the dative alternation is in principle.

A second binary logistic regression model was then fitted (Table 6), this time with only VSM-semantic predictors (i.e., those that we distributionally modeled). These predictors are basically binary dummy variables indicating if in a given dative observation, the recipient and theme head lemmas are or are not members of distributional-semantic clusters as discussed in Section 3.1. As in the "traditional" model, the response variable is dative variant choice (predicted odds are for the ditransitive dative). The maximal model contained 30 predictors (15 recipient clusters and 15 theme clusters, as discussed in Section 3.1). After model trimming, the minimal adequate contains four predictors, as shown in Table 6. The minimal adequate model has a *C* score of 0.991, and its condition number kappa is 6.804.

The model in Table 6 may be interpreted as follows. Predicted odds are for the ditransitive dative variant. The model, being minimally adequate, only shows

Table 3: Description of recipient clusters. Exemplification is provided in the supplementary materials.

Medoid of the cluster	Cluster description
Advance	Lemmas related to geopolitics, e.g., *nation, army, Europe.* Also contains the recipient *government*, as in (2a).
Member	Lemmas related to grouping and parts of groups, e.g., *association, team, citizen.*
Company	Lemmas related to the economy and the world of economics, e.g., *employee, taxpayer, management*.
Bush	The medoid refers to the former president of the United States, George Bush Senior, and the cluster attracts lemmas related to politics, such as *president, campaign, house*.
Court	Law terminology lemmas, e.g., *plaintiff, judge, jury*.
Theory	Lemmas related to abstract concepts (e.g., *matter, topic*) and research (e.g., *environment, aids*).
Irvin	The medoid is a family name, attracting some proper nouns, but also lemmas such as *Washington* (the politician) and *district.*
Teacher	Includes lemmas such as *student*, *artist* and the proper noun *Rashad*.
People	This is a semantically coherent cluster attracting lemmas such as *folk, peasant, public.*
Man	Lemmas related to the individual, like *person, woman, guy.*
Myself	This is a cluster featuring a collection of pronouns, such as *him, someone, them.*
Anything	Similarly to the cluster *myself*, this cluster mostly contains pronouns, e.g., *everyone, anybody, it*.
Kid	Lemmas related to childhood, e.g., *scout, boy, grown-up*.
Husband	This contains a blend of proper nouns and lemmas related to individuals, such as *actress, boss, friend.*
Mother	This cluster mostly consists of kinship terms, such as *grandparent, dad, sister*, and the recipient of the example in (2b).

significant VSM-semantic effects. If the recipient is in the *advance* cluster, which mostly contains lemmas about geopolitics (see previous subsection for discussion and exemplification), the odds for the ditransitive dative variant decrease by a factor of 0.142, i.e., by approximately 85 %. If the recipient is in the *anything* cluster (which mostly contains pronouns), the odds for the ditransitive dative increase by a factor of almost 7. If the recipient is in the *myself* cluster (which also mostly contains pronouns), the odds for the ditransitive dative increase substantially by a factor of almost 129. The *anything* and *myself* effects overlap with traditional modeling in that these clusters are highly pronominal in nature, and recipient pronominality has been shown to be a strong predictor of ditransitive dative usage in traditional dative alternation research (see, e.g., Bresnan et al. 2007). We also find one significant theme predictor: if the theme is in the *one* cluster (lemmas relating to general referents such

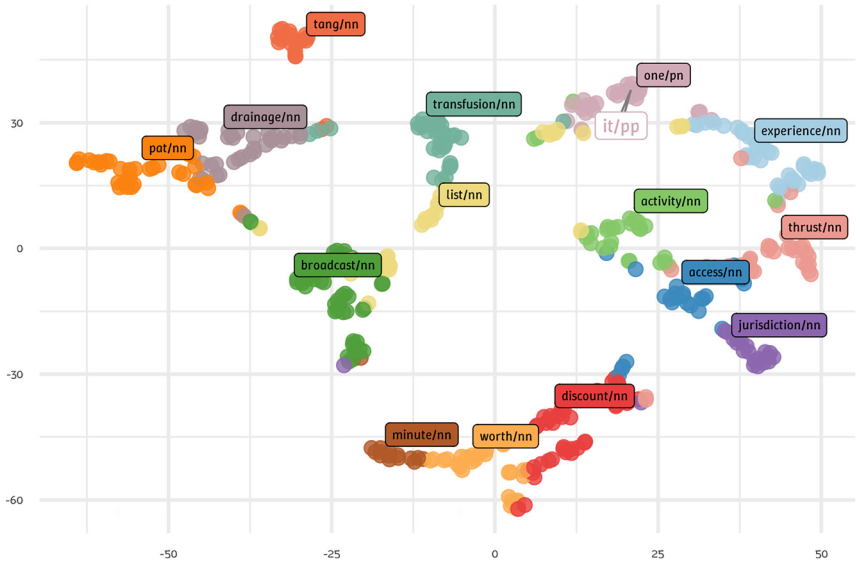

Figure 2: Scatterplot of theme clusters. Color coding indicates cluster memberships. Clusters are labeled by their central medoid. The item in the white rectangle with the arrow refers to the pronoun *it*, a key lemma in the dative alternation dataset under analysis.

as the pronouns *something, anything and,* but also nouns such as *item* and *thing*), the odds for the ditransitive dative decrease by a factor of 0.097. Figure 3 visually depicts the effect sizes of the predictors in the VSM-semantic model in Table 6.[15]

Now that we have the traditional model (Table 5) and the innovative VSM-semantic model (Table 6), it is time to revisit the question of which model has more explanatory power. One clear way of assessment is grounded on the difference between full effects (fixed and random) and fixed effects only that are inherent in the two models. Table 7 presents the *C*-values for both kinds of effects in the two models.

Both models have the same random effect structure, which includes lexical lemma effects (as intercept adjustments) for both the recipient and theme. It has become an industry standard to include such lexical effects (so-called by-item effects) in regression models in alternation research (see, e.g., Röthlisberger et al. 2017 and many other recent studies on the dative alternation and beyond). We also include lexical effects in the VSM-semantic model because this allows us to better tease apart

15 When interpreting the results of a variationist binomial regression model, it is important to understand that these results are probabilistic, not deterministic. In other words, the model reveals and explains general patterns in the use of the two variants, rather than identifying absolute or categorical differences.

Table 4: Description of theme clusters. Exemplification is provided in the supplementary materials.

Medoid of the cluster	Cluster description
Tang	Lemmas related to food, e.g., *flavor, fruit, tartness.*
Drainage	Lemmas related to the household (e.g., *shower, plant, bedroom*) and repair tools (e.g., *equipment, tub, oil*).
Pat	Large cluster of general household and hobby items, e.g., *ball, coat, knife.*
Transfusion	Fairly coherent cluster focused on health and sport terminology, e.g., *therapy*, *workout*, *backpack*.
Activity	Lemmas related to the workplace and education, e.g., *instruction, guidance, training.*
Experience	Lemmas related to more abstract concepts related to the sphere of emotions and (self-)consciousness, e.g., *ability, dignity, grief.*
Thrust (noun)	A hard-to-interpret cluster, containing lemmas such as *advantage*, *ambition*, *attention*.
Access (noun)	This cluster comprises theme lemmas to which one can have access, such as *care*, *aid, scholarship.*
Jurisdiction	Lemmas related to law terminology, e.g., *felony, law, permission.*
Discount	Terminology related to payments and, more generally, finance (e.g., *payment, raise, profit-sharing).*
Worth	Lemmas related to money and numbers (e.g., *billion, buck*, *penny*), but also tokens such as *much* and *more.*
Minute	Time and temporal terminology, among which numbers and lemmas such as *half, time*, *month.*
One	An interesting cluster: Rather incoherent from a semantic point of view, comprising lemmas relating to general referents such as the pronouns *something, anything,* and also nouns such as *item* and *thing.* This cluster also includes the most frequent theme lemma in the dataset, *it*, with 68 occurrences.
List	A cluster with rather weak semantic structure that groups words from the family and daily-life lexicon such as *husband, kiss, letter.*
Broadcast	Entertainment terminology, such as *comedy, entertainment, headline.*

the power of semantic effects from the power of purely lexical collocation/collostructional effects (à la Gries and Stefanowitsch 2004), and we are primarily interested in the power of semantics. The point is that particular lexemes, such as the recipient *mother* in (2b), may favor particular dative variants, but the question that we are primarily interested in is whether such effects can be generalized to lexemes with similar meanings. According to our analysis, the lexical effect of *mother* is not generalizable because the cluster with the medoid *mother* is not selected as significant in regression analysis.

Table 7's "*C*-value fixed effects only" scores are calculated using the *somers2()* function in the *Hmisc* R package: somers2(), in combination with the *predict()*

Table 5: Fixed effects in a minimal adequate model of the dative alternation based on traditional predictors only. Predicted odds are for the ditransitive dative variant. Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1. Random effects structure: Theme.lemma.pruned (Intercept) – variance: 4.8194, Std. Dev.: 2.1953; Speaker.pruned (Intercept) – variance: 0.3392, Std. Dev.: 0.5824; Recipient.lemma.pruned (Intercept) – variance: 0.2618, Std. Dev.: 0.5116; Number of obs: 1,164, groups: Theme.lemma.pruned, 156; Speaker.pruned, 88; Recipient.lemma.pruned, 38. Model formula: glmer(Response.variable ~ Semantics + Recipient.type.bin + Recipient.definiteness.bin + Theme.definiteness.bin + Recipient.animacy.bin + Length.difference + (1|Speaker.pruned) + (1|Recipient.lemma.pruned) + (1|Theme.lemma.pruned).

Fixed effects	Odds ratio (exp(*b*))	Standard error (SE)	Significance
(Intercept)	1.018	0.731	
Semantics	0.507	0.449	
Abstract → communication			
Semantics	0.264	0.146	*
Abstract → transfer			
Recipient.type.bin	18.912	12.818	***
Nonpronominal → pronominal			
Recipient.definiteness.bin	0.235	0.161	*
Definite → indefinite			
Theme.definiteness.bin	13.442	7.877	***
Definite → indefinite			
Recipient.animacy.bin	0.073	0.045	***
Animate → inanimate			
Length.difference	0.127	0.049	***

function. This row essentially tells us how explanatory the fixed effects (traditional or VSM-semantic) are, counting out purely lexical effects.[16] What Table 7 shows, then, is that the goodness-of-fit of both the traditional model and the VSM-semantic models is practically identical, scoring outstandingly high *C*-values (0.990 vs. 0.991). However, the two models work differently "under the hood" regarding the division of labor between fixed effects and random effects (including, crucially, purely lexical effects). In the traditional model, the fixed effects do the lion's share of the explanatory work (0.957 out of 0.990). In the VSM-semantic model, the fixed effects (i.e., VSM-semantic effects) also do a lot of explanatory work (0.782 out of 0.991), but comparatively more variability is left over to be accounted for by the random effects

16 As to fixed versus random effects, note that random effects account for variation specific to the dataset under analysis (e.g., individual speakers having particular preferences), which can not necessarily be generalized to other datasets. Fixed effects ought to be repeatable and generalizable to other datasets.

Table 6: Fixed effects in a minimal adequate model of the dative alternation based on VSM-semantic predictors only. Predicted odds are for the ditransitive dative variant. Significance codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ “ 1. Random effects structure: Theme.lemma.pruned (Intercept) – variance: 8.646, Std. Dev.: 2.9404; Speaker.pruned (Intercept) – variance: 0.224, Std. Dev.: 0.4733; Recipient.lemma.pruned (Intercept) – variance: 1.503, Std. Dev.: 1.2260; Number of obs: 1,164, groups: Theme.lemma.pruned, 156; Speaker.pruned, 88; Recipient.lemma.pruned, 38. Model formula: glmer(Response.variable ~ Recipient.advance.bin + Recipient.anything.bin + Recipient.myself.bin + Theme.one.bin + (1|Speaker.pruned) + (1|Recipient.lemma.pruned) + (1|Theme.lemma.pruned).

Fixed effects	Odds ratio (exp(*b*))	Standard error (SE)	Significance
(Intercept)	9.393	6.691	**
Recipient.advance.bin Recipient is not in the ADVANCE cluster → recipient is in the ADVANCE cluster	0.142	0.136	*
Recipient.anything.bin Recipient is not in the ANYTHING cluster → recipient is in the ANYTHING cluster	6.832	6.437	*
Recipient.myself.bin Recipient is not in the MYSELF cluster → recipient is in the MYSELF cluster	129.454	126.489	***
Theme.one.bin Theme is not in the ONE cluster → theme is in the ONE cluster	0.097	0.092	*

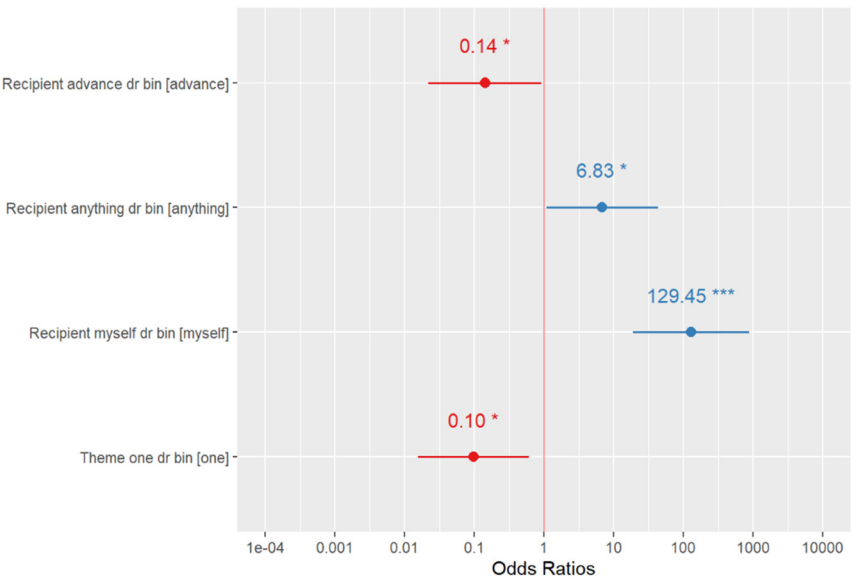

Figure 3: Odds ratios in the VSM-semantic model reported in Table 4.

Table 7: *C*-values of regression models.

	Regression model with only traditional predictors	Regression model with only VSM-semantic predictors
C-value fixed and random effects	0.990	0.991
C-value fixed effects only	0.957	0.782

structure (including lexical random effects) via intercept adjustments. In other words: traditional predictors do more heavy explanatory lifting than VSM-semantic predictors. The latter leave more unaccounted variability for the random effects structure to deal with.

4 Discussion and conclusion

Our point of departure in this paper was that skepticism about grammatical variation and optionality is common. The variationist literature demonstrates that grammatical variation between "alternate ways of saying 'the same' thing" (Labov 1972: 188) is by no means exceptional. Yet many theorists believe that the existence of different ways of saying the same thing is unpredicted because variation and optionality without meaning distinctions is allegedly dysfunctional and "unmotivated" (De Smet 2019), a gut feeling that has been referred to as the "Doctrine of Form-Function Symmetry" (Poplack 2018: 7) in the variationist community. In this study, we endeavored to investigate the extent to which we can reconcile form-function symmetry à la Haiman (1980: 516) and Goldberg (1995: 67) with the idea that we often do find optionality between different ways of saying the same thing. Specifically, we tested the extent to which meaning differences play a role in grammatical choice-making after all via the semantics of the materials in the argument slots of competing constructions.

As a case study, we utilized a corpus-based research design to revisit an in principle extremely well-studied case of grammatical optionality in language, viz. the dative alternation after the verb *to give* in English (*they give the guy a job* versus *they give a job to the guy*). Crucially, we know since Bresnan et al. (2007) that dative variants have fundamentally the same meaning and/or the same function. The empirical question we addressed in this study, then, was how adequately we can predict dative choices as a function of the semantics of dative constituents (as opposed to the semantics of dative constructions).

In this spirit, we reanalyzed a dataset that is largely identical to the Switchboard corpus-derived dative dataset investigated in Bresnan et al.'s seminal (2007) study. To assess the semantics of the dative constituents, we relied on Semantic Vector Space modeling (trained on the Corpus of Contemporary American English) as a customary technique in distributional semantics (Lenci 2018). The meaning distinctions that we thus established were subsequently submitted to variationist modeling. The distributional-semantic dative alternation model was then benchmarked against a traditional dative alternation model considering mostly higher-level formal predictors (such as, e.g., weight differences).

In what follows, we summarize key findings (Section 4.1), then discuss theoretical implications (Section 4.2), and finally sketch directions for future research (Section 4.3).

4.1 Key findings

Vector space modeling involves various parameter settings. We tested a range of parameters and parameter combinations and in so doing generated a total of 432 models. We then assessed the models according to predefined, principled criteria, and eventually selected the model with the lowest negative token silhouette in both recipient and theme clusters and the highest C-value of a conditional random forest model predicting dative choices. This optimal model generates 15 semantic clusters for the recipient slot in dative constructions and 15 semantic clusters for the theme slot (and this is by the way also what the next 19 top-ranked models do). There appears to be a richer variety of meanings in the theme slot, ranging from economics, family, daily-life items, to more abstract referents like feelings and numbers. The recipient slot in dative constructions has a strong preference for animate subjects and pronouns, as well as lexical items related to general referents like *matters, thing*.

We then moved on to variationist modeling: to what extent can the semantic clusters predict dative choices? To set a baseline, we initially calculated a regression model containing in its fixed-effect structure only traditional predictors, and enriched with by-subject and purely lexical by-item effects in the random effects structure (Table 5), in the spirit of Bresnan et al. (2007). The predictors that were selected as significant were semantics of the verb *give*, recipient pronominality, recipient definiteness, and theme definiteness, recipient animacy, and end-weight effects. The model thus does contain some semantic information (semantics of the verb *give*, recipient animacy), which, however, is standardly included in state-of-the-art modeling.

Next, we calculated an alternative regression model (Table 6) including only distributional-semantic predictors in its fixed effects structure while retaining the

exact same random effects structure as in the first model. The semantic predictors are specifically dummy variables indicating whether a given recipient or theme lemma is a member of one of the clusters identified in semantic vector space modeling. Four clusters were identified as significant predictors:

- the *advance* cluster relating to recipients (which mostly contains lemmas about geopolitics)
- the *anything* cluster relating to recipients (which mostly contains pronouns)
- the *myself* cluster relating to recipients (which also mostly contains pronouns)
- the *one* cluster relating to themes (lemmas relating to general referents such as the pronouns *something, anything and,* but also nouns such as *item* and *thing*)

The *anything* and *myself* clusters favor the ditransitive dative, while the *advance* and *one* clusters favor the prepositional dative. We note that the *anything* and *myself* clusters are strongly pronominal in nature and thus overlap with traditional pronominality effects as modeled in the traditional model. To some extent, pronominality also plays a role in the *one* cluster. We would like to remind the reader that in spoken English in particular it is hard to find dative constructions without pronominal slots, so excluding pronominal tokens from analysis was not an option.

How do the two variationist models compare to each other? The crucial criterion for our purposes is goodness-of-fit, which we measured by calculating the Index of Concordance *C* (Levshina 2015: 259). As shown in Table 7, both the traditional model and the alternative, VSM-semantic model, have outstanding and very comparable fits (0.990 vs. 0.991). But in the traditional model, the traditional fixed-effect predictors do considerably more work than the semantic fixed-effects predictors in the VSM-semantic model. In other words, there is more unaccounted variability in the VSM-semantic model, which is then "mopped up" as it were by the random-effects structure (including purely lexical effects). The conclusion is that traditional predictors are more explanatory than VSM-semantic predictors.

4.2 Theoretical implications

We note, first, that the semantic clusters identified by vector space modeling are overall lexically coherent, demonstrating that the method works in principle. But showing that vector space modeling works was not the aim of the paper. Rather, the research question we asked in the introduction section was the following: how adequately can we predict dative choices as a function of the lexical semantics of dative *constituents* (as opposed to the semantics of dative *constructions*)?

The upshot is that we *can* adequately model the dative alternation using information about the lexical semantics of dative constituents combined with by-subject and lexical by-item effects in the random effects structure. A model with this design accomplishes outstanding goodness-of-fit that is equal to goodness-of-fit of traditional models. The problem is that in this alternative model, only 4 out of 30 VSM-semantic clusters are selected as significant predictors of dative variant choice, and outstanding goodness-of-fit is achieved by letting purely lexical random effects do a lot of the work. The semantic predictors *per se* are less explanatory than traditional higher-order predictors. In a nutshell, the best of in all 432 semantic space models creates predictors that are outperformed by traditional, state-of-the-art predictors.

This verdict does not lend support to the idea that semantics is a strong factor fueling the dative alternation, not even through the detour via the constituent slots. Let us recapitulate: we knew before that dative variants are functionally equivalent and normally interchangeable (Bresnan et al. 2007). We now also know that variant choice is to some extent constrained by the semantics of the materials in the constituent slots, but this conditioning is weaker than conditioning by traditional predictors (note here that unlike some construction grammarians, we do not assume that probabilistic conditioning comes under the remit of the "meaning" of dative variants – see fn. 2). Had our exercise in dative alternation modeling shown that inclusion of VSM-semantic predictors improves dative alternation modeling (e.g., by reducing the share of explanatory work that random effects have to do), we would have been in a position to conclude that the idea of form-function symmetry has (some) empirical back up. However, inclusion of VSM-semantic did not buy us much. Therefore, the idea of form-function symmetry in the dative alternation retains the status of an empirically unconfirmed conjecture.

We also add that from the effect of VSM-semantic predictors in regression modeling, we do not see any indications whatsoever that change-of-state or change-of-possession contexts favor the ditransitive dative variant, or that change-of-place or movement-to-a-goal contexts favor the prepositional dative variant, as they should if advocates of the multiple-meaning approach had a point (see, e.g., Green 1974; Oehrle 1976). With that being said, we acknowledge that in the "traditional" model, we report in Table 5 there is a significant effect such that *give*-semantics of "transfer" decrease the odds for the ditransitive dative variant and increase the odds for the prepositional dative variant. Thus, one could interpret our results as indicating that, first, there is a semantic effect that bears some resemblance to the multiple-meaning claim, and that, second, the traditional model picks up traces of this effect, but the VSM-semantic-model does not. Hence, one might conclude that (a) there are traces of multiple-meaning effects, but they are less strong than the

effects of other factors, and they certainly are not strong enough to fully explain or fully motivate the alternation, and that (b) the VSM-semantic predictors are simply not good at picking up these traces. The extent to which it is easy to predict the meaning of *to give*, in a particular observation, on the basis of the VSM-semantics of what is in the theme and recipient slots is an empirical question whose exploration is reserved for another occasion.

4.3 Directions for future research

The analysis we conducted is clearly just a first step and ought to be extended to other alternations in the grammar of English or of other languages. By choosing the dative alternation after the verb *to give* as a test case for the VSM-semantic approach, we sought to establish a gold standard: this alternation is, after all, extremely well understood, and variationist linguists have had decades to fine-tune traditional modeling. The issue is, however, that state-of-the-art traditional modeling of the dative alternation after *give* achieves outstanding goodness-of-fit scores that are hard to surpass by any model, so the English dative alternation after *give* is perhaps a particularly tough nut to crack for alternative approaches. But the point is that our methodology paves the way for reanalyzing the dative alternation after verbs other than *give*, and also for taking a fresh look at (perhaps less well understood) alternations in which semantics potentially plays a larger role. In this context, it would be particularly interesting to reinvestigate grammatical alternations that are less likely (compared to the dative alternation) to have pronominal constituents – pronouns, after all, do not straightforwardly carry semantic meaning, although of course Semantic Vector Space will assign pronouns to particular clusters. Be that as it may, what this paper would seem to have provided is a robust methodology for including distributional-semantic information in variationist modeling.

Acknowledgments: The authors would like to thank two anonymous reviewers for constructive feedback.
Research funding: This work was supported by FWO under the grant no. G044721N.

References

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. *Predicting the dative alternation. Cognitive foundations of interpretation*, 69–94. Amsterdam: KNAW.

Bresnan, Joan, Anette Rosenbach, Benedikt Szmrecsanyi, Sali Tagliamonte & Simon Todd. 2017. Syntactic alternations data: Datives and genitives in four varieties of English. Dataset. Stanford digital Repository. Available at: https://purl.stanford.edu/qj187zs3852.

Bullinaria, John A. & Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* 39(3). 510–526.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.

Curzan, Anne, Robin M. Queen, Kristin VanEyk & Rachel Elizabeth Weissler. 2023. Language standardization & linguistic subordination. *Dædalus* 152(3). 18–35.

Davies, Mark. 2019. The corpus of contemporary American English (COCA): One billion words, 1990–2019. https://www.english-corpora.org/coca/ (accessed 9 March 2023).

De Smet, Hendrik. 2019. The motivated unmotivated: Variation, function and context. In Kristin Bech & Ruth Möhlig-Falke (eds.), *Grammar – discourse – context: Grammar and usage in language variation and change*, 305–332. Berlin, Boston: De Gruyter.

De Smet, Hendrik, Frauke D'hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2). 197–234.

Garside, Roger & Nicholas Smith. 1997. A hybrid grammatical tagger: CLAWS4. In Roger G. Garside, Geoffrey Leech & Anthony Mark McEnery (eds.), *Corpus annotation: Linguistic information from computer text corpora*, 102–121. London, New York: Longman.

Gerwin, Johanna. 2014. *Ditransitives in British English dialects* (Topics in hEnglish Linguistics 50.3). Berlin: Mouton de Gruyter.

Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *ICASSP-92: 1992 IEEE international conference on acoustics, speech, and signal processing*, vol. 1, 517–520.

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

Green, Georgia. 1974. *Semantics and syntactic regularity*. Bloomington: Indiana University Press.

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

Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56(3). 515.

Heylen, Kris, Thomas Wielfaert, Dirk Speelman & Dirk Geeraerts. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua (Polysemy: Current Perspectives and Approaches)* 157. 153–172.

Jurafsky, Dan & James H. Martin. 2023. Speech and language processing, 3rd edn. draft. https://web.stanford.edu/~jurafsky/slp3/ (accessed 18 December 2023).

Kaufman, Leonard & Peter J. Rousseeuw. 1990. *Finding Groups in data: An Introduction to cluster analysis (Wiley Series in Probability and Statistics)*, 1st edn. Hoboken, New Jersey: John Wiley and Sons, Inc.

Krijthe. 2015. *Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation*. R. Available at: https://github.com/jkrijthe/Rtsne.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

Langacker, Ronald W. 1990. *Concept, image, and symbol: The cognitive basis of grammar* (Cognitive Linguistics Research 1). Berlin, New York: Mouton de Gruyter.

Leclercq, Benoît & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. In Lotte Sommerer & Stefan Hartmann (eds.), *Constructions. Constructions 2023: Special issue "35 Years of constructions"*.

Lenci, Alessandro. 2018. Distributional models of word meaning. *Annual Review of Linguistics* 4(1). 151–171.
Lenci, Alessandro & Magnus Sahlgren. 2023. *Distributional semantics* (Studies in Natural Language Processing). Cambridge: Cambridge University Press.
Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
Levshina, Natalia & Kris Heylen. 2014. A radically data-driven construction grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the scope of construction grammar*, 17–46. Berlin, Boston: De Gruyter Mouton.
van der Maaten, Laurens & Hinton Geoffrey. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9(86). 2579–2605.
McInnes, Leland, John Healy & James Melville. 2020. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*. https://doi.org/10.48550/arXiv.1802.03426.
Montes, Mariana. 2021. *Cloudspotting*. Leuven: KU Leuven PhD thesis. Available at: https://cloudspotting.marianamontes.me/.
Oehrle, Richard Thomas. 1976. *The grammatical status of the English dative alternation*. MIT Doctoral Dissertation.
Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel & Mathieu Blondel. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12. 2825–2830.
Perek, Florent & Martin Hilpert. 2017. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22(4). 490–520.
Poplack, Shana. 2018. Categories of grammar and categories of speech: When the quest for symmetry meets inherent variability. In Naomi L. Shin & Daniel Erker (eds.), *Studies in functional and structural linguistics*, vol. 76, 7–34. Amsterdam: John Benjamins Publishing Company.
QLVL. 2021. *nephosem*. Zenodo.
Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). https://doi.org/10.1515/cog-2016-0051.
Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65.
Sankoff, David. 1988. Sociolinguistics and syntactic variation. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey*, 140–161. Cambridge: Cambridge University Press.
Szmrecsanyi, Benedikt & Jason Grafmiller. 2023. *Comparative variation analysis: Grammatical alternations in world Englishes* (Studies in Language Variation and Change). Cambridge New York, NY: Cambridge University Press.
Szmrecsanyi, Benedikt, Jason Grafmiller, Joan Bresnan, Anette Rosenbach, Sali Tagliamonte & Simon Todd. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa: A Journal of General Linguistics* 2(1). https://doi.org/10.5334/gjgl.310.
Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.
Yáñez-Bouza, Nuria. 2006. Prescriptivism and preposition stranding in eighteenth-century prose. *Historical Sociolinguistics and Sociohistorical Linguistics* 6. https://www.let.leidenuniv.nl/hsl_shl/preposition%20stranding.htm.
Zuu, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York, NY: Springer-Verlag New York.

Bionotes

Chiara Paolini
Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium
chiara.paolini@kuleuven.be
https://orcid.org/0000-0001-7961-3624

Chiara Paolini is a PhD student at the Quantitative Lexicology and Variationist Linguistics research group of KU Leuven. Her research focuses on how much the choice between variants in grammatical alternations depends on the lexical semantics characteristics of the context in which those variants are embedded, making use of analytical techniques from variationist linguistics and corpus-based distributional semantics. She is also involved in the Wikidata Gender Diversity project, which analyze the linguistic habits and approaches of the users toward gender diversity and inclusivity using corpus linguistics methods.

Hubert Cuyckens
Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium

Hubert Cuyckens (°1956) is emeritus professor of English language and linguistics at KU Leuven. While retaining an interest in cognitive semantics, his main research focus lies in the domain of historical syntax of English, specifically in the grammaticalization phenomena, its interface with construction grammar, and diachronic developments in the system of verbal clause complementation. Recently, he has turned his attention to historical variation and competition of nonfinite versus finite complementation, thus marrying the variationist with the diachronic syntactic research agenda. His theoretical orientation has always been that of cognitive-functional and usage-based linguistics.

Dirk Speelman
Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium

Dirk Speelman (°1965) is Professor at the Department of Linguistics at the KU Leuven. His main research interest lies in the fields of corpus linguistics, computational lexicology, and variational linguistics in general. His work focuses on methodology and on the application of statistical and other quantitative methods to the study of language. Some topics of interest: distributional semantics and its application to language variation, quantitative measures in support of aggregate level analyses of lexical variation, statistical methods, and machine learning in support of the analysis of syntactic variation.

Benedikt Szmrecsanyi
Quantitative Lexicology and Variationist Linguistics (QLVL) - KU Leuven, Leuven, Belgium

Benedikt Szmrecsanyi (°1976) is associate professor of linguistics at KU Leuven. Benedikt's preferred theoretical framework is usage-based/experience-based linguistics. He views linguistic variation as a window into the hidden structure of human language and the nature of linguistic knowledge. His research interests specifically include variation studies (synchronic & diachronic), probabilistic grammar, sociolinguistics and register analysis, language complexity, geolinguistics, dialectology, dialectometry, and dialect typology, learner language.