CHAPTER 10

Enabling and allowing in Hebrew

A Usage-Based Construction Grammar account

Elitzur Dattner
Tel Aviv University

Three-Argument Dative constructions in Hebrew include two sub constructions, each with a specific form-meaning correlation, a specific usage pattern, and a particular discursive context. Syntactically, the two sub-constructions differ in that the THEME argument can be either (i) a noun phrase, or (ii) an infinitival predicate. In particular, the verb *natan* 'give' represents a prototypical construal in both sub-constructions. That is, with a NP THEME argument the verb *natan* 'give' has its literal meaning. With an infinitival THEME argument, however, the verb has the meaning of 'allow, enable.' Analyzing corpus data of the Infinitival complement Construction (ii), I account for these uses of the verb *natan* 'give' in a Usage-Based Construction Grammar perspective (Bybee, 2010). Utilizing the exploratory statistics techniques of Multidimensional Scaling (Borg and Groenen, 2005), Multiple Correspondence Analysis (Greenacre, 2010), and Hierarchical Classification on Principal Components (Husson et al., 2011) on corpus data, I show that multiple factors play a role in structuring an Argument Structure Construction. Furthermore, based on the bottom-up statistical analysis I argue that the verb's morphological paradigm is one of the main cues for interpreting the relevant event frame and Dative participant roles in Hebrew.

1. Introduction: Hebrew Three Argument Dative constructions

Three-Argument Dative constructions in Hebrew include two sub constructions, each with a specific form-meaning correlation, a specific usage pattern, and a particular discursive context. Syntactically, the two sub-constructions differ in that the THEME argument can be either (i) a noun phrase, or (ii) an infinitival predicate. In particular, the verb *natan* 'give' represents a prototypical construal in both sub-constructions. That is, with a NP THEME argument the verb *natan* 'give' has its literal meaning. With an infinitival THEME argument, however, the verb has

DOI 10.1075/slcs.167.10dat

the meaning of 'allow, enable.' Example (1) is a prototypical representation of the Nominal Complement Construction; the Infinitival Complement Construction is exemplified in (2):

(1) *im* ***titnu*** ***lahem*** ***takciv*** *hem yaasu et ze be-kecev*
if you.will.give to.them budget they will.do ACC this in-pace
od yoter mahir.
even faster.
'If you'll **give** them the budget, they'll do it even faster.'

(2) ***tnu lanu*** *beynataym* ***latet*** *lahem darga zmanit.*
give to.us meanwhile to.give to.them position temporary.
'For now, **let us** give them a temporary position' (lit. 'give to us to give to them').

Thus, it seems that the same verb, *natan* 'give,' construes two different event frames, depending on the specific type of three argument sub-construction. In the present paper I focus on the Infinitival Complement Construction (2), which attributes an 'allowing/enabling' meaning to a verb such as *natan* 'give.' Analyzing corpus data, I account for these uses of the verb *natan* 'give' in a Usage-Based Construction Grammar perspective (Bybee, 2010). First, I argue that the different functions of the two sub-construction are best explained by analysing them as belonging to a network of Dative constructions. Then, I show that accounting for the Hebrew Dative constructions as a category of different but related constructions allows us to explain special uses of verbs such as *natan* 'give' in different syntactic environments.

In order to arrive at a coherent representation of a construction one needs to account for the construction's category structure. That is, one needs to analyse similarities and dissimilarities within a set of tokens belonging to the same construction, measuring, quantifying, and defining family resemblance between the tokens in the category. Thus, one can arrive at a detailed description of the construction's usage patterns. Taking a usage-based point of view, I examined family resemblance between tokens based on linguistic and extra-linguistic characteristics. Each token in the corpus was coded for multiple parameters from multiple sources of information: lexical, morphological, syntactic, semantic, and discursive.

Exploratory Multivariate Statistics analysis allows one to account for multiple parameters in order to uncover hidden patterns in a database. In the present paper I use three Exploratory Multivariate Analysis tools: Multidimensional Scaling (Cox and Cox, 2001; Borg and Groenen, 2005; Levshina, 2012), Multiple Correspondence Analysis (Greenacre, 2010; Husson et al., 2011), and Hierarchical Classification on Principal Components (Husson et al., 2011). Using these methods I present an extensive account of the usage patterns of the Infinitival Complement

Construction, revealing the parameters that constitute an organizing principle for the construction's cognitive category, thus providing a comprehensive description of the construction. Mainly, I show that multiple factors play a role in structuring the Infinitival Complement Construction's category structure. Furthermore, based on the bottom-up statistical analysis I argue that the unique Semitic feature of verbal paradigm (i.e. Binyan) is one of the main cues for interpreting the relevant event frame and Dative participant roles in Hebrew, combining morphological, syntactic and transitivity parameters (Hopper and Thompson, 1980) together.

Like all Hebrew Three Argument Dative Constructions (see Dattner (In progress) for a survey of Hebrew Dative Constructions), the Infinitival Complement Construction construes a state of affairs realized by the Subj-V-InfV as affecting the Dative-marked participant. That is, while affectedness is commonly referred to as a Direct Object characteristic, the present construction (like all Three Argument Dative Constructions in Hebrew) profiles a different affected entity, namely, the Dative-marked participant. With a NP Object this construal can be of creation: Subject creates Object for the benefit of the Dative-marked participant, or of motion: Subject moves Object into the Dative-marked participant possession, see (3). With an Infinitival Object the construal is of enabling or allowing: Subject enables/allows the Dative-marked participant to perform the action denoted by the infinitival clause, see (4):

(3) *hu sholeax li meyl.*
He sends to.me mail.
'He sends me an email.'

(4) *ha-texnologia meafsheret lanu ha-yom lehacig meyda […]*
the-technology enables to.us today to present information […]
'Thanks to the technology, we can present our data today (lit. 'the technology enables us the presentation of the data').'

Thus, while the nominal construction prototypically construes transfer or benefiting events, the Infinitival Complement Construction construes enabling and allowing scenarios.

The verb *natan* 'give' is one of the most frequent in both the Nominal Construction and the Infinitival Construction. However, in each of these constructions it has a different meaning. While in the Nominal construction it denotes a transfer event (5), in the Infinitival construction it portrays an allowing/enabling situation (6):

(5) *misrad ha-shikun natan li milyon shekel.*
Ministry of Construction and Housing gave to.me milion Shekels.
'The Ministry of Construction and Housing gave me a million Shekels.'

(6) *ten li ledaber.*
Give to.me to.talk.
'Let me talk.'

Thus, it is an interesting case, first for pointing out the differences between these two closely related constructions, and second, for accounting for the construction's form and function. One of the main claims of the present paper is that the verb *natan* 'give' in (6) gains a non-lexical meaning from the Argument Structure Construction it participates in. That is, the argument structure

(7) [Subj V DatNP Infinitival-clause]

construes an allowing/enabling situation, and a certain category of verbs (including *natan* 'give') inserted into its verbal slot alter their meaning to fit such a construal.

The paper is structured as follows: In Section 2 I lay out the theoretical framework in which the present research is located. Section 3 presents the corpus data for the present paper and the statistical methods used to analyze it, and Section 4 discusses the results of these analyses. The paper ends with a conclusion in Section 5.

2. Usage-Based Construction Grammar

The main framework in which I analyze the data in this study is Usage-based Construction Grammar. Construction Grammar has been described and developed in the works of Fillmore and Kay (for example, Fillmore et al., 1988; Kay, 1997; Kay and Fillmore, 1999), Goldberg (e.g. Goldberg, 1995, 2003, 2005b, 2006; Goldberg and Jackendoff, 2004), Croft (e.g. Croft, 2001) and Bybee (e.g. Bybee and Eddington, 2006; Bybee, 2006) among many others. Constructions, in Construction Grammar theories, are the basic units of grammar (Diessel, 2004): conventionalized form-meaning correspondences. The existence – or the necessity – of these form-meaning parirings is a subject of debate between different branches of the construction grammar framework. Goldberg states that constructions are "posited whenever there is evidence that speakers cannot predict some aspect of their form, function, or use from other knowledge of language (i.e. from other constructions already posited to exist)" (Goldberg, 2005a, p. 17). On the other hand, in usage-oriented models of Construction Grammar constructions are a function of usage. That is, a particular linguistic sequence is a construction (theoretically and presumably cognitively) based on usage parameters such as frequency and discourse pattern, whether it is predictable from other constructions or not (Ariel, 2008; Bybee, 2010).

A special case of constructions is the Argument Structure Construction (Goldberg, 1995): the form-meaning pairings in the language that concern schematic clausal expression, rather than fully or partially fixed constructions such as idioms or prefabs (Fillmore et al., 1988). The English Caused Motion Construction is an example for such a basic, schematic, Argument Structure Construction:

(8) Pat sneezed the napkin off the table.

Goldberg's innovative approach to Argument Structure Constructions shows that the meaning associated with (8) cannot be analyzed as stemming from the particular verb (an intransitive, non-motion related verb in this case); it has to be searched for elsewhere. Goldberg puts forward the proposition that it is the specific Argument Structure Construction in (9) that bears the 'causing to move' meaning, such that any verb (with several functional limitations) inserted into the V slot of the construction will 'gain' such a meaning:

(9) Subj V Obj Obl
'X causes Y to move Z' (Goldberg, 1995, p. 3).

One claim about Argument Structure Constructions is of critical importance for the present analysis; that is, that simple clause constructions are associated with construals of basic human experience (Goldberg, 1995, p. 5). In the account of Hebrew Dative Constructions advocated in the present study I argue that the Hebrew Dative is related to basic event types such as TRANSFER, MOVING, AFFECTING, and EXPERIENCING. Furthermore, I show that on top of these basic event types, Dative Argument Structure Constructions are associated with basic *discursive* functions as well, thus emphasizing the role of *construing* events and scenes rather than merely reflecting them.

Usage-based Construction Grammar is one of the most recent developments in Construction Grammar theories. A usage-based model of language seeks explanations for linguistic phenomena in terms of domain-general cognitive processes, and accounts for the ways experience with language affects its representation. Frequency effects on structure are emphasized, with reference to both type and token frequency. A usage-based model of representation assumes that "grammar is the cognitive organization of one's experience with language" (Bybee, 2006, p. 711); thus, comprehension and production are integral to the linguistic system. Since linguistic representations emerge from usage patterns, usage, synchronic variation, and diachronic change are related (Bybee, 2006), and discourse function has a critical role: a cognitive distinction will become useful based on its discourse function (Ariel, 2008).

One interesting approach within the usage-based framework advocates an exemplar-based representation (and organization) of linguistic categories. This branch has been developed mostly with regard to phonetics (e.g. Pierrehumbert, 2001), but the 2006 special issue of The Linguistic Review (Gahl and Yu, 2006), devoted to exemplar-based models in linguistics, shows that exemplar-based accounts are conquering more and more areas, such as language evolution (Wedel, 2006), or syntactic acquisition (Abbot-Smith and Tomasello, 2006).

In an exemplar-based model each instance of use is said to have an impact on the cognitive representation of language. Exemplar representations store every available information for each instance of use. Phonetic, lexical, and constructional information, morpho-syntax, semantics, context, and discourse function, are all part of an exemplar's cognitive representation (Bybee, 2010, p. 14). As in other usage-based models, frequency is of major importance. High frequency clusters of exemplars are stronger than lower frequency ones in that they are cognitively entrenched: each usage token strengthens the representation of a particular construction in the language.

Recently, Construction Grammar has been claimed to be the most adequate linguistic representation considering the notions behind exemplar-based models of cognitive organization (Bybee and Eddington, 2006; Bybee, 2006, 2010). This offshoot of Construction Grammar theories is the context in which I account for the data in the present paper.

Since a usage-based model assumes a rich memory representation, representation of particular instances of constructions can be accessed for analogical extensions or the creation of new constructions through family resemblance. Constructions are particularly appropriate for an exemplar representation; they are amalgams of several sources of information: morpho-syntactic, semantic, and discursive. Cognitive representations of grammar are organized into constructions which are partially schematic, conventionalized sequences of morphemes with a direct semantic representation. That is, constructions are grounded in lexical items. Exemplars of phrases that are similar on different dimensions (i.e., have a high degree of family resemblance) are grouped together in cognitive representation. From such a grouping a construction can emerge. If there are similarities (in particular, semantic similarities) among the items occurring in the open slot, a category for these items would begin to develop (Bybee 2006). A string of words or morphemes that is used with some frequency would be considered a construction even if its form and function or meaning are entirely predictable (Bybee, 2010), and the categories that define the open slots in such constructions constitute the grammatical categories of the language.

3. Data and method

3.1 Data

The corpus serving as a database for the present research is an approximately 1,760,000 words corpus of spoken Hebrew. The corpus is a collection of transcriptions of 198 meetings of committees of the Knesset, the Israeli parliament, composed of multiple registers of language, both formal and colloquial.

From this corpus, all occurrences of a Dative (*le-*) marked pronoun were extracted, resulting with 16,575 tokens of Dative uses. The process of downloading the files from the Knesset database and extracting only clauses containing the Dative was automated using an R-script I modified to fit my needs for web-crawling and text analysis (following Gries, 2009). Each token was then coded by hand for 18 parameters (some were relevant only for a subset of the tokens). Table 1 presents a fragment of the coding parameters, the ones relevant for the current study. The construction we focus on in the present paper, the Three Argument Infinitival Complement Construction, occurs 664 times in the corpus.

Table 1. Variables and values of the corpus data

Variable	Interpretation	Values
AFFIRM	Affirmation of the clause	2 values: Affirmative, Negative
DAT.FUNC	Dative function: participant role of the Dative-marked participant	4 values: Human Endpoint, Addressee, Experiencer, Recipient
DAT.PRSN	Person of the Dative-marked participant	3 values: First, Second, Third
MODE	Mode of the clause	2 values: Irrealis, Realis
P.LEX.CAT	Lexical category of the main predicate	2 values: Transitive Verb, Intransitive Verb
P.SEM.TYPE	Semantic type of the main predicate	7 values: Making, Giving, Speaking, Value, Deciding, Transitive.Motion, Wanting
S.AGE.INDIV	Agency of Transitive Subject, Individuation of an Intransitive Subject	3 values: S.High, S.Low, S.Mid
V.BIN	Verbal paradigm of the main verb (the Hebrew Binyan system)	3 values: Kal, Piel, Hifil
Total: 8		26

3.2 Method

The analysis presented in this paper is based on exploratory statistics, and specifically on Multidimensional Scaling, Multiple Correspondence Analysis and Hierarchical Classification on Principal Components. Multidimensional scaling (MDS) is a statistical technique suitable for the consideration of similarities (Borg and Groenen, 2005). MDS represents the distance between pairs of observations as distance between two points on a map. This map is a geometric display of the data, allowing a visual analysis of the data's structure. Thus, using MDS one can analyze similarities between sets of observations as regularities in the data. The course of action using MDS is converting the corpus data into a distance matrix that considers the distance between a row in the data and all other rows, based on their coded categorical variables. Such a distance matrix resembles the diagonal figure representing distance in kilometres or miles between cities one can find in a road-map. Then, these distances are graphically visualized on a map, each point on the map represents a token in the database, in the same way each point on a road-map represents a city. That is, with a database of formal patterns taken out of a linguistic corpus, for instance, using MDS one can observe the similarities and dissimilarities between particular tokens of use, as well as between sets of tokens. Working within an exemplar-based approach, each token of use is considered to be an exemplar, and sets of geometrically related tokens can be analyzed as clusters.

Multiple Correspondence Analysis (MCA) is a technique for uncovering patterns in a multivariate database (Greenacre and Blasius, 2006; Husson et al., 2011). That is, it is an exploratory method used to detect patterns and correlations in data with multiple categorical variables. For example, collecting data such as answers to a questionnaire , the researcher can produce a table with columns representing the questions and rows representing individuals answering these questions. Each cell in the table, then, corresponds to an individual's answer to a particular question. Having this type of table, one can ask about correspondences in the data, and search for patterns. For instance, what is common between all individuals answering the first, third, an fifth questions the same. Or, another possible question would be what is common between the answers to the questions. That is, if, for instance, question A has three possible answers (x, y, and z), and question B has two possible answers (j, and k), are there correspondences between the possible answers, such that A(x) and B(k) correspond with each other. The visual representation of individuals and categorical variables on the MCA factor map aids in identifying such correlations. Individuals, in MCA, will be represented as points on the map. Simply put, two individuals answering exactly the same answers for the same questions will appear on the same point on the two-dimensional map. The distance between categorical variables is calculated based on the number of individuals carrying each variable (Husson et al., 2011). Note that although MCA is an extension of Correspondence Analysis, distance between points is less meaningful as in correspondence analysis, due to the conflation of multiple dimensions into a two-dimensional map. However, quadrants and approximation do have meanings, and interpreting these plots is rather straightforward (Rencher, 2012; Glynn, 2014). In the present paper I used the R program for statistical computing (R Core Team, 2013) with the package FactoMineR (Husson et al., 2013) for Multiple Correspondence Analysis and Hierarchical Classification on Principal Components, with the MCA done on the Burt matrix. This package was chosen due to its visualizations options and the possibility to add Hierarchical Classification on the basis of the MCA. While the explained inertia (explained variation in the data) are usually lower using this package than using the ca package (Nenadic and Greenacre, 2007) due to Greenacre's inertia adjustment, the amount of variables in the present research renders the inertia almost uninterpretable (Glynn, 2014). Thus, better visualization options of the correlations and the possibility to compute the Hierarchical Classification on the MCA motivated my decision.

In linguistic terms, MCA is an "exploratory technique that reveals frequency based associations in corpus data" (Glynn, 2014), visualizing these associations in the form of a map. Highly associated forms, for instance, appear closer on the map than two forms with no association. That is, a suitable linguistic database for MCA would have a row for every hit in the corpus, and a column for every coded category. Each cell would represent a token's behaviour relative to a coded category. Individuals in this case would be tokens of use: every hit of a searched formal pattern in the corpus, for instance, would be represented as a point on the map, relative to its coded attributes.

Hierarchical Classification on Principal Components (HCPC) is another tool for representing similarities or correlations (Husson et al., 2011). HCPC outputs a hierarchical tree in which similar observations sprout from the same branch. In the present paper the hierarchical classification is done on the Principal Components. That is, it is based on the dimensions of the Multiple Correspondence Analysis. Thus, HCPC helps us detect relevant clusters of exemplars based on the MCA and a particular level of explained inertia gained. Moreover, beyond the graphical output, HCPC allows one to consider the most central objects in a cluster, and the objects that belong to a cluster and are placed the furthest from other cluster in the data. That is, using HCPC we can find a cluster, its prototypical exemplar, and the relevant features that provide cue validity.

The following section presents the results of applying these methods on the Infinitival Construction's set of tokens in the corpus.

4. Results and discussion

The Infinitival Complement Construction occurs with 23 types of main verbs, spreading over 664 tokens. The distribution of verbs in the verbal slot of the construction is not even: the two most frequent verbs capture more than 50% of the tokens, and 12 types occur in less than 1% of the data each. This uneven distribution is presented in Table 2.

The most frequent verb in the construction is *ifsher* 'enable' with 182 tokens (182/664 = 27.41%). The fact that the most frequent verb of the construction denotes an enabling event suggests that this is indeed the construction's meaning. The verb *natan* 'give' is the second most frequent verb in the construction, occurring 169 times (25.45%). As a reference for comparison, the verb *natan* 'give' in the Nominal Complement Constructions presented in Section 1 (in which it denotes its lexical meaning of transfer) is the most frequent, and appears 1,034 times

Table 2. Main verb frequency in the Infinitival Complement Construction

Verb type	Tokens	Relative frequency
ifsher, 'enable'	182	27.41%
natan, 'give'	169	25.45%
azar, 'help'	79	11.90%
hicia, 'suggest'	52	7.83%
hirsha, 'allow'	45	6.78%
kara, 'call'	34	5.12%
garam, 'cause'	28	4.22%
amar, 'say'	20	3.01%
isher, 'approve'	12	1.81%
siyea, 'aid'	12	1.81%
hitir, 'allow'	8	1.20%
himlic, 'recommend'	5	0.75%
hora, 'instruct'	5	0.75%
hifria, 'disturb'	3	0.45%
taram, 'donate'	2	0.30%
gamal, 'recompense'	1	0.15%
higid, 'tell'	1	0.15%
hikciv, 'allocate'	1	0.15%
hishir, 'leave'	1	0.15%
hivtiax, 'promise'	1	0.15%
ixel, 'wish'	1	0.15%
shilem, 'pay'	1	0.15%
sider, 'organize, fix'	1	0.15%
Total: 23	664	100.00%

which are 32% of the Nominal Construction's tokens. In the Nominal Construction the distribution is highly skewed; the second most frequent verb appears only 144 times (4.5% of the construction's occurrences).

Taking this distribution into account, a possible hypothesis would be to assume that the Infinitival Complement Construction's category is organized in few clusters, each centred around prototypical exemplars (Bybee and Eddington, 2006). This hypothesis is verified using Multidimensional Scaling (Section 4.1) and Multiple Correspondence Analysis (Section 4.2). Finally, the nature of the clusters is examined using Hierarchical Classification on Principal Components (Section 4.2).

4.1 Multidimensional Scaling analysis of the Infinitival Complement Construction

For the Multidimensional Scaling investigation, I follow the steps presented in Levshina (2012). First, I converted the data into a distance matrix. The average distance between all exemplars is 0.55 (minimum distance 0.00, maximum 1.00). That is, the category does not have a homogeneous structure. Then, the average distance from each exemplar to every other exemplars was calculated. The exemplars that are placed with minimum distance from all the rest are the ones with relatively high Transitivity: An affirmative clause, the Transitive Verb *ifsher*,'enable', a Subject high in Agency and a Dative-marked participant in the third person. The mode of the clause, however, is irrealis in these cases. The complement infinitival verb in these exemplars is related to high Transitivity activities as well: to do something, to finish something, or to present something, for instance. Consider the following example, which has the smallest distance from all other exemplars in the data:

(10) *teafsher* *lo* *lesayem.*
enable to.him to.finish.
'Let him finish.'

This exemplar may well represent the core meaning of the construction, as was defined in Section 1: enabling an action.

The next step was to perform a Multidimensional Scaling analysis on the distance matrix, represented in a two-dimensional map. The points on the map represent the tokens of use. The closer they are to each other, the more similar they are in terms of the coded variables. The Multidimensional Scaling map reveals the heterogeneous structure of the Infinitival Complement Construction's exemplar space (Levshina, 2012). Figure 1 presents the two-dimensional solution of the

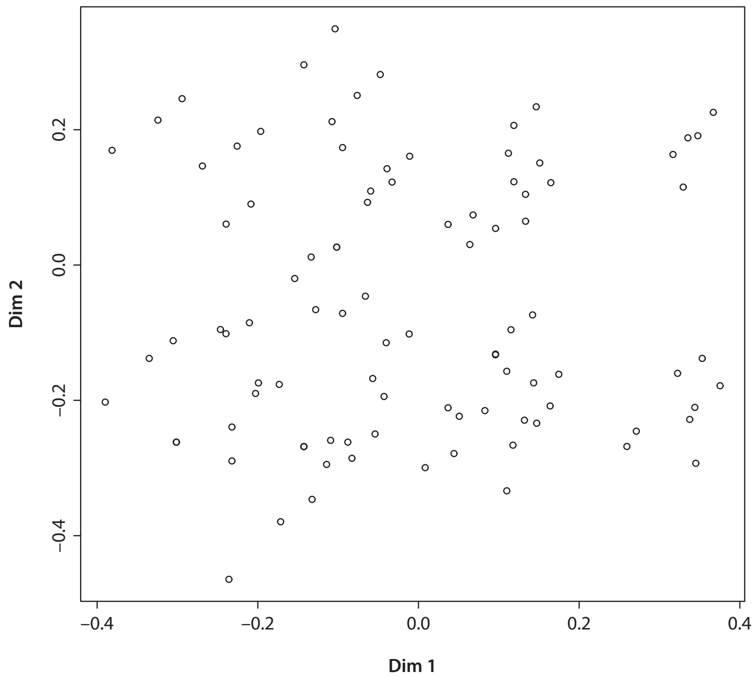

Figure 1. Multidimensional Scaling map: Conceptual space of the Infinitival Construction. Each circle represents a token of use

Multidimensional Scaling analysis. On a first glance, the exemplar space seems to have no specific organization. We can notice several small clusters, but there is no distinguishable cluster that pops out of the map, nor can we see a clear central prototyopical cluster for the category.

However, projecting the data's variables on the map we can see two things. First, in Figure 2, we see that the vertical dimension is correlated with the Mode of the clause: Realis vs. Irrealis. Second, in Figure 3, we see three clusters, correlated with the verbal paradigm of the main verb (annotated as V.BIN in Table 1). The exemplars of the *piel* paradigm appear on the right hand of the map, the exemplars of the *hifil* paradigm appear on the central-bottom part of the map, and the *kal* paradigm appears to be scattered mainly in the top-left part of the map. Note that the fact that two tokens belong to the same verbal paradigm is not enough, by itself, to establish similarity. In order to be placed adjacently on the map, these tokens must share a cluster of categories. That is, beyond the fact that a set of tokens is related to a particular morpho-syntactic paradigm, it seems that it shares other

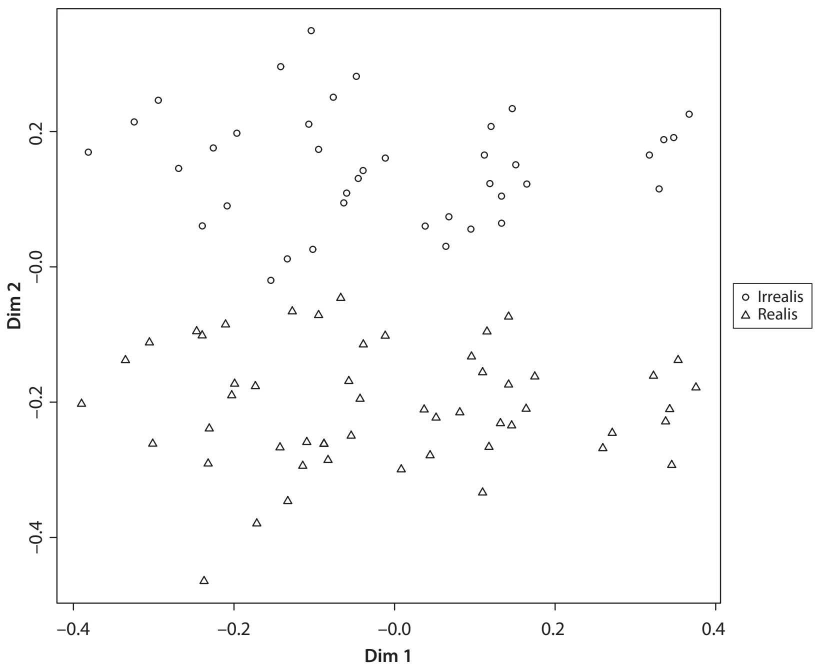

Figure 2. Multidimensional Scaling map: Conceptual space of the Infinitival Construction. Points are labelled according to the token's Mode of the clause

features as well, thus displayed as a close group of points on the map. This suggests that the verbal paradigm is not merely a morpho-syntactic system, but rather a generalization over multiple features from multiple sources of information.

Each of the clusters of tokens in Figure 3 is associated with a different event frame. That is, even though both *hifil* and *piel* belong to the transitive paradigm, we can conclude that their discourse pattern is different: the set of exemplars belonging to each of these paradigms seem to have other parameters in common as well. These paradigm-related clusters, however, are not dense. Rather, they are composed of smaller, denser, clusters. Moreover, the borders between the clusters are not at all clear in the two-dimensional MDS map in Figure 3. Thus, in order to better understand these discourse patterns and clusters, and to place them in a constructional space context, I used Multiple Correspondence Analysis (MCA) and Hierarchical Classification on Principal Components (HCPC).

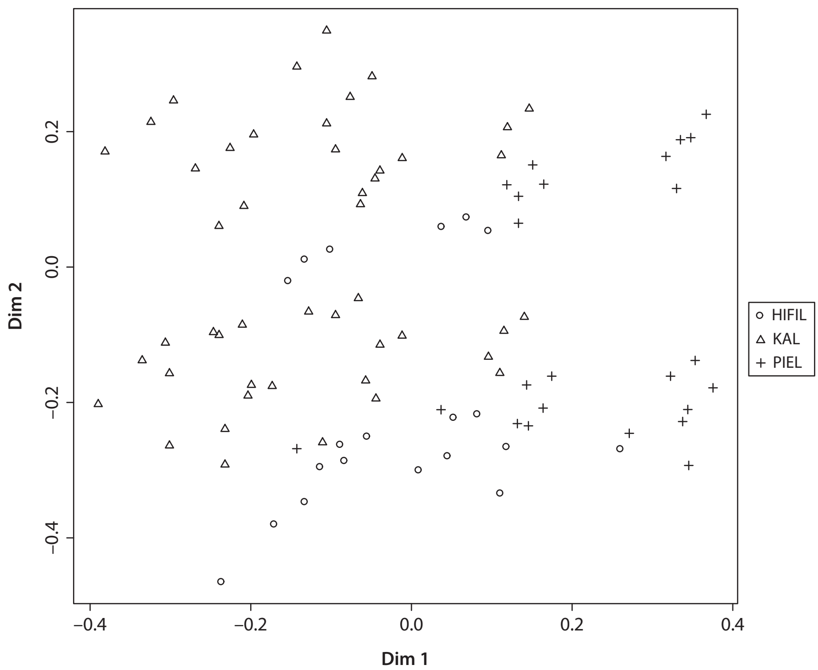

Figure 3. Multidimensional Scaling map: Conceptual space of the Infinitival Construction. Points are labelled according to the Verbal paradigm of the token

4.2 Multiple Correspondence Analysis and Hierarchical Classification on Principal Components

Multiple Correspondence Analysis (MCA) is a tool for uncovering patterns in a database characterized by multiple categorical variables. Thus, the present database is a perfect candidate for such an analysis. Each token of the Infinitival Construction is a row in the original database, and the columns are the various coding parameters presented in Section 3.1. The graphical output aids us in recovering correlations of variables, as well as in discovering closely associated data points (i.e. tokens of use). Figure 4 represents the MCA (computed on the Burt matrix; see Husson et al., 2011), showing associations between categorical variables. The percentage in brackets indicate the percentage of variation in the data explained by the MCA (and see Section 3).

In the top-right quadrant we can see that the verbal paradigm *hifil* is associated with the semantic types of Speaking (e.g. *higid* 'tell'), Deciding (one token, with the verb *hikciv* 'allocate'), and Transitive Motion (e.g. *hish'ir* 'leave') in the realis Mode, with the Dative-marked participant in the second person. The top-left

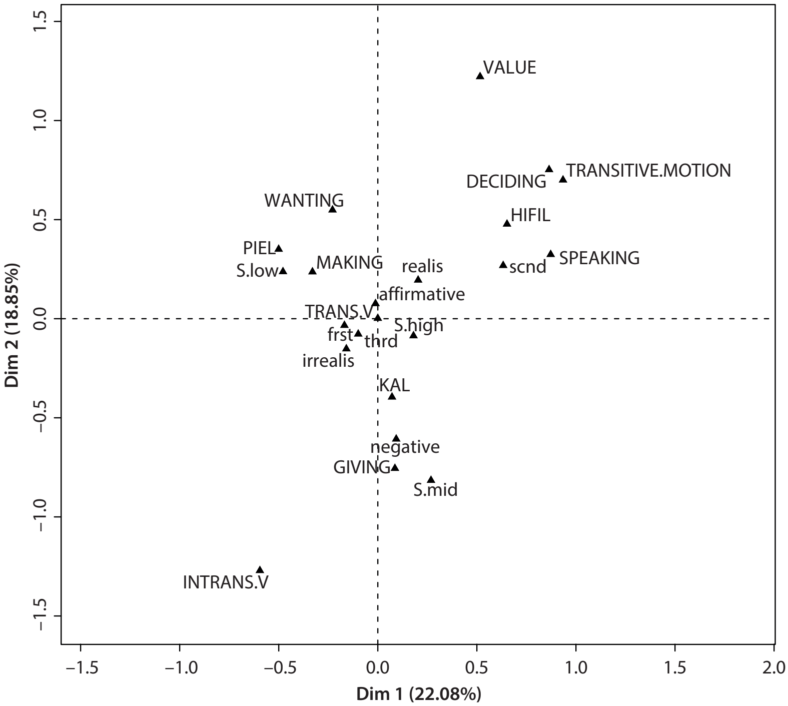

Figure 4. Multiple Correspondence Analysis: associations between categorical variables

quadrant contains the verbal paradigm *piel* and the semantic type of Making with verbs such as *ifsher* 'enable' (and one token of the semantic type Wanting, with the verb *wishing*). We can also see that the *piel* paradigm is associated with low Agentivity (annotated as S.low). While the bottom part of the map is less organized and consists of the verbal paradigm *kal* and a scatter of many features, thus far we can see a clear cut between the verbal paradigms of *piel* and *hifil*: the first is associated with *enabling* scenarios, and the second is associated with *allowing* ones. Moreover, as can be seen from the map, the enabling scenario is associated with less agentive agents than the allowing scenario, while the latter is associated with the realis mode. That is, we can conclude that allowing construals in Hebrew are related to higher transitivity than enabling construals.

An interesting perspective can be drawn from the representation of the data points themselves on the MCA map. This method resembles the MDS map presented above in that the tokens are directly displayed, with distance between data points representing dissimilarity between tokens. This map is presented in Figure 5.

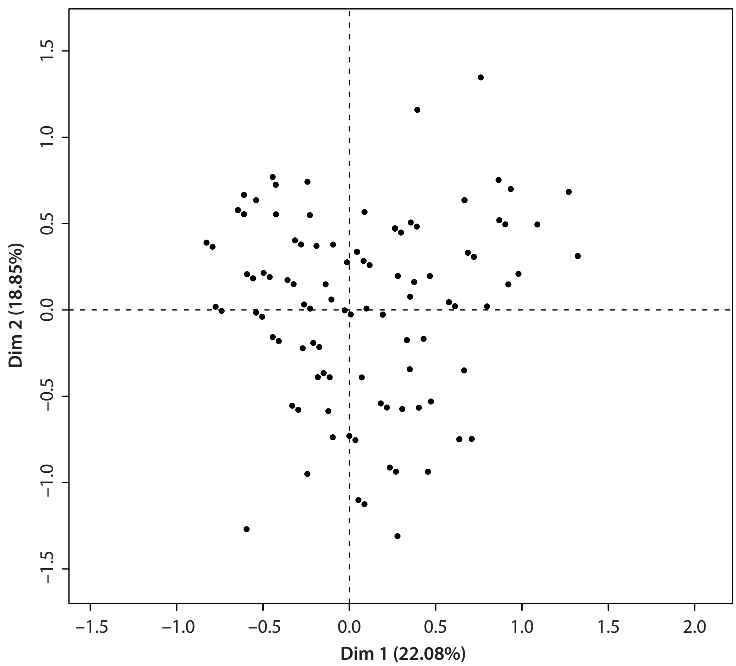

Figure 5. Multiple Correspondence Analysis: associations between tokens

Here we can see a triangle of data-points, resembling the triangle of categories in Figure 4 surrounding the verbal paradigms. Thus, we can hypothesize that the same clusters we saw in the MDS map in Figure 3 are evident here as well. In order to consider data-points relative to their verbal paradigm, the MCA datapoints can be represented according to their respective paradigm. This is shown in Figure 6.

The structure of the data relative to the verbal paradigms is better represented here than in the MDS map. We can see that the triangular map is neatly divided into three sub-parts, each corresponding with a particular verbal paradigm together with other categories as has been analyzed above.

However, the question remains, are these sub-parts correspond to bottom-up built clusters, or is it merely an optical illusion. In order to answer this question, I performed Hierarchical Classification on the Principal Components (HCPC). HCPC has two merits. First, a hierarchical tree is produced, showing the clusters in the data as a graphical display of these clusters on the map. Second, the output of HCPC includes reference to individual data-points as either close to the center of a cluster, or unique to a cluster in that these data-points are located at a maximum distance from other clusters in the data.

First, the graphical output of the HCPC is shown in Figures 7 and 8.

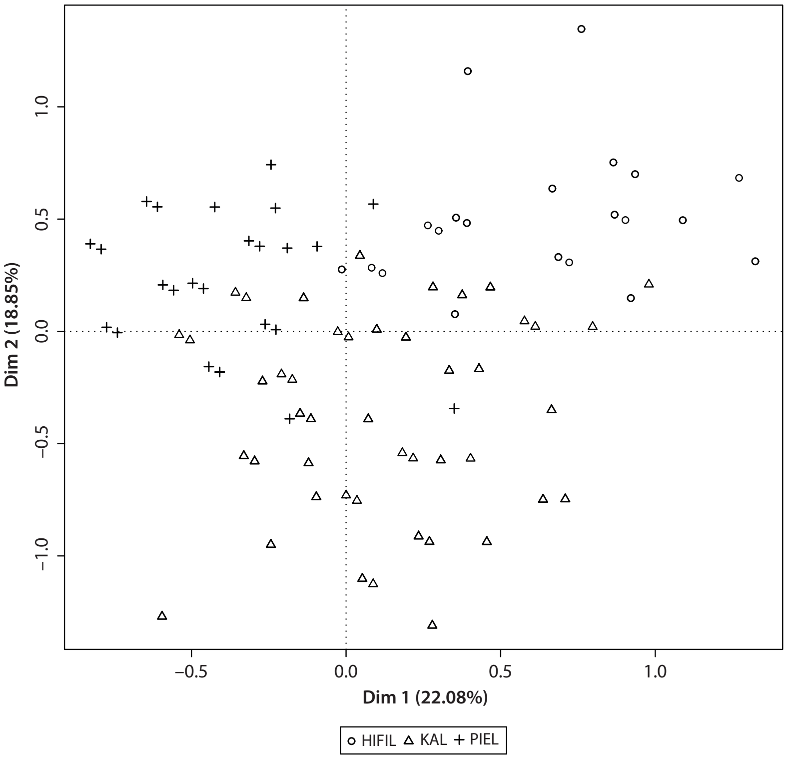

Figure 6. Multiple Correspondence Analysis: associations between tokens, according to verbal paradigm

We can see that the data can be characterized as composed of three clusters, corresponding to the clusters described above on the basis of the verbal paradigm.

As a second step, the numerical output of the HCPC, indicating the strength of the link between a category and a cluster, confirms this interpretation of the map, revealing the construction's conceptual space. Cluster one is related to the *piel* verbal paradigm, the Making semantic type with verbs such as *ifsher* 'enable,' a Subject referent low in Agentivity, and Irrealis mode. Cluster two is linked to the *kal* paradigm, the Giving semantic type with the verb *natan* 'give,' mid-high Agentivity of the Subject referent, and Irrealis mode. Cluster three is linked to the *hifil* paradigm, the Speaking semantic type, Realis mode, and high Agentivity of the Subject referent. We can see, then, that the *piel* paradigm corresponds to lower transitivity parameters compared to the *hifhil* paradigm, with the *kal* paradigm located between the two. Moreover, we can see that enabling construals correspond to lower transitivity (cluster one), than allowing construals (in cluster three).

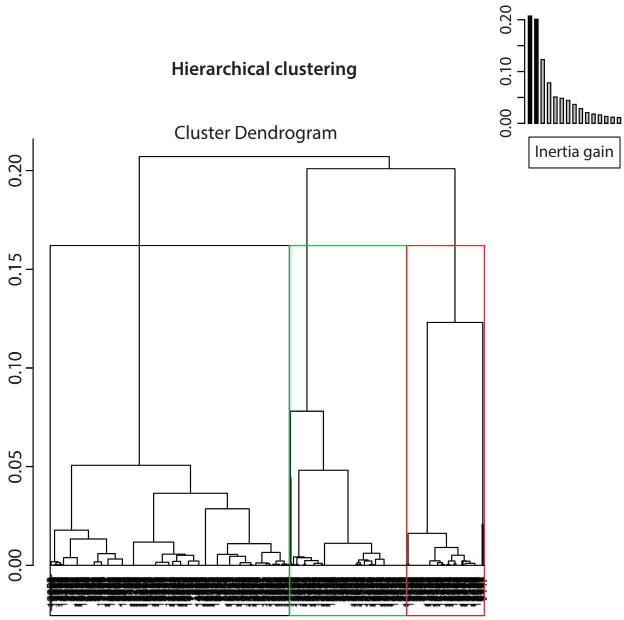

Figure 7. Hierarchical Classification on Principal Components: hierarchical tree

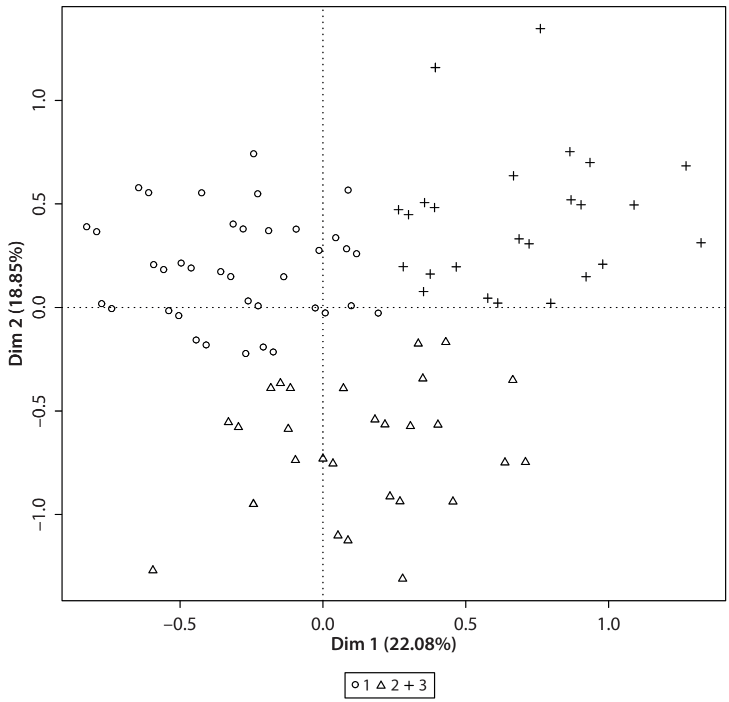

Figure 8. Hierarchical Classification on Principal Components: tokens labeled according to clusters

An interesting question emerges from the HCPC analysis regarding the similarity between allowing construals and 'suggesting an action' construals, both belong to cluster three. The fact that both construals are related to high transitivity parameters suggests that they involve high Agentivity and high degree of affectedness of the Dative-marked participant. However, due to lack of space, I leave this question to future research.

Now that we have laid down the construction's structure and its exemplar space, we can deepen our understanding of the construction. In order to better understand the construction's meaning, let us consider the mixed cluster, cluster 5, and look at the verb *natan* 'give' as a case study. The verb *natan* 'give' belongs to the *kal* paradigm and does not portray enabling or allowing situations in its lexical meaning. However, it is one of the most frequent verbs in all of the Three-Argument Hebrew Dative sub-constructions. Thus, we can learn more about the Infinitival Construction from its behaviour. Looking at the numbers, *natan* 'give' seems to be in between categories: 67% of its tokens in the present construction show mixed transitivity-related features with irrealis on the one side (low transitivity, related to the enabling space), and high Agentivity on the other (high transitivity, related to the allowing space), for instance. Looking closely at the *natan* 'give' exemplars, however, this particular verb aids us in defining the exact difference between enabling and allowing in Hebrew: a difference in transitivity, and particularly, in Agentivity. While allowing scenarios involve a Subject argument with high Agentivity, enabling construals do not present such a demand. Consider the difference between the following examples:

(11) *xashavti she-ulay ba-gvul lo* ***yitnu*** ***li*** ***lacet***.
I.thought that-maybe at.the-border not **they.will.give** **to.me** **to.get.out**.
'I thought that maybe **they won't allow me to cross over** at the border.'

(12) *hi ne'elecet lehaskim le-**tna'im** she-lo* ***notnim*** ***la*** ***lacet***.
she is.forced to.agree to.conditions that-not give to.her to.exit
'She is forced to agree with conditions that **won't let her go.**'

Both (11) and (12) are composed of the same main verb, *natan* 'give', and the same Infinitival complement, *lacet* 'to exit, to go outside, to leave'. However, their interpretation in Hebrew is slightly different. Consider the following paraphrases, replacing *natan* 'give' with either *hirsha* 'allow' (in the (a) sentences) or *ifsher* 'enable' (in the (b) sentences):

(13) a. *xashavti she-ulay ba-gvul lo* ***yarshu***
I.thought that-maybe at.the-border not **they.will.allow**
li lacet.
to.me to.get.out.
'I thought that maybe **they won't allow me to cross over** at the border.'

b. *xashavti she-ulay ba-gvul lo* ***ye'afsheru***
I.thought that-maybe at.the-border not **they.will.enable**
li lacet.
to.me to.get.out.
'I thought that maybe **they wouldn't enable me to cross over** at the border (= I won't be able to cross).'

(14) a. ??*hi ne'elecet lehaskim le-****tna'im*** *she-lo* ***marshim la lacet.***
she is.forced to.agree to.conditions that-not allow to.her to.exit
'She is forced to agree with conditions that won't allow her to go.'

b. *hi ne'elecet lehaskim le-****tna'im*** *she-lo* ***me'afsherim la***
she is.forced to.agree to.conditions that-not enable to.her
lacet.
to.exit
'She is forced to agree with conditions that won't let her go.'

The Subject argument in (11) is characterized with high Agentivity. Thus, the sentence interpretation can accommodate both an allowing construal and an enabling one: (11) can be paraphrased with both *isher* 'allow' (13a) and *ifsher* 'enable' (13b). Conversely, the Subject argument in (12) is characterized with low Agentivity. Thus, the only suitable paraphrase for (12) is the one with the verb *ifsher* 'enable' (14b).

Summing up the discussion about the Infinitival Construction's clusters, we can say that the investigation into the MCA's and HCPC's outputs, and into the behaviour of *natan* 'give' as a case study, teaches us that the structure of the Infinitival Construction's exemplar space reflects differences in interpretation. That is, differences in Transitivity correspond with differences in the verbal paradigm, and with different types of closely related construals. Thus, the degree of Transitivity and the morpho-syntactic verbal paradigm function as organizing principles for the Infinitival Construction's cognitive category, allowing the use of verbs which, lexically, do not depict allowing or enabling events. These verbs, in turn, gain the constructional meaning of the Argument Structure Construction.

5. Conclusion

The present paper aimed at presenting a unique structure of an Argument Structure Construction, showing that lexically unrelated verbs, when used in a particular environment, can gain constructional meaning through similarity in other parameters to other exemplars of the construction. Moreover, similarity was shown to be relevant in different levels, including semantic type of verbs, verbal paradigm (i.e. Binyan) and Transitivity related features such as Agentivity. These levels converge together to create different discourse patterns that correspond with different construals, within the same construction. Only an analysis that assumes rich memory and redundancy in representation can account for the structure of the data presented in this study. An exemplar-based model of categorization allows us to consider each exemplar and its particular features, together with abstracting over clusters of exemplars. Such a bottom-up account of both the local and the global, and the uncovering of patterns in the construction's category structure was made possible using exploratory statistics, specifically Multidimensional Scaling, Multiple Correspondence Analysis, and Hierarchical Clustering on Principal Components.

References

Abbot-Smith, Kirsten & Tomasello, Michael. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23(3): 275–290. DOI: 10.1515/TLR.2006.011

Ariel, Mira. 2008. *Pragmatics and Grammar*. Cambridge: CUP. DOI: 10.1017/CBO9780511791314

Borg, Ingwer & Groenen, Patrick J. F. 2005. *Modern Multidimensional Scaling: Theory and Applications*, 2nd edn. Berlin: Springer.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4): 529–551.

Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: CUP. DOI: 10.1017/CBO9780511750526

Bybee, Joan L. & Eddington, David. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2): 323–355. DOI: 10.1353/lan.2006.0081

Cox, Trevor F. & Cox, Michael A. A. 2001. *Multidimensional Scaling*, 2nd edn. Boca Raton FL: Chapman & Hall.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: OUP. DOI: 10.1093/acprof:oso/9780198299554.001.0001

Dattner, Elitzur. In progress. *Mapping Hebrew Dative Constructions*. PhD dissertation, Tel Aviv University.

Diessel, Holger. 2004. *The Acquisition of Complex Sentences*. Cambridge: CUP. DOI: 10.1017/CBO9780511486531

Fillmore, Charles J., Kay, Paul & O'Connor, Mary Catherine. 1988. Regularity and idiomaticity in grammatical constructions: The case of 'let alone'. *Language* 64(4): 501–538. DOI: 10.2307/414531
Gahl, Susanne & Yu, Alan C. L. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23(3): 213–216. <http://www.reference-global.com/doi/abs/10.1515/TLR.2006.007> DOI: 10.1515/TLR.2006.007
Glynn, Dylan. 2014. Correspondence analysis. Exploring data and identifying patterns. In *Corpus Methods in Cognitive Semantics. Quantitative Studies in Polysemy and Synonymy* [Human Cognitive Processing 43], Dylan Glynn & Justyna A. Robinson, 443–486. Amsterdam: John Benjamins.
Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: University of Chicago Press.
Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7(5): 219–224. DOI: 10.1016/S1364-6613(03)00080-9
Goldberg, Adele E. 2005a. Argument realization: The role of constructions, lexical semantics and discourse factors. In *Construction Grammars. Cognitive Grounding and Theoretical Extensions* [Constructional Approaches to Language 3], Jan-Ola Östman & Mirjam Fried (eds), 17–43. Amsterdam: John Benjamins. DOI: 10.1075/cal.3.03gol
Goldberg, Adele E. 2005b. Constructions, lexical semantics, and the correspondence principle. In *The Syntax of Aspect: Deriving Thematic and Aspectual Interpretation*, Nomi Erteschik-Shir & Tova Rapoport (eds). Oxford: OUP.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.
Goldberg, Adele E. & Jackendoff, Ray. 2004. The English resultative as a family of constructions. *Language* 80(3): 532–568. DOI: 10.1353/lan.2004.0129
Greenacre, Michael. 2010. *Correspondence Analysis in Practice*, 2nd edn. Boca Raton FL: Chapman and Hall/CRC.
Greenacre, Michael & Blasius, Jörg. 2006. *Multiple Correspondence Analysis and Related Methods*. Boca Raton FL: Chapman & Hall/CRC. DOI: 10.1201/9781420011319
Gries, Stefan Thomas. 2009. *Quantitative Corpus Linguistics With R: A Practical Introduction*. New York NY: Routledge. DOI: 10.1515/9783110216042
Hopper, Paul J. & Thompson, Sandra A. 1980. Transitivity in Grammar and Discourse. *Language* 56(2): 251–299. DOI: 10.1353/lan.1980.0017
Husson, François, Lê, Sébastien & Pagès, Jérôme. 2011. *Exploratory Multivariate Analysis by Example Using R*. Boca Raton FL: Chapman & Hall/CRC.
Husson, Francois, Josse, Julie, Lê, Sébastien et al. 2013. FactoMineR: Multivariate *Exploratory Data Analysis and Data Mining with R*, version 1.25 <http://CRAN. R-project.org/package=FactoMineR>
Kay, Paul. 1997. *Words and the Grammar of Context*. Stanford CA: CSLI.
Kay, Paul & Fillmore, Charles J. 1999. Grammatical constructions and linguistic generalizations: The what's X doing Y? construction. *Language* 75(1): 1–33. DOI: 10.2307/417472
Levshina, Natalia. 2012. Comparing constructicons: A usage-based analysis of the causative construction with doen in Netherlandic and Belgian Dutch. *Constructions and Frames* 4(1): 76–101. DOI: 10.1075/cf.4.1.04lev
Nenadic, Oleg & Greenacre, Michael. 2007. Correspondence Analysis in R, with two and three-dimensional graphics: The ca package. *Journal of Statistical Software* 20(3): 1–13. <http://www.jstatsoft.org>

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the Emergence of Linguistic Structure* [Typological Studies in Language 45], Joan L. Bybee & Paul J. Hopper (eds), 137–158. Amsterdam: John Benjamins. DOI: 10.1075/tsl.45.08pie

R Core Team. 2013, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>

Rencher, Alvin. 2012. *Methods of Multivariate Analysis*. Hoboken NJ: Wiley. DOI: 10.1002/9781118391686

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3): 247–274. DOI: 10.1515/TLR.2006.010