

Learning Meaning without Primitives: Typology Predicts Developmental Patterns

Barend Beekhuizen

Leiden University Centre for Linguistics
Leiden University
(b.f.beekhuizen@hum.leidenuniv.nl)

Afsaneh Fazly

Department of Computer Science
University of Toronto
(afsaneh@cs.toronto.edu)

Suzanne Stevenson

Department of Computer Science
University of Toronto
(suzanne@cs.toronto.edu)

Abstract

Does the cognitive naturalness of concepts affect the acquisitional path of meaning? In this paper, we explore the use of crosslinguistically elicited data to approximate cognitive naturalness, following Gentner and Bowerman's (2009) Typological Prevalence Hypothesis. Using the domain of topological spatial relations as a case study, we show how this kind of data allows us to simulate developmental patterns of order of acquisition and overgeneralization in Dutch. This result suggests that the Typological Prevalence Hypothesis can be computationally operationalized and evaluated, that modeling semantic acquisition without hand-coded semantic primitives is possible, and finally, that crosslinguistic data provides a good source of information to do so.

The acquisition of meaning

How does a child acquire the language-specific ways of conceptualizing the world that are necessary to form the meaning constructs of her language? Does non-linguistic maturation of the conceptual system play a central role, or is it language that is in the driver's seat? Gentner and Bowerman (2009) (henceforth: GB) argue that both are true to some extent: language plays a role in constructing subspaces of the conceptual space that constitute the meanings of the language a child learns (Bowerman & Choi, 2001), but the conceptual system is not a blank-slate when language learning commences. Some dimensions of categorization come more natural to the child, being perhaps cognitively more basic, whereas others are harder for the child to grasp (Casasola & Cohen, 2002). Bridging these two insights is GB's Typological Prevalence Hypothesis (TPH):¹

All else being equal, within a given domain, the more frequently a given way of categorizing is found in the languages of the world, the more natural it is for human cognizers, hence the easier it will be for children to learn. (Gentner & Bowerman, 2009, 467)

In this paper, we present a computational model that acquires the meaning of linguistic expressions using crosslinguistically elicited data. The goals of this endeavor are twofold: first, to operationalize the TPH, and second, to explore a novel method for studying meaning without hand-coded semantic primitives. Using only crosslinguistic distributional patterns, the approach is similar to distributional semantic approaches to meaning, which use the textual distribution of linguistic items in a corpus to approximate their meaning, often from a cognitive perspective (e.g. Mitchell & Lapata, 2010; Louwerse, 2011). Rather than using corpus data from a single

language, we use crosslinguistically elicited data for a fixed set of stimulus situations.

Developing methods circumventing the use of manual features is desirable: although we agree with Xu and Kemp (2010) that discovering crosslinguistically valid semantic primitives is indispensable, this approach also has a fundamental methodological problem. The selection of a finite set of discrete primitives is bound to reflect a coder's bias. As long as we are not sure how the coder's culturally informed conceptualization influences coding practices, there is no independent ground truth of what the correct primitives are. Using the same set of stimuli for all informants and letting the elicited data speak for itself largely obviates this problem.²

Using crosslinguistic data furthermore provides a novel basis for computational approaches to meaning, in particular the acquisition of form-meaning pairings. Most acquisitional modeling studies pair utterances with primitives that are derived from the language itself (Fazly, Alishahi, & Stevenson, 2010), taken from resources like WordNet (Alishahi, Fazly, Koehne, & Crocker, 2012), or hand-coded on the basis of video data (Fleischman & Roy, 2005; Beekhuizen, Fazly, Nematzadeh, & Stevenson, 2013). Given the crosslinguistic variation noted in papers like Brown (1994) and Bowerman and Choi (2001), all of these methods are very prone to reflect the subdivisions of the conceptual space English makes. The learner's subdivision of the conceptual hypothesis space is thus already fixed and set to the target language in these approaches. In this study, we show how this can be avoided and more language-neutral primitives can be selected.

We apply our model to the domain of topological spatial relations and simulate GB's findings about the acquisition of prepositions in Dutch: namely that the crosslinguistically common grouping constituting the meaning of the prepositions *op* ('horizontal, stable support', e.g. a cup on a table) and *in* ('containment', e.g. an apple in a bowl) are acquired earlier than and overgeneralized to two prepositions reflecting crosslinguistically less common groupings, *aan* ('tenuous support by being attached by one or more points', e.g. a coat on a rack) and *om* ('encircling (support)', e.g. a ring on a finger). We show how the model is able, first, to learn the extensional meaning of Dutch prepositions reasonably well, and, most importantly, that it simulates the order of acquisition as well as the developmental pattern of overgeneralization GB observed. Finally, we show that this effect is not merely due to the frequencies of the prepositions.

¹Bowerman (1993) first proposed the TPH, but not by that name. See Jakobson (1971) for an analogous idea for phonology.

²Any bias is then moved to the construction of a set of stimuli for which linguistic material is elicited (cf. Lucy, 1997), which is less severe, as the *categorization* of situations is left to the informants.

Typological data as a proxy for meaning

The use of crosslinguistic data to explore the conceptual space onto which languages map their linguistic forms, having its roots in work on color vocabulary (Berlin & Kay, 1969), is a recent technique that has been applied to several semantic phenomena, such as expressions of cutting and breaking events (Majid, Boster, & Bowerman, 2008) and markers of topological space (Levinson, Meira, & The Language and Cognition Group, 2003; Regier, Khetarpal, & Majid, 2013), the phenomenon studied here. In these approaches, informants speaking different languages are asked to describe a fixed set of stimuli. Mathematical techniques for extracting latent information in the data are applied to this data in order to explore the main loci of variation. These techniques show information in the data that would be hard to find ‘by hand’. We use similar techniques, but go beyond their exploratory use, by employing the extracted dimensions directly as a conceptual space within which a learner constructs the semantic categories of her language.

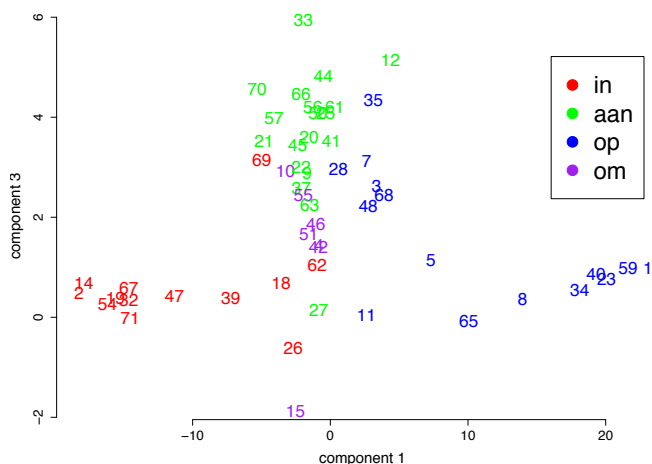
We use the dataset of Levinson et al. (2003), which consists of elicited markers of topological relations for a dataset of 71 pictures of such relations developed by Bowerman and Pederson (1992). The elicitation was done in 9 genetically unrelated languages,³ for a varying number of participants (between 1 and 26) per language, where each participant was asked to label each situation in his or her own language. This dataset constitutes a matrix in which the rows are the 71 situations and each of the 120 columns is a pair of a language and a linguistic marker in that language. The cells are filled with the counts of how many informants used that marker for that situation. We used all responses to the stimuli, including secondary ones, so as to get as much within-language variation as possible. As this matrix contains crosslinguistic counts of situation-marker mappings, the similarity between pairs of situations (on the rows) reflects how often certain situations are described with the same linguistic marker across languages. Following the TPH, this information is thought to reflect the cognitive naturalness of the groupings.

We use Principal Component Analysis (PCA; Hotelling (1933)) to extract underlying dimensions (components) from the data matrix, where each component represents a combination of the dimensions in the original matrix such that crosslinguistic patterns of similarity in the classification of situations will surface.⁴ PCA iteratively extracts a vector of coordinates defining a line in the high-dimensional space of the data matrix (an eigenvector or component) that has the largest variance or eigenvalue given all previously extracted components, until all variance is covered. Applying PCA, we

³Basque, Dutch, Ewe, Lao, Lavukaleve, Tiriyo, Trumai, Yéfi Dnye, Yukatek.

⁴Unlike Levinson et al. (2003), who use dimension-reduction techniques to explore the conceptual space, we do not normalize or filter the data by using only modal responses, because by doing so, we lose information about the within-language variation. Also, by minimizing the number of operations on the data, we minimize possible alternative explanations of the results.

Figure 1: Values on components 1 and 3 for situations with modal responses *in*, *aan*, *op* and *om*. Numbers correspond to the numbering of the situation in Levinson et al. (2003).



extract 70 components, where the first component accounts for 24% of the variance in the 120 dimensions of the original data matrix, the second for 19% and so on.

What does the PCA-transformed space look like? We focus here on the four Dutch prepositions GB studied, viz. *in*, *aan*, *op* and *om*. We define the notion **observed modal response** (to a situation) as the single preposition that is given most often by Dutch-speaking informants when presented with that situation. Figure 1 presents, for each situation for which the observed modal response is one of the four prepositions, the position of that situation on the first and third component.⁵ The colors code the observed modal response.

The PCA-transformed space displays a similar structure to the space Levinson et al. (2003) obtain: component 1 reflects Bowerman and Choi’s (2001) insight that one of the primary conceptual dimensions is a continuum from containment (left) to support (right). The leftmost situations are prototypical containment situations like an apple in a bowl (2) and a box in a bag (14), whereas the rightmost ones instantiate prototypical support situations: a cup on a table (1) and a pencil on a desk (59). Looking at the colors, we can see that situations with *in* as observed modal responses are at the left and those with *op* at the right, whereas the *aan* and *om* situations are in the middle, thus visualizing the prototypicality of the *op* and *in* situations (Bowerman & Choi, 2001; Levinson et al., 2003; Gentner & Bowerman, 2009).

Whereas *in* situations are primarily extended along component 1, *aan* and *om* are mainly extended along component 3. The *aan* category is on top, with the *om* category located below it. Interestingly, we also find instances of *op* with slightly higher values on component 3. That is: parts of the conceptual subspace paired with *op* are located close to the *aan* and *om* situations. All of these *op* situations are not prototypical

⁵Component 2 is not very informative for the four Dutch prepositions: it sets *aan* and *om* situations apart from ones with *in* and *op*, which component 1 already did.

horizontal support relations: we find inscriptions and prints on surfaces (3, 28, 68), a spider on a ceiling (7), a bandaid on a leg (35), and raindrops on a window (48). We will return to the relevance of this observation in the discussion of the experiment. Situations 15 (fence around house) and 27 (apple on tree), finally, are noteworthy because they are outliers for the *aan* and *om* groups. Both stand out in their conceptual groups: for the fence around the house, the figure is adjacent but not in contact with the ground (which all of the other *om* cases are), and an apple on a tree is not just a case of tenuous, mechanistic support, but also an organic, almost meronymic relation, which may have caused greater crosslinguistic variance in the description of this situation.

The transformed matrix reflects the most informative dimensions of the crosslinguistic commonalities in describing situations. By hypothesis, the make up of the space defined by the components represents the cognitive naturalness of groupings: situations that are close to each other constitute more natural categories than situations that are far apart.

Case study: Dutch prepositions

GB discuss the acquisition of (a part of) the Dutch system of prepositions, namely those expressing relations of containment, contact, and support between a figure and a ground. Like English *in*, Dutch *in* expresses containment. Where English *on* covers the full range of contact and support relations, Dutch uses three prepositions: *op* for stable support, *aan* for tenuous support, and *om* for encirclement.

The Dutch subdivision of the conceptual space (with the designated markers *aan* and *om* for less prototypical situations of support) is crosslinguistically rare, GB remark: common linguistic systems include lumping the full space of contact and support under one linguistic item, using general locative markers combined with specialized expressions for prototypical cases of stable support, and expressing the less prototypical cases of support with markers of containment. Because of the rarity of the semantic categories *aan* and *om* signify, they should, according to the TPH, be cognitively less natural, and therefore hard to acquire.

GB make two specific predictions about the Dutch prepositions. First, *op* should be acquired earlier than *aan* and *om*, because the conceptual category to be paired with *op* is crosslinguistically more common than those to be paired with *aan* and *om* (prediction 3). Second, the errors made by Dutch children in encoding relations in the subdomain of contact and support should not be random, but rather involve overgeneralization of *op* towards situations where adult speakers would prefer *aan* and *om*, because the meaning of *op* reflects a cognitively more natural grouping, being a prototypical support relation (prediction 5).

In an experiment, GB asked Dutch and English children (ages 2–6) to describe situations reflecting one of four semantic categories (‘containment’, ‘stable support’, ‘tenuous support’, ‘encirclement’). The target responses were respectively *in*, *op*, *aan* and *om* (Dutch), and *in* and *on* (English). Dutch

Table 1: Overgeneralization patterns for Dutch children (ages 2–6). Based on table 34.2 of Gentner & Bowerman (2009)

children’s response	correct response		
	<i>op</i>	<i>aan</i>	<i>om</i>
no preposition	.12	.23	.18
<i>op</i>	.73	.23	.15
<i>aan</i>	.04	.42	.07
<i>om</i>	.02	.03	.55

children were indeed slower in acquiring *aan* and *om* than *op* (and *in*), and *op* was overgeneralized to situations with *aan* and *om* as target responses. English speaking children acquired *in* and *on* rapidly and correctly. Table 1 presents the overgeneralization results for Dutch (over all ages).

Experiment: Predicting the acquisition pattern

If the TPH is correct, the conceptual space based on crosslinguistically elicited data should reflect cognitively more natural ways of grouping situations and thereby explain GB’s observed behavior, viz. that *op* is acquired earlier and overgeneralized to *aan* and *om*. In this experiment, we train a simple classifier on the PCA-transformed data to see if it does so.

Experimental set-up

As the set of situations is small, and the situations are not deterministically associated with a single Dutch preposition, we generate input on the basis of the PCA-transformed data to train the model. We do so by sampling situation-preposition pairs on the basis of the estimated probability distribution over the joint events w, s , where w is a Dutch preposition,⁶ and s a situation from the set S of 71 situations. The data point has w as its label and the values on the PCA components for situation s as its values. We can decompose the joint probability $P(w, s)$ into $P(w, s) = P(s|w)P(w)$, where:

$$P(s|w) = \frac{N(w, s)}{\sum_{s' \in S} N(w, s')} \quad (1)$$

where $N(w, s)$ is the number of responses with preposition w to situation s . $P(s|w)$ is thus the proportion of responses with preposition w that occurred with situation s . The prior probability $P(w)$ of the preposition is given by an estimate of the relative frequency of its spatial use (Figure 3).⁷

Using this generation method, we can train a Gaussian Naïve Bayes classifier on generated batches of different sizes

⁶Out of the 14 in Levinson et al.’s elicited data. The other ten are *round* ‘around’, *voor* ‘in front of’, *tegen* ‘against’, *binnen* ‘inside of’, *naast* ‘besides’, *door* ‘through’, *boven* ‘above’, *onder* ‘under’, *over* ‘over’, *achter* ‘behind’.

⁷As many prepositions are also used non-spatially, we sampled one hundred cases of each preposition from the child-directed language in the Groningen Corpus (Bol, 1996), annotated whether they were used spatially and as a preposition (rather than, e.g. a verb particle or adverb) and used the proportions of these samples to infer the total amount of spatially used prepositions in the whole corpus.

to simulate the categorization behavior of the learner (henceforth: the model) over time. The model estimates the mean and variance for every preposition on every dimension on the basis of the generated data, and uses a Gaussian distribution defined by the mean and variance to calculate the likelihood, which, together with the preposition’s prior probability (its relative frequency in the generated data) determines the posterior probability of each of the prepositions given the situation.

We evaluate whether the model predicts GB’s two observations by looking at the behavior of the classifier given a variable amount of generated training data. As the data generation is done randomly, we run thirty simulations. Within every simulation, we increment the size of training data with 20 newly generated items at a time, up to the point where it has seen 1000 items, and at each iteration classify each situation on the basis of the datapoints associated with the other 70 situations. This “leave-one-out” method over the situations effectively tests how well the model can generalize to unseen situations. At every iteration of every simulation, and for every situation, this results in a posterior probability distribution over the prepositions. Like we defined the **observed modal response** for the elicited data, we can also define the model’s **expected modal response** to be the preposition with the highest posterior probability given the situation. Comparing the expected and observed modal responses, we can find out what categories are overgeneralized. Apart from evaluating GB’s developmental hypotheses, we would also like to know how well the model approaches adult behavior. We do so by taking, for all situations, the proportion of expected modal responses that is identical to the observed modal response for that situation, which we call the **accuracy**.

The classifier can in principle use all components of the PCA-transformed data. However, later components explain less variance and thereby dilute the probability mass, smoothing out the posterior probability towards the prior probability of the prepositions. To avoid this, we use only the first c components, for a c such that the addition of component $c+1$ does not result in a significant increase in accuracy when the model is trained on 1000 data points. How actual learners determine the salience of each component in the course of acquisition, is a topic we are looking into. Significance is measured, here and elsewhere, with a t -test over the 30 simulations.

Results

Global accuracy With 6 components, the model reaches an optimal accuracy of 0.74, averaged over 30 simulations ($\sigma = 0.03$), that is: after having seen 1000 training items, 74% of the situations are classified with the correct preposition. The ceiling level of performance is 0.94, or $\frac{67}{71}$, as two situations do not have a single observed modal response and can thus not be classified correctly, and another two have an observed modal response preposition that occurs only for that single situation, so that can’t be classified correctly with our training/testing regime. As a simple but informed baseline, we assumed a learner that estimated the probability of the un-

seen preposition on the basis of the corpus frequency of the preposition: it predicts the preposition correctly in 37% of the cases, which is significantly lower than the 74% of the learner (two-sided t -test, $p < .001$).

Developmental pattern Next, we want to see if the model learns *op* correctly before *aan* and *om*, and overgeneralizes the former preposition to situations where the latter two would apply, as observed by GB. We can define four groups of situations based on whether the observed modal response for that situation is one of the four prepositions *in*, *aan*, *om* or *op*. For each of these groups, Figure 2b gives the percentage of expected modal responses for the four prepositions over time.⁸ We can see that for *in* situations and *op* situations, the model reaches an optimal performance very early on. For *aan* and *om* situations, however, we see that it takes the model some time to reach ceiling level, confirming the prediction regarding the order of acquisition.

Moreover, we see that the preposition *op* is overgeneralized to the situations with both *aan* and *om* as well. That is, in situations with observed modal responses of *aan* and *om*, we initially find many expected modal responses with *op*, but not the other way around, thus confirming GB’s prediction regarding the direction of overgeneralization.

Discussion

The model is able to learn the meaning of the Dutch prepositions reasonably well, especially when considering that the model classifies situations it was not trained on, it only had 71 distinct situations to learn from, and, importantly, it had no explicit semantic features for determining the similarity with seen exemplars: the model only relies on information extracted from the crosslinguistic patterns of categorization of these situations. Moreover, the learner shows a developmental pattern conforming to GB’s observations: *op* and *in* are acquired first, within very few iterations, and before *aan* and *om*. *Op* is overgeneralized, mainly to *om* and *aan*, but also, slightly, and against the prediction of the TPH, to *in*, whereas *in*, *aan* and *om* do not spill over to other categories.

How do the methodology and subsequent results in this paper relate to the TPH? We believe that the conceptual space extracted from the crosslinguistic data reflects crosslinguistically common ways of cutting up the domain of topological spatial relations. The prototypes will therefore be at the ends of low-numbered components: it is there that languages maximally agree among each other. Situations with values on the middle of those scales will inevitably be drawn towards both edges and need to have a very strong own ‘profile’ to supersede this attraction. Recall that on component 1, the containment-support continuum, the *op* situations are at one end and the *in* ones at the other. The situations with an observed modal response with *aan* and *om* fall somewhere in the middle on this component, thus being drawn to the attractors on both poles. However, on component 3, *aan* and

⁸Note that none of the stacked areas reaches 1, as some expected modal responses are prepositions other than the target four.

Figure 2: Expected modal responses for the four prepositions over the course of training, summarizing over 30 simulations.

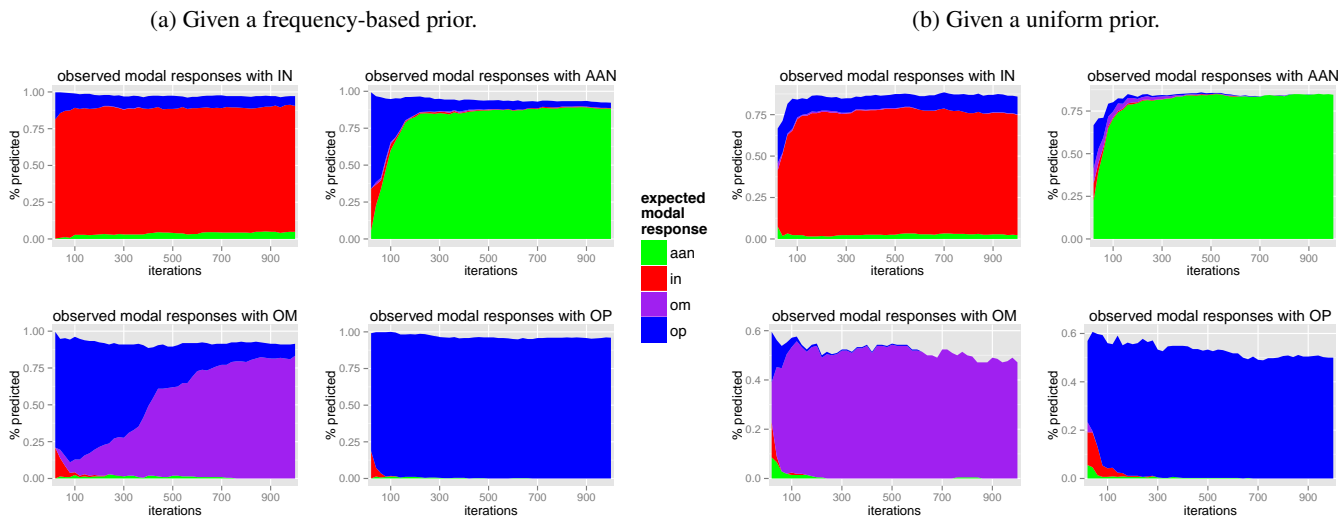
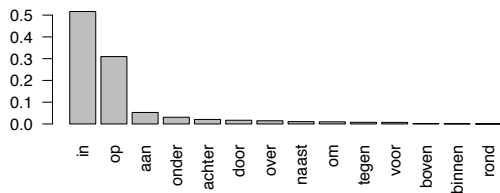


Figure 3: Relative frequency of the spatial uses of the Dutch prepositions in the Groningen corpus



om situations can be differentiated reasonably well from each other and from the other two categories, but not completely, likely because, as can be seen in Figure 1, there are also some *op* situations with values in the range of the *om* and *aan* situations. Because the crosslinguistic frequency of categorization directly shapes the extracted conceptual space, the effects of the make-up of the space we observed can be regarded as an effect of typological prevalence, thus corroborating the Typological Prevalence Hypothesis.

An alternative explanation: Frequency effects

GB also remark that it may be the frequency of the adpositions in the child-directed language causing the developmental pattern. To test this alternative explanation, we also ran the simulation with a different generation regime. Instead of using the relative frequencies from the corpus as the prior probability, we assigned a uniform prior distribution to all prepositions. As can be seen in Figure 3, *op* is far more frequent than *aan* and *om*, and this may also be the cause of the earlier acquisition and overgeneralization.

Using this generation regime, we obtain a global accuracy score (after 1000 training items) of 0.58 ($\sigma = 0.05$), which is significantly lower than the score obtained with the generation regime based on corpus frequency (two-sided *t*-test, $p < 0.001$). However, it remains interesting to see if the pre-

dictions following from the TPH still hold.

Figure 2a gives the expected modal response over time for the four categories. Here, we do not see a strong overgeneralization of *op* to the other two categories, and only a slight undergeneralization of *aan* and *om*, and both *aan* and *om* are quickly learned correctly. This, however, does not mean that the overgeneralization of *op* is only due to frequency: when we generated by corpus frequency, the most frequent preposition, *in*, was hardly overgeneralized to *aan* and *om*. This means that, although the frequency of *op* plays a role, it only does so together with the part of the conceptual space it covers: at the far edge of component 1, as an attractor, and encroaching on the *aan* and *om* situations on component 3. That is: if the meaning of *op* were not as close to that of *aan* and *om*, it would be *in*, the most frequent preposition, being overgeneralized on the basis of its prior probability, and not *op*. Finally, it is hard to completely disentangle a frequency effect and an effect of cognitive naturalness in this case: after all, it is likely that speakers often verbalize frequently occurring topological relations, and that frequently occurring topological relations are more natural *because* of their frequency.

Conclusion

In this paper, we explored the possibility of using crosslinguistic data to learn the meaning of linguistic expressions. To this end, we used Principle Component Analysis to extract the most informative latent dimensions from a dataset of crosslinguistic elicitation of topological spatial relations. We went beyond the exploratory use of techniques like PCA by using the components directly as a space on which a computational classifier can be trained. Without any explicit, contentful semantic features, but only trained on the data from the crosslinguistic elicitation, the model was able to learn the meaning of Dutch topological relational markers reasonably well. More importantly, we showed how this model simulates the developmental patterns of order of acquisition and

overgeneralization in Dutch found by Gentner and Bowerman (2009). This research corroborates their insight that crosslinguistic data can be used to approximate cognitive naturalness. In our case, cognitively more natural groupings will show up close to each other in the PCA-transformed space, and cognitively prototypical situations will, by virtue of being often categorized together and not with other situations, be found on the edge of dimensions.

This result fits in with the optimism about distributional semantic approaches to cognition. We applied the model to one specific, intricate pattern of acquisition, but the approach can, given the right data, be extended to other semantic phenomena. Some care is in order, however. Meaning is more than a multi-label classification task. As Levinson et al. (2003, 488–489) note, the topological relation markers are organized in a system of implicational hierarchies, with some being superordinate to others (e.g. *inside* being a case of *in*). Secondly, compositionality is not so easy to achieve using continuous subspaces of a conceptual space as an approximation of the meaning of a word. This problem becomes especially clear in the case of the topological relation markers, where in languages like Dutch the prepositions interact with positional verbs that express other parts of the topological relation.

Finally, this result is a stepping stone towards more realistic computational models of the acquisition of meaning. Most models use semantic features that are directly or indirectly based on one particular language – a completely unbiased, universal set of features seems impossible to come up with, and even if it were, there would still be the epistemological problem of knowing that it is the correct set. The approach taken in this paper circumvents this problem, only loosely associating the latent structures in the data with features, but essentially letting the crosslinguistic data speak for themselves.

Acknowledgements

We thank Folgert Karsdorp, Elena Tribushinina, Rens Bod, and Arie Verhagen for stimulating discussions and helpful suggestions, and Stephen Levinson and Asifa Majid for courteously making available the data. We gratefully acknowledge the funding of BB through NWO of the Netherlands (322.70.001) and AF and SS through NSERC of Canada, and the Faculty of Arts & Science, University of Toronto.

References

- Alishahi, A., Fazly, A., Koehne, J., & Crocker, M. W. (2012, January). Sentence-based attentional mechanisms in word learning: Evidence from a computational model. *Frontiers in Psychology, 3*, 1–16.
- Beekhuizen, B., Fazly, A., Nematzadeh, A., & Stevenson, S. (2013). Word learning in the wild: What natural data can tell us. In *Proceedings CogSci*.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.
- Bol, G. (1996). Optional subjects in Dutch child language. In C. Koster & F. Wijnen (Eds.), *Proceedings of the Groningen Assembly on Language Acquisition*.
- Bowerman, M. (1993). Typological perspective on language acquisition: Do crosslinguistic patterns predict development. In E. V. Clark (Ed.), *Proceedings of the twenty-fifth annual Child Language Research Forum*.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of semantic categories. In *Language acquisition and conceptual development*. Cambridge Univ. Press.
- Bowerman, M., & Pederson, E. (1992). *Cross-linguistic studies of spatial semantic organization*.
- Brown, P. (1994). The INs and ONs of Tzeltal locative expressions: The semantics of static descriptions of location. *Linguistics, 32*, 743–790.
- Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science, 5*(2), 247–264.
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science, 34*(6), 1017–1063.
- Fleischman, M., & Roy, D. K. (2005). Why verbs are harder to learn than nouns. Initial insights from a computational model of intention recognition in situated word learning. In *Proceedings CogSci*.
- Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others. The typological prevalence hypothesis. In J. Guo et al. (Eds.), *Crosslinguistic approaches to the psychology of language. Research in the tradition of Dan Isaac Slobin*. Psychology Press.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–441, 498–520.
- Jakobson, J. R. (1971). *Studies on child language and aphasia*. Mouton.
- Levinson, S. C., Meira, S., & The Language and Cognition Group. (2003). ‘Natural concepts’ in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language, 79*(3), 485–516.
- Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273–302.
- Lucy, J. A. (1997). The linguistics of ‘color’. In L. Hardin & L. Maffi (Eds.), *Color categories in thought and language*. Cambridge University Press.
- Majid, A., Boster, J. S., & Bowerman, M. (2008). The crosslinguistic categorization of everyday events: a study of cutting and breaking. *Cognition, 109*(2), 235–50.
- Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science, 34*(8), 1388–1429.
- Regier, T., Khetarpal, N., & Majid, A. (2013). Inferring semantic maps. *Linguistic Typology, 17*, 89–105.
- Xu, Y., & Kemp, C. (2010). Constructing spatial concepts from universal primitives. In *Proceedings Cogsci*.