Same same but different:
Type and typicality in a distributional model of complement coercion

Alessandra Zarcone[1]
Universität des Saarlandes
Saarbrücken, Germany

Sebastian Padó[2]
Universität Stuttgart
Stuttgart, Germany

Alessandro Lenci[3]
Università degli Studi di Pisa
Pisa, Italy

[1]zarcone@coli.uni-saarland.de, [2]pado@ims.uni-stuttgart.de,
[3]alessandro.lenci@unipi.it

Abstract

We aim to model the results from a self-paced reading experiment, which tested the effect of semantic type clash and typicality on the processing of German complement coercion. We present two distributional semantic models to test if they can model the effect of both type and typicality in the psycholinguistic study. We show that one of the models, without explicitly representing type information, can account both for the effect of type and typicality in complement coercion.

1 Introduction: Complement Coercion

Complement coercion (*The author began the book* → <u>*reading*</u> *the book*) has been shown to cause an increase in processing cost (Pylkkänen and McElree, 2006; Katsika et al., 2012), which has been ascribed to a *type clash* between an event-selecting verb (*begin*) and an entity-denoting object (*book*). The increase in processing costs is found in comparison with a baseline condition, where the same verb is combined with an event-denoting object (*journey*), which does not trigger a type clash.

A second influence on processing cost is the *thematic fit* or *typicality* of the fillers of the verb's argument slots (Bicknell et al., 2010; Matsuki et al., 2011): high-typicality combinations are processed more quickly than low-typicality ones (*the mechanic checked* <u>*the brakes*</u> */ the spelling*).

Distributional semantic models (DSMs) can successfully model a range of psycholinguistic phenomena, including the effect of typicality on complement coercion (Zarcone et al., 2012). However, they generally do not include a notion of type. Can a DSM account for effects both of type and typicality?

In this paper, we consider experimental results from a study on complement coercion in German that manipulates both type and typicality. We discuss the performance of existing DSMs and a novel DSM combination. We also discuss how type information can be emerge from distributional information.

2 Manipulating Type and Typicality

In a self-paced reading study on German complement coercion (Zarcone et al., in preparation), we have manipulated both type and typicality. The dataset consists of 20 pairs of subjets (S) and aspectual verbs (V). Each pair is combined with four nominal objects (O) in SOV order:

[$_S$ Das **Geburtstagskind**] hat [$_O$ mit den Geschenken / der Feier / der Suppe / der Schicht] [$_V$ **angefangen**].
[$_S$ The **birthday boy**] has [$_O$ with the presents / party / soup / work shift] [$_V$ **begun**].

The objects are: a high-typicality entity (*presents*); a high-typicality event (*party*); a low-typicality entity (*soup*); and a low-typicality event (*work shift*). The low-typicality objects are drawn from the high-typicality objects of other S-V pairs.

The self-paced reading study yielded the following significant effects: (1) an effect of typicality on reading times ($t = 2.28, p = .02$) at the object region (indicating subject-object integration), (2) an effect of object type on reading times ($t = -2.5, p = .01$) at the verb region (the region of the type clash), (3) an interaction of type and thematic fit at the verb region ($t = 2.04, p = .04$). Mean reading times per condition are reported in Table 1. In sum, the study shows that complement coercion involves both type and typicality. Thus, computational models of complement coercion need to account for both.

3 Modeling the Experimental Results

Distributional semantic models (DSMs) represent word meaning as high-dimensional vectors recording co-occurrences with elements of their

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

	Object region *mit den Geschenken* *with the presents*	**Verb region** *angefangen* *began*
high-fit entity	642	819
high-fit event	655	736
low-fit entity	667	802
low-fit event	710	806

Table 1: Mean reading times per condition (in ms) in the self-paced reading study.

usage contexts. Semantic similarity is defined in terms of a vector similarity metric such as cosine.

Distributional Memory (DM, Baroni and Lenci (2010)) is a DSM that includes syntactic knowledge into the word representations. More concretely, the TypeDM version of DM records word-relation-word tuples $\langle w_1 \ \mathtt{r} \ w_2 \rangle$. The tuples are weighted by *Local Mutual Information* (Evert, 2005), which can be employed to model predicate-argument typicality. For example, the weight of ⟨*book* `obj` *read*⟩ is higher than ⟨*label* `obj` *read*⟩, which in turn is higher than ⟨*elephant* `obj` *read*⟩. TypeDM has been shown to be versatile and effective in several semantic tasks, including predicting verb-argument plausibility.

3.1 Complement Coercion and DSMs.

DM has been extended into the Expectation Composition and Update model (ECU, Lenci (2011)), a family of procedures that can be used to predict the typicality of one sentence part given other sentence parts. E.g., to model the typicality at the verb region in a German sentence with SOV word order (e.g. *Das Geburtstagskind hat mit dem Geschenk angefangen / The birthday boy has with the present begun*), ECU determines the thematic fit for the verb given subject and object:

- compute an expectation for the verb given the subject s, as the distribution over verbs v defined by the weights of the tuples $\langle s \ \mathtt{subj} \ v \rangle$
- compute an expectation for the verb given the object o, as the distribution over verbs v defined by the weights of the tuples $\langle o \ \mathtt{obj} \ v \rangle$.

To combine the subject and object expectations, we combine the two distributions component by component, typically either by sum or products. This distribution is then represented in a vector space by computing the centroid or prototype of the vectors of the 20 most expected verbs. Finally, the thematic fit for a verb v given the subject s and the object o is its cosine similarity to the centroid.

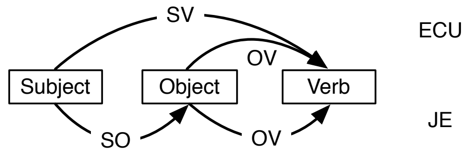

Figure 1: ECU vs. Joint Expectations for the verb

ECU. We call the models following the ECU procedure *SOV+* and *SOV**, depending on their combination operation (sum and product, respectively). Simpler models only consider the influence of subject or object on the verb (*SV* and *OV* respectively), just by leaving out the combination step. These models can successfully account for reading time results on a dataset of complement coercion in German that manipulates typicality but not type (Zarcone et al., 2012).

In order to test ECU on a dataset which manipulates both type and typicality, we evaluate the following ECU models on the complement coercion data in (Zarcone et al., in preparation): *SO* to model effects at the object given the subject; *SOV+*, *SOV** and *OV* to model effects at the verb. We expect these models to account for the typicality effect at the object (1), but not for the type effects at the verb (2,3).

The results are summarized in Table 2 (left and middle). In accordance with our prediction, *SO* correctly yields the typicality effect at the object ($F = 7.38$, $p < 0.01$). Neither *SOV+*, *SOV**, nor *OV* can model the type-typicality interaction at the verb (3). Surprisingly, though, *SOV** and *OV* yield (2), an effect of type at the verb ($F = 5.3228$, $p < 0.05$ and $F = 20.388$, $p < 0.001$, respectively).

Joint Expectations. The reading time study found that the subject-object typicality effects linger at the verb, interacting with type. The main shortcoming of ECU is its inability to model the typicality effects at the verb. This is due to the architecture of the *SOV* models (cf. Fig. 1, top): they compute the expectations for the verb first from the subject (*SV*) and update them with the object's expectations (*OV*). They ignore the interaction between subject and object (SO) – the source of typicality effects (1,3) – corresponding to the assumption that this interaction should only matter at the object. In order to account for this, we draw an analogy to the concept of *joint probability*:

$$P(S, O, V)$$

	non-compos.		*ECU*		*JE*	
	SO	*OV*	*SOV+*	*SOV**	SO+OV	*SO*OV*
(1) effect of typicality at the object region (SO interaction)	✓	×	×	×	✓	✓
(2) effect of type at the verb region (type clash)	×	✓	×	✓	×	✓
(3) type x thematic fit interaction at the verb region	×	×	×	×	×	×

Table 2: Overview of the results of the different DSMs: non-compositional, ECU and JE.

which is equivalent (by the chain rule), to

$$P(S)P(O|S)P(V|O)$$

Treating the first term as a constant prior, we obtain

$$P(O|S)P(V|O)$$

which we can interpret distributionally as motivation to *reweight* the typicality of the verb given the object with the typicality of the object given the subject, thus re-introducing the subject-object interaction into the verb prediction (cf. Figure 1, bottom).

In the *Joint Expectation (JE)* model, the thematic fit score assigned to the target verb is influenced both by the verb's thematic fit with the object (the verb's initial thematic fit score, equivalent to the ECU weight for the ⟨*object* `obj` *verb*⟩ tuple) and by the object's thematic fit with the subject (equivalent to the ECU weight for the ⟨*subject* `verb` *object*⟩ tuple), which in turn is used to reweight the verb's score.

Similar to ECU, there is a choice of combination operations in JE (sum or product). Since JE can be formulated as a simple wrapper around ECU, ECU can be used to compute the individual components (e.g. SO, OV, or more complex ones) and these then just need to be combined additively (SO+OV) or multiplicatively (SO*OV).

The right-hand side of Table 2 shows the results for JE. *SO+OV* yields an effect of typicality ($F = 6.777$, $p < 0.05$) but no effect of type (2) or interaction (3). *SO*OV* yields two main effects of (2) type ($F = 7.2359$, $p < 0.05$) and typicality ($F = 7.2359$, $p < 0.01$), although no interaction (3).

Comparing the two models, we see that ECU *SO* accounts for the results obtained at the object (1), but the *SOV* models cannot explain the interaction with typicality on the verb (2,3). JE (*SO * OV*) models the effects of both type (2) and typicality at the verb, but does not (yet) account for their interaction (3).

4 Discussion: Type and Typicality

We found that the *SO* model successfully accounts for the effect of typicality at the object. This is not surprising: one of the most typical tasks successfully performed by distributional models such as ECU is predicting verb-argument plausibility, and ECU had already been successful in modeling effects of typicality on reading times in German complement coercion (Zarcone et al., 2012).

On the other hand, the ECU *SOV* models were not able to account for the type–typicality interaction at the verb. The JE model (*SO * OV*), which we presented as an alternative to the ECU model to better account for the typicality effects at the verb, yielded effects of both type and typicality at the verb, but did not account for their interaction.

Our most surprising result is that the *OV*, *SOV**, and *SO*OV* models explain the effect of type. As DSMs do not represent this concept explicitly, a possible interpretation suggested by our results is that type and typicality are not distinct categories, but capture properties of predicate-argument combinations at different granularity levels.

Distributional models can account for types because they emerge from the observed corpus distributions. Specifically, for the aspectual verbs used in the present data set, the distribution over their objects – namely that event nouns occur much more frequently that object nouns (Zarcone et al., 2013) – corresponds more naturally to an interpretation in terms of types than of typicality. A compositional distributional model where semantic types emerge as patterns of behavior has the advantage of relying on minimal assumptions regarding the granularity of the type ontology, which is intriguing, as pattern recognition is a key aspect of human cognition (Rumelhart and McClelland, 1987; Saffran et al., 1996; Tomasello, 2009).

In conclusion, the picture that emerges from our experiments is one where (1) expectations for predicate-argument combinations have a hierarchical structure, with types as a high-level distinction and typicality as a low-level distinction, (2) both levels are different, but interact early during processing, influencing reading times, and (3) both type and typicality can emerge from the "same same" distributional model.

Acknowledgments

This research was funded by the German Research Foundation (DFG) as part of SFB 732 "Incremental Specification in Context" and SFB 1102 "Information Density and Linguistic Encoding".

References

Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: a general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Klinton Bicknell, Jeffrey L Elman, Mary Hare, Ken McRae, and Marta Kutas. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63:489–505.

Stefan Evert. 2005. *The statistics of word cooccurrences*. Ph.D. thesis, Universität Stuttgart.

Argyro Katsika, David Braze, Ashwini Deo, and Maria Mercedes Piñango. 2012. Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon*, 7(1):58–76.

Alessandro Lenci. 2011. Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 58–66, Portland, OR.

Kazunaga Matsuki, Tracy Chow, Mary Hare, Jeffrey L Elman, Christoph Scheepers, and Ken McRae. 2011. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4):913–934.

Liina Pylkkänen and Brian McElree. 2006. The syntax-semantics interface: On-line composition of sentence meaning. In M. Traxler and M. A. Gernsbacher, editors, *Handbook of Psycholinguistics*, pages 539–579. Elsevier, Amsterdam, The Netherlands, 2nd edition.

David E. Rumelhart and James L. McClelland. 1987. Learning the past tenses of English verbs. Implicit rules or parallel distributed processing. In *Mechanisms of language acquisition*, pages 249–308. Lawrence Erlbaum Associates, Hillsdale, NJ.

Jenny R Saffran, Richard N Aslin, and Elissa L Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928.

Michael Tomasello. 2009. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press, Cambridge, MA.

Alessandra Zarcone, Jason Utt, and Sebastian Padó. 2012. Modeling covert event retrieval in logical metonymy: probabilistic and distributional accounts. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, pages 70–79, Montréal, Canada.

Alessandra Zarcone, Alessandro Lenci, Sebastian Padó, and Jason Utt. 2013. Fitting, not clashing! a distributional semantic model of logical metonymy. In *Proceedings of the 10th International Conference on Computational Semantics*, Potsdam, Germany.

Alessandra Zarcone, Alessandro Lenci, Ken McRae, and Sebastian Padó. in preparation. Type and thematic fit in logical metonymy resolution.