Vector Semantics: Lecture 8

András Kornai

SZTAKI Computer Science Research Institute

May 22 2024

Summary

- During the semester, we discussed the direct motivation (information theoretical) Kornai, 1996; Kornai, 1998; Kornai, 1999a; Kornai, 2002; Kornai, 2008; Kornai, 2011
- And the strategic/longterm motivation (keeping it down to linear and quadratic terms)
- We also discussed much of the technology (polytopes, graphs-vectors-formulas-automata) VS1.3-6 poly.pdf
- The key innovation: homogeneous formalism that covers both the compositional and the non-compositional cases (direct v. subdirect products) VS2.2
- Algebraic outlook: no search for primitives, circularity in definitions is a fact of life S19:4.5 Kornai, 2010
- Resultant change in outlook: we don't see the vectors/polytopes as the result of an iterative process minimizing some divergence, we have a set of equations we need to solve (even if solved iteratively, this is a very different process) VS9.5

LINGUISTIC TECHNIQUES IN 4lang

- Phonology by string patterns VS2.2
- Morphology by mark and simple pattern substitution
- Syntax by the same means including frames (Fillmore and Atkins, 1998)
- Negation by lack and gen. VS4 dyaneg.pdf No variable binding
- Constructions are big (Nemeskey et al., 2013), we only discussed accusativus cum infinitivo (Kálmán and Kornai, 1985) VS1.5,VS8.3
- We use the absolute minimum of linkers VS2.4 minsem.pdf
- IsA is not a linker, it is for organizing the lexicon
- Temporal issues handled by before,after VS3.2

SEMANTIC COMPUTATION IN 4lang

- Unification (extent of overlap with "merge" remains to be seen, especially as unification is a perfectly well-defined operation, whereas merge is still not, in spite of MCB, MBC, et seq.)
- Coercion (slot filling, "variable binding", but more like equalizers in category theory) Pāṇini-style substitution
- Spreading activation (substitution *salva veritate*) VS7.3
- Relies on schemas (finite automata with temporal behavior) *place, bound, exchange*
- Schemas also used for verbal expression of natural laws S19:3
- Schemas not good for differential equations (current framing of natural laws)
- Somewhat useable for world knowledge VS8.1
- Semantic mechanism includes pragmatics VS9.2 dare.pdf

Links to linguistic theories

- Wave our hands toward autosegmental phonology (Goldsmith, 1976; Goldsmith, 1990; Kornai, 1995)
- Provide formalisation for cognitive linguistics (Jackendoff, Fauconnier, Langacker, Lakoff) and particularly from force dynamics (Talmy, 1988; Talmy, 2000)
- Strong ties to Berkeley Construction Grammar (Fillmore and Kay, 1997)
- Weaker but not nonexistent ties to Unification Grammar (Shieber, 1986) and DUG (Hellwig, 1986)
- Conceptual ties to AMR (Banarescu et al., 2013; Kornai and Kracht, 2015)
- Practical ties to DG (Tesniére, 1959) universaldependencies.org (Recski, 2016; Recski, 2018)
- Deep ties to computational lexicography beginning with CED, keeping LDOCE in focus (Procter, 1978; Boguraev and Briscoe, 1989)

WHAT WE HAVEN'T DISCUSSED

- System of lexical categories (N,V) important for syntax but VS is about semantics. Some discussion in VS2.5
- Adjectives, grading VS7.1-2
- Systematic coverage of semantic fields VS5.3
- Probability VS5.1-2
- Dynamic embeddings S19:2.7, VS8.3
- Modality S19:7.3, VS6
- Other linkers, 3-tensors explored in Makrai (2024)

How about LLMs?

- Connection with vector semantics very strong
- Bridge paper is Smolensky, 1990
- State vectors come in two varieties: stable (eigenvectors of transition matrix), these span the linguistic subspace Little, 1974
- Meaning representation (momentarily knowledge space) uses state matrixes (there are only a dozen matrices treated as primitive)
- Morphology by LLMs explored in Ács (2014)

Banarescu, Laura et al. (2013). "Abstract Meaning Representation for Sembanking". In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 178–186. URL: https://www.aclweb.org/anthology/W13-2322.

Boguraev, Branimir K. and Edward J. Briscoe (1989). *Computational Lexicography for Natural Language Processing*. Longman.

Fillmore, Charles and Sue Atkins (1998). "FrameNet and lexicographic relevance". In: *Proceedings of the First International Conference on Language Resources and Evaluation*. Granada, Spain.

Fillmore, Charles and Paul Kay (1997). *Berkeley Construction Grammar*. URL: http://www.icsi.berkeley.edu/%5C~%7B%7Dkay/bcg/ConGram.html.

Goldsmith, John A. (1976). *Autosegmental Phonology*. PhD thesis MIT.

📄 Goldsmith, John A. (1990). *Autosegmental and Metrical Phonology*. Cambridge, MA: Blackwell.

📄 Hellwig, P. (1986). "Dependency unification grammar". In: *Proceedings of the 11th coference on Computational linguistics*. Association for Computational Linguistics, pp. 195–198.

📄 Kálmán, László and András Kornai (1985). *Pattern matching: a finite state approach to generation and parsing*.

📄 Kornai, András (1995). *Formal Phonology*. New York: Garland Publishing. ISBN: 0815317301.

📄 — (1996). "Analytic models in phonology". In: *Current Trends in Phonology: Models and Methods*. Ed. by J. Durand and B. Laks. Vol. 2. Paris: CNRS, ESRI, Paris X, pp. 395–418.

📄 — (1998). "Quantitative Comparison of Languages". In: *Grammars* 1.2, pp. 155–165.

📄 — (2002). "How many words are there?" In: *Glottometrics* 2.4, pp. 61–86.

📄 Kornai, András (2008). *Mathematical Linguistics*. Advanced Information and Knowledge Processing. Springer. ISBN: 9781846289859. DOI: 10.1007/978-1-84628-986-6.

📄 — (2010). "The algebra of lexical semantics". In: *Proceedings of the 11th Mathematics of Language Workshop*. Ed. by Christian Ebert, Gerhard Jäger, and Jens Michaelis. LNAI 6149. Springer, pp. 174–199. DOI: 10.5555/1886644.1886658.

📄 — (2011). "Probabilistic grammars and languages". In: *Journal of Logic, Language, and Information* 20, pp. 317–328.

📄 — (2018). "Truth or *dare*". In: *Tokens of Meaning: Papers in Honor of Lauri Karttunen*. Ed. by Cleo Condoravdi and Tracy Holloway King, pp. 511–521. URL: http://kornai.com/Drafts/dare.pdf.

📄 — (1999a). "Zipf's law outside the middle range". In: *Proceedings of the Sixth Meeting on Mathematics of Language*. Ed. by J. Rogers. University of Central Florida, pp. 347–356.

📄 Kornai, Andras (2024). "What is the simplest semantics imaginable?" In: *From fieldwork to linguistic theory: A tribute to*

Dan Everett. Ed. by Edward Gibson and Moshe Poliak. Language Science Press, pp. 247–259. URL:
https://langsci-press.org/catalog/book/434.

📄 Kornai, András and Marcus Kracht (July 2015). "Lexical Semantics and Model Theory: Together at Last?" In: *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 14)*. Chicago, IL: Association for Computational Linguistics, pp. 51–61.

📄 Little, W. A. (1974). "The existence of persistent states in the brain". In: *Mathematical Biosciences* 19, pp. 101–120.

📄 Nemeskey, Dávid et al. (2013). "Spreading activation in language understanding". In: *Proceedings of the 9th International Conference on Computer Science and Information Technologies (CSIT 2013)*. Yerevan, Armenia: Springer, pp. 140–143. URL:
https://hlt.bme.hu/media/pdf/nemeskey_2013.pdf.

📄 Procter, Paul (1978). *Longman Dictionary of Contemporary English*. 1st ed. Longman.

📄 Recski, Gábor (2016). "Computational methods in semantics". PhD thesis. Eötvös Loránd University, Budapest. DOI:

📄 10.15476/ELTE.2016.126. URL: http://sandbox.hlt.bme.hu/~recski/to_print/dissz_recski_gabor_nyelvtud.pdf.

📄 Recski, Gábor (2018). "Building concept definitions from explanatory dictionaries". In: *International Journal of Lexicography* 31 (3), pp. 274–311. DOI: 10.1093/ijl/ecx007.

📄 Shieber, Stuart M. (1986). *An Introduction to Unification-Based Approaches to Grammar*. Stanford, CA: CSLI.

📄 Smolensky, Paul (1990). "Tensor product variable binding and the representation of symbolic structures in connectionist systems". In: *Artificial intelligence* 46.1, pp. 159–216. DOI: 10.1016/0004-3702(90)90007-M.

📄 Talmy, L. (1988). "Force dynamics in language and cognition". In: *Cognitive science* 12.1, pp. 49–100. DOI: 10.1207/s15516709cog1201_2.

📄 Talmy, Leonard (2000). *Toward a cognitive semantics*. MIT Press.

📄 Tesniére, Lucien (1959). *Élements de syntaxe structurale*. Paris: Klincksieck.