Competence in lexical semantics

András Kornai[1], Judit Ács, Márton Makrai, Dávid Nemeskey, Katalin Pajkossy, Gábor Recski

[1]Research Institute for Computer Science, Hungarian Academy of Sciences, andras@kornai.com

Problem statement

Characterizing what it means for a speaker to be competent in lexical semantics remains perhaps the most significant stumbling block in reconciling the two main threads of semantics, Chomsky's cognitivism and Montague's formalism. In the paper, three broad families of semantic theories are compared and contrasted from the competence standpoint:

MG Montague Grammar and its lineal descendants, including Discourse Representation Theory (DRT), dynamic predicate logic, etc.

ACR Algebraic Conceptual Representations starting with [Quillian1969], and KL-ONE [Brachman and Schmolze1985, Woods and Schmolze1992], today best represented by Natural Semantic Metalanguage [Wierzbicka1972, Goddard2002], Abstract Meaning Representation [Banarescu et al.2013], Hobbs' system [Gordon et al.2011] and 4lang [Kornai2010, Kornai2012, Recski and Ács2015]

CVS Continuous Vector Semantics [Osgood et al.1975, Schütze1993, Bengio et al.2003, Turian et al.2010, Mikolov et al.2013, Pennington et al.2014]

The standard view [Partee1979, Partee2013] is that psychological reality requires compositionality, otherwise an infinite amount of information would need to be memorized. MG assumes that words are characterized by intensions, formal objects that require an infinite amount of information to specify.

Main questions of competence

- Learnability
- Replicating traditional notions of lexical relatedness such as synonymy, antonymy, ambiguity, polysemy, etc.
- Interface to compositional semantics
- Language-specificity and universality

Basic vocabularies

Two main approaches: [Thorndike1921] based on term and document frequency, and [Ogden1944] based on the ability to define other words. We have inspected the directed graph that can be obtained from any dictionary by taking the words as nodes and adding edges from a to b if a appears in the definition of b. The minimal set required for defining all words is called a Feedback Vertex Set (FVS). Finding the FVS is NP-complete, we used heuristics (dropping rare words first).

Properties of four different dictionaries

Dictionary	#words	FVS
4lang (all senses)	31,192	1,008
4lang (first senses)	3,127	361
LDOCE (all senses)	79,414	1,061
LDOCE (first senses)	34,284	376
CED (all senses)	154,061	6,490
CED (first senses)	80,495	3,435
`en.wiktionary` (all senses)	369,281	2,504
`en.wiktionary` (first senses)	304,029	1,845
formal	2,754	129

MG Learnability of intension/extension – infinite task, huge problem.

ACR Vocabulary learning reduces to learnability of basic vocabulary. The dominant portion of the vocabulary is not connected to direct sensory signals but is learned from context [McKeown and Curtis1987].

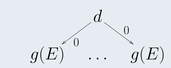
CVS Embeddings are already learned from corpora, but data requirements are currently unrealistically large.

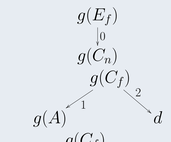
4lang syntax

$D \to E(,E)^*$

$E \to E_n \mid E_f$

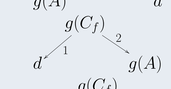
$E \to C_n(E_f)$



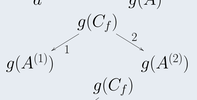


$E_f \to AC_f$

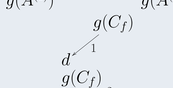
$E_f \to C_f A$



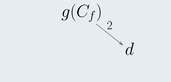
$E_f \to A^{(1)}C_f A^{(2)}$



$E_f \to C_f$'



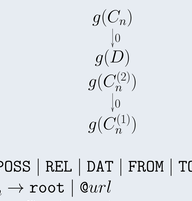
$E_f \to$ 'C_f



$E_f \to E_n$

$E_n \to C_n$

$E_n \to C_n[D]$



$E_n \to C_n^{(1)}(C_n^{(2)})$

$A \to E_n \mid$ `[D]`

$C_n \to$ `AGT` | `PAT` | `POSS` | `REL` | `DAT` | `FROM` | `TO` | `AT`

$C_n \to$ `root` | `@`url

$E \to$ <C>, $A \to$ <A>

$C_n \to$ `[a-z-]+(/[0-9]+)`

$C_f \to$ `[A-Z_]+(/[0-9]+)`

4lang, AMR, and other ACR systems be formulated in terms of hyperedge replacement grammars [Chiang et al.2013].

Lexical relatedness

MG Word similarity =? similarity of meaning postulates. Only a handful of meaning postulates exist, generally restricted to temporals, intensionals, quantifiers i.e. for those words where the underlying logic has the resources [Zimmermann1999].

ACR Similarity of words = similarity of graphs. Used in a competitive SemEval 2015 system [Recski and Ács2015].

CVS Widely used, best results so far.

Interface with compositional semantics

MG Smooth interface except for issues of contextuality, which force as many derivation trees at the disambiguation stage as there are distinct word meanings. With the average English word 3-ways ambiguous, this requires 3^{15} trees for a sentence of length 15.

ACR Built from content words by spreading activation, early disambiguation, good handling of fragmentary input [Nemeskey et al.2013].

CVS Approaches to compositionality broadly investigated, several competing methods, currently with Context Vector Grammars [Socher et al.2013] in the lead.

Deep cases in 4lang

freq	abbr	comment
487	`AGT`	agent
388	`PAT`	patient
34	`DAT`	dative
82	`REL`	root or adpositional object
70	`POSS`	default for relational nouns
20	`TO`	target of action
15	`FROM`	source of action
3	`AT`	location of action

Universality, language-specificity

MG No discussion of deeper universals e.g. in case linking

ACR Universal deep cases, surface realization separated out

CVS Universals may be present in embeddings, but we cannot recover them for now.

Conclusions

MG Hopeless as a theory of lexical competence

ACR Fits well with Pāṇini, generative semantics, and several versions of post-TG syntax such as LFG, RG, RRG, Berkeley Construction Grammar [Goldberg1995]

CVS Farther from direct linguistic/psycholinguistic concerns

References

[1] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.

[2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155.

[3] Ronald J. Brachman and James G. Schmolze. 1985. An overview of the kl-one knowledge representation system*. *Cognitive science*, 9(2):171–216.

[4] David Chiang, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, Bevan Jones, and Kevin Knight. 2013. Parsing graphs with hyperedge replacement grammars. In *ACL*.

[5] Cliff Goddard. 2002. The search for the shared semantic core of all languages. In Cliff Goddard and Anna Wierzbicka, editors, *Meaning and Universal Grammar – Theory and Empirical Findings*, volume 1, pages 5–40. Benjamins.

[6] Adele E. Goldberg. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

[7] Andrew S. Gordon, Jerry R. Hobbs, and Michael T. Cox. 2011. Anthropomorphic self-models for metareasoning agents. In Michael T. Cox and Anita Raja, editors, *Metareasoning: Thinking about Thinking*, pages 295–305. MIT Press.

[8] András Kornai. 2010. The algebra of lexical semantics. In Christian Ebert, Gerhard Jäger, and Jens Michaelis, editors, *Proceedings of the 11th Mathematics of Language Workshop*, LNAI 6149, pages 174–199. Springer.

[9] András Kornai. 2012. Eliminating ditransitives. In Ph. de Groote and M-J Nederhof, editors, *Revised and Selected Papers from the 15th and 16th Formal Grammar Conferences*, LNCS 7395, pages 243–261. Springer.

[10] Margaret G. McKeown and Mary E. Curtis. 1987. *The nature of vocabulary acquisition*. Lawrence Erlbaum Associates.

[11] Tomas Mikolov, Wen-tau Yih, and Zweig Geoffrey. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT-2013*, pages 746–751.

[12] Dávid Nemeskey, Gábor Recski, Márton Makrai, Attila Zséder, and András Kornai. 2013. Spreading activation in language understanding. In *Proc. CSIT 2013*, pages 140–143, Yerevan, Armenia. Springer.

[13] C.K. Ogden. 1944. *Basic English: A General Introduction with Rules and Grammar*. Psyche miniatures: General Series. Kegan Paul, Trench, Trubner.

[14] Charles E. Osgood, William S. May, and Murray S. Miron. 1975. *Cross Cultural Universals of Affective Meaning*. University of Illinois Press.

[15] Barbara H. Partee. 1979. Semantics - mathematics or psychology? In R. Bäuerl, U. Egli, and A. von Stechow, editors, *Semantics from Different Points of View*, pages 1–14. Springer-Verlag, Berlin.

[16] Barbara Partee. 2013. Changing perspectives on the 'mathematics or psychology' question. In *Philosophy Wkshp on "Semantics – Mathematics or Psychology?"*.

[17] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*.

[18] M. Ross Quillian. 1969. The teachable language comprehender. *Communications of the ACM*, 12:459–476.

[19] Gábor Recski and Judit Ács. 2015. Mathlingbudapest: Concept networks for semantic similarity. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 543–547, Denver, Colorado, June. Association for Computational Linguistics.

[20] Hinrich Schütze. 1993. Word space. In SJ Hanson, JD Cowan, and CL Giles, editors, *Advances in Neural Information Processing Systems 5*, pages 895–902. Morgan Kaufmann.

[21] Richard Socher, John Bauer, Christopher D. Manning, and Andrew Y. Ng. 2013. Parsing with compositional vector grammars. In *The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*.

[22] Edward L. Thorndike. 1921. *The teacher's word book*. New York Teachers College, Columbia University.

[23] Joseph Turian, Lev Ratinov, and Yoshua Bengio. 2010. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394. Association for Computational Linguistics.

[24] Anna Wierzbicka. 1972. *Semantic Primitives*. Athenäum, Frankfurt.

[25] William A. Woods and James G. Schmolze. 1992. The KL-one family. *Computers & Mathematics with Applications*, 23(2):133–177.

[26] Thomas E. Zimmermann. 1999. Meaning postulates and the model-theoretic approach to natural language semantics. *Linguistics and Philosophy*, 22:529–561.