Explicit world-knowledge and distributional semantic representations

Asad Sayeed, Ph.D. and Alessandra Zarcone, Ph.D.
Universität des Saarlandes, Fachrichtung 4.7 Allgemeine Linguistik
Postfach 15 11 50, 66041 Saarbrücken, Saarland, Germany
{asayeed,zarcone}@coli.uni-sb.de
http://www.coli.uni-saarland.de/~{asayeed,zarcone}

Course type Advanced

Abstract This is an interdisciplinary course intended to bring together students from psycholinguistic and computational backgrounds and explore the question of world-knowledge in distributional semantics through lectures on recent published research. Distributional semantics exploits co-occurrences in corpus data in order to represent semantic knowledge implicitly through statistics about word context, but the extent to which this can serve as a proxy for semantic grounding in some form of world-knowledge is still an unresolved question. What we currently understand and how to think about the boundary between distributionally-represented knowledge and explicit world-knowledge will be the main topic of the course.

1 Introduction

This advanced course provides students with the opportunity to learn about an issue of increasing importance in semantics: the interaction of pragmatic knowledge with distributional representation. Distributional semantics has become a very successful computational approach to semantic representation, with applications both in typical end-user natural language processing tasks but also in psycholinguistic modeling. Because of the great progress made in distributional semantics, the field is only barely starting to become aware of the practical limits of these representations. This course is designed to give students the requisite background to be able to understand the issue and ask the requisite questions in their own research. Our course will present perspectives from computational linguistics, psycholinguistics, and formal lexical semantics, both foundational and very recent research; our goal is interdisciplinary breadth with a particular focus on class discussion of fundamental semantics problems in this area.

The intuition behind distributional semantics is the distributional hypothesis (Sahlgren, 2008; Rubenstein and Goodenough, 1965): the meaning of a word can be characterized, at least indirectly, by the company it keeps. The advent of high-performance computing that can efficiently process very large volumes of data and evaluate very large statistical models has meant that this hypothesis can be exploited in a way that produces large performance gains on language tasks that previously might have required explicit knowledge (Baroni and Lenci, 2010; Mikolov et al., 2013). The encoding of explicit knowledge, however, is highly costly as well as fragile to the frequent exceptions that characterize real-world language use (Sowa, 1993). Whether or not the distributional hypothesis is "true" in a psychological sense is immaterial for many applications; it is sufficient that a large body of statistics allows us to characterise meaning relations between words in a flexible way.

As we describe in the next sections, a growing body of psycholinguistic evidence is showing, however, that humans *do* to some extent rely on learned expectations of word contexts to characterize (in)congruity and (un)expectedness in meaning juxtaposition. If this is so, this leads to an interesting conundrum: how is semantics—as a layer of human processing—grounded at all in the characteristics of referenced objects and concepts? Is there any necessary role in semantic representation for the encoding of knowledge other than the merely distributional? It may seem obvious to many linguists that there must be, but the apparent success of distributional approaches even in cognitive modeling raises the question of the point at which other forms of semantic knowledge regain psychological necessity.

2 World-knowledge in the lexicon

The linguistics literature has attempted to draw boundaries between lexical knowledge (systematic, amenable to generalization) and world knowledge (situated, culture-dependent), with the former being considered an approachable object of analysis and the latter feared to elude a systematic characterization and analysis. The Generative Lexicon (Pustejovsky, 1995) proposed to enrich the lexicon with complex information (e.g. qualia structure) specifying our "understanding of an object or a relation in the world", in order to account for compositionality and gain in generative power, systematicity and productivity, while establishing the boundaries and constraints of the lexicon through rich but still concise and compact representations. An attempt to keep lexical knowledge (activated first) and world knowledge (activated later in processing) separate has also been made in psycholinguistic research (see Bornkessel and Schlesewsky, 2006; Warren and McConnell, 2007). The key question remains open, however: can we place this distinction on a better empirical footing, rather than dismiss the robust computational characterization of world-knowledge as presently impractical (Jackendoff, 2002; Hobbs, 2009).

3 Psycholinguistic evidence

Recent work generalized event knowledge (Ferretti et al., 2001; McRae and Matsuki, 2009; Bicknell et al., 2010), has challenged this architectural distinction, showing that people use rich syntactic, lexical, semantic, situational and pragmatic information associated with words at each point in processing to build expectations about upcoming input. Work on Grounded Cognition (Barsalou, 1999; Pecher and Zwaan, 2005) has argued that multimodal representations (acquired during sensorimotor experience) are reactivated in the form of simulations during cognitive processes (including language) and should inform our representations of knowledge and experience. Other studies provide evidence for early activation of scenario-relevant objects (Metusalem et al., 2012), object-shape Sato et al. (2013) and perceptuomotor (Amsel et al., 2015) knowledge early during online language processing, and for integration of perceptuomotor features during expectation-based processing of visual scenes (Sitnikova et al., 2008).

Building on such experimental results, Elman (2011) wonders to what extent we can just keep enriching the lexicon with the complex knowledge we associate with words and proposes a "lexical knowledge without a lexicon", moving away from a model of the lexicon in the narrow linguistic sense and towards rich and context-sensitive lexical knowledge stored in memory as a dynamic system. Capturing rich lexical knowledge information is a natural challenge for distributional semantic models (e.g. Padó and Lapata, 2007; Erk, 2010; Lenci, 2012), including, more recently, multimodality and grounded representations, for example by exploiting visual information extracted from images to build distributional and perceptually grounded models of word meaning (e.g. Feng and Lapata, 2010; Bergsma and Goebel, 2011; Bruni et al., 2012). The most recent revisions of the Generative Lexicon (Pustejovsky, 2012, 2013) also strive to enrich the lexicon by imposing structure on a domain (world knowledge and sensomotory experience) which was considered too elusive for rigorous analysis. This may encompass more situated, spatial knowledge, see e.g., the most recent revisions of the Generative Lexicon (Pustejovsky, 2012, 2013), where qualia structures transition into habitats, i.e., frames depicting generalizations about a situation which arise from world-knowledge and on which compositional process can operate, reaching to closer to affordances and to perceptual and motor capacities, which the cognitive scientist would hesitate to call lexical, because it is relevant to other cognitive processes besides language.

4 Computational approaches

Present-day computational approaches to semantics are designed to use large volumes of data to solve the problem of non-robustness in formal representation. Linguistic intuitions are expected to fall out implicitly from an automatically-derived, usually vector-based account of corpus collocations. The quantity of literature on this is now immense and generated constantly in the growing computational literature, but in addition to some of the works cited above, Wang et al. (2016) and Flanigan et al. (2016) are two recent examples out of many works published this year alone on using distributional representations induced via neural networks.

While biologically-inspired "neural" models are the dominant means to construct these representations, the question remains open: how appropriate are these representations to the task of psycholinguistic modeling? Insofar as psycholinguistic models have scientific interest in and of themselves (e.g., Sorodoc et al., 2016) as well as potential applications in areas such as (increasingly realistic and consumer-focused) dialogue systems and natural language generation (e.g., Serban et al., 2016), the boundaries of world-knowledge become increasingly important as an object of semantic study. How far is it true that words that are distributionally similar across large and representative samples of text are always semantically similar? (For example, "knife" and "hammer" may be distributionally similar, being tools, but "knife" and "wire" despite distributional dissimilarity, have real-world features in common, such as the ability to cut certain objects accurately.) This is the type of conflict which students will be encouraged to consider in this course.

References

Amsel, B. D., DeLong, K. A., and Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of memory and language*, 82:118–132.

Baroni, M. and Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–660.

Bergsma, S. and Goebel, R. (2011). Using visual information to predict lexical preference. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 399–405, Hissar, Bulgaria.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., and Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63:489–505.

Bornkessel, I. and Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4):787–821.

Bruni, E., Boleda, G., Baroni, M., and Tran, N.-K. (2012). Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 136–145, Jeju Island, South Korea.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, 6(1):1–33.

Erk, K. (2010). What is word meaning, really? (and how can distributional models help us describe it?). In *Proceedings of the 2010 Workshop on Geometrical Models of Natural Language Semantics*, pages 17–26, Uppsala, Sweden.

Feng, Y. and Lapata, M. (2010). Visual information in semantic representation. In *Proceedings of the 11th Meeting of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, pages 91–99, Los Angeles, CA.

Ferretti, T. R., McRae, K., and Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44:516–547.

Flanigan, J., Dyer, C., Smith, N. A., and Carbonell, J. (2016). Generation from abstract meaning representation using tree transducers. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 731–739, San Diego, California. Association for Computational Linguistics.

Hobbs, J. R. (2009). Word meaning and world knowledge. In Maienborn, C., von Heusinger, K., Portner, P., and van Leusen, N., editors, *Semantics: An International Handbook of Natural Language Meaning*, pages 740–761. Mouton de Gruyte, The Hague.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, Oxford, UK.

Lenci, A. (2012). Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, pages 58–66, Montréal, Canada.

McRae, K. and Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, 3(6):1417–1429.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of memory and language*, 66(4):545–567.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Padó, S. and Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.

Pecher, D. and Zwaan, R. A. (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press, Cambridge, UK.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge, MA.

Pustejovsky, J. (2012). The semantics of functional spaces. In Schalley, A., editor, *Practical Theories and Empirical Practice: Facets of a Complex Interaction*. John Benjamins Publishing, Amsterdam, The Netherlands.

Pustejovsky, J. (2013). Dynamic event structure and habitat theory. In *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013)*, pages 1–10, Pisa, Italy.

Rubenstein, H. and Goodenough, J. B. (1965). Contextual correlates of synonymy. *Commun. ACM*, 8(10):627–633.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, 20(1):33–54.

Sato, M., Schafer, A. J., and Bergen, B. K. (2013). One word at a time: mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, 5(4):345–373.

Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., and Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., and Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of cognitive neuroscience*, 20(11):2037–2057.

Sorodoc, I., Lazaridou, A., Boleda, G., Herbelot, A., Pezzelle, S., and Bernardi, R. (2016). look, some green circles!: Learning to quantify from images. In *Proceedings of the 5th Workshop on Vision and Language at ACL*.

Sowa, J. F. (1993). Building large knowledge-based systems: Representation and inference in the cyc project. *Artificial Intelligence*, 61(1):95 – 104.

Wang, W. Y., Mehdad, Y., Radev, D. R., and Stent, A. (2016). A low-rank approximation approach to learning joint embeddings of news stories and images for timeline summarization. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 58–68, San Diego, California. Association for Computational Linguistics.

Warren, T. and McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14(4):770–775.

Tentative outline

- Day 1: Foundations of representation
  - Historical perspective on formal lexical semantics
  - Scripts and pragmatic knowledge
  - Cognitive foundations of abstract world representation
  - Motivation: finding the boundaries between distributional, formal, and pragmatic knowledge
- Day 2: World-knowledge in the lexicon
  - Generalized event knowledge (e.g. thematic fit)
  - Debates on the contents of the lexicon
  - Qualia and habitats
- Day 3: Distributional semantics
  - Distributional hypothesis: underlying intuitions
  - Corpus-based approaches to building distributional models
  - Applications in natural language processing
  - Hands-on demo of existing computational models
- Day 4: Representational conflicts in models of the lexicon
  - Incongruities between distributional knowledge and world-knowledge in event structure representation
  - Attributive vs. predicative adjectives and other very abstract semantic distinctions
  - Psycholinguistic evidence for distinct processing modalities
- Day 5: Modeling the distinctions
  - Multimodal approaches to feature extraction (image processing, etc).
  - Unexplored experimental avenues
  - Exploitation of human-coded and rule-based processing in distributional contexts
  - Student discussion of research directions and open relevant research questions to be formulated in a published list

Prerequisites

This is an advanced but interdisplinary course intended to provoke thought and discussion between computational linguists and psycholinguists, among others, so the prerequisites cannot be very strict. Encouraged would be previous training in formal and lexical semantics. Exposure to either psycholinguistic experimental methodologies or computational distributional semantics would enhance a student's capacity for participation.

Funding

Organizers are presently located within Germany (Saarbrücken). Some travel support may be available from the Multimodal Computing and Interaction Cluster of Excellence or from the DFG project SFB1102 Information Density and Linguistic Encoding.