CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Representation of sentence meaning
(A JNLE Special Issue)

Ondřej Bojar[1,*], Raffaella Bernardi[2] and Bonnie Webber[3]

[1]Charles University, Faculty of Mathematics and Physics, ÚFAL, Malostranské nám 25, Praha 1, 118 00, Czech Republic, [2]DISI and CIMeC, University of Trento, Via Calepina, 14, 38122 Trento TN, Italy and [3]School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK
*Corresponding author. Email: bojar@ufal.mff.cuni.cz

Abstract
This paper serves as a short overview of the JNLE special issue on representation of the meaning of the sentence, bringing together traditional symbolic and modern continuous approaches. We indicate notable aspects of sentence meaning and their compatibility with the two streams of research and then summarize the papers selected for this special issue.

Keywords: continuous representation of sentences; sentence meaning; vector space

1. Introduction

Sentence meaning has been studied for centuries, offering up representations that reflect properties (or theories) of the syntax–semantic boundary (e.g. Functional Generative Description (Sgall, Hajičová, and Panevová 1986), Meaning-Text Theory (Mel'čuk 1988; Kahane 2003), Abstract Meaning Representation (Banarescu *et al.* 2013, Xue *et al.* 2014) as well as representations with the properties of complex, but expressive logics (e.g. intensional logic)).

Recent years have seen the emergence of a rather novel technical means for capturing sentences: points in highly dimensional continuous spaces, that is, vectors of real numbers. Such representations have been successfully used for the representation of individual *words*, and these colloquially called *word embeddings* serve now in a large portion of applications of Natural Language Processing. In contrast to words, sentences have a much more intricate internal structure, so the corresponding '*sentence embeddings*' are comparably hard to define in a more or less universally acceptable way.

Multiple workshops have explored this area in the past few years, for example, Workshop on Representation Learning for NLP[a] (2016–2018), Workshop on Evaluating Vector Space Representations for NLP[b] (2016–2018), Representation Learning[c] or Dagstuhl seminars.[d] Seeking a way to summarize this work, we put together the current special issue of *Natural Language Engineering*.

2. Structuralist and continuous approaches to meaning

First studies of the meaning of sentences go back to the 6th century BC: Pāṇini's study on the classification of word classes such as nouns and verbs; see, for example, Bod (2013) for a more complete account.

---

[a]https://sites.google.com/site/repl4nlp2019/

[b]https://repeval2019.github.io/

[c]https://simons.berkeley.edu/workshops/machinelearning2017-2

[d]For example, http://www.dagstuhl.de/17042

Table 1. Aspects of meaning handled by symbolic theories and continuous representations. The symbols indicate which aspects are inherently or easily captured ($\checkmark$) in methods of the given approach, hard or impossible to capture ($\times$) or possible to capture in some ways ($\sim$), although not really straightforward or elegant. With '?', we indicate that the aspect could be perhaps handled, but no such approach is readily available or widely known

Aspect of meaning	Symbolic theories	Continuous representations
Abstraction	$\checkmark$	$\times$
Compositionality	$\checkmark$	$\sim$
Learnability	?	$\checkmark$
Relatability (similarity, operations)	$\sim$	$\sim$
Vagueness of meaning	$\times$	$\checkmark$
Ambiguity of expressions	$\checkmark$	$\times$

In linguistics, the study of sentence meaning has been strongly influenced by the approach of structuralists, such as Tesniere and members of the Prague linguistic circle, that divided this study into separate disciplines of phonology, morphology, syntax and semantics. In this *layered view*, as exercised in a number of linguistic theories (e.g. Functional Generative Description, Sgall *et al.* 1986; Meaning-Text Theory, Mel'čuk 1988, Kahane 2003 and others), *meaning* of a deeper layer is defined on the shallower layer below it.

Structuralist theories invariably bring in the concept of *unit of meaning*, similar to the atomicist view in physics, where complex physical objects can be decomposed down to smaller indivisible, units. So too, the layered view opens the possibility for further decomposition: units of the deeper layer correspond to (compounds of) units of the shallower layer.

The structuralist or symbolic view is in sharp contrast with continuous representations of meaning, where the common representation is a fixed-sized vector in a highly dimensional vector space.

2.1 Aspects of meaning

Empirically, a number of *aspects of meaning* have been observed and targeted across many studies stemming from either of the two approaches. Table 1 summarizes the most salient aspects, suggesting the extent to which the particular aspect is captured in symbolic theories and in continuous representations.

Abstraction The notion of *abstraction* is natural within symbolic theories: deeper layers abstract away features that are not important for 'understanding' of the unit, which entails its proper use within expressions of the deeper layer, proper drawing of conclusions based on this more abstract representation of the meaning and others. In terms of the mathematical theory of sets, deeper layers of representations which capture meaning in a more abstract way correspond to coarser partitions of the space of meanings. As a shortcut, we could say that 'meaning is a coarsening', mapping the complex expression in a natural language onto a considerably smaller set of actions demanded or statements about the world.

In a symbolic theory at a given layer of representation, synonymy can be easily defined as the equivalence relation of units that share the same representation at the deeper, more abstract layer. In a continuous representation, the closest approximation of this concept is the similarity of the items in the vector space, but it lacks the capability of drawing hard borders.

Compositionality of meaning is one of the cornerstones of symbolic meaning representation. Idiomatic expressions breach the rules of compositionality, but overall, compositionality is assumed to play a critical role in the language users' ability to understand novel expressions.

In continuous representations, on the other hand, compositionality seems far from being automatically available. Networks of particular structures can embrace compositionality by design, while other may or may not develop hierarchical modelling of meaning in an unsupervised fashion.

Learnability from corpora or various (linguistic) tasks is the key benefit of continuous representations. With symbolic approaches, such learning is possible with complex system architectures and extensive computation – cf. NELL, the never-ending language learner (Mitchell *et al.* 2018) – and thus not in common use.

Relatability (Similarity, Operations): For many applications to work, the representations of the meaning across sentences must be made comparable, which requires a measure be defined across the space of meanings. In other applications, it would be highly useful to be able to reformulate a given sentence to express a desired shift in its meaning. This requires operations that are relevant for the formally captured meanings and how they can be carried out.

Formally, (sentence) embeddings are comparable simply because they are vectors in the same continuous space. However, measures commonly used in vector spaces, for example, the Euclidean or angular distance, need not correspond to the naturally observed types of relations among sentences. The interesting observation by Mikolov, Chen, Corrado, and Dean (2013) that simple arithmetic operations like linear transposition in their word embedding (word2vec) space can correspond to semantically relevant shifts (e.g. '*king*'→'*queen*') will be probably harder to find in the sentence embedding spaces.

We expect that more advanced techniques for relating sentence embeddings will need to be identified, perhaps not satisfying the conditions of a measure in the mathematical sense. We thus prefer to use the more general term 'relatability' to refer to this aspect of the meaning.

Finding a vector space for the meaning of a sentence, in which we can move along various axes (e.g. tense, perspective and politeness) by simple arithmetic operations, would be extremely interesting from several perspectives.

A further desired property of a continuous space of representations could be *self-reference*. In a self-referring space, the vector corresponding to a desired operation would be obtained simply by embedding the *plain language instruction* into the same space. With the word2vec example, embedding the phrase 'the feminine version of' would lead to a vector $\vec{o}$. This vector would be then generally applicable to find the vector of a feminine noun (e.g. '*queen*') given its masculine counterpart: $\text{emb}('king') + \vec{o} \approx \text{emb}('queen')$.

Relatability and operations on symbolic representations seem even less stabilized. Straightforward definitions like edits of deep syntactic trees and their overall number as a measure of similarity are surely possible, but we would hope for a more principled and encompassing repertoire of relevant alternations of sentences. A good starting point could be, for example, verb alternations (Levin 1993). In the end, a self-referring system would be capable of processing books about linguistics, discussing about the language in the language itself.

Ambiguity and vagueness are well studied in the symbolic approaches, with ambiguity corresponding to the situation that one expression has more than one distinct semantic representations and vagueness shown in even very specific and verbose text remaining underspecified in many features observable in the real world. We observe that, unlike in symbolic representations, the ambiguity of expressions is not directly and visibly captured in continuous representations of sentence meaning. Most approaches assume that the 'embedding' function is deterministic and that one expression gets one particular vector representation. While the ambiguity may be implicitly captured in this representation, we are not aware of an easy way of extracting this information from the vector.

On the other hand, vagueness is at least a partially accounted for in continuous representations, simply by the (theoretical) infinite precision in the vector space. By embedding a (naturally vague) expression into the vector space, we can assume that the point we obtained is only one such

possible representative, and that in its close neighbourhood, there are points corresponding to the many possible 'exact contents' of the expression, that is, that we are getting the meaning representation *only approximately*. By including additional information, for example, visual input, we could constrain the vagueness in one way or another.

Symbolic representations provide us with one particular formula encoding the assumed meaning, while simply giving up on vagueness. The fact that the same formula is applicable to many situations is beyond what the formula expresses.

3. Selected papers

The papers in this special issue investigate three main aspects that are under debate in the current research: (a) the capability of end-to-end trained neural networks (NNs) to learn sentence representations with no *a priori* assumption as to the existence or specifics of the interface between syntax and semantics in natural languages; (b) the need to combine NNs with formal structures defined *a priori*, following some theoretical assumptions on language syntax and its interplay with semantics; and (c) the importance of developing explainable models that are transparent in their findings and whose decisions are traceable.

3.1 Unsupervised learning of sentence representations

Maillard, Clark, and Yogatama (2019); Merkx and Frank (2019); Talman, Yli-Jyrä, and Tiedemann (2019) study the extent to which a neural model can learn sentence representations in an end-to-end fashion. In particular, Maillard *et al.* (2019) let the model start from word embeddings to learn syntax and semantic structures jointly through a downstream task and without ever seeing gold standard parse trees. The binary trees learned by the systems differ from those that linguistic theories would argue for, but the model is shown to perform quite well on the downstream task, namely natural language inference. The syntactic ambiguity of the learned representations is only briefly mentioned, but the proposed models clearly have the capacity of learning to model it.

Merkx and Frank (2019) put the emphasis on not having *a priori* assumptions about lexical meaning: the model learns sentence representations by learning to retrieve the captions from images and vice versa. It is then evaluated on Semantic Textual Similarity tasks shown to correlate with human judgement quite well, but it does not reach high performance on the entailment task.

Natural language inference is again taken as the evaluation task in Talman *et al.* (2019), where the authors touch upon another interesting issue, namely transfer learning, studying whether the model can learn generic sentence embeddings that are suitable for other tasks than the one the model has been trained from.

3.2 Applying neural networks within formal linguistic theories

A second track of papers has focused on exploiting powerful neural network models within formal linguistic theories. Chersoni *et al.* (2019) proposes a Structured Distributional Model using Discourse Representation Structures that represent sentences as events. Inspired by psycholinguistic research on the role of semantic memory in sentence comprehension, the authors propose a model that dynamically activates new word embedding lists for each pre-defined syntactic role, while a sentence is being processed.

Karlgren and Kanerva (2019) import into the NNs models ideas from Construction Grammar, extending standard word embedding representations to much higher dimensions so to include syntactic features like tense, the presence of negation, adverbs, but also semantic roles and sequence labels. Aside from clearly illustrating how compositionality could be established in

continuous-space models, they also briefly discuss the empirical capacity of n-dimensional vectors for storing a particular number of feature vectors.

3.3 Building explainable models

Karlgren and Kanerva (2019) also touch on a interesting feature that Croce, Rossini, and Basili (2019) focus upon – namely, the importance of developing explainable models – which was also the focus of the EMNLP Workshop on BlackBox NLP. Alishahi, Chrupała, and Linzen (2019) give an overview of the workshop itself.

Croce *et al.* (2019) propose to exploit kernel spaces to produce an explainable justification of the inferences carried out by the model. The network activates nodes in the space that are used as either positive or negative examples; readable sentences are compiled out of such examples. The authors show the benefit of such linguistic explanations, proposing an evaluation method based on information theory.

4. Conclusion

This short summary of aspects of the meaning of the sentence and overview of the papers that were selected for this special issue are still far from providing a complete and unified picture of the methods for modelling sentence meaning. A complete marriage of symbolic and continuous approaches has yet to happen. Nevertheless, we hope that this special issue contributes to a clearer picture of what users of language might have in their minds when speaking, and what we linguists demand from the corresponding formal representations thereof.

Happy reading and deliberating on the meaning of sentences.

References

Alishahi A., Chrupała G. and Linzen T. (2019). A report on the first BlackboxNLP workshop. *Natural Language Engineering* **25**(4), 543–557.

Banarescu L., Bonial C., Cai S., Georgescu M., Griffitt K., Hermjakob U., Knight K., Koehn P., Palmer M. and Schneider N. (2013). Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop*, Sophia, Bulgaria.

Bod R. (2013). *A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present*. New York: Oxford University Press.

Chersoni E., Santus E., Pannitto L., Lenci A., Blache P. and Huang C.-R. (2019). A structured distributional model of sentence meaning and processing. *Natural Language Engineering* **25**(4), 483–502.

Croce D., Rossini D. and Basili R. (2019). Neural Embeddings: accurate and readable inferences based on semantic kernels. *Natural Language Engineering* **25**(4), 519–541.

Kahane S. (2003). The meaning-text theory. In Agel V., Eichinger L., Eroms H.-W., Hellwig P., Heringer H. and Lobin H. (eds), *Dependency and Valency. An International Handbook of Contemporary Research. 1. Halbband*, vol. 1. Berlin, Boston: De Gruyter Mouton, pp. 546–570.

Karlgren J. and Kanerva P. (2019). High-dimensional distributed semantic spaces for utterances. *Natural Language Engineering* **25**(4), 503–517.

Levin B.C. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.

Maillard J., Clark S. and Yogatama D. (2019). Jointly learning sentence embeddings and syntax with unsupervised Tree-LSTMs. *Natural Language Engineering* **25**(4), 433–449.

Mel'čuk I.A. (1988). *Dependency Syntax - Theory and Practice*. Albany: State University of New York Press.

Merkx D. and Frank S. (2019). Learning semantic representations from visually grounded language without lexical knowledge. *Natural Language Engineering* **25**(4), 451–466.

Mikolov T., Chen K., Corrado G. and Dean J. (2013). Efficient estimation of word representations in vector space. *CoRR abs/1301.3781*.

Mitchell T., Cohen W., Hruschka E., Talukdar P., Yang B., Betteridge J., Carlson A., Dalvi B., Gardner M., Kisiel B., Krishnamurthy J., Lao N., Mazaitis K., Mohamed T., Nakashole N., Platanios E., Ritter A., Samadi M., Settles B., Wang R., Wijaya D., Gupta A., Chen X., Saparov A., Greaves M. and Welling J. (2018). Never-ending learning. *Communications of the ACM* **61**(5), 103–115.

Sgall P., Hajičová E. and Panevová J. (1986). *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Prague, Czech Republic/Dordrecht, Netherlands: Academia/Reidel Publishing Company.

Talman A., Yli-Jyrä A. and Tiedemann J. (2019). Sentence embeddings in NLI with iterative refinement encoders. *Natural Language Engineering* **25**(4), 467–482.

Xue N., Bojar O., Hajič J., Palmer M., Urešová Z. and Zhang X. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In Calzolari N., Choukri K., Declerck T., Loftsson H., Maegaard B. and Mariani J. (eds), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavík, Iceland. European Language Resources Association, pp. 1765–1772.