

Proceedings of the IWCS Workshop *Vector Semantics for Discourse and Dialogue*

Workshop organizers: Mehrnoosh Sadrzadeh, Matthew Purver, Arash Eshghi, Julian Hough, Ruth Kempson and Patrick G. T. Healey.

Vector models of meaning—both strictly distributional models derived directly from co-occurrence statistics, and embeddings learned using representations such as those learnt by a neural network—have revolutionised computational linguistics via their ability to reflect semantic similarities and regularities while providing flexibility to model dynamics and change. However, while there has been much recent interest in extending these models from the level of words to that of larger phrases and sentences, with its own scalability and transparency problems, and there is relatively little progress in understanding how they might apply beyond the sentence: in the realm of discourse and dialogue.

This requires a shift in perspective, moving beyond the static word/sentence view of language as a jigsaw of pieces, to a dynamic perspective seeing language as a set of mechanisms for interaction in real time, encompassing a whole range of actions both sub- and supra-sentential. At a sub-sentential level, dialogue is highly incremental: individuals can interrupt, extend, correct, or request clarification mid-turn, in effect constructing joint utterances without any sense of breakdown in the dialogue exchange. And at the suprasentential level, we have the challenges not only of establishing coreference but of modelling the vast array of speech act effects in dialogue, and rhetorical effects in text discourse, and how they evolve.

The inaugural Vector Semantics for Discourse and Dialogue workshop brings together researchers using vector space methods for distributional semantics, word and sentence embeddings, and dialogue and discourse, to discuss these challenges and fill this gap. We received 13 submissions, each receiving at least two reviews, and all contributions were accepted.

Invited Talks

1 **Gemma Boleda (Universitat Pompeu Fabra) “Talking about you: Deep Learning models of linguistic reference.”**

We use language to talk about things. For instance, within the TV series Friends, “the brother of Monica Geller” can be used to refer to the character Ross Geller. Yet, most computational work on language lacks this connection to the reality that it is about: in Machine Translation, we get for instance “le frère de Monica Geller”, with no link to the person it refers to. Instead, we want to model language in context. Our hypothesis is that jointly learning to represent language and the entities referred to will improve computational models of both. I will report on ongoing research testing this hypothesis with Deep Learning models.

2 **Raquel Fernández (University of Amsterdam) “Analysing Language in Use with Vector Representations.”**

Distributed vector representations have become ubiquitous in computational semantics. Yet their application to aspects related to language interaction and dialogue is still limited. In this talk, I will present recent work on two lines of research connected to this. In the first part, I will focus on lexical meaning within online communities of practice. In the second part, I will zoom into dyadic interaction and present a case study on using Recurrent Neural Networks on synthetic dialogue data to investigate incremental language understanding.

3 **Julie Weeds (University of Sussex) “On Composition vs Contextualisation of Distributed Word Representations: Whats the Point of Sentence Representations?”**

It is now common practice to represent words as points in some high dimensional space where proximity in the space is taken to correlate with similarity in usage and, following the distribution hypothesis (Harris, 1954), often also in meaning. Researchers have also long been interested in how such representations can be combined to form representations of larger units of meaning including phrases, sentences and documents. However, despite a large body of work in this area, at many benchmark evaluation tasks, it has been hard to beat simple models of composition which typically add, or sometimes multiply word representations (Mitchell and Lapata, 2010). There have been recent advances in neural methods (Peters et al., 2018; Devlin et al. 2018) where the representation of a word is first effectively contextualised given the sentence in which it appears. However, typically, the final representation of a sentence is still related to the sum of word, or other intermediate layer, embeddings.

In this talk, I will address the question of whether it makes sense to represent a sentence as a single point in space or whether it is better represented as a collection of contextualised points. The answer is of course application dependent and we must turn to tasks in discourse and dialogue analysis to discover which is the best approach. I will highlight the dangers of reducing sentence meaning to a single point in space and also discuss tasks where contextualisation rather than composition is crucial. I will also ask how we can better contextualise words for use in sentence representations and subsequent discourse and dialogue analysis. I will demonstrate the importance of syntax sensitivity and discuss the growing body of work on dependency-aware embeddings, including our own efforts to create a space which enables syntax-sensitive contextualisation based on the Anchored Packed Tree (APT) framework (Weir et al., 2016).

Contributed Presentations

1 **Jussi Karlgren (Gavagai and KTH Royal Institute of Technology). “High-dimensional distributed semantic spaces for utterances.”**

High-dimensional distributed semantic spaces have proven useful and effective for aggregating and processing visual, auditory, and lexical information for many tasks related to human-generated data. The model proposed here represents both lexical and structural linguistic items in a common framework. This framework is a high-dimensional representation for utterance and text level data based on a mathematically principled and behaviourally plausible approach, which allows the configurations that a lexical item is observed in to be treated similarly to the lexical data. The implementation of the representation is a straightforward extension of Random Indexing models previously used for lexical linguistic items. The model is computationally habitable, and suitable as a bridge between symbolic representations such as dependency analysis and continuous representations such as classifiers or further machine-learning approaches.

2 **Mehrnoosh Sadrzadeh¹, Matthew Purver¹, Gijs Wijnholds¹, Julian Hough¹ and Ruth Kempson² (¹Queen Mary University of London, ²King’s College London). “Incremental Semantic Judgements.”**

We show how one develops a structure-preserving vector space semantic mapping for Dynamic Syntax trees. In order to do so, we use the tensor contraction mechanisms used in compositional distributional semantics, e.g. by Maillard et al. We then implement a word-by-word incremental vector semantics on a verb disambiguation task. The results show that as the context increases, i.e. as we increment the utterances from subject to subject-verb to subject-verb-object, their ambiguous verbs disambiguate better. The results also show that the copy-subject tensor model works best.

3 **Lorenzo Bertolini, Julie Weeds and David Weir (University of Sussex). “CoDE: Learning Composable Dependency Embeddings.”**

Through the years, a restricted number of authors have tackled the non-trivial problem of encoding syntactic information in distributional representations by injecting dependency-relation knowledge directly into word embeddings. Although such representations should bring a clear advantage in complex representations, such as at the phrasal and sentence level, these models have been tested mainly through word-word similarity benchmarks or with rich neural architecture. Outside the embeddings' domain, the APT model has offered an effective resource for modelling compositionality via syntactic contextualization. In this work, we present a novel model, built on top of GloVe, to reduce APT representations to low-dimensionality dense dependency-based vectors, that showcase APT-like composition ability. We then propose a detailed investigation of the nature of these representations, as well as their usefulness and contribution in semantic composition.

4 Bill Noble and Vladislav Maraev (University of Gothenburg). “Neural dialogue act recognition with transformer pre-training.”

BERT, a multi-layer attention-based transformer, uses language model pre-training to achieve state of the art results on a variety of NLP tasks. To assess its potential for dialogue applications, we propose a series of dialogue act recognition experiments with various utterance encoders, including BERT.

5 Ruth Kempson¹, Julian Hough², Christine Howes³, Matthew Purver², Patrick G.T. Healey², Arash Eshghi⁴ and Eleni Gregoromichelaki⁵ (¹King's College London, ²Queen Mary University of London, ³University of Gothenburg, ⁴Heriot-Watt University, ⁵Heinrich Heine Universität Düsseldorf). “Why natural language models must be partial and shifting: a Dynamic Syntax with Vector Space Semantics perspective.”

This paper brings together Dynamic Syntax (DS) and Vector Space semantics (VSS) with current work in cognitive neuroscience and evolution to argue that natural language (NL) models need to be defined in partial and shifting terms. We argue that by combining process-oriented language perspectives from DS and the open-endedness of meaning from VSS, the possibility of developing an integrated account of language behaviour and how it might have emerged becomes a more nearly realisable goal.

6 Henk Zeevat (University of Amsterdam/Heinrich Heine Universität Düsseldorf). “Using neurophysiological features for learning truth-conditional lexical meanings.”

The paper sketches a path under exploration that might lead from flat text to symbolic descriptions of lexical meaning. It offers an alternative to vector based representations by using these same vectors to abduce symbolic representations. The path starts from the observed similarities between the primitives of frame semantics and the semantic features that can be located by neuro-imaging on brain areas in Binder et al (2012). Associations with Binder features can be estimated well from flat text and for each of the Binder features it is possible to find a frame template that explains the association with the Binder feature for a particular class of words. The method seems to perform well on first attempts, but it is obvious that it needs further techniques to become comparable with human knowledge of concepts.

7 Ibai Roman, Roberto Santana, Alexander Mendiburu and Jose A. Lozano (The University of the Basque Country). “Evaluation of sentence embeddings transformations for estimating translation editing effort with Gaussian kernels.”

We focus on a practical NLP regression problem that is related to automatic translation of texts. In this domain, post-editing work is frequently required, and an estimation of the cost of the editing process (in terms of time, effort, and editing distance) is essential. Instead of manually defining the features, as in previous work, our approach relies on sentence embeddings, vector representations of the source and the automatically translated texts, to predict the post-editing effort that leads to the final text. Effort prediction is made by using Gaussian Processes (GP) with different choices of the kernel.

8 Alexis Toumi (University of Oxford). “Discourse Complexity in Categorical Compositional Relational Semantics.”

Categorical compositional distributional (DisCoCat) models give a semantics to sentences by turning their syntactic structure into a linear map for composing vector representations of words. We look at the variant RelCoCat (categorical compositional relational) models, where linear maps are replaced by relations, and show how it yields a Montague semantics for the fragment of natural language expressible in regular logic, which underlies conjunctive queries in database theory. RelCoCat allows to go beyond the sentence boundary into the realm of large-scale discourse: we use tools from query complexity to study the asymptotic behaviour of natural language processing. We show how question answering and text summarisation reduce to conjunctive query containment and minimisation respectively: they are in fact NP-complete. We discuss two ways of taming the complexity of discourse semantics. The first is to restrict the structure of our discourse to tractable fragments of conjunctive queries such as bounded tree-width, which is well-motivated from a cognitive perspective. The second is to consider probabilistic databases and relax question answering to an approximation problem, this yields a principled way of going from Boolean to distributional semantics.

9 Adam Ek, Jean-Philippe Bernardy and Shalom Lappin (University of Gothenburg). “The Influence of Semantic and Syntactic Tags on Sentence Acceptability Judgments.”

In this paper, we investigate the effect of enhancing LSTM language models (LM) with syntactic and semantic tags, whose vectors are combined with the lexical embedding vectors of the words to which they are assigned. We evaluate the effect on language modeling by comparing the enhanced LM’s perplexity to that of a plain LM. Additionally, we use LMs to predict sentence acceptability judgments. The results show that syntactic tags lower the perplexity in LMs while semantic tags increase the perplexity. We also show that neither syntactic or semantic tags improve the sentence acceptability predictions compared to human judgments.

10 Saba Nazir, Mehrnoosh Sadrzadeh, Julian Hough and Patrick G.T. Healey (Queen Mary University of London). “Towards a multimodal vector analysis for verbal and non-verbal data.”

We provide an audio-text vectorial analysis of a toy dataset of contents of BBC programmes and an audio vectorial analysis of laughter with the future goal of fusing the latter with text and employing to derive dialogue semantics.

11 Matthew Purver, Mehrnoosh Sadrzadeh and Julian Hough (Queen Mary University of London). “Vectors Under Discussion.”

We propose a vector-space approach to the Questions Under Discussion (QUD) model of Ginzburg (2012), which promises to fill a gap in recent implementations (Maraev et al, 2018) by providing a direct measure of question resolution while allowing fine-grained encoding of expectations about the answer.

12 Gijs Wijnholds and Mehrnoosh Sadrzadeh (Queen Mary University of London). “Evaluating Composition Models for VP-Elliptical Sentence Embeddings.”

Ellipsis is a natural language phenomenon where part of a sentence is missing and its information must be recovered from its surrounding context, as in “Cats chase dogs and so do foxes.”. Formal semantics has different methods for resolving ellipsis and recovering the missing information, but the problem has not been considered for distributional semantics, where words have vector embeddings and combinations thereof provide embeddings for sentences. In elliptical sentences these combinations go beyond linear as copying of elided information is necessary. In this paper, we develop different models for embedding VP-elliptical sentences. We extend existing verb disambiguation and sentence similarity datasets to ones containing elliptical phrases and evaluate our

models on these datasets for a variety of non-linear combinations and their linear counterparts. We compare results of these compositional models to state of the art holistic sentence encoders. Our results show that non-linear addition and a non-linear tensor-based composition outperform the naive non-compositional baselines and the linear models, and that sentence encoders perform well on sentence similarity, but not on verb disambiguation.

- 13 **Michael Moortgat¹, Mehrnoosh Sadrzadeh² and Gijs Wijnholds²** (¹**Utrecht University,** ²**Queen Mary University of London**). “**A Frobenius Algebraic Analysis for Parasitic Gaps.**”

We provide syntactic types in the language of modal Lambek Calculus and semantic types in vector spaces with Frobenius algebras to be able to present a vector space semantics for parasitic gapping. To our knowledge, this is the first time that parasitic gaps have been treated in a vector semantic setting. We develop string diagrams, category-theoretic Frobenius algebraic types, and normal and closed forms for this phenomena. Experimenting with the results is left to future work.

Programme Committee

Raffaella Bernardi	Ruth Kempson
Arash Eshghi	Matthew Purver
Patrick G. T. Healey	Stephen McGregor
Hannah Gibson	Mehrnoosh Sadrzadeh
Eleni Gregoromichelaki	David Schlangen
Aurelie Herbelot	Ye Tian
Julian Hough	Gijs Wijnholds
Christine Howes	