

GKR: Bridging the gap between symbolic/structural and distributional meaning representations

Aikaterini-Lida Kalouli
University of Konstanz

aikaterini-lida.kalouli@uni-konstanz.de

Richard Crouch
Chegg

{dick.crouch, valeria.depaiwa}@gmail.com

Valeria de Paiva
University of Birmingham

Abstract

Three broad approaches have been attempted to combine distributional and structural/symbolic aspects to construct meaning representations: a) injecting linguistic features into distributional representations, b) injecting distributional features into symbolic representations or c) combining structural and distributional features in the final representation. This work focuses on an example of the third and less studied approach: it extends the Graphical Knowledge Representation (GKR) to include distributional features and proposes a division of semantic labour between the distributional and structural/symbolic features. We propose two extensions of GKR that clearly show this division and empirically test one of the proposals on an NLI dataset with hard compositional pairs.

1 Introduction

Can one combine distributional and structural (symbolic) aspects to construct expressive meaning representations? Three broad approaches have been attempted. First, there is work where linguistic features are used as additional input to systems that create distributional representations, e.g. [Padó and Lapata \(2007\)](#); [Levy and Goldberg \(2014\)](#); [Bowman et al. \(2015b\)](#); [Chen et al. \(2018\)](#). Second, there are approaches where distributional features are used as input to systems that create symbolic representations, e.g. [Banarescu et al. \(2013\)](#); [van Noord et al. \(2018\)](#). Third, and less represented, is the approach attempting to bridge the gap between the other two by combining structural and distributional features in the final representation, e.g. [Lewis and Steedman \(2013\)](#); [Beltagy et al. \(2016\)](#). This paper describes an example of the third approach, and extends the Graphical Knowledge Representation (GKR) ([Kalouli and Crouch, 2018](#)) to include distributional features.

We argue for a division of semantic labour. Distributional features are well suited for dealing with conceptual aspects of the meanings of words, phrases, and sentences, such as semantic similarity, and conceivably hypernym and antonym relations ([Mikolov et al., 2013a](#); [Pennington et al., 2014](#); [Devlin et al., 2018](#)). But they have yet to establish themselves in dealing with Boolean and contextual phenomena like modals, quantifiers, implicatives, or hypotheticals ([Zhu et al., 2018](#); [Dasgupta et al., 2018](#); [Naik et al., 2018](#); [Shwartz and Dagan, 2019](#)). These are phenomena to which more symbolic/structural approaches are well suited. But these approaches have struggled to deal with the more fluid and gradable aspects of conceptual meaning ([Beltagy et al., 2016](#)).

Unlike most symbolic meaning representations, GKR does not attempt to push all aspects of meaning into a single uniform logical notation. Nor does it attempt to push all aspects of meaning into a single vector representation, as most distributional meaning representations do. Instead it allows for the separation of, and controlled interaction between, different levels of meaning. In this respect it borrows heavily from the projection architecture of Lexical Functional Grammar ([Kaplan, 1995](#)), where constituent and functional structure are seen as two separate but related aspects of syntax, each with their own distinct algebraic characteristics. GKR posits a number of distinct layers of semantic structure, the two principal ones being conceptual, predicate-argument structure, and contextual, Boolean structure. This paper discusses how conceptual structure can be enriched with a distributional sub-layer, while still allowing the contextual layer to continue doing the heavy lifting of dealing with modals, quantifiers, booleans, and the like. Our contributions in this paper are three-fold: Firstly, we briefly describe the construction principles of GKR and show why

it is suitable for bridging the gap between structural and distributional approaches. Secondly, we propose two extensions of GKR that allow for the proposed division of semantic labour. Thirdly, we show how one of the proposals can work in practice, by testing it on a subset of the inference dataset of Dasgupta et al. (2018) containing hard compositional pairs.

2 Relevant Work

Symbolic frameworks for meaning representations such as Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), Minimal Recursion Semantics (MRS) (Copestake et al., 2005; Oepen and Lønning, 2006) or Abstract Knowledge Representation (AKR) (Bobrow et al., 2007) were developed with the goal of supporting natural language inference (NLI) and reasoning, and took special care of complex semantic phenomena such as quantification, negation, modality, factivity, etc. More recent meaning representations such as the Abstract Meaning Representation (AMR) (Banarescu et al., 2013) and the Tectogrammatical Representation (TR) from the Prague Dependency Treebank (Hajič et al., 2012), focus more on lexical semantic aspects, such as semantic roles and word senses, on entities and on relations between them. Automatic parsing of text into these different meaning representations has gained great attention, from early, more rule-based systems like Boxer (Bos, 2008) parsing sentences into DRSs, to more recent, statistical or deep learning systems parsing sentences to AMR e.g. (Flanigan et al., 2014; Wang and Xue, 2017; Ballesteros and Al-Onaizan, 2017) or even to DRSs (van Noord et al., 2018). However, to facilitate annotation and parsing, some of the later automated systems have glossed over many of the more complex semantic phenomena. This has raised questions about their expressive power for hard tasks like NLI, as already critiqued for AMR by Bos (2016) and Stabler (2017).

Distributional meaning representations of sentences range from models that compose representations by operating over word embeddings (Mitchell and Lapata, 2010; Mikolov et al., 2013b; Wieting et al., 2016; Pagliardini et al., 2018) to approaches integrating linguistic/structural features into a learning process (Padó and Lapata, 2007; Levy and Goldberg, 2014; Bowman et al., 2015b) to end-to-end neural network architectures like

SkipThoughts (Kiros et al., 2015) and InferSent (Conneau et al., 2017). Already White et al. (2015) and Arora et al. (2017) showed that the more complex architectures do not always outperform simpler vector operations of the former kind, while recently Zhu et al. (2018), Dasgupta et al. (2018) and Naik et al. (2018) argued that current distributional representations fail to capture important aspects of what they call “semantic properties”, “compositionality” or “complex semantic phenomena”, respectively.¹ This was evaluated based on the task of NLI: the researchers created inference pairs requiring complex semantic knowledge and showed that current sentence representations struggle with them. It could be argued that this can be solved by training on data with more instances of such phenomena. But in the absence of the right kinds of annotation in sufficient volumes, this remains an open question.

Fewer approaches have attempted to bridge the gap between the two ends. Lewis and Steedman (2013) attempted to learn a CCG lexicon which maps equivalent words onto the same logical form, e.g. *author* and *write* map to the same logical form. This is done by first mapping words to a deterministic logical form, using a process similar to Boxer, and then clustering predicates based on their arguments as found in a corpus. The resulting lexicon is used to parse new sentences. Beltagy et al. (2016) present a 3-component system that first translates a sentence to a logical form, also based on Boxer, and then integrates distributional information into the logical forms in the form of weights, e.g. the rule “if *x* is grumpy, then there is a chance that *x* is also sad” is weighted by the distributional similarity of the words *grumpy* and *sad*. As a last step, the system draws inferences over the weighted rules using Markov Logic Networks (Richardson and Domingos, 2006), a Statistical Relational Learning (SRL) technique (Getoor and Taskar, 2007) that combines logical and statistical knowledge in one uniform framework, and provides a mechanism for coherent probabilistic inference. Both approaches integrate distribution by clustering or weighting *logical* representations but are still further from the goal to represent the sentence predicate-argument structure as a *distributional* representation suitable for further processing.

¹“Compositionality” is something of a misnomer: basic predicate-argument structure can be compositionally driven by sentence structure.

3 A brief presentation of GKR

The Graphical Knowledge Representation was introduced by Kalouli and Crouch (2018) as a *layered* semantic graph, produced by the open-source semantic parser the researchers make available online.² GKR is inspired by Abstract Knowledge Representation (AKR) (Bobrow et al., 2007), the semantic component of the XLE/LFG framework, which was decoupled from XLE/LFG by Crouch (2014) and then revisited in an explicitly graphical form in Boston et al. (2019). Despite important differences between these approaches, the two main principles are common: first, the sentence information is separated in layers/subgraphs/levels and second, there is a strict separation between the conceptual/predicate-argument structure and the contextual/Boolean structure of the sentence.

These two main principles are exactly how GKR lends itself to the blending of structural/symbolic and distributional features. On the one hand, the separation in layers, analogously to the separation into levels in the LFG architecture (Kaplan, 1995), allows for the formulation of modular linguistic generalizations which govern a given level independently from the others. This explicit organization of information exactly allows for the combination of multiple logics and styles of representations, i.e. structural/linguistic and distributional, and contrasts with the “latent” representations used in end-to-end deep learning approaches to sentence representations and in other graph-based approaches like AMR. On the other hand, the division between conceptual and contextual structure already means that boolean, quantificational, and modal structures do not have to be shoe-horned into predicate argument structures. Likewise, there is no reason to try to shoe-horn boolean, quantification, and modal aspects, or predicate argument structure into a distributional vector. The structures can live alongside one another. This still leaves some latitude for how much predicate-argument and contextual structure needs to be injected into vector representations, depending on the task.

The GKR representation, just like its predecessors, is specifically designed for the task of NLI. But the efficacy of layered graphs has also been shown in dialogue management systems by Shen et al. (2018). Precisely, GKR is a rooted,

node-labelled, edge-labelled, directed graph. It currently consists of five sub-graphs, layered on top of a central conceptual (predicate-argument) sub-graph: a dependency sub-graph, a properties sub-graph, a lexical sub-graph, a coreference sub-graph and a contextual sub-graph.

The dependency graph of GKR is straightforwardly rewritten from the output of the Stanford CoreNLP parser (Chen and Manning, 2014) to fit the GKR format. More precisely, the output is obtained from the enhanced++ dependencies of Schuster and Manning (2016). The conceptual graph is the core of the semantic graph and glues all other sub-graphs together. It contains the basic predicate-argument structure of the sentence: what is talked about; the semantic subject or agent, the semantic object or patient, the modifiers, etc. In other words, this graph expresses the basic propositional content of the utterance and thus already captures the “basic”, predicate-argument compositionality of the sentence. The graph nodes, which correspond to all *content* words of the dependency graph, assert the existence of the *concepts* described by these words, but do not make claims about the existence of *instances* of those concepts. This means that the nodes represent *concepts* and not individuals and given that, no judgments about truth or entailment can be made from this graph. The edges of the graph encode the semantic relationship between the nodes, as this is translated from the dependency label to a more general “semantic” label.

The properties graph associates the conceptual graph with morphological and syntactical features such as the cardinality of nouns, the kind of quantifiers, the verbal tense and aspect, the finiteness of specifiers, etc., so that crucial information required for tasks like NLI is kept in place. For now, this information is gathered from the surface forms and the POS tags provided by CoreNLP in a rule-based fashion. The lexical graph carries the lexical information of the sentence. It associates each node of the conceptual graph with its disambiguated sense and concept, its hypernyms and its hyponyms, making use of the disambiguation algorithm JIGSAW (Basile et al., 2007), WordNet (Fellbaum, 1998) and the knowledge base SUMO (Niles and Pease, 2001). The coreference graph resolves coreference and anaphora phenomena between words of the sentence, based on the output of CoreNLP. The edges of this graph model the

²Available under https://github.com/kkalouli/GKR_semantic_parser

coreferences between the concept nodes.

The contextual graph is also built on top of the conceptual graph and it provides the existential commitments of the sentence: since the conceptual graph only deals with *concepts* and not individuals and thus is incapable on its own to make existential claims and support the attribution of truth and validity, the contextual level is necessary for making such existential commitments and thus support inference. It is also not reducible to some variation of the conceptual layer, because it is exactly this strict separation between the two layers that allows GKR the division of the semantic labour, as it will be shown in the following. The contextual graph introduces a top context (or possible world) which represents whatever the author of the sentence takes the described world to be like; in other words, whatever her “true” world holds, what concepts are instantiated and what are not. Additional contexts can be added, corresponding to any alternative possible worlds introduced in the sentence. Such contexts can be introduced by negation, disjunction, modals, clausal contexts of propositional attitudes (e.g. belief, knowledge, obligation), implicatives and factives, imperatives, questions, conditionals and distributivity. These phenomena are extracted from the sentence in a rule-based manner and their exact conversion into the context graph is defined by a dictionary-like look-up; see [Kalouli and Crouch \(2018\)](#) for more details. This means that the contexts correspond to what we called contextual/Boolean phenomena and what the literature often calls “hard compositionality phenomena”. Each of these embedded contexts makes itself commitments about its own state of affairs, also by stating whether a specific concept is instantiated in it or not. As the logic behind this graph is central to our proposal, we show the conceptual and contextual graph of the sentence *The boy faked the illness*, taken from [Kalouli and Crouch \(2018\)](#), in Figure 1. The conceptual graph in blue contains the concepts involved in the sentence and their semantic relations: there is a concept of faking of a concept of illness by a concept of boy. The contextual graph in grey goes further than this to make commitments about the instances of those concepts. The implicative verb *fake* causes the introduction of an additional context (*ctx(illness)*). The top context has an edge (*ctx_hd*) linking it to its head *fake*, which shows

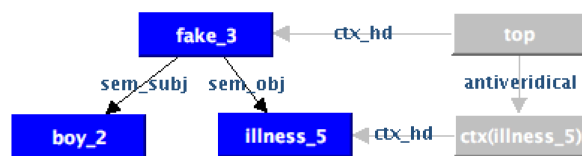


Figure 1: The conceptual graph (left) and the contextual graph (right) of *The boy faked the illness*.

that there is an instance of *faking* in this top context. The top context has a second, *anti-veridical* edge linking it to the context *ctx(illness)* which has *illness* as its head. This head edge asserts that there is an instance of *illness* in this contrary-to-fact context *ctx(illness)*. But since *ctx(illness)* and *top* are linked with an anti-veridical edge, it means that there is no instance of *illness* in the top world which is accurate as the illness was faked.

Similar graphs are produced for sentences with negation, e.g. *The dog is not eating the food*: the concepts of *dog*, *food* and *eating* are included in the conceptual graph and the contextual graph contains a top context linking to the embedded context introduced by the negation. The linking is again through an anti-veridical edge, so that the concept of *eating* is not instantiated in the context *top*. This setting means that negation does not have an impact on the conceptual graph; it is the *contextual* graphs of the positive and negative versions of the sentence that differ. This will prove a very useful feature for our purposes.

An equally useful feature is the treatment of disjunction and conjunction, allowed by the layered nature of GKR. Disjunction and conjunction do have an impact on the conceptual graph. Both introduce an additional *complex* concept that is the combination of the individual disjoined/conjoined concepts (Figure 2, left). The concept graph marks with the edges *is_element* each component concept, of which the complex concept consists (Figure 2, left). However, the difference between conjunction and disjunction is mirrored in the context graph: there, disjunction introduces one additional context for each component of the complex concept (Figure 2, right). These contexts say that in one arm of the disjunct the *walking* concept is instantiated, while in the other arm it is the *driving* concept that is instantiated. The conjunction would instead only contain one top context, in which both concepts are instantiated.

A similar treatment is undertaken for phenomena like modals or quantification. For modals, we

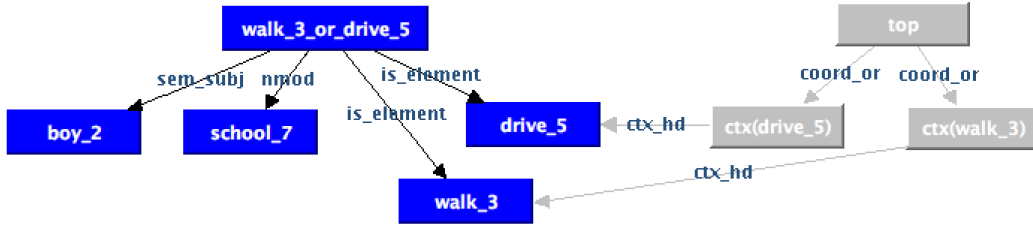


Figure 2: The conceptual graph (left) and the contextual graph (right) of *The boy walked or drove to school*.

can look at the example *Negotiations might prevent the strike* shown in Figure 3. The modal *might* introduces an extra context which is in a “might” relation to top.³ The implicative *prevent* also introduces an extra context in which the concept of *strike* is not instantiated (anti-veridical relation) because in this context the strike does not take place – since in this context the strike was prevented. If we decide to translate *might* to the averidical relation and by transitive instantiability, we can then conclude that the strike is averidical in top, because in the top world we do not know whether there is a strike or not, which is what the modal *might* conveys.

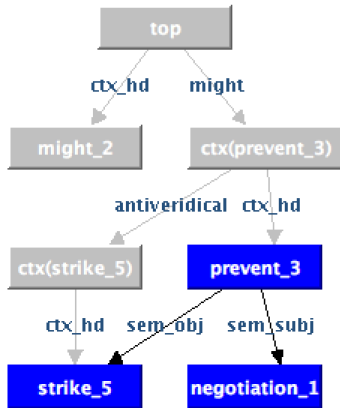


Figure 3: The conceptual graph (bottom) and the contextual graph (top) of *Negotiations might prevent the strike*.

In fact, the interaction between the concept and context graphs implements the “naming” technique of Named Graphs (Carroll et al., 2005), discussed by the creators of GKR in Crouch and Kalouli (2018). A Named Graph, a small extension on top of RDF, associates an extra identifier with a set of triples. For example, a propositional attitude like *Fred believes John does not like Mary*

³We can choose to translate each modal to a specific veridicality relation, e.g. *might* to averidical, but the initial graph makes no such translation so that no crucial information gets lost.

could be represented as follows:

```
:g1 { :john :like :mary }
:g2 :not :g1
:fred :believe :g2
```

where :g1 is the name given to the graph expressing the proposition *John likes Mary*, and :g2 to the graph expressing its negation. But this is also how the context graph works: the contexts are the “names” and the concepts (and their children) associated with them are the “triples” identified by them. For example, in Figure 2, *ctx(drive_5)* is the name given to the subgraph expressing the proposition {boy: drive : school} and *ctx(walk_5)* is the name given to the subgraph expressing the proposition {boy: walk : school}. *top* is the name given to the graph expressing the disjunction between the two contexts *ctx(drive_5)* and *ctx(walk_5)*. This shows how the “basic” predicate-argument compositionality (concept graph) and the “harder” compositionality (context graph) can be kept apart and foreshadows our proposals: the method of factoring out the “harder” compositionality can lead to better performance for both the symbolic/structural and the distributional systems.

For a more detailed discussion of how the distinct graphs are constructed and how other Boolean/contextual cases can be handled, see Kalouli and Crouch (2018).

4 Our proposal for extension of GKR

The two core principles of GKR, i.e. the strict separation of concepts and contexts, with sentence words representing concepts and not individuals, and the modularity and layer separation of the information, allow us to formulate our proposal for a hybrid meaning representation with symbolic/structural and distributional features.

In this section we show how GKR allows for two different ways of combining symbolic/structural and distributional meaning features, each way involving a different degree of the

contribution of each kind of feature and thus being freely select-able based on the needs of the researcher and of the given application. We present these solutions based on the task of NLI, which has been one of the mostly used tasks for the training and evaluation of meaning representations and is the driving force for the design of GKR.

4.1 More symbolic

This proposal is the closest to the original proposal of Kalouli and Crouch (2018) because it only expands the current lexical graph of GKR but keeps all other linguistic structures in place. In that sense, it is more symbolic/structural than it is distributional: it exploits the distributional strengths for the conceptual meaning of the words but builds both the “basic” (predicate-argument) compositionality as well as the “harder” compositionality phenomena in a symbolic/structural way.

The current GKR lexical graph connects its nodes to hand-curated resources like WordNet and SUMO but it could easily be expanded to also contain links to word embeddings. Given the great success of contextualized word embeddings like ELMo (Peters et al., 2018) and BERT (Devlin et al., 2018), it is promising to expand the graph with such embeddings. With this, each concept node would be further connected to its contextualized word embedding. These contextualized word embeddings can be calculated based on the sentence which is currently modelled or, in the case of NLI, based on both sentences of the pair for a more accurate context.

With such an expanded lexical graph in place, we can proceed to do inference in a similar fashion as the one originally proposed by Kalouli and Crouch (2018): each sentence of the pair is parsed into a GKR graph and then the concepts of the two graphs are matched through specificity relations like the ones proposed in Natural Logic systems (cf. MacCartney and Manning (2007) and Crouch and King (2007)), e.g. that *dog* of the premise is a subclass of *animal* of the hypothesis. So far these relations can only be established based on the human-curated resources, which means that some relations will fail to be captured either because they do not exist in the resources or because the strict, logic-based resources do not allow their associations. For example, as discussed in Kalouli et al. (2018), for a pair like $A = \textit{The dog is catching a black frisbee}$

$B = \textit{The dog is biting a black frisbee}$, the words *catch* and *bite* will not be found related in human-curated resources but given that we are talking about dogs, they should be related. With our proposed extension, such similarities can be captured by *contextualized* word embeddings. By integrating relevant literature attempting to define hypernymy/hyponymy relations between embeddings (e.g. see Yu and Dredze (2015) and Nguyen et al. (2017)), we could even define the exact relation (hypernymy, hyponymy) between two similar embeddings instead of defaulting them to “similar” and thus “entailing”. Then, the established specificity judgments are updated with further restrictions imposed by the properties and conceptual graphs. Specifically, the conceptual graph imposes constraints concerning the semantic roles of the concepts, i.e. the “basic” predicate-argument composition, and is thus defining what specificity matches are “compatible” and which have to be removed, e.g. the subject of the one sentence cannot be matched with the object of the other (note that GKR solves active/passive voice and produces the same semantic graph for the active and passive version of a given sentence). Given enough training data, the plausibility of a given match can be estimated through a learning process. After the update of the concept matches, the context graph can determine which of those matched concepts are (un-)instantiated within which contexts, i.e. we now deal with “hard” compositionality cases. This is possible due to the “naming” role that the contexts play: for each concept which we have matched and updated with restrictions, we can find the context it is the head of and look up its instantiation. As a final step for inference, instantiation and specificity are combined to determine entailment relations. A preliminary, experimental version of this proposal is under implementation but its detailed presentation is beyond the scope of this paper.

4.2 More distributional

The previous approach attempts to inject distributional features on the lexical layer of GKR, thus restricting it to the simple contribution of word embeddings. It also integrates a learning process in the match update, but in its core, it solves the “basic” predicate-argument as well as the “harder” boolean/contextual compositionality with symbolic/structural methods, namely through

the use of the concept and context graphs. However, for a given application it might be more beneficial to have a stronger distributional effect than the previous approach allows. For this we can still benefit from GKR factoring out the contextual structure, i.e. dealing separately with the “harder” compositionality cases that distributional approaches struggle with, and use the concept graph only in an assisting way.

So, in this approach the merit of the “naming” technique implemented in the context graph shows itself more clearly: we go through the context graph and we collect all contexts being introduced. For each of them we find its head (*ctx_hd*), which leads us back to the node of the concept graph (see Figure 1 and 2). For this node of the concept graph and all of its children (arguments, modifiers), i.e. for the subgraph with this node as the root, we compute a distributional representation with whichever (neural net) approach we want. Now, each context of the context graph, i.e. each “name”, is associated with a distributional representation and within the context graph these distributed representations are linked with each other with veridical, anti-veridical or averidical edges, based on the original context graph. After doing this computation for each of the sentences of the inference pair, the resulting “named” graphs can be fed into a subsequent layer function, which matches some or all the representations across graphs/sentences based on a computed similarity. Finally, by look-up of the instantiability of each of the matched representations and, if required, by computation of the result of subsequent instantiabilities, the inference relation is decided.

This simple “trick” of factoring out the “hard” compositionality cases, i.e. packing this information in the context graph, allows us the flexibility of using a variety of options for how word vectors can be composed into phrase vectors. In other words, in this approach the “basic” predicate-argument structure compositionality can be achieved in any (distributional) way a given application requires – independently from the concept graph and not necessarily as a logical form as relevant literature (Lewis and Steedman, 2013; Beltagy et al., 2016) has attempted so far. For example, the researcher could choose a more end-to-end deep architecture, like the one used by Conneau et al. (2017) in InferSent, or train a tree-structured recursive neural model as it is done by

Bowman et al. (2015b), where the tree on which the model is based, is built considering the compositionality principles applying to constituents parsing. No matter the predicate-argument composition approach and the final distributional representation, what is crucial is that Boolean and contextual phenomena can be treated outside this representation and thus distributional approaches can benefit from the precision that symbolic/structural methods achieve in such phenomena. A sample implementation of this proposal is described in Section 5.

5 Proof-of-concept for the “more distributional” approach

Recently, Dasgupta et al. (2018) (DS from now on) experimented with the compositionality of the InferSent (Conneau et al., 2017) embeddings. They created different NLI test sets which contain pairs that cannot be solved with world-knowledge but instead involve some more complex semantic phenomena. They trained a classifier on the inference corpus SNLI (Bowman et al., 2015a), using the state-of-the-art InferSent embeddings, and found that the performance on all of their created sets reaches around 50%, thus proving that such embeddings do not yet capture aspects of “basic” predicate-argument and “harder” compositionality. After including the created test sets into the training data of the classifier, DS show that performance improves. With our “more distributional” proposal, we show that it is not necessary to attempt to adequately include all possible linguistic phenomena in the training data: we choose two of the test sets of DS⁴ containing a total of 4800 pairs, where sentence A involves a conjunction of a positive sentence with a negative sentence and sentence B contains one of the conjunct sentences either in its positive or its negative version, as shown below, resulting into entailment or contradiction.

A= The boy does frown angrily, but the girl does not frown angrily.

B= The boy does not frown angrily.

CONTRADICTION

For this subset, DS report a performance of 53.2% and 53.8% for *subjv_long* and *subjv_short*, respectively, on the original SNLI trained model.

⁴Available from <https://github.com/ishita-dg/ScrambleTests>. Chosen sets: *subjv_long* and *subjv_short*.

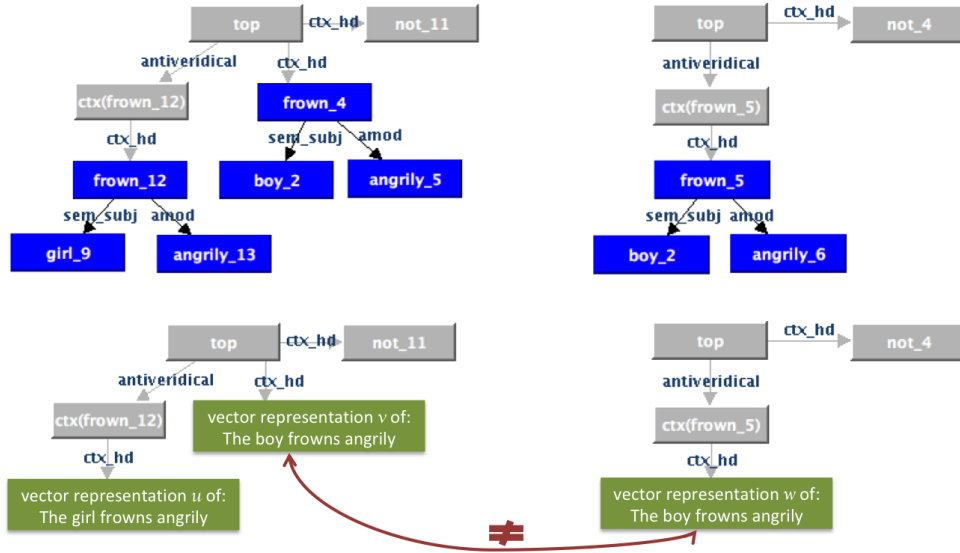


Figure 4: Computation of the “more distributional” proposal. Top: GKR concept and context graphs of the sentences *The boy does frown angrily, but the girl does not frown angrily*. (left) and *The boy does not frown angrily*. (right). Bottom: Injection of the distributional representations in the context graphs for the two sentences, respectively. The red arrow is matching the two similar representations

This set was chosen for three reasons: a) it has one of the lowest performances among DSs’ sets, b) it combines two of the most challenging compositionality phenomena contained in DSs’ sets altogether, i.e. it requires both the treatment of negation and the distinction between the conjunct sentences/events, and c) the phenomena it deals with are of the type for which GKR’s division of semantic labor can show its value and offer a direct solution. Future work can apply the proposed method to the other sets, some of which however, e.g. the scrambled word order sets, might need a stronger symbolic/structural component as presented in our first proposal in Section 4.1.

To test our “more distributional” proposal, we proceed as described in 4.2. We first process both sentences of each pair with GKR and then we go through each sentence to match it to its distributional representation: for each context introduced in the context graph (Figure 4, top, in grey), we retrieve its *ctx.head*, which is a node of the concept graph (Figure 4, top, in blue). For the phrase/sentence consisting of this concept node and all its children, we compute the InferSent representation (Figure 4, bottom, in green). Now, within the context graph, every context (“name”) is associated with such a representation, which means that we have the instantiability of each representation. For each pair, we attempt to match

one of the representations of sentence A with the representation of sentence B. In this test set, simple cosine similarities are enough to compute this, because we know that representation B exactly matches one of the A representations. For more complex cases, a trained function should be responsible for the matching, as described above. After a match is found (Figure 4, bottom, red arrow), we look up the instantiability of each of the matched representations in the top context: if one of them is anti-veridical and the other one veridical, there is a contradiction; if both of them have the same veridicality, then we have an entailment. In our example of Figure 4 we have one match between vectors v and w . Vector v is in a veridical relation with the top context (it is in fact the head of the context, thus it is veridical in it), while vector w is in an anti-veridical relation to top. This means that there is a contradiction between the matched representations and thus the whole pair is labelled contradictory.

This process allowed us to achieve 99.5% accuracy on the two test sets. The 24 wrongly labelled pairs were caused by the wrong output of the Stanford Parser, which led to the wrong dependency graph, wrong conceptual graph and finally wrong contextual graph. In fact, there were more cases where the output of the Stanford Parser was incorrect, but if the assignment of concepts to contexts

is correct, i.e. a partially wrong conceptual graph is matched to a valid context, those weaknesses might not be crucial for the final result. This additional merit shows how we combine the best of both worlds: the computation can succeed even if the concept graph is erroneous, as long as the contexts assigned to the concepts and the matching between the distributional representations of A and B are good enough. In an erroneous concept graph the concepts acting as context heads might be associated with wrong concepts (children), which in turn means that the distributional representation will also not encode the subgraph that we would ideally want. However, given the robustness of such representations and the fact that they encode world knowledge, the matching between the representations across the two sentences can still succeed if the trained similarity function can recognize two representations as more similar. Then, if the contexts assigned to the concepts and thus the computed representations are correct, the system can still predict the correct relation because it can use the matched representations of the distributional approach and their instantiability of the symbolic/structural approach. This means that we benefit from the robustness of the distributional approaches without sacrificing the precision of the symbolic/structural ones.

Nevertheless, we should also note that the two test sets are artificially simple so that the simple trick of factoring out the contextual structure, i.e. the “hard” compositionality phenomena, performs extremely well in comparison to the purely distributional approaches. Firstly, in this test set, there is little variation between the predicate-argument structures of the sentences of the pairs so that we cannot fully check how the Stanford Parser would perform in other cases and how well the GKR concept and context graphs would then be able to “repair” the mistakes of the parser. Furthermore, in this test set we know that sentence B has only one representation which definitely matches with one of the representations of A. This makes the simple cosine similarity as metric for the matching of the representations efficient enough; however, in a harder data set with no such “patterns”, the performance would strongly depend on the quality of the trained matching function, which would have to be more complex than simply the “match with the highest cosine similarity” and thus more error-prone. Despite this grain-of-salt caution, this ap-

proach is expected to perform well for many other complex phenomena apart from negation and conjunction. For example, it will work reasonably well for implicatures such as $A = \textit{The boy forgot to close the door. B= The boy closed the door.}$ For sentence A the distributional representations of the subgraph *The boy close the door* will be anti-veridical in the top context of *forget*, while in B the representation of the whole sentence will be veridical in top. These two representations will have the highest similarity in the matching procedure and will thus match. Considering the instabilities of this match, the pair will be deemed a contradiction.

Testing this approach with further datasets of complex examples can show potential weaknesses of using GKR in this way and particularly highlight other aspects where the distributional or the symbolic/structural strengths should be used more or less. For example, as indicated above, testing with sets with scrambled word order pairs (e.g. *The dog is licking the man* vs. *The man is licking the dog*) might show the need for a stronger symbolic/structural component where the predicate-argument structure is considered more, as it is done in the first proposed approach in 4.1. Additionally, it would be interesting to compare this approach to a purely symbolic/structural one to highlight differences in performance. However, to the best of our knowledge, there is no openly-available, purely symbolic NLI system to which we could straight-forwardly compare our results.

6 Conclusions

In this paper we combine symbolic/structural and distributional features for meaning representations and propose that each of them be used in what it is best at: for complex phenomena like quantification, booleans and modality, use structural meaning and for robust, world-knowledge-informed lexical representations, use distributional semantics. We show how GKR could fulfill this role in two different ways and implement one of them to empirically test its adequacy in the setting of simple, but hard problems for distributional approaches. The good performance results make us confident that there is indeed value in combining the merits of distributional and symbolic approaches. Future work will show how the current proposals can be extended to larger scale systems, maybe also in a combined manner.

References

- Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but tough-to-beat baseline for sentence embeddings. In *International Conference on Learning Representations (ICLR)*.
- Miguel Ballesteros and Yaser Al-Onaizan. 2017. [AMR parsing using stack-LSTMs](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1269–1275, Copenhagen, Denmark. Association for Computational Linguistics.
- 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 & Interoperability with Discourse*, pages 178–186.
- Pierpaolo Basile, Marco de Gemmis, Anna Lisa Gentile, Pasquale Lops, and Giovanni Semeraro. 2007. [UNIBA: JIGSAW algorithm for Word Sense Disambiguation](#). In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 398–401, Prague, Czech Republic. Association for Computational Linguistics.
- I. Beltagy, Stephen Roller, Pengxiang Cheng, Katrin Erk, and Raymond J. Mooney. 2016. [Representing Meaning with a Combination of Logical and Distributional Models](#). *Computational Linguistics*, 42(4):763–808.
- Daniel G. Bobrow, Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Valeria de Paiva, Charlotte Price, and Annie Zaenen. 2007. PARCs Bridge and Question Answering System. In *Proceedings of the Grammar Engineering Across Frameworks Workshop (GEAF 2007)*.
- Johan Bos. 2008. [Wide-Coverage Semantic Analysis with Boxer](#). In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 277–286. College Publications.
- Johan Bos. 2016. Expressive Power of Abstract Meaning Representations. *Computational Linguistics*, 42(3):527–535.
- Marisa Boston, Richard Crouch, Erdem Özcan, and Peter Stubbley. 2019. Natural language inference using an ontology. In Cleo Condoravdi and Tracy Holloway King, editors, *Lauri Karttunen Festschrift*. CSLI Publications.
- Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015a. [A large annotated corpus for learning natural language inference](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal. Association for Computational Linguistics.
- Samuel R. Bowman, Christopher Potts, and Christopher D. Manning. 2015b. [Recursive Neural Networks Can Learn Logical Semantics](#). In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 12–21, Beijing, China. Association for Computational Linguistics.
- Jeremy Carroll, Christian Bizer, Pat Hayes, and Patrick Stickler. 2005. [Named Graphs](#). *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(4).
- Danqi Chen and Christopher Manning. 2014. [A Fast and Accurate Dependency Parser using Neural Networks](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750. Association for Computational Linguistics.
- Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Diana Inkpen, and Si Wei. 2018. [Neural Natural Language Inference Models Enhanced with External Knowledge](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2406–2417, Melbourne, Australia. Association for Computational Linguistics.
- Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017. [Supervised learning of universal sentence representations from natural language inference data](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, Copenhagen, Denmark. Association for Computational Linguistics.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. [Minimal Recursion Semantics: An introduction](#). *Research on Language and Computation*, 3(2):281–332.
- Richard Crouch. 2014. [Transfer Semantics for the Clear Parser](#). In *Proceedings of NLCS 2014*.
- Richard Crouch and Aikaterini-Lida Kalouli. 2018. [Named Graphs for Semantic Representation](#). In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 113–118, New Orleans, Louisiana. Association for Computational Linguistics.
- Richard Crouch and Tracy Holloway King. 2007. *Systems and methods for detecting entailment and contradiction*. US Patent 7,313,515.
- Ishita Dasgupta, Demi Guo, Andreas Stuhlmüller, Samuel J. Gershman, and Noah D. Goodman. 2018. [Evaluating compositionality in sentence embeddings](#). *CoRR*, abs/1802.04302.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). *CoRR*, abs/1810.04805.

- C. Fellbaum. 1998. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.
- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. [A Discriminative Graph-Based Parser for the Abstract Meaning Representation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland. Association for Computational Linguistics.
- L. Getoor and B. Taskar. 2007. *Introduction to Statistical Relational Learning*. MIT Press.
- Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. [Announcing Prague Czech-English dependency treebank 2.0](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3153–3160, Istanbul, Turkey. European Language Resources Association (ELRA).
- Aikaterini-Lida Kalouli and Richard Crouch. 2018. [GKR: the Graphical Knowledge Representation for semantic parsing](#). In *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*, pages 27–37, New Orleans, Louisiana. Association for Computational Linguistics.
- Aikaterini-Lida Kalouli, Livy Real, and Valeria De-Paiva. 2018. WordNet for “Easy” Textual Inferences. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.
- Ronald M. Kaplan. 1995. The formal architecture of lexical-functional grammar. In *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, Stanford University.
- Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. [Skip-Thought vectors](#). *CoRR*, abs/1506.06726.
- Omer Levy and Yoav Goldberg. 2014. [Dependency-Based Word Embeddings](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308, Baltimore, Maryland. Association for Computational Linguistics.
- Mike Lewis and Mark Steedman. 2013. [Combined Distributional and Logical Semantics](#). *Transactions of the Association for Computational Linguistics*, 1:179–192.
- Bill MacCartney and Christopher D. Manning. 2007. [Natural logic for textual inference](#). In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 193–200, Prague. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. [Distributed Representations of Words and Phrases and their Compositionality](#). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.
- Jeff Mitchell and Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34(8):1388–1429.
- Aakanksha Naik, Abhilasha Ravichander, Norman Sadeh, Carolyn Rose, and Graham Neubig. 2018. [Stress Test Evaluation for Natural Language Inference](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2340–2353, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Kim Anh Nguyen, Maximilian Köper, Sabine Schulte im Walde, and Ngoc Thang Vu. 2017. [Hierarchical Embeddings for Hypernymy Detection and Directionality](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 233–243, Copenhagen, Denmark. Association for Computational Linguistics.
- Ian Niles and Adam Pease. 2001. Toward a Standard Upper Ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pages 2–9.
- Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018. [Exploring neural methods for parsing discourse representation structures](#). *Transactions of the Association for Computational Linguistics*, 6:619–633.
- Stephan Oepen and Jan Tore Lønning. 2006. Discriminant-based MRS banking. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006)*, Genoa, Italy. European Language Resources Association (ELRA).
- Sebastian Padó and Mirella Lapata. 2007. [Dependency-Based Construction of Semantic Space Models](#). *Comput. Linguist.*, 33(2):161–199.

- Matteo Pagliardini, Prakhar Gupta, and Martin Jaggi. 2018. [Unsupervised learning of sentence embeddings using compositional n-gram features](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 528–540, New Orleans, Louisiana. Association for Computational Linguistics.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. [Glove: Global Vectors for Word Representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Matthew Richardson and Pedro Domingos. 2006. [Markov logic networks](#). *Machine Learning*, 62(1):107–136.
- Sebastian Schuster and Christopher D. Manning. 2016. [Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- Jiaying Shen, Henk Harkema, Richard Crouch, Ciaran O’Reilly, and Peng Yu. 2018. [Layered semantic graphs for dialogue management](#). In *Proceedings of the 22nd workshop on the Semantics and Pragmatics of Dialogue (SemDial)*.
- Vered Shwartz and Ido Dagan. 2019. [Still a pain in the neck: Evaluating text representations on lexical composition](#). *CoRR*, abs/1902.10618.
- Ed Stabler. 2017. [Reforming AMR](#). In *Formal Grammar 2017. Lecture Notes in Computer Science*, volume 10686. Springer.
- Chuan Wang and Nianwen Xue. 2017. [Getting the Most out of AMR Parsing](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1257–1268, Copenhagen, Denmark. Association for Computational Linguistics.
- Lyndon White, Roberto Togneri, Wei Liu, and Mohammed Bennamoun. 2015. [How well sentence embeddings capture meaning](#). In *Proceedings of the 20th Australasian Document Computing Symposium, ADCS ’15*, pages 9:1–9:8, New York, NY, USA. ACM.
- John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2016. [Towards Universal Paraphrastic Sentence Embeddings](#). *CoRR*, abs/1511.08198.
- Mo Yu and Mark Dredze. 2015. [Learning Composition Models for Phrase Embeddings](#). *Transactions of the Association for Computational Linguistics*, 3:227–242.
- Xunjie Zhu, Tingfeng Li, and Gerard de Melo. 2018. [Exploring semantic properties of sentence embeddings](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 632–637, Melbourne, Australia. Association for Computational Linguistics.