Appendix: Preposition Sense Disambiguation and Representation

A Preposition Sense Representation

A.1 Word similarity task

We learn preposition sense representations, and explore the semantic information they carry. A straight-forward approach is to examine the nearest neighboring words, corresponding to each sense of each preposition. Our results on this exploration for two exemplar prepositions are here: *in* (Table 9) and *with* (Table 10). We enumerate senses for a preposition in each table, and also provide their closest words, TPP sense types and example sentences for each sense.

sense number	1	2	3	4
closest words	backwards, reverse, angles, diagonal, between, forward	wearing, dress, hats, dresses, trousers, sleeves, pants, jacket	back, inside, underneath, from, into, where, onto	where, near, from, at, southern, northern, during
example	in all directions, move in, differ in	dress in black, in leather, in size	in the mail, in most cases, in confined space	in military aircraft, in the UK, in Argentina
TPP sense	Manner_or_Degree	VariableQuality	ThingEntered	ThingEnclosed
sense number	5	6	7	8
closest words	until, during, subsequently, following, after, late, since	university, graduate, college, teaching, faculty, school	economic, systematic, growth, technological	wearing, dressed, costume, wears, clothes, jacket
example	in 1978, in may 1993, in 2002, in the weeks	in a lecture, in graduate studies, in college	focus in science, growth in sales, vocals in her pieces	in the costume, in the jacket, in a gown
TPP sense	Timeframe	ProfessionAspect	Attribute	Garment
sense number	9	10	11	12
closest words	explicitly, interpretation, discourse, fundamental, notion, principles	prosecutor, prosecution, criminal, judicial, justice	onwards, for, wherein	violent, betrayal, bloody, brutal, bitter, fearful
example	in a diagram, in this process, in different ways, in the work	in a constitution, in military justice, in court	in computer graphics, in engineering projects, in the war	result in, in custody, involved in, participate in
TPP sense	Medium	Activity	FramingEntity	Condition

Table 9: Senses of preposition "in".

In Table 9, we notice that the nearest neighbors tend to be words semantically related to the given sense. For example, *until*, *during* and *since* are close to TimeFrame sense of *in*. The nearest words might also be the governors of a preposition. For example, when *in* carries the sense of Garment, verbs such as *dressed* are close to it. Also, the nearest words can be complements of this preposition. For example, nouns such as *university*, *college* and *school* are neighbors of *in* in the sense of ProfessionAspect.

The senses of *with* are listed in Table 10. A nearest neighbor of the sense MeansName is *using*, and based on this the phrase "treatment with laser" can be rewritten as "treatment using laser". A nearest neighbor of sense Accompanier is *alongside*, and "partner *with* systems integrators" can be understood as "partner *alongside* systems integrators". *Against* is synonymous to the sense Opponent, and "collided *with* a bus" can be paraphrased as "collided *against* a bus". Besides semantically similar words, governors and complements also tend to occur as nearest words. *Interviewing* is a governor in "interviewing me with questions", and is one neighbor of sense Message. *News* is a complement in "contact me with recent news", and also close to *of*'s sense Message.

These tables of sense representations show us that preposition sense-specific embedding carries non-trivial lexical semantics. The nearest neighbors listed serve as a qualitative evaluation of these representations in that semantically-similar words are included as nearest neighbors. Thus, the nearest neighbors

sense number	0	1	2	3
closest words	pair, featuring, twisted, assorted	using, stacked, resembling, molded, mechanically, adding	signed, contract, professional, manager, career, full-time	switches, microphones, setup, installing, radios, audio
example	rubble with bare hands, nudged Graham with her elbow	healed them with our doctor 's hand, treatment with laser	stint with Somerset, manager with The Northern Echo	a wooden cart with small wheels, the envelope with her resignation
TPP sense	MeansName	MeansName	Employer	Accountrement
sense number	4	5	6	7
closest words	scholar, studies, professor, doctoral	treasurer, leader elected, deputy	while, alongside, mutual, befriend, interpersonal	community, voluntary, facilitate, implementing
example	studies literature with the Open University	the value of benefits rises with income, fantastic with the day	partner with systems integrators, conspire with enemy	complied with their obligation, conform with the legislation
TPP sense	Partner	Coresultant	Accompanier	Harmonizer
sense number	8	9	10	11
closest words	express, emotional, jealousy, fearful	against, teammate, throwing, punching	news, reporter, press, interviewing, announcing	collar, wears, shoulders, waist, belly, neck
example	glistened with dew, shimmers with crystals	the showdown with his father, collided with a bus	contact me with recent news, interview me with questions	people with disabilities, a lady with a pale face
TPP sense	FeatureCause	Opponent	Message	Attribute
sense number	12	13	14	
closest words	mutual, mutually, resulting, when, thus, meanwhile	symptoms, prognosis, syndrome, abnormalities	dazed, furiously, relentlessly, taunt	
example	reason with her, compatible with autonomy	woke with a heavy head, awoke with a start	forecast with certainty, express it with passionate intensity	
TPP sense	Concomitant	Malady	Manner&Mood	

Table 10: Senses of preposition *with*.

can be treated as 'definitions' of the specific preposition sense. We also find that these nearest words include governors and complements of the given preposition.

A.2 Preposition senses aid paraphrasing

In the experiment on phrasal verb paraphrasing, we use the global preposition embedding, the simplex verb embedding and our sense-specific preposition embedding to approximate the representation of phrasal verbs. For a given phrase, the nearest verbs of the phrasal representation are used (excluding the verb in the phrase) as its paraphrase.

For each approximate phrasal embedding ($v_{\text{vp}}^{\text{sense}}$, $v_{\text{vp}}^{\text{global}}$, $v_{\text{vp}}^{\text{simplex}}$), we list the nearest three verbs (excluding the verb in the phrase) as candidate paraphrases. Here, distance is measured in terms of the cosine similarity between the word vectors.
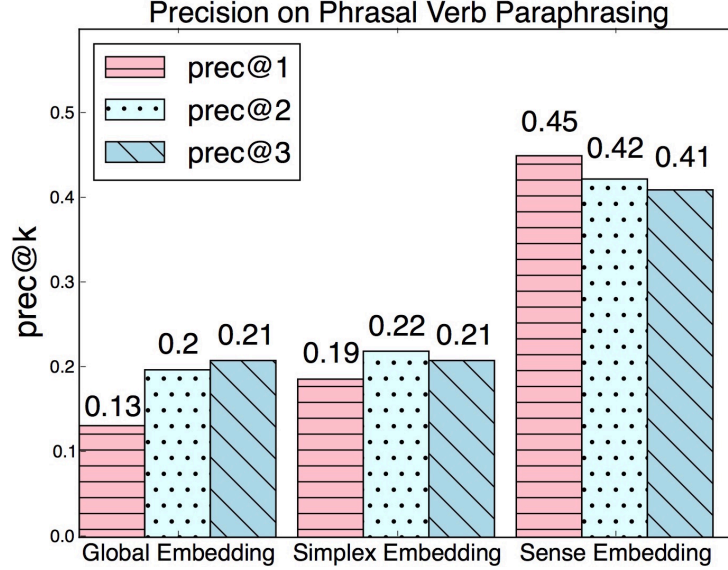

Figure 2: Precision on phrasal verbs paraphrasing.

Since we listed the top three candidate paraphrases for a phrasal verb and consider which of them are valid, we choose the metric of precision at k (prec@k), defined as:

$$\text{prec@k} = \frac{1}{N} \sum_{i=1}^{N} \text{Precision}(i, k),$$

where N is the number of phrasal verbs, and $\text{Precision}(i, k)$ is the percent of good paraphrases among the top k paraphrases for phrase i. The precision metrics for each method are reported in Fig. 2. As we can see, our sense-specific preposition embedding has a significantly better performance than global and simplex embeddings, in terms of all the three prec@1, prec@2 and prec@3 metrics.

We notice that paraphrasing is closely related to the nature of phrasal verbs. A three way classification of phrasal types is adopted in (Dehé, 2002; Jackendoff, 2002; Emonds, 1985; Villavicencio, 2006), where verb particle compounds (VPC) are classified into compositional, idiomatic or aspectual. For the compositional VPCs, the meaning of the construction is determined by the literal interpretations of the particle and the verb (e.g., throw out). Idiomatic VPCs, however, cannot have their meaning determined by their component words (e.g., *get through* meaning 'manage to deal with'). The third class, aspectual VPCs, have the particle providing the verb with an endpoint, describing the action in more detail (e.g., tear up).

In the investigation described here we focus on compositional and aspectual VPCs. Verb particle compounds (VPC) are classified into three types based on the semantic relation between phrases and component words (Villavicencio, 2006):

1. Compositional: the phrasal meaning is determined by the literal interpretations of the verb and the particle;

2. Aspectual: the particle provide the verb with an endpoint, suggesting that verb action is performed thoroughly;

3. Idiomatic: phrasal meaning can not be inferred from the literal meaning of its component words.

The dataset of English phrasal verbs consisted of 91 phrases, in which there were 54 compositional phrases, 16 aspectual phrases and 21 idiomatic phrases in our dataset. Here we report the precision@k (k=1,2,3) of different methods on these three types of verb phrases respectively.

As is shown in Table 11, the precision of paraphrasing with preposition sense embedding is higher than those of the baselines with the global preposition embedding and simplex verb embedding on all three types of phrases. As we can see from the table, the precision improvement of sense embedding over global embedding is larger on aspectual phrases than on compositional phrases. The reason could be that the preposition plays a more important semantic role in aspectual phrases than in compositional phrases.

We also observe that the precision achieved by the simplex embedding is close to the precision by global embedding. It could be interpreted to mean that the preposition global embedding does not capture the semantic information of prepositions.

Sense embeddings outperform simplex ones in instances where: (a) prepositions are important in aspectual phrases (where the particle provides the verb with an endpoint, suggesting that the action described by the verb is performed completely, thoroughly or continuously), e.g., "go against"; (b) prepositions disambiguate the verb, e.g., "carried" has multiple senses: (i) support the weight of something; (ii) assume or accept (responsibility or blame). In vector representation, "carried" is close to "laid", "wiped" and "phased", while sense "on" drives "carry on" closer to "conducted".

Sense embeddings outperform global ones since the latter only represent the semantics of dominant sense while sense embedding is better at capturing the true sense. For example, the global embedding of *down* is close to *destroyed* and *crashed*, and thus in phrase *put down*, global method gives paraphrases such as *slammed* and *snapped*. Sense embedding provides its sense of "downward direction", and gives the paraphrases *laid* and *tossed*.

Sense embeddings encode phrasal verb semantics even though the preposition in the phrasal verb has lost its functional aspect; we see that computationally (and in a vector space), the sense-tagged preposition remains inside a phrasal verb. This is more pronounced in compositional phrasal verbs and in aspectual ones, and less so in idiomatic ones (Villavicencio, 2006).

we also observe that the precision ranking for phrase types are:

Compositional > Aspectual > Non-compositional.

phrase type	Compositional			Aspectual			Idiomatic		
embedding	global	simplex	sense	global	simplex	sense	global	simplex	sense
prec@1	0.125	0.232	0.482	0.0625	0.0625	0.5	0.190	0.143	0.333
prec@2	0.196	0.277	0.429	0.125	0.0625	0.406	0.238	0.167	0.429
prec@3	0.220	0.268	0.417	0.188	0.042	0.438	0.190	0.159	0.381

Table 11: Precision on verb phrase paraphrasing.

In addition, we find that the paraphrasing of compositional or aspectual phrasal verbs is better than that of idiomatic ones. This is because in the case of the idiomatic phrasal verbs, the component words do not give much information about the semantics. Hence the sum of components is not a good approximation of an idiomatic phrasal verb representation.

Empirically we observe that a phrasal verb approximation using addition of verb and preposition sense gives good paraphrasing mainly in the following cases:

1. The verb dominates the phrasal meaning, e.g., focus on ($\sim$ focus), carry in ($\sim$ carry);

2. The preposition dominates the phrasal meaning, e.g., go against ($\sim$ against), keep from ($\sim$from, one sense of 'from' is close to 'stop' and 'prevent');

3. The verb is polysemous, and the preposition helps disambiguate the verb. For example, 'headed down' where the verb 'headed' has two senses: 'chaired/led' and 'approached'. The phrase 'headed down' prompts that 'headed' should have the sense 'approached'.