Composing and Updating Verb Argument Expectations: A Distributional Semantic Model

Alessandro Lenci
University of Pisa, Department of Linguistics
via S. Maria, 36
56126 Pisa (Italy)
alessandro.lenci@ling.unipi.it

Abstract

The aim of this paper is to present a computational model of the dynamic composition and update of verb argument expectations using Distributional Memory, a state-of-the-art framework for distributional semantics. The experimental results conducted on psycholinguistic data sets show that the model is able to successfully predict the changes on the patient argument thematic fit produced by different types of verb agents.

1 Introduction

A number of studies using different experimental paradigms (priming, self-paced reading, etc.) have shown that verbs (*eat*) activate expectations about nouns occurring as their arguments (*cheese*)(McRae et al., 1998), and vice versa (McRae et al., 2005). Nouns also activate expectations about other nouns occurring as co-arguments in the same event (*key* – *door*)(Hare et al., 2009). These behavioral effects support the hypothesis that in the mental lexicon verbs and their arguments are arranged into a web of *mutual expectations*. Verb argument expectations encoded in lexical representations are exploited by subjects to determine the plausibility of a noun as an argument of a particular verb (*thematic fit*, or *selectional preferences* in the linguistic literature), which has been proved to have important effects on human sentence processing (McRae et al., 1998).

In a recent work, Bicknell et al. (2010) bring evidence suggesting a more complex view of the organization and on-line use of verb argument expectations. In fact, the expectations about the likely fillers of a given verb argument (e.g., the patient role) depend on the way another verb argument (e.g., the agent role) is filled. For instance, if the agent noun is *journalist*, the most likely patient for the verb *check* is *spelling*, while if the agent noun is *mechanic*, the most likely patient for the same verb is *brakes*. As a consequence, thematic fit judgments are also sensitive to the way other roles of the same verb are filled. Bicknell et al. (2010) show that this fact has clear consequences for sentence processing, and argue that subjects dynamically compute and update verb argument expectations and thematic fit during on-line sentence comprehension, by integrating various types of knowledge about events and their arguments.

The aim of this paper is to present a computational model of the dynamic composition and update of verb argument expectations using *Distributional Memory* (DM)(Baroni and Lenci, 2010), a state-of-the-art Distributional Semantic Model (DSM). DSMs (aka *vector space models*) represent word meaning with vectors encoding corpus-based co-occurence statistics, under the assumption of the so-called *Distributional Hypothesis* (Miller and Charles, 1991; Sahlgren, 2006): Words occurring in similar contexts are also semantically similar (Landauer and Dumais, 1997; Padó and Lapata, 2007; Turney and Pantel, 2010). Thematic fit judgments have already been successfully modeled with DSMs (Erk et al., 2010), but to the best of our knowledge the problem of how thematic fit is dynamically updated depending on the way other arguments are filled has not been addressed yet. The core of our proposal is that Distributional Memory

Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics, pages 58–66,
Portland, Oregon, June 2011.

can be used to represent the subject's expectations about the most likely words co-occurring in given syntactic role. We will add to the original Distributional Memory framework a model for verb argument expectation composition called ECU, *Expectation Composition and Update*. Specifically, we will show how the expectations of an agent-verb pair about their patient noun argument can be compositionally derived from the DM representation of the verb and the DM representation of its agent. ECU is evaluated on the data set used in Bicknell et al. (2010), and the experimental results show that it is able to successfully predict the changes on the patient thematic fit with a verb, depending on different agent fillers. More generally, we want to argue that the ECU model proposed here can represent a general and viable approach to address compositionality in distributional semantics.

After reviewing some related work in section 2, we present Distributional Memory (section 3) and its use to model verb-argument composition (section 4). Experiments and evaluation are reported in section 5.

2 Background and related work

Elman (2009) argues that the information on preferred fillers of one verb argument depends on what the filler is of one of the other arguments. For instance, the most likely patient of *cut* is *wood*, when the agent is *lumberjack*, but it is *meat*, when the agent is *butcher*. This claim finds an empirical confirmation in the experiments reported by Bicknell et al. (2010), in which subjects are presented with sentence pairs like the following ones:

(1) The **journalist**$_{AG}$ checked the **spelling**$_{PA}$ of his latest report. (*congruent condition*)

(2) The **mechanic**$_{AG}$ checked the **spelling**$_{PA}$ of his latest report. (*incongruent condition*)

Each pair contains the same verb and patient argument, while differing for the agent argument. In the congruent condition, the patient is a preferred argument of the verb, given the agent, e.g. spelling is something which is typically checked by a journalist. In the incongruent condition, the patient is not a preferred argument of the verb, given its agent, e.g. spelling is not something that is typically checked by a mechanic, who rather checks brakes, engines, etc. Thematic fit judgments used to determine congruent and incongruent agent-verb-patient tuples were collected in an off-line norming study. Bicknell et al. (2010) report that self-paced reading times were shorter at the word directly following the patient for the congruent than the incongruent items. Similar results were obtained in an event-related brain potential (ERP) experiment, in which an N400 effect was observed immediately at the patient noun in the incongruent condition. In eye-tracking experiments, Kamide et al. (2003) also demonstrated that the thematic fit of an object depended on the other verb argument fillers.

The conclusion drawn by Bicknell et al. (2010) is that verb argument expectations and thematic fit are not simply stored in the lexicon, but are rather dynamically updated during sentence comprehension, by integrating various types of knowledge. In fact, if the verb expectations about an argument role depend on the nouns filling its other arguments, the hypothesis that they are compositionally updated is highly plausible, since, "it is difficult to envision how the potentially unbounded number of contexts that might be relevant could be anticipated and stored in the lexicon" (Elman 2009: 21).

Thematic fit judgments have been successfully modeled in distributional semantics. Erk et al. (2010) propose the Exemplar-Based Model of Selectional Preferences, in turn based on Erk (2007). The thematic fit of a noun n as an argument of a verb v is measured with the similarity in a vector space between n and a set of noun exemplars occurring in the same argument role of v. A related approach is adopted by Baroni and Lenci (2010), the main difference being that the thematic fit of n is measured by comparing its vector with a "prototype" vector obtaining by averaging over the vectors of the most typical arguments of v. In both cases, the distributional measure of thematic fit is shown to be highly correlated with human plausibility judgements. Their success notwithstanding, these models fall short of accounting for the dynamical and context-sensitive nature of thematic fit. In the next section, we will extend the Baroni and Lenci (2010) approach with a model for verb-argument composition, which is able to account for the argument interdependency phenomena shown by the experiments in Bicknell et al (2010).

If verb argument expectations are likely to be dynamically computed integrating knowledge of the verb with information about its fillers, modeling thematic fit with DSM requires us to address compositional representations. DSMs have mostly addressed semantic issues related to the representation of the content of single words. However, growing efforts have recently been devoted to the problem of how to build distributional semantic representations of complex expressions (e.g., phrases, sentences, etc.) by composing the distributional representations of their component lexical items (Kintsch, 2001; Clark and Pulman, 2007; Widdows, 2008; Mitchell and Lapata, 2010). Different proposals to address compositionality in DSM exist, but the most common approach is to model semantic composition as vector composition. Mitchell and Lapata (2010) systematically explore various vector composition functions (e.g., vector addition, vector product, and other more sophisticated variants thereof), which are used to build distributional vector representations for verb-noun and adjective-noun phrases. The various models for vector composition are then evaluated in a phrase similarity task.

Erk and Padó (2008) address a partially different and yet crucial aspect of compositionality, i.e., the fact that when words are composed, they tend to affect each other's meanings. The meaning of *run* in *The horse runs* is in fact different from its meaning in *The water runs* (Kintsch, 2001). Erk and Padó (2008) claim that words are associated with various types of expectations (typical events for nouns, and typical arguments for verbs)(McRae et al., 1998; McRae et al., 2005) that influence each other when words compose, thereby altering their meaning. They model this context-sensitive compositionality by distinguishing the lemma vector of a word w_1 (i.e. its out-of-context representation), from its vector in the context of another word w_2. The vector-in-context for w_1 is obtained by combining the lemma vector of w_1 with the lemma vectors of the expectations activated by w_2. For instance, the vector-in-context assigned to *run* in *The horse runs* is obtained by combining the lemma vector of *run* with the lemma vectors of the most typical verbs in which *horse* appears as a subject (e.g. *gallop*, *trot*, etc.). Like in Mitchell and Lapata (2010), various functions to build vectors in contexts are tested. Erk and Padó (2008) evaluate their model for context-sensitive vector representation to predict verb similarity in context (e.g. *slump* in the context of *shoulder* is more similar to *slouch* than to *decline*) and to rank paraphrases.

Our model draws close inspiration from Erk and Padó (2008), with which it shares the importance of verb argument expectations. However, differently from them, we want to model how the combination of a verb with an argument affects its expectations about the likely fillers of its other arguments. While Erk and Padó (2008) test their model on a standard word similarity task (i.e. they measure the similarity between the vector-in-context of a verb with the vector of another "landmark" verb), we evaluate our model for compositionality in distributional semantics in a thematic fit task. Indeed, to the best of our knowledge this is the first time in which the issues of thematic fit and compositionality in DSMs are addressed together.

3 Distributional Memory

Distributional Memory (DM) (Baroni and Lenci, 2010) is a framework for distributional semantics aiming at generalizing over different existing typologies of semantic spaces. Distributional Memory represents corpus-extracted distributional facts as a *weighted tuple structure* T, a set of weighted word-link-word tuples $\langle\langle w_1, l, w_2\rangle, \sigma\rangle$, such that w_1 and w_2 belong to W, a set of content words (e.g. nouns, verbs, etc.), and l belongs to L, a set of syntagmatic co-occurrence links between words in a text (e.g. syntactic dependencies, lexicalized patterns, etc.). For instance, the tuple $\langle\langle book, \texttt{obj}, read\rangle, \sigma\rangle$ encodes the piece of distributional information that *book* co-occurs with *read* in the corpus, and `obj` specifies the type of syntagmatic link between these words, i.e. direct object. The score σ is some function of the co-occurrence frequency of the tuple in a corpus and is used to characterize its statistical salience.

Distributional Memory belongs to the family of so-called *structured DSMs*, which take into account the crucial role played by syntactic structures in shaping the distributional properties of words. To qualify as context of a target item, a word must be linked to it by some (interesting) lexico-syntactic relation, which is also typically used to distinguish the type of this co-occurrence (Lin, 1998; Padó and Lapata, 2007). Differently from other structured DSMs, the tuple structure T is formally represented as a 3-way geometrical object, namely a *third order labeled tensor*. A tensor is a multi-way array (Turney, 2007; Kolda and Bader, 2009), i.e. a generalization of vectors (first order tensors) and matrices (second order tensors). Different semantic spaces are then generated "on demand" through *tensor matricization*, projecting the tuple tensor onto 2-way matrices, whose rows and columns represent semantic spaces to deal with different semantic tasks.

For instance, the space $W_1 \times LW_2$ is formed by vectors for words and the dimensions represent the attributes of these words in terms of lexico-syntactic relations with lexical collocates, such as $\langle$*obj, read*$\rangle$, or $\langle$*use, pen*$\rangle$. Consistently, this space is most suitable to address tasks involving the measurement of the "attributional similarity" between words (Turney, 2006), such as synonym detection or modeling selectional preferences. Instead, the space $W_1W_2 \times L$ contains vectors associated with word pairs, whose dimensions are links between these pairs. This space is thus suitable to address tasks involving the measurement of so-called "relational similarity" (Turney, 2006), such as analogy detection or relation classification (cf. Baroni and Lenci 2010 for more details about the Distributional Memory spaces and tasks). Crucially, these spaces are now alternative "views" of the same underlying distributional memory formalized in the tensor. Many semantic tasks (such as analogical similarity, selectional preferences, property extraction, synonym detection, etc.), which are tackled in the literature with different, often unrelated semantic spaces, are addressed in DM with the same distributional tensor, harvested once and for all from the corpus. This is the reason why Distributional Memory is claimed to be a general purpose resource for semantic modeling.

Depending on the selection of the sets W and L and of the scoring function σ, different DM models can be generated. The Distributional Memory instantiation chosen for the experiments reported in this paper is *TypeDM*, whose links include lexicalized dependency paths and lexico-syntactic shallow patterns, with a scoring function based on pattern type frequency.[1] We have chosen TypeDM, because it is the best performing DM model across the various semantic tasks addressed in Baroni and Lenci (2010). The TypeDM tensor contains about 130M non-zero tuples automatically extracted from a corpus of about 2.83 billion tokens, obtained by concatenating the the Web-derived ukWaC corpus (about 1,915 billion tokens),[2] a mid-2009 dump of the English Wikipedia (about 820 million tokens),[3] and the British National Corpus (about 95 million tokens).[4] The resulting concatenated corpus was tokenized, POS-tagged and lemmatized with the TreeTagger[5] and dependency-parsed with the MaltParser.[6]

The TypeDM word set (W_{TypeDM}) contains 30,693 lemmas (20,410 nouns, 5,026 verbs and 5,257 adjectives). These are the top 20,000 most frequent nouns and top 5,000 most frequent verbs and adjectives in the corpus, augmented with lemmas in various standard test sets in distributional semantics, such as the TOEFL and SAT lists. The TypeDM link set (L_{TypeDM}) contains 25,336 direct and inverse links formed by (partially lexicalized) syntactic dependencies and patterns. This is a sample of the links in L_{TypeDM}:

- **obj**: *The journalist is checking his article* → $\langle$*article,* obj*, check*$\rangle$
- **verb**: *The journalist is checking his article* → $\langle$*journalist,* verb*, article*$\rangle$
- **sbj_tr**: *The journalist is checking his article* → $\langle$*journalist,* sbj_tr*, check*$\rangle$
- **preposition:** *I saw a journalist with a pen* → $\langle$*pen,* with*, journalist*$\rangle$
- **such_as:** "*NOUN such as NOUN*" and "*such NOUN as NOUN*": *animals such as cats* → $\langle$*animal,* such_as+*ns*+*ns, cat*$\rangle$

[1] The TypeDM tensor is publicly available at http://clic.cimec.unitn.it/dm

[2] http://wacky.sslmit.unibo.it/

[3] http://en.wikipedia.org/wiki/Wikipedia:Database_download

[4] http://www.natcorp.ox.ac.uk

[5] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

[6] http://w3.msi.vxu.se/~nivre/research/MaltParser.html

The first two links above are the most relevant ones for the purposes of the present paper: `obj` links a transitive verb and its direct object, and `verb` is a lexically underspecified link between a subject noun and a complement noun of the same verb.

The scoring function σ is the *Local Mutual Information* (LMI) (Evert, 2005) computed on link type frequency (negative LMI values are raised to 0):

$$\mathrm{LMI} = O_{ijk} \log \frac{O_{ijk}}{E_{ijk}} \quad (1)$$

O_{ijk} and E_{ijk} are respectively the observed and expected frequency of a triple $\langle w_i, l_j, w_k \rangle$.

4 Composing verb argument expectations

In this section, we address the fact that the information on preferred fillers of one verb argument depends on the filler of its other arguments by proposing a model for *Expectation Composition and Update* (ECU), which will then be computationally formalized with Distributional Memory.

ECU relies on the hypothesis that nouns and verbs are linked in a web of mutual expectations. Verbs are associated with expectations about their likely arguments, and nouns have expectations about the events they are involved with and also about other nouns co-occuring in the same events (cf. section 1). We argue that, when words compose (e.g. a verb and a noun), their expectations are integrated and updated. Specifically, we focus here on how the composition of a verb and its agent argument determines an update of the verb expectations for its patient argument. Let $EX_{PA}(v)$ be the expectations of a verb v about its patient arguments, i.e. a set of nouns likely to occur as verb patients. For instance, $EX_{PA}(check)$ = *mistakes*, *engines*, *books*, etc. Let $EX(n_{AG})$ be the expectations about typical events performed and typical entities acted upon by the agent noun. For instance, $EX(mechanic)$ = *mechanics fix cars*, *mechanics check oil*, etc. ECU is formally defined as follows:

$$EX_{PA}(\langle n_{AG}, v \rangle) = f(EX(n_{AG}), EX_{PA}(v)) \quad (2)$$

f is some function for expectation composition and update (cf. below). ECU assumes that the result of semantically composing the verb and its agent argument is an update of the verb expectations about its patient argument. $EX_{PA}(\langle n_{AG}, v \rangle)$ are the updated expectations of v about its patient arguments, resulting from the composition of its original expectations with the agent's expectations. For instance, the result of composing *check* with the agent argument *mechanic* is a new set of expectations $EX_{PA}(\langle mechanic, check \rangle)$ formed by objects that are likely checked by mechanics, such as *cars*, *engines*, *wheels*, etc. These updated expectations are a function of the typical patients of checking events, and of the typical patients of events performed by mechanics.

4.1 Modeling ECU with Distributional Memory

The tuple structure of the DM tensor is well suited to represent the web of mutual expectations in which lexical items are arranged. In fact, given a word w, the expectations of w, $EX(w)$, can be defined as the subset of the DM tensor formed by the tuples $\langle \langle w_1, l, w_2 \rangle, \sigma \rangle$, such that $w = w_1$ or $w = w_2$. The tuple score σ determines the statistical salience and typicality of a particular expectation.

To model ECU with TypeDM, we approximate the patient semantic role with the syntactic dependency DM link of `obj` (cf. section 3). The expectations about the typical patient arguments of a verb v ($EX_{PA}(v)$) thus correspond to the set of TypeDM tuples $\langle n_i$, `obj`, $v\rangle$: e.g, $EX_{PA}(check)$ = ⟨*mistake*, `obj`, *check*⟩, ⟨*engine*, `obj`, *check*⟩, etc. We model $EX(n_{AG})$ with the set of DM tuples $\langle n_{AG}$, `verb`, $n_j \rangle$, which characterize the typical patients (i.e., direct objects) of events performed by the agent noun n_{AG}: e.g, $EX(mechanic)$ = ⟨*mechanic*, `verb`, *car*⟩, ⟨*mechanic*, `verb`, *oil*⟩, ⟨*mechanic*, `verb`, *engine*⟩, etc.

The expectation composition function f of equation 2 is modeled as a tensor updating function: f modifies the TypeDM tensor by updating the scores of the relevant subset of tuples. Following current compositionality models in distributional semantics (cf. Mitchell and Lapata 2010), we focus here on two alternative versions of f:

check	$\langle journalist, check\rangle$	$\langle mechanic, check\rangle$
site	article	car
page	book	tyre
website	information	work
box	question	price
detail	fact	vehicle
link	report	job
list	site	system
file	source	bike
record	content	value
information	account	problem

Table 1: Original TypeDM expectations for *check* and their compositional updates obtained with f = **PRODUCT**

SUM
For each tuple $\langle\langle n_i, \texttt{obj}, v\rangle, \sigma_i\rangle \in EX_{PA}(v)$, $\langle\langle n_i, \texttt{obj}, v\rangle, \sigma_u\rangle \in EX_{PA}(\langle n_{AG}, v\rangle)$, and

$$\sigma_u = \begin{cases} \sigma_i + \sigma_j & \text{if } \langle\langle n_{AG}, \texttt{verb}, n_j\rangle, \sigma_j\rangle \\ & \in EX(n_{AG}) \text{ and } n_i = n_j \\ \sigma_i & \text{otherwise} \end{cases} \quad (3)$$

PRODUCT
For each tuple $\langle\langle n_i, \texttt{obj}, v\rangle, \sigma_i\rangle \in EX_{PA}(v)$, $\langle\langle n_i, \texttt{obj}, v\rangle, \sigma_u\rangle \in EX_{PA}(\langle n_{AG}, v\rangle)$, and

$$\sigma_u = \begin{cases} \sigma_i * \sigma_j & \text{if } \langle\langle n_{AG}, \texttt{verb}, n_j\rangle, \sigma_j\rangle \\ & \in EX(n_{AG}) \text{ and } n_i = n_j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The idea underlying both types of tensor updating functions is that the verb expectations about its likely patients are modified with the score of the tuples representing objects that are typical patients of events performed by the agent noun. With **SUM**, expectation composition is a linear function of the score of the tuples in $EX_{PA}(v)$ and in $EX(n_{AG})$: $EX_{PA}(\langle n_{AG}, v\rangle)$ contains all the tuples belonging to $EX_{PA}(v)$, but their score is added to the score of the tuples in $EX(n_{AG})$ sharing the same object noun. With the **PRODUCT** function, the updated expectations only include the tuples in $EX_{PA}(v)$ sharing the same objects with a tuple in $EX(n_{AG})$. The score of these tuples is the product of the original tuple score, while the score of the other tuples in $EX_{PA}(v)$ is set to 0.

Table 1 reports a sample output of the application of ECU to the TypeDM tensor. The left column contains the objects of the top-scoring tuples of $EX_{PA}(check)$ in the original TypeDM tensor (ordered by decreasing values of σ). The central column contains the top-scoring nouns in $EX_{PA}(\langle journalist, check\rangle)$, compositionally derived by updating $EX_{PA}(check)$ with $EX(journalist)$: the nouns are ordered by decreasing value of σ modified according to the **PRODUCT** composition function. We can notice that the updated verb argument expectations include nouns consistent with what journalists typically check. The right column instead contains the top-scoring nouns in $EX_{PA}(\langle mechanic, check\rangle)$, derived from updating $EX_{PA}(check)$ with $EX(mechanic)$: the composition function is still the **PRODUCT**. The difference with the central column is striking: the top-scoring nouns in the updated verb argument expectations are now related to what mechanics typically check.

5 Experiments and evaluation

The ECU model for the compositional update of verb-argument expectations has been evaluated by measuring the thematic fit between an agent-verb pair ($\langle n_{AG}, v\rangle$) and a patient noun argument (n_{PA}) of the same verb. Thematic fit is computed with the verb expectations in $EX_{PA}(\langle n_{AG}, v\rangle)$, which in turned have been obtained by composing $EX(n_{AG})$ and $EX_{PA}(v)$ with either of the two functions described in section 4. In the following subsections, we illustrate the data sets used for the experiments, the procedure to compute the compositional thematic fit in TypeDM, and the results of the experiments.

5.1 Data sets

Two data sets of agent-verb-patient triples from Bicknell et al. (2010) have been used to test ECU:

- **bicknell.64** - 64 test triples used in the self-paced reading and ERP experiments in Bicknell et al. (2010);
- **bicknell.100** - 100 test triples, a superset of bicknell.64.

Triples are organized in pairs, each sharing the same verb, but differing for the agent and patient nouns:

- *journalist*$_{AG}$ - *check* - *spelling*$_{PA}$
- *mechanic*$_{AG}$ - *check* - *brake*$_{PA}$

Patients in each triple were produced by 47 subjects as the prototypical (*congruent*) arguments of the verbs, given a certain agent. The patient noun in one triple is *incongruent* for the other triple with the same verb: e.g., *brake* is the incongruent patient for the *mechanic*$_{AG}$ - *check* pair. The bicknell.100 dataset contains all the triples produced in the original norming study. The bicknell.64 data set is a subset of the normed triples selected by Bicknell et al. (2010) after removing test items that were potentially problematic for the behavioral experiments.

5.2 Procedure

The thematic fit of a noun n_{PA} as the patient of $\langle n_{AG}, v\rangle$ is measured with the cosine between the vector of n_{PA} in the TypeDM $W_1{\times}LW_2$ space and the "prototype" vector in the same space built with the vectors of the top-k expected objects belonging to $EX(n_{AG}, v)$. This is an extension of the approach to selectional preferences modeling presented in Baroni and Lenci (2010) (in turn inspired to Erk 2007). These are the steps used to compute the compositional thematic fit in the TypeDM $W_1{\times}LW_2$ space:

1. we select a set of k of prototypical patient nouns n_{PA} for $\langle n_{AG}, v\rangle$ (in the reported experiments we set $k = 20$). The selected nouns are the n_i in the k tuples $\langle\langle n_i,\ \texttt{obj},\ v\rangle,\ \sigma_u\rangle \in EX_{PA}(\langle n_{AG}, v\rangle)$ with the highest score σ. The patient nouns in the datasets are excluded;

2. the vectors in the $W_1{\times}LW_2$ TypeDM space of the selected nouns are normalized and summed. The result is a centroid vector representing an abstract "patient prototype vector" for $\langle n_{AG}, v\rangle$;

3. for each n_{AG} - v - n_{PA} test triple (e.g., *journalist*$_{AG}$ - *check* - *spelling*$_{PA}$), we measure i.) the cosine between n_{PA} and the "patient prototype vector" for the congruent $\langle n_{AG}, v\rangle$ pair, (e.g., *journalist*$_{AG}$ - *check*) and ii.) the cosine between n_{PA} and the "patient prototype vector" for the incongruent $\langle n_{AG}, v\rangle$ pair, belonging to the other triple with the same verb v (e.g., *mechanic*$_{AG}$ - *check*).

For each test triple, we score a "hit" if n_{PA} has a higher thematic fit (i.e., cosine) with the congruent $\langle n_{AG}, v\rangle$ pair, than with the incongruent one. For instance, if $cosine(\langle journalist, check\rangle, spelling) > cosine(\langle mechanic, check\rangle, spelling)$, we score a "hit", otherwise we score a "fail".

5.3 Results

Experiments to model the verb-argument compositional thematic fit have been carried out with the two ECU functions, **SUM** and **PRODUCT**, each tested on both datasets. Model performance has been evaluated with "hit" accuracy, i.e. the percentage of "hits" scored on each data set. As a baseline, we have simply adopted the random accuracy. The results of the ECU models are reported in table 2.

We can notice that when the verb-argument expectations are compositionally updated with the **PRODUCT** function, the model is able to significantly outperform the baseline accuracy with both data sets. Conversely, **SUM** is never able to go beyond the baseline. This is remindful of the results reported by Erk and Padó (2008) and Mitchell and Lapata (2010), in which multiplicative vector composition achieves better performance in the (verb in context or phrase) similarity tasks than (at least simple) additive functions. In fact, the advantage of the multiplicative function is that it allows the composition process to highlight the dimensions shared by the vectors of the component words, thereby emphasizing context effects. Something similar can be argued to explain the results of the current experiments. With **PRODUCT** the expectations of $EX_{PA}(\langle n_{AG}, v\rangle)$ are a non-linear function of the expectations about patient nouns shared by v and n_{AG}. Therefore, the objects that are likely to be checked by a mechanic depend on the things that are both typical patients of checking events and typical patients of actions performed by a mechanic. This results in a stronger thematic fit in the congruent condition than in the incongruent one.

We also carried out experiments to investigate whether the choice of the parameter k (the number of nouns selected to build the "prototype patient vector") affects the model performance. However, we obtained no significant difference with respect to the values reported in table 2.

data set	*ECU function*	*accuracy*	*p-value*
bicknell.64	SUM	40.62%	
bicknell.64	PRODUCT	84.37%	3.798e-08 ***
bicknell.100	SUM	37.5%	
bicknell.100	PRODUCT	73%	4.225e-06 ***
baseline		50%	

Table 2: Results of the thematic fit experiments (p-values computed with a χ^2 test).

6 Conclusions and further directions of research

Psycholinguistic evidence has proved that verb argument thematic fit is highly context-sensitive. In fact, subjects' sensitivity to the likelihood of a noun as a verb argument strongly depends on the nouns filling other arguments of the same verb. These data hint at a dynamic process underlying verb argument expectation and thematic fit computation, resulting from the compositional integration of the verb expectations with those activated by its arguments. In this paper, we have presented ECU, a distributional semantic model for the compositional update of verb argument expectations. ECU has been applied to Distributional Memory, a state-of-the-art Distributional Semantic Model, whose core tensor of corpus-derived tuples is particularly suited to represent word expectations. ECU has been tested succesfully in an experiment to measure the thematic fit between an agent-verb pair ($\langle n_{AG}, v\rangle$) and a patient noun argument (n_{PA}) of the same verb, with the data set used in the psycholinguistic experiments reported in Bicknell et al. (2010). The good results we have obtained prove that DSMs can provide interesting computational models of the compositional update of thematic fit. Of course, other factors besides verb-argument knowledge may also contribute to the context-sensitive nature of thematic fit. However, it is worth noticing that one of the hypotheses advanced by Bicknell et al. (2010) to explain their experimental results is indeed that subjects use their knowledge of statistical linguistic regularities. This is exactly the type of knowledge that is represented in the Distributional Memory tensor structure and is exploited by ECU.

Starting from the experimental results in Bicknell et al. (2010) on sentence on-line processing, in this paper we have addressed the issue of how the agent of a verb modulates the subjects' expectations about its patients. On the other hand, there is broad evidence that the meaning of a verb is predominantly modulated by its object. This suggests that ECU should also be applied to model how the preferences about the agent argument are determined by the choice of the verb object. We leave this issue for future research.

Besides being a computational model for thematic fit, we also claim that the ECU approach has a more general relevance for the issue of how to address compositionality in DSMs. In fact, let us assume that part of the semantic content of a word consists of expectations about likely co-occurring words, which in turn can be modeled with subsets of a distributional tuple tensor. We can therefore claim that (at least part of) the effect of the semantic composition of words is to update their expectations about other co-occurring words, like ECU does. We have seen here that this hypothesis finds a nice confirmation with verb-argument composition. We believe that an interesting empirical question is to investigate to what extent this hypothesis can be generalized to other cases of compositionality.

In the future, we also plan to experiment with other types of expectation composition functions. Moreover we will extend the ECU model to tackle context-sensitive effects in the thematic fit with respect to other types of verb argument relations, besides agent and patient ones. In fact, Matsuki et al. (submitted) have reported that patient and instrument verb arguments show interdependency effects similar to the ones between agents and patients that we have addressed in this paper.

Acknowledgments

I am very grateful to Ken McRae for providing the data set used in this paper. I also thank Marco Baroni - the co-author of Distributional Memory - and the two anonymous reviewers for their helpful comments. The usual disclaimers apply.

References

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

Klinton Bicknell, Jeffrey L. Elman, Mary Hare, Ken McRae, and Marta Kutas. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4): 489–505.

Stephen Clark and Stephen Pulman. 2007. Combining symbolic and distributional models of meaning. In *Proceedings of the AAAI Spring Symposium on Quantum Interaction*: 52–55.

Jeffrey L. Elman. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4): 547–582.

Katrin Erk. 2007. A simple, similarity-based model for selectional preferences. In *Proceedings of ACL*: 216–223.

Katrin Erk and Sebastian Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of EMNLP 08*: 897–906.

Katrin Erk, Sebastian Padó, and Ulrike Padó. 2010. A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, 36(4): 723–763.

Stefan Evert. 2005. *The Statistics of Word Cooccurrences*. Ph.D. dissertation, Stuttgart University.

Mary Hare, Michael Jones, Caroline Thomson, Sarah Kelly,and Ken McRae. 2009. Activating event knowledge. *Cognition*, 111(2): 151–167.

Yuki Kamide, Gerry T.M. Altmann, and Sarah L. Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49: 133–156.

Walter Kintsch. 2001. Predication. *Cognitive Science*, 25(2): 173–202.

Tamara Kolda and Brett Bader. 2009. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500.

Thomas Landauer and Susan Dumais. 1997. A solution to Plato's problem: The latent semantic analysis. theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2): 1–31.

Dekang Lin 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL*: 768–774.

Kazunaga Matsuki, Tracy Chow, Mary Hare, Jeffrey L. Elman, Christoph Scheepers, and Ken McRae. submitted for publication. Event-based plausibility immediately influences on-line language comprehension

Ken McRae, Michael J. Spivey-Knowlton, and Michael K.Tanenhaus. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension *Journal of Memory and Language*, 38: 283–312.

Ken McRae, Mary Hare, Jeffrey L. Elman, and Todd Ferretti. 2005. A basis for generating expectancies for verbs from nouns *Memory & Cognition*, 33(7): 1174–1184.

George Miller and Walter Charles. 1991. Contextual correlates of semantic similarity *Language and Cognitive Processes*, 6: 1–28.

Jeff Mitchell and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8): 1388–1429.

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

Magnus Sahlgren. 2006. *The Word-Space Model*. Ph.D. dissertation, Stockholm University.

Peter D. Turney. 2006. Similarity of semantic relations. *Computational Linguistics*, 32(3): 379–416.

Peter D. Turney. 2007. Empirical evaluation of four tensor decomposition algorithms. *Technical Report ERB-1152, NRC*.

Peter D. Turney, Patrick Pantel. 2010. From frequency to meaning : Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37: 141–188.

Dominic Widdows. 2008. Semantic vector products: Some initial investigations. In *Proceedings of the Second AAAI Symposium on Quantum Interaction*: 1–8.