Ágoston Tóth

The Distributional Compatibility Relation[*]

Abstract

The present paper discusses the nature of the lexical relation exploited in automatized, corpus-based, statistical explorations of word meaning. This relation captures and quantifies the distributional similarity of lexical items. For reasons presented in this paper, I call it the Distributional Compatibility Relation (DCR). I argue that DCR is a *fuzzy* relation and I compare it to selected lexical relations known from the linguistic literature to see if – and to what extent – their basic properties are similar.

Keywords: distributional semantics, lexical semantics

1   Introduction

An important field of computational linguistics is the measurement of the similarity of words. Emerging *vector-space model* solutions implement this task by collecting word co-occurrence frequency information from large text corpora. Preselected words of a corpus (the *target words*) are characterized by the frequency of certain co-occurrence phenomena, usually the appearance of one or more of the many *context words* in the close vicinity (a "window") of the target word, but other linguistic phenomena, including part of speech information and syntactic features, can also be considered for pattern analysis. Co-occurrence statistical information extracted in this way can be treated as empirical evidence of a word's potential for replacing another word, which is an approach to *measuring similarity* (cf. Miller & Charles 1991). According to the *distributional hypothesis*, this similarity is a semantic phenomenon. More details about distributional semantics and the distributional hypothesis (including their precursors in the linguistic literature) can be found in Lenci (2008).

In computational linguistics, distributional semantics is seen as an alternative to measuring semantic similarity by seeking shared hyperonyms (e.g. *car* and *van* share the same hyperonym: *vehicle*; their similarity can be quantified, too, cf. Resnik 1995). That type of analysis requires the use of ontologies (*is-a* hierarchies) and also the identification of the right concept in the ontology before the similarity measurement can be carried out. Distributional similarity measurement, however, works with words (rather than concepts) and corpora

---

(rather than precompiled databases), which makes it an important, readily available alternative.

The next section of this paper provides details about the vector space model used in measuring distributional similarity. As an illustration, I include a case study that returns data on a common Hungarian adjective. Section 3 investigates the nature of the Distributional Similarity Relation, the relation exploited in my case study and in all vector-space investigations of word similarity. Section 4 highlights a few areas where distributional similarity is used and adds my concluding remarks.

2   The vector space model

Systems designed to collect distributional information about words[1] usually rely on a geometrical interpretation of the empirical data. Each target word is represented by a *context vector*. Each position of the vector is responsible for counting the number of co-occurrences of the given target word with one of the context words. For example, if the word *drink* is a target word, and the word *tea* is among the context words, and *tea* occurs 23 times in the close vicinity (in the context window) of *drink*, then the vector element corresponding to the word *tea* (in the context vector describing the word *drink*) will be set to 23. In most cases, we work with a few target words (typically 10-100) and a much larger number of context words (e.g. 10,000 words or more). The result is a multi-dimensional vector space in which each context word has its own dimension.

Vectors can be collected into a matrix in which each row is a context vector for a single target word. These matrices are useful for illustrative purposes, too (figure 1).

$$C_{t,x} = \begin{bmatrix} 0, & 0, & 23, & 8 & \cdots & 0 \\ 0, & 1, & 18, & 9 & & 0 \\ & & \vdots & & \ddots & \vdots \\ 3, & 5, & 0, & 0 & \cdots & 3 \end{bmatrix}$$

Figure 1: A context matrix

Large corpora (20-50-100 million words or even more) are necessary for this type of investigation. "Raw", unprocessed corpora may be suitable for the task. In the presence of linguistic annotation, we can take additional details into consideration (part of speech labels, syntactic category labels, etc.) – in this case, we can make the feature vectors more directly useful in finding linguistic patterns.

As a next phase, the values in the context matrix can be weighted so that unusual or "surprising" events become more salient in our large collection of co-occurrence events. An

---

[1] It is possible to use *word forms* or *lemmas* as target and context words. This choice is usually treated as one of the many parameters of vector space experiments. In Bullinaria and Levy's (2012) paper on parameter setting, lemmatization and stemming did not consistently improve precision. In what follows, I will default to the *word form* interpretation when referring to "words" in this paper.

effective way of normalizing the vectors is replacing positive pointwise mutual information (pPMI) scores for the raw frequency values (Turney 2001).

At this point, an optional dimension-reduction step may be carried out (see, for instance, Landauer & Dumais 1997).

We can now compare the distribution of the target words by comparing their context vectors. There are two basic methods for comparing context vectors: we can measure vector distances (figure 2) or the cosine of the angle between vectors (figure 3). The latter promises the advantage of being able to avoid problems arising from vector length differences, which is useful, since length depends on the frequency of context words and, because of this, it also depends on the frequency of the target word itself, which is a problem if we try to detect a relation between a frequent and a rare word. More information about the geometrical background of distributional semantics can be found in Widdows (2004).
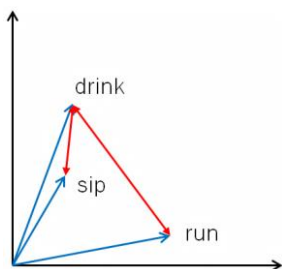
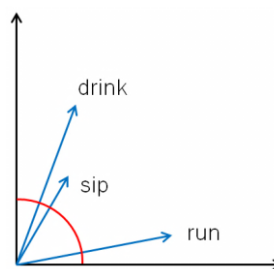*Figure 2: Vector similarity: distance*

Figure 3: Vector similarity: cosine

Testing the results is the usual last phase of vector space experiments. The steps are the following: *1)* a semantic task is solved, *2)* the performance of the system is measured (through computing precision and recall) and compared to a known baseline and to the performance of similar systems, and *3)* the parameters of the system are fine-tuned so that the performance indicators are maximized.[2] In vector-space investigations, the evaluation task can be a similarity-related multiple choice test: for an input word, the system selects the most "similar" word from a list of candidates, then the automatically selected answer is compared to a key. A variation of this evaluation method is the TOEFL test, in which the system answers TOEFL exam questions.

As an illustration of what kind of "raw" results are returned in a vector-space investigation, I have set up an experiment for a brief qualitative case study.

My experiment has been based on the analysis of a 80-million-word subcorpus of the Hungarian Webcorpus (Kornai et al. 2006). A high number of target words have been examined: 15,000 words (the most frequent words of the corpus) have been characterized by co-occurrence data with 15,000 context words (again, the 15,000 most frequent words of the corpus). The resulting context matrix has had 15,000 x 15,000 (225 million) elements. I have used pPMI weighting on the values before comparing the context vectors. Comparison was carried out using cosine vector comparison.

I have chosen the adjective *kis* (English equivalents include 'small', 'little' and 'short') for this case study.[3] Table 1 shows the distributionally most similar words (out of the 15,000

---

2   Computational linguistic research tends to have a very strong quantitative character.
3   This adjective is very frequent and quite general, but it is not completely unaffected by selectional restrictions and it is not free from lexical ambiguity, either. Further investigations are required to see if and to what extent these properties influence the results. Note that nouns and verbs may behave differently, too.

words examined and ranked) and their measured distributional similarity grades. Figure 4 visualizes the scores in a chart.

Rank	Similar word	Typical English equivalents	Similarity score
1	nagy	big, large	0.413
2	kisebb	smaller	0.376
3	nagyobb	bigger, larger	0.347
4	hatalmas	huge, enormous, vast	0.32
5	apró	tiny, minuscule	0.3
6	sok	many, much	0.296
7	egy	a, an, one	0.291
8	a	the	0.282
9	kicsi	tiny, small, little	0.265
10	olyan	such, so	0.264
11	legnagyobb	biggest, largest	0.258
12	szép	nice, pretty, beautiful	0.257
13	ilyen	such a(n), so	0.253
14	másik	other	0.25
15	kevés	little	0.242
16	két	two	0.241
17	egész	all, whole, complete	0.237
18	óriási	gigantic, giant, enormous	0.237
19	legtöbb	most	0.223

Table 1: Words distributionally most similar to kis
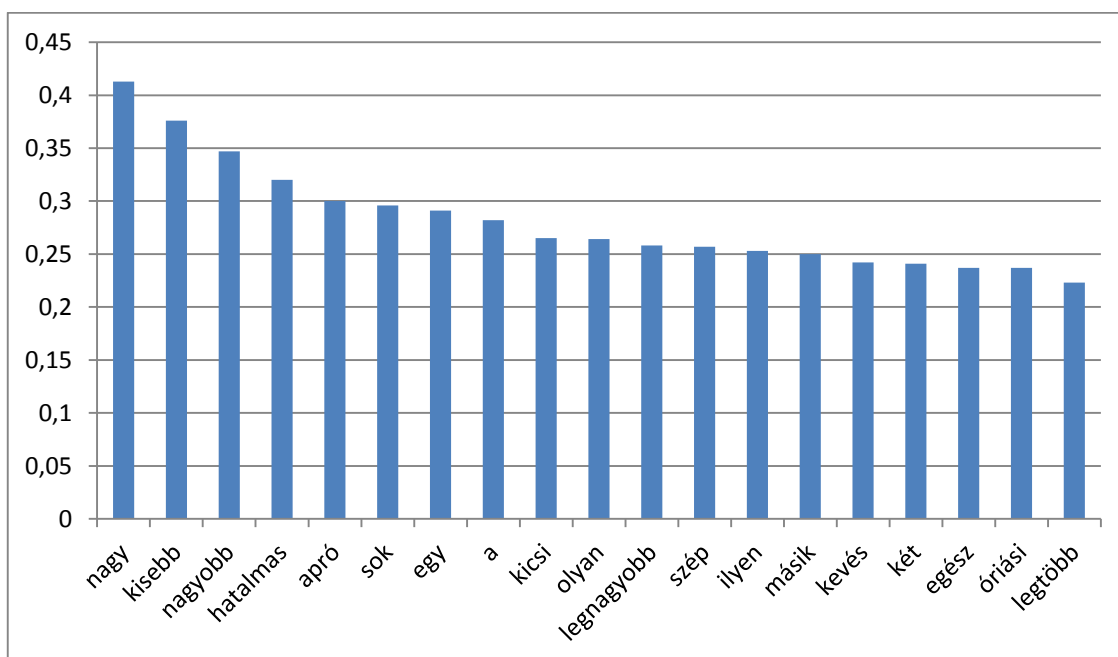

Figure 4: Words most similar to kis

In this experiment, the distribution of the adjective *kis* is found most similar to the distribution of the adjective *nagy* ('big', 'large'). The top 20 include other antonyms, too (*hatalmas, sok, óriási)*. Synonyms are also on the list (*kicsi, kevés),* as well as the comparative form of *kis* (*kisebb).* The superlative form, *legkisebb,* is the 26th item on the list and therefore it is not shown above, although its score is still relatively high. Separating synonymy from antonymy is virtually impossible in this approach. In the case of nouns, it is equally difficult to tell hyponymy/hyperonymy apart from synonymy. It is a general observation that words that can be related to the target word through the established lexical semantic relations do appear in the results, but we cannot distinguish among these relations using distributional vector-space calculations.

Given this situation, we may wonder about the nature of the connection established between lexical items in vector-space investigations.

3   The distributional compatibility relation

Kiefer (2007: 13-36) distinguishes three ways of describing meaning:
- focusing on reference and denotation and using a logical calculus (in formal semantics),
- factoring in the cognitive aspects of our experiencing the world (cognitive semantics), and
- focusing on language-internal facts, attributing meaning to relations between linguistic expressions (structuralist semantics).

Distributional studies collect statistical information about the use of words and try to measure the relatedness of lexical items; therefore, they belong to the realms of structuralist semantics.

Technically, we can build a relation for any two words of a language. Consider Cruse's proposal, the *dogbananomy* relation (Cruse 2011: 129), which connects *banana* and *dog*. This entertaining (and satirical) idea leaves us to wonder what kind of regularity lexical relations are supposed to capture. In general, Cruse argues that the following criteria must be met for a relation to be significant for semantic investigations (ibid.):
- sense relations must *recur* and relate items in a way that expresses a generalization,
- *discrimination*: relations must also exclude a number of pairs,
- the "significance" of a relation should correspond to a concept that we can name.

The distributional relation has a tendency to become very powerful: in the case study described in the previous section, the word *kis* showed a nonzero similarity value to 97% of the 15,000 target words – *recurrence* is not a problem. Discriminatory power depends on the choice of target words, vector weighting and vector comparison method: usually, a similarity value of 0 is rare. A distributional relation is very general and less specific than most lexical semantic relations. Notice, however, that this relation also returns a *grade* of relatedness.

As far as *significance* is concerned, distributional similarity (the similarity of the contexts in which the words are found to occur) is a useful concept. For people working on real-world tasks such as finding a word/sentence/document similar to a query word/sentence or document (in applications involving information retrieval from a database or from the World Wide Web) there is no denying that such a relation is useful and is worth researching. Other applications will be listed in section 4.

Let us accept the standpoint that our relation qualifies as a lexical relation and let me call this relation the *distributional compatibility relation* (DCR) for reasons clarified later in this paper.

3.1 From crisp to fuzzy relations

A relation (including the relations of lexical semantics) usually represents the presence or absence of interconnectedness between the elements of two or more sets. In a simple binary relation we have two sets (X and Y), and the relation R(X,Y) will tell us whether an element of X is related to an element of Y.

Mothers / *Sons*	Harry	Oliver	Jack
Olivia	*1*	*1*	*0*
Amelia	*0*	*0*	*1*
Jessica	*0*	*0*	*0*

Table 2: Mothers – sons relation represented in a table

Consider the data in table 2 as an example. Olivia has two sons: Harry and Oliver; Amelia has one son: Jack; Jessica has no sons. The first set contains the mothers; the second set contains the sons. By introducing the *mothers-sons* relation on these two sets, we get ordered pairs that define the resulting *mothers-sons* set: {(Olivia, Harry), (Olivia, Oliver), (Amelia, Jack)}.

In some cases, more than two sets are involved in a relation. More importantly for our purposes, we can also relate the elements of the *same* set using a relation; in lexical semantic relations, we have a single set of words, which contains the lexical items of a language. Let us suppose that we label this set of words using W. A lexical relation will be a subset of the Cartesian product W x W. By constructing the Cartesian product W x W, we create a set of all ordered pairs that contain two words of the language (e.g. (*cold, cold*), (*cold, cool*), (*cold, hot*), etc.) [4] An important feature of these relations is that a pair is either an element of the result set (when the relation holds) or it is not.

A striking, but poorly documented feature of vector-space comparisons is that they result in *fuzzy sets*. The members of a binary fuzzy set are pairs that belong to this set to a certain degree. The *degree of membership* (also known as the *membership grade* or *membership value)* is in the closed interval [0,1]. In general, an element with a membership of 1 will be the best representative (or one of the best representatives) of the fuzzy set (cf. prototype theory). Elements with a 0.9, 0.75, 0.0001 etc. membership values are also elements of the fuzzy set, whereas in a traditional "crisp" set, x is either a member of set X (x∈X) or not.

The distributional relation is a fuzzy relation. The distributionally most similar pairs will have a higher degree of membership in the fuzzy set. We define this set by listing all R(W,W) pairs together with membership values, where the elements of set W are the target words of the experiment. For example, I have measured the compatibility of the words *cool*, *hot* and *cold* to be {*0.318*/(cool,cold), *0.259*/(cool,hot), *0.321*/(cold,hot)} in a vector space experiment using a 100 million word sample of Wikipedia, a 1+1 word context window, 8000 context words (the 8000 most frequent words of the corpus except for function words), pPMI weighting and cosine similarity measurement. Notice that the *degree of membership* values prefix the list of word pairs in this notation. A tabulated format is another convenient way of representing fuzzy relations, as shown in table 3: the columns and rows represent the target

---

[4] Due to this ordering, (*cold, cool*) and (*cool, cold*), for instance, are different pairs. In the case of symmetric relations (see below), ordering becomes less obvious and semantically irrelevant, but it remains a formal issue and it is also an inherent feature of all relations, whether they are symmetric or not.

items and the entries in the table show the degree of membership of the corresponding pair in the fuzzy set created by the DCR.

lexical items	cool	hot	cold
cool	1	0.259	0.318
hot	0.259	1	0.321
cold	0.318	0.321	1

Table 3: Pairwise DCR values for all possible pairs that contain cool, hot *and* cold

3.2 Reflexivity

A traditional ("crisp") relation R(X,X) is *reflexive iff* (if and only if) $(x,x) \in R$ for each $x \in R$, i.e. each element is related to itself.

In lexical semantics, synonymy is reflexive (e.g. *book* is a synonym of *book*). Antonymy is not reflexive (e.g. *cold* is not an antonym of *cold*). Hyperonymy, hyponymy and meronymy are not reflexive: they are built on the notion of *inclusion* rather than *identity* (using Cruse's terminology), which requires a semantic difference to exist, and this difference will also function as a contrasting meaning element when we test for reflexivity (therefore, *furniture* cannot serve as a hyperonym of *furniture*, and *leg* cannot be a meronym of *leg*).

Klir and Yuan (1995: 130) point out that reflexivity can be extended to fuzzy relations in the following way: R(X,X) is reflexive *iff* R(x,x)=1 for all $x \in X$, i.e. each element is maximally related to itself. The distributional compatibility relation is reflexive: the measured similarity of a vector to the same vector is maximal. From this perspective, DCR resembles synonymy, a relation expressing identity.

3.3 Symmetry

A crisp relation R(X,X) is *symmetric iff* from $(x,y) \in R$ it follows that $(y,x) \in R$, where $x \in X$ and $y \in X$. Klir and Yuan (1995: 130) argue that reflexivity should be extended to fuzzy relations in the following way: a fuzzy relation is symmetric *iff* R(x,y)=R(y,x) for all $x,y \in X$.

Synonymy is symmetric since if (and only if) *cool* is a synonym of *cold* then *cold* is also a synonym of *cool*. Antonymy is symmetric, too. Hyponymy, hypernymy and meronymy are not symmetric.

The distributional compatibility relation is symmetric: comparing *cold* to *cool* and comparing *cool* to *cold* result in the same degree of membership values. In this respect, DCR resembles synonymy and antonymy.

3.4 Transitivity

A crisp relation R(X,X) is *transitive iff* $R(x,z) \in R$ whenever $R(x,y) \in R$ and $R(y,z) \in R$ for at least one y, where $x,y,z \in X$. In other words, if a transitive relation holds between *x* and *y* and also between *y* and *z*, then the relation also holds between *x* and *z*.

I start with synonymy again. In a synonym dictionary, we find groups of words that contain synonymous items. Transitivity is supposed to hold together the entire group. If *cold* is a synonym of *cool* and *cool* is a synonym of *icy* then *cold* becomes a synonym of *icy*. These

groupings (classes) have a special role in capturing meaning. In WordNet (a large, publicly available lexical database, a major resource for Natural Language Processing; see Fellbaum 1998), the main organizing relation is synonymy, too: synonyms are grouped together into synonym sets or "synsets". The compilers' idea has been to represent senses by synonym sets; synonym sets are supposed to be the descriptions of word meaning. In WordNet, synonymy is a lexical relation (it holds between lexical items rather than concepts) and it is transitive: if x is synonymous with y and, at the same time, y is a synonym of z, then x,y and z will all be listed in the same synonym set, which also means that x and z are synonyms (and co-listing words in the same synonym set is the only way to make them synonymous). Both in the case of thesauri and WordNet, the transitivity of synonymy is part of their system design: in the former case, the user can find information by taking and considering the information found in the similarity classes, in the latter case, semantic relations are defined over synonym sets.

This approach to synonymy does not automatically follow from the treatment of synonymy in the linguistic literature, however. A basic approach would be to treat synonymy as a relation between words that have the same meaning in some or all contexts (which is the approach most relevant to distributional studies, too). The next definition is from Cruse: synonyms are the "lexical items whose senses are identical in respect of 'central' semantic traits, but differ, if at all, only in respect of what we may provisionally describe as 'minor' or 'peripheral' traits" (Cruse 1986: 267). In his 2011 book, Cruse highlights that synonyms have "construals whose semantic similarities are more salient than their differences" (Cruse 2011: 142). Lyons (1981: 50-51) argues that synonymy can be *full* (if and only if all their meanings are identical), *total* (when they are synonymous in all contexts) and *complete* (if and only if they are identical in all relevant dimensions of meaning). *Absolute* synonyms are full, total and complete while partial synonyms satisfy one or any two of the above criteria. In distributional experiments, a DCR value of 1 would probably indicate total synonymy.

Cruse (2011) uses the notion of *normality* in defining absolute synonymy: "for two lexical items X and Y, if they are to be recognized as absolute synonyms, in any context in which X is fully normal, Y is, too" (Cruse 2011: 142). Later he adds that "absolute synonyms are vanishingly rare, and do not form a significant feature of natural vocabularies" (Cruse 2011: 143). Whether at least near synonymy holds is determined by the presence of semantic differences. Minor differences that do *not* destroy synonymy include (Cruse 2011: 145):
- neighbouring values on a scale of degree,
- adverbial specializations of verbs (e.g. *drink – quaff*),
- aspectual differences,
- differences in prototype (e.g. *brave* – physical vs. *courageous* – intellectual/moral).

Transitivity is not guaranteed by the very nature of the synonymy relation: if x is synonymous with y, and y is a synonym of z, then x and z may or may not be evaluated as synonyms, because the difference between x and y will be combined with the difference between y and z, and the resulting distance between x and z may be beyond what we consider a 'minor' difference.

Hyponymy is transitive as far as we treat it as a logical notion (if x is a hyponym of y and y is a hyponym of z then x is a hyponym of z), and it is also transitive in *WordNet* (where we can build *is-a* hierarchies by following hyponymy links). As far as natural language examples are concerned, transitivity is not guaranteed, however (Cruse 2011: 136):
a) A hang-glider is a type of glider.
b) A glider is a type of aeroplane.

If hyponymy were transitive, it would follow from a) and b) that a hang-glider is an aeroplane, but we do not agree with that. Cruse suggests the following explanation: "for

informants to assent to statements like *A is a type of B*, it is sufficient that a prototypical A should fall within the category boundaries of B: it is not necessary that all As should be Bs" (Cruse 2011: 136). On this account, transitivity is absent because a hang-glider is not a prototypical aeroplane.

We also expect meronymy to be transitive: if *x* is part of *y,* and *y* is part of *z*, then *x* should be part of *z*. Real-life examples are sobering again; the following is from Cruse (2011: 141):

a) Fingers are parts of the hand.
b) The hand is a part of an arm.

It would be a logical conclusion that *fingers* are parts of an *arm*, too – but we would never say that. This breakdown of transitivity is connected to the notion of *attachment:* x (*fingers*) is immediately attached to *y* (*hand*), but not to *z* (*arm*) (Cruse 2011: 141).

We can argue that antonymy is not transitive without much explanation: if *x* is an antonym of *y,* and *y* is an antonym of *z*, then we do not want to argue that *x* is an antonym of *z* by the nature of this relation.

Let us return to the examination of the distributional compatibility relation. Transitivity is extended to fuzzy sets through the notion of "max-min transitivity", which is based on max-min compositions as shown in Klir and Yuan (1995: 125-130): the degree of membership produced by the fuzzy R(x,z) relation must be equal or greater than the degree of membership produced by R(x,y) or R(y,z), whichever is greater for all x,y,z∈X. The distributional compatibility relation is *not* (max-min) transitive. Let us just take a single example shown in table 3 above. Consider the *0.318*/(cool,cold) element and the *0.321*/(cold,hot) element of the DCR set. If DCR were transitive, a *cool-hot* relation would return a value greater than or equal to *0.321* – whereas the actual membership grade for (*cool*,*hot*) is 0.259. This situation can easily be repeated with other, randomly chosen words.

3.5   Overview of properties

Table 4 summarizes the properties of lexical relations (also including their WordNet versions) and those of the distributional compatibility relation (DCR).

	type	reflexivity	symmetry	transitivity
synonymy	*crisp / ?*	✓	✓	*x*
WN synonymy	*crisp*	✓	✓	✓
antonymy	*crisp / ?*	*x*	✓	*x*
WN antonymy	*crisp*	*x*	✓	*x*
hyponymy/hyperonymy	*crisp / ?*	*x*	*x*	*x*
WN hyponymy/hypernymy	*crisp*	*x*	*x*	✓
meronymy	*crisp / ?*	*x*	*x*	*x*
WN meronymy	*crisp*	*x*	*x*	✓
DCR	*fuzzy*	✓	✓	*x*

Table 4: Properties of selected lexical relations

Reflexive, symmetric and transitive relations, including WordNet's crisp transitive synonymy relation, are "equivalence" relations. If distributional similarity resulted in a reflexive, symmetric and transitive fuzzy relation, it would be an equivalence relation, too, and we could call it a fuzzy "similarity" relation – a term reserved for fuzzy relations that exhibit all of the

above properties (cf. Klir & Yuan 1995: 133). DCR lacks transitivity, however, which also affects the classification of this relation: it belongs to the group of fuzzy *compatibility* relations.

4   Conclusion

The earliest vector-space models of representing semantic information were used for finding relevant documents in Information Retrieval (Salton 1971). Question answering (e.g. Tellex et al. 2003) and document clustering (e.g. Manning et al. 2008) may be implemented in a similar way. Systems designed around the distributional compatibility relation as portrayed in this paper are used for word clustering and disambiguation (Schütze 1998), thesaurus generation through automatized discovery and clustering of word senses (Crouch 1988, Pantel & Lin 2002), named-entity recognition (Vyas & Pantel 2009), etc. Pennacchiotti et al. (2008) use Distributional Semantics in a cognitive semantic context: they propose a method for extending FrameNet's scope by covering more (potentially: frame-evoking) lexical items through distributional lexical unit induction. It is also interesting to see that a vector-space tool can be better at solving a TOEFL test than humans are (Rapp 2003). All the above applications are based on the Distributional Compatibility Relation, although the authors do not identify and analyse the underlying lexical relation exploited for their computational needs.

DCR is a very general relation and connects much more lexical items than lexical semantic relations usually do. As a fuzzy relation, DCR does so in a quantified manner, however. The type of (paradigmatic) distributional relation captured by DCR corresponds well to people's intuitive notion of word similarity, too: human subjects' decisions on the degree of word similarity correlate with the DCR values returned in vector-space experiments (see Rubenstein & Goodenough 1965, Miller & Charles 1991 for English data and Tóth 2013 for Hungarian results). In Tóth (2013) I argue that synonymy is a major factor of judging the degree of 'similarity' in human experiments, while other types of association between words may also play a role in the absence of (near) synonymy. I also note that lexical ambiguity plays a role and decreases the average human similarity scores – the same phenomenon can also be seen in vector-space experiments.

The lack of transitivity in the case of DCR has at least two consequences. A methodological consequence is that planning, interpretation and further processing require attention since nothing is granted about the R(x,z) DCR relation even when R(x,y) and R(y,z) are known. A terminological consequence is that the relation should be called a *compatibility* relation.

Finally, as DCR is a *fuzzy* relation, calculations with it are more complex and demanding than operations with crisp relations. Klir and Yuan (1995) offer a solid mathematical background. Regrettably, linguistic research involving fuzzy relations is extremely rare while many natural language phenomena seem ideal targets for (re)interpretation in a fuzzy framework.

References

Bullinaria, J.A. & Levy, J.P. (2012): Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming and SVD. *Behavior Research Methods* 44, 890-907.

Crouch, C.J. (1988): A cluster-based approach to thesaurus construction. In *Proceedings of the 11th Annual International ACM SIGIR Conference,* 309-320.

Cruse, D.A. (1986): *Lexical Semantics*. Cambridge, UK: Cambridge University Press.

Cruse, D.A. (2011): *Meaning in language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Fellbaum, C. (ed.) (1998): *WordNet: An Electronic Lexical Database*. Cambridge, Massachusetts: MIT Press.

Kiefer, F. (2007): *Jelentéselmélet* [Semantic Theory] (2nd edition). Budapest: Corvina.

Klir, G.J & Yuan, B. (1995): *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Upper Saddle River, New Jersey: Prentice Hall.

Kornai A., Halácsy P., Nagy V., Oravecz Cs., Trón V. & Varga D. (2006): Web-based frequency dictionaries for medium density languages. In: Kilgarriff, A. & Baroni, M. (ed.): *Proceedings of the 2nd International Workshop on Web as Corpus,* 1-8.

Landauer, T.K. & Dumais, S.T. (1997): A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104(2), 211-240.

Lyons, J. (1981): *Language, Meaning and Context.* London: Fontana.

Lenci, A. (2008): Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics* 20(1), 1-31.

Manning, C.D., Raghavan, P. & Schütze, H. (2008): *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.

Miller, G.A. & Charles, W.G. (1991): Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1), 1-28.

Pantel, P. & Lin, D. (2002): Discovering word senses from text. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 613-619.

Pennacchiotti, M., Cao, D.D., Basili, R., Croce, D. & Roth, M. (2008): Automatic induction of FrameNet lexical units. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-08)*, 457-465.

Rapp, R. (2003): Word sense discovery based on sense descriptor dissimilarity. In: *Proceedings of the Ninth Machine Translation Summit*, 315-322.

Resnik, P. (1995): Using information content to evaluate semantic similarity. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* San Francisco, CA: Morgan Kaufmann Publishers Inc., 448-453.

Rubenstein, H. & Goodenough, J.B. (1965): Contextual correlates of synonymy. *CACM* 8(10), 627-633.

Salton, G. (1971): *The SMART retrieval system: Experiments in automatic document processing*. Upper Saddle River, New Jersey: Prentice-Hall.

Schütze, H. (1998): Automatic word sense discrimination. *Computational Linguistics* 24(1), 97-124.

Tellex, S., Katz, B., Lin, J., Fern, A. & Marton, G. (2003): Quantitative evaluation of passage retrieval algorithms for question answering. In: *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 41-47.

Tóth, Á. (2013): Az ember, a korpusz és a számítógép: Magyar nyelvű szóhasonlósági mérések humán és disztribúciós szemantikai kísérletben [Humans, corpora and computers: Hungarian word similarity measurements in human and distributional semantic experiments]. *Argumentum* 9, 301-310.

Turney, P.D. (2001): Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In: *Proceedings of the Twelfth European Conference on Machine Learning (ECML-01)*, Freiburg, 491-502.

Vyas, V. & Pantel, P. (2009): Semi-automatic entity set refinement. In: *Proceedings of NAACL-09*, 290-298.

Widdows, D. (2004): *Geometry and Meaning*. Stanford, California: Center for the Study of Language and Information.

Ágoston Tóth
University of Debrecen
Institute of English and American Studies
H-4010 Debrecen
Pf. 73
toth.agoston@arts.unideb.hu