Does Latent Semantic Analysis Reflect Human Associations?

Tonio Wandmacher
Institute of Cognitive Science
University of Osnabrück
Albrechtstr. 28,
49069, Germany
twandmac@uos.de

Ekaterina Ovchinnikova
Institute of Cognitive Science
University of Osnabrück
Albrechtstr. 28,
49069, Germany
eovchinn@uos.de

Theodore Alexandrov
Center for Industrial Math.
University of Bremen
Bibliothekstr. 2,
28359, Germany
theodore@math.uni-bremen.de

Abstract

In the past decade, Latent Semantic Analysis (LSA) was used in many NLP approaches with sometimes remarkable success. However, its abilities to express semantic relatedness have been not yet systematically investigated. In this work, the semantic similarity measures as provided by LSA (based on a term-by-term matrix) are compared with human free associations. Three tasks have been performed: (i) *correlation* with human association norms, (ii) *discrimination* of associated and unassociated pairs and (iii) *prediction* of the first human response. After a presentation of the results a closer look is taken to the statistical behavior of the data, and a qualitative (example-based) analysis of the LSA similarity values is given as well.

1 Introduction

In its beginnings, Latent Semantic Analysis aimed at improving the vector space model in information retrieval. Its abilities to enhance retrieval performance were remarkable; results could be improved by up to 30%, compared to a standard vector space technique (Dumais, 1995). It was further found that LSA was able to retrieve documents that did not even share a single word with the query but were rather semantically related.

This finding was the headstone for many subsequent researches. It was tried to apply the LSA approach to other areas, such as automated evaluation of student essays (Landauer et al., 1997) or automated summarization (Wade-Stein and Kintsch, 2003). In (Landauer and Dumais, 1997), even an LSA-based theory of knowledge acquisition was presented.

Many researches have made claims on the analytic power of LSA. It is asserted that LSA does not return superficial events such as simple contiguities, but is able to describe semantic similarity between two words (cf. Wade-Stein and Kintsch, 2003). The extracted word relations are referred to as latent, hidden or deep (cf. Landauer et al. 1998), however, only few articles address the nature of this deepness.

Some steps in this direction were taken by Landauer and Dumais (1997) and later by Rapp (2003). In these works, LSA-based similarities were used to solve a synonym test, taken from the TOEFL[1]. However, the results achieved can only be seen as a first indication for the capacity of LSA.

We try to make a little step further. The main objective of this work is therefore not improvement, but evaluation and a better understanding of the method. The present investigation is carried out in the framework of the *Lexical Semantics Workshop: Bridging the gap between semantic theory and computational simulations* at ESSLLI'08[2], which is devoted to discovering of the relationships between word spaces computed by corpus-based distributional models and human semantic spaces. In this paper, we concentrate on exploration of the correlation between the LSA semantic similarity measures and human free associations[3].

[1] Test Of English as a Foreign Language

[2] *http://wordspace.collocations.de/doku.php/esslli:start*.

[3] See *http://wordspace.collocations.de/doku.php/data:correlation_with_free_association_norms*.

The paper is structured as follows. In section 2 we briefly introduce the LSA method. We then (section 3) give an overview on related work exploring the semantic and associative capacities of LSA. In section 4 we describe the workshop tasks on free associations and provide the results that we have obtained. In section 5 we present a detailed quantitative and qualitative analysis of the achieved results. In the final section we draw conclusions and discuss open issues.

2 Latent Semantic Analysis: Method

LSA is based on the vector space model from information retrieval (Salton and McGill, 1983). Here, a given corpus of text is first transformed into a term$\times$context matrix A, displaying the occurrences of each word in each context. The decisive step in the LSA process is then a *singular value decomposition* (SVD) of the matrix which enables the mapping of this matrix to a subspace. The resulting lower-dimensional matrix is the best reduced-rank least-squares approximation of the original matrix. According to the proponents of LSA this reduction plays an important role for the uncovering of important relations which are hidden (or '*latent*') in the original matrix.

In its original form (cf. Deerwester et al. 1990), LSA is based on a co-occurrence matrix of terms in documents; such a matrix is normally extremely sparse[4], and it is obvious that this matrix grows with the number of documents in the training corpus. Moreover, the notion of document varies strongly over different corpora: a document can be only a paragraph, an article, a chapter or a whole book, no hard criteria can be defined. Therefore, another type of matrix can be used, as described by (Schütze, 1998) and (Cederberg and Widdows, 2003), which is not based on occurrences of terms in documents but on other co-occurring terms (term$\times$term-matrix). The two sets of terms need not be identical, one can also define a (usually smaller) set of *index terms* $I = (i_1, ..., i_m)$. The size of the matrix is then independent of the size of the training data, so that much larger corpora can be used for training.

After applying SVD, each word is represented as a vector of k dimensions, and for every word pair w_i, w_j of the vocabulary we can calculate a similarity value $\cos(w_i, w_j)$, based on the $cosine$ between their respective vectors.

Figure 1 summarizes the processing steps involved in training an LSA-based semantic space.

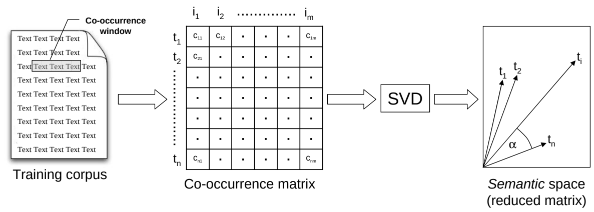

Figure 1: Schematic overview on the generation of an LSA-based semantic space

In the following we will apply this kind of model, based on an SVD-reduced term-by-term co-occurrence matrix, to the different tasks, and we will compute term similarity by measuring the cosine of term vectors in the reduced space.

3 Related Work

Considering the large number of works applying LSA for various purposes, it is a surprising matter of fact that only little research was done in order to better understand the kind of relatedness that distributional approaches like LSA are able to reflect.

In (Landauer and Dumais, 1997) a theory of knowledge acquisition and representation is presented, assuming that the meaning induction mechanisms performed by LSA are very similar to those of humans. As an example task, LSA is applied to solve the TOEFL synonym test, and it could be shown that the results of LSA are the same as those of the average foreign student passing the *TOEFL* (LSA: 64.4%; human participants: 64.5%). In (Rapp, 2003), an LSA model, based on a term$\times$term matrix and trained on much more data, was able to solve even 92.5% of the synonym questions.

In (Wandmacher, 2005) term relations (nearest neighbors) generated by LSA as well as a first-order co-occurrence approach are systematically analyzed and compared. It could be shown that only a small part of the relations are systematically related (e. g. by hyponymy or synonymy), the largest part of the nearest neighbors of a term were loose associations.

[4] In (Wandmacher, 2005) a matrix was used that had less than 0.08% non-zero elements.

While the error rate for LSA was lower than for the first-order approach, no substantial differences between the results of the two methods could be determined. It could however be observed that a crucial factor for the quality of the nearest neighbors is the specificity of a term.

The correspondence of human association and first-order co-occurrence was investigated in (Wettler et al., 2005). Here, 100 stimulus words from the Kent-Rosanoff word association test with associations selected from the *Edinburgh Associative Thesaurus* (cf. the following section) were predicted with the help of associationist learning theory. This theory states that the associative strength between two events i and j increases by a constant fraction of the maximally possible increment whenever these two events cooccur. This idea was applied to the coocurence of terms in the British National Corpus. The achieved results appear to be very promising. For 29 of the 100 stimulus words the model produced the primary associative response.

4 Tasks and Results

The main goal of our analysis was to find out to what extent free associations can be explained and predicted by statistical similarity measures computed by LSA. In order to address this issue, the workshop organizers have proposed the three tasks described below. Different training and test data sets containing association pairs were provided for each of the three tasks[5].

Free associations are the first words that come to the mind of a native speaker when he or she is presented a stimulus word. The degree of free association between a stimulus (*cue*) and a response (*target*) is quantified by the percentage of test subjects who produced *target* when presented with *cue*.

4.1 Method

For training we used 108M words from two British newspapers (*The Times*, *The Guardian*) of the years 1996 to 1998. Using the *Infomap* NLP toolkit[6], developed at Stanford University's CSLI, we generated a term×term co-occurrence matrix of size 80.000×3.000, closed-class words not occurring in the test data were disregarded. The vocabulary ($|V| = 80.000$) as well as the index terms ($|I| = 3.000$) were determined by corpus frequency, and terms occurring less than 24 times in the corpus were excluded from the vocabulary. We calculated several spaces for co-occurrence windows of $\pm 5, \pm 25, \pm 50, \pm 75$ words, respectively; the window did not cross article boundaries. The results presented in the following are obtained using the ± 75-window space, if not mentioned otherwise. The matrix was reduced by SVD to 300 dimensions; term similarity was determined by measuring the truncated cosine of the angle between the corresponding term vectors. Since negative cosine values can occur but are meaningless for similarity measurements (i. e. terms having a negative similarity value are not more dissimilar than those having a value of 0), negative values are set to 0.

4.2 Discrimination

This task consists in discrimination between three classes of association strengths:

- the `FIRST` set – strongly associated cue-target pairs given by more than 50% of test subjects as first responses,
- the `HAPAX` set – cue-target pairs that were produced by a single test subject,
- the `RANDOM` set – random combinations of headwords from EAT that were never produced as a cue-target pair.

For each of the cue–target pairs, excluding those which contained terms not being present in our vocabulary, we have computed LSA similarity values. We obtained results for 300 of the 301 suggested pairs of the test data set, using a discrimination threshold of 0.23 between `FIRST` and `HAPAX`, and a threshold of 0.02 for discrimination between `HAPAX` and `RANDOM`, which showed to be optimal for the training data set. The following table shows the discrimination results for all classes considered[7]:

[5] The data sets are based on a database of English association norms, the Edinburgh Associative Thesaurus (EAT). Cf. also: *http://www.eat.rl.ac.uk/*.

[6] *http://infomap-nlp.sourceforge.net/*

[7] `HoR` stands for `HAPAX` or `RANDOM`;
Accuracy = Right * 100 / (Right+Wrong).

	Right	Wrong	Accuracy
FIRST (th=0.23)	50	50	50%
HAPAX (th=0.02)	63	32	68%
RANDOM	68	17	78.2%
Total (F/H/R)	181	119	60.33%
HoR	189	11	94.5%
FIRST/HoR	239	61	79.66%

4.3 Correlation

The task is to predict free association strength (ranging from 0 to 1) for a given list of cue-target pairs, quantified by the proportion of test subjects that gave this target as a response to the stimulus cue. Pairs in the training and test set have been selected by stratified sampling so that association strength is uniformly distributed across the full range.

We have computed LSA similarity values for 239 of the 241 suggested pairs, achieving the *Pearson* correlation of 0.353 between the human scores and the LSA values; the *Kendall* correlation coefficient is 0.263. Both are significant with a p-value$<$0.01.

4.4 Response Prediction

In this task, models have to predict the most frequent responses for a given list of stimulus words. The data sets contain cue-target pairs with the association strength of the target response and the association strength of the second (unknown) response. The cues were selected from the EAT in such a way that the association strength of the dominant response must be $\geqslant$0.4, and at least three times as high as that of the second response. For the first response prediction we have computed the LSA similarity between cues and all terms in our vocabulary for 199 pairs from 201. The resulting average rank of the correct response is 51.89 (if the correct response is not among the suggested candidates, it is assigned rank 100 regardless of the number of suggestions). The distribution of the target ranks is as follows:

Target rank	1	2	3	4	5	6	7-99	100
Frequency	31	10	7	5	6	7	43	89

4.5 Co-occurrence Window

The size of the co-occurrence window on which the input matrix is based is a crucial factor establishing relatedness. Previous works using term$\times$term matrices employed rather small windows: Lund and Burgess (1996) used windows of ± 8 words, Cederberg and Widdows (2003) used ± 15 words and Rapp (2003) used a window of ± 2 words only.

To get a better understanding of this parameter, we calculated models for different window sizes ($\pm 5, \pm 25, \pm 50, \pm 75$ words) and tested them on the above described tasks[8].

	± 5	± 25	± 50	± 75
Correlation (*r*)	0.254	0.326	0.347	0.354
Disc. (Acc.)	54.67	55.67	58.67	60.33
Pred. (Av. Rank)	62.61	54.11	52.69	51.89

The results for all three tasks are quite univocal: The performance improves with the size of the co-occurrence window. This is of course only a first and rather coarse-grained observation, but it indicates that this parameter deserves more attention in the application of distributional models.

5 Analysis of the Results

In this section, we will analyse our results by comparing the similarity values produced by LSA with the human scores.

5.1 Quantitative Analysis

5.1.1 Correlation Analysis

As the reliability of a statistical analysis depends on the size of considered sample, in this section we examine not only the test set (of size 239) but the test and training sets altogether (of size 278). Since the distributions of human values both of the training and test sets are the same, the training values can be regarded as sampled from the same general population.

The calculated Pearson and Kendall correlation coefficients are close to those reported for the training set (see section 4), and are 0.353 and 0.263, correspondingly. Both are significant with $p < 0.01$. The Spearman correlation is 0.384 and is also significant. This confirms a significant monotone and, moreover, linear dependence between the human and LSA values.

As an initial step, let us visually examine figure 2 which depicts for each pair of terms (i) its human and LSA values against its ordinal number (rank),

[8]Due to computational restrictions we were not able to calculate co-occurrence matrices for windows larger than ± 75.

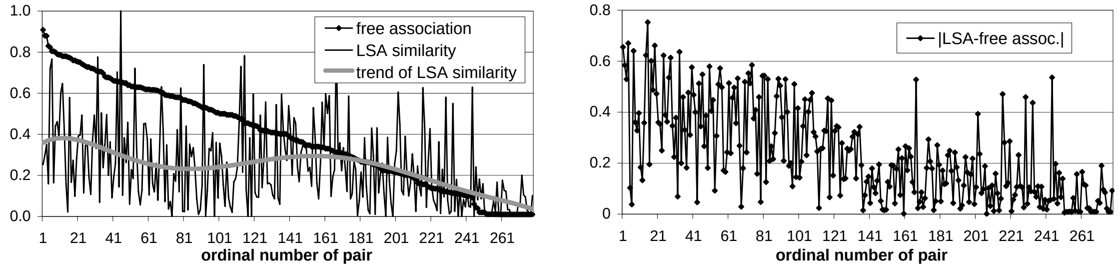

Figure 2: The human and LSA values (left) and their absolute differences (right).

(ii) the absolute difference between these two values, where the pairs are sorted by their human values. The behavior of the observed characteristics seems to differ at around the 136'th pair (human value $\approx$ 0.4). For the pairs with higher ranks (i.e. ≥ 0.4) the LSA values approximate human values on average and the difference between the LSA and human values looks like noise with constant variance. For the pairs with lower ranks the averaged LSA values show no clear dependence on the human values.

Based on these observations, we state the hypothesis of separation of the whole data set into two groups: *high human association group* G_1 with the human values >0.4 and *low human association group* G_2 with the values <0.4, where there is no correlation between LSA and human values in the first group G_1 in contrast to G_2.

For testing the stated hypothesis, we calculated the following characteristics between the human and the LSA values in each group and for the whole data set: (i) mean absolute difference, (ii) Pearson and Kendall correlation coefficients, and their significance and, furthermore, (iii) in each group we tested the hypothesis of randomness of the LSA values (using the *Abbe criterion*). The results are given in table 1; they show that in the high human association group G_1 there is no dependence between the human and the LSA values; moreover the mean absolute difference between these values is large (0.35), and it considerably exceeds the mean difference over the whole data set (0.23). At the same time, the results for the low human association group G_2 indicate a significant linear correlation producing small mean absolute difference (0.12).

Thus, we confirmed our hypothesis of difference between the groups G_1 and G_2. The existence of these groups demonstrates the fact that low association can be easily established, whereas correct estimation of high association strength seems to be complicated (cf. section 5.2). This observation conforms with the good discrimination results reported for the `RANDOM` group and bad results for the `FIRST` group. We would like to note that the Pearson and Kendall correlations between the LSA and human values calculated for the prediction data set (where all human values $\geq$0.4) are insignificant, which additionally confirms our hypothesis of independence between the LSA similarity and the human association values for pairs with a high latter value.

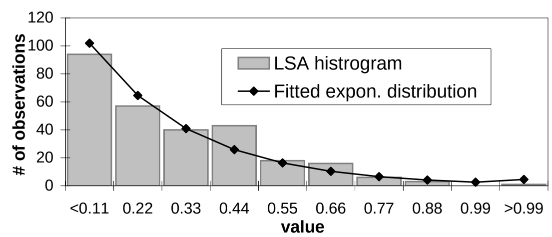

Figure 3: Histogram of the LSA values and the fitted exponential density.

The next interesting conclusion can be derived considering the histogram of LSA values (see figure 3; recall that the human values are uniformly distributed). Though the hypothesis of an exponential distribution of the LSA values is rejected with p-value<0.01, it becomes obvious that LSA underestimates association strength as compared with human scores. Moreover, all but one of the 12 pairs with the highest LSA values ($\geqslant$0.63) have high human values ($\geq$0.45), see table 2. Thus, it is possible that the pairs with high LSA values also have high human values but not vice versa.

5.1.2 Prediction Analysis

In the following a closer look is taken on the results of the prediction task. First, though for each

Group	Mean abs. diff.	Pearson corr.	Kendall corr.	Randomness of LSA values
G_1	0.35	0.211 ($-$)	0.172 ($+$)	Not rejected (p-value=0.43)
G_2	0.12	0.514 ($+$)	0.393 ($+$)	Rejected (p-value=0.00)
$G_1 \cup G_2$ (whole data set)	0.23	0.353 ($+$)	0.263 ($+$)	Not rejected (p-value=0.07)

Table 1: Intragroup properties, the signs $-$ or $+$ indicate significance of the correlation coefficients with p-value<0.01.

cue	target	human value	LSA value
ha	ha	0.66	1.00
inland	revenue	0.31	0.84
four	five	0.45	0.78
question	answer	0.71	0.78
good	bad	0.80	0.77
grammar	school	0.53	0.74
below	above	0.47	0.73
daughter	son	0.63	0.72
vehicle	car	0.82	0.72
parish	church	0.66	0.70
boy	girl	0.78	0.65
sing	song	0.60	0.63

Table 2: The 12 pairs with the highest LSA values.

cue the first association value is at least three time larger than the second association value (see section 4), we do not detect the same effect for LSA. The first and second LSA nearest neighbor values differ in only 1.1 times on average (vs. 8.6 times for the human values). It means that for every cue, the LSA similarity values of the most strongly related terms are very close. Second, it is interesting to note that in the human data when the large first association values ($\geq$0.65) increase, the second association values decrease, see figure 4. For LSA values no such effect is observed. A possible interpretation of this fact is that for humans a first strong association suppresses the others.

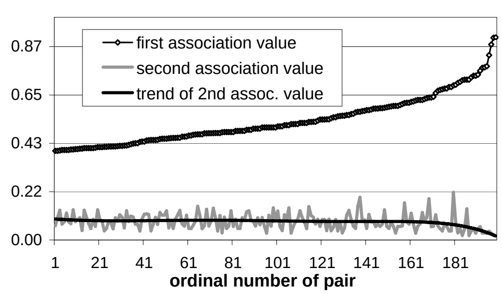

Figure 4: The association values for the prediction task.

5.1.3 Parts of Speech and Lexical-Semantic Relations

Wandmacher (2005) indicates that the quality of the term relations for a given cue may depend on its part of speech, e.g. LSA has found far more meaningful relations for nouns then for adjectives and verbs.[9] We have made different observations: For the correlation task the best result was achieved with adjectives (for adjectives the Pearson correlation is 0.68, for nouns 0.33, and for verbs 0.14) and for the prediction task there is no significant difference.[10]

We have also tagged relations for the prediction task test data set.[11] For syntagmatic relations the standard classification (cf. Cruse, 1986) was used: near-synonymy (association highlights a common part of the meaning, e.g. (*incorrect, wrong*)), opposition (association highlights an opposite part of the meaning, e.g. (*female, male*)), hypo-/hyperonymy (e.g. (*finch, bird*)), co-hyponymy (e.g. (*july, august*)), mero-/holonymy (e.g. (*deck, ship*)).[12] In order to estimate relations between terms belonging to different parts of speech we have distinguished following relations: collocation (e.g. (*wizard, oz*)), attribute-class relation (e.g. (*sugar, sweet*)), predicate-argument (e.g. (*eating, food*)), unsystematic association (which mostly express connection of terms via an implicit predicate, e.g. (*prefect, school*)). The information about the corresponding classes is given in table 3. We acknowledge that any tagging of this kind is highly subjective. Moreover, the number of pairs in some of our classes is definitely not enough to perform an analysis. Nevertheless, we decided to present these results, since the LSA values for the class of oppositions show dis-

[9] These results refer to German.

[10] Morphologically ambiguous words (e.g. *sting* or *shotgun*) were excluded from this analysis.

[11] The prediction data set was chosen because it contains meaningful associations only (cf. section 4).

[12] We do not mention relations that occurred less than 5 times in the data set, e.g. causation, presupposition etc.

tinctively better performance than others.

relation	average rank	number of pairs
n.-syn.	46.98	47
oppos.	24.42	31
hypo.	53.32	22
mero.	58.43	21
co-hyp.	40.50	6
colloc.	77.59	17
attr.-cl.	85.86	7
pred.-arg.	49	13
assoc.	62.65	31

Table 3: Average rank of targets and number of pairs in every class of relations for the prediction task data set.

5.2 Qualitative Analysis

In order to get a better understanding of what kind of information is reflected by LSA, we will take a look at some specific examples. First, we consider the term pairs that have got the highest LSA values ($\geq$0.63, see table 2). Obviously, LSA assigns a similarity of 1 to the pairs where cue and target are identical (e.g. (*ha, ha*)), whereas for human subjects such an association is not necessarily preferential. Then, LSA strongly associates oppositions, e.g. (*question, answer*), (*good, bad*), (*daughter, son*).[13] High LSA estimates for other semantic relations, such as collocations (e.g. (*inland, revenue*)), hyponyms (e.g. (*vehicle, car*)), co-hyponyms (e.g. (*four, five*)) etc., are found to be less regular and more corpus dependent.

The widest range of disagreements between LSA and human evaluations seems to be corpus-related. Since we have used a newspaper corpus, LSA extracted rather specific semantic neighbors for some of the terms. For example, terms from the food domain seem to stand out, possibly because of numerous commercial statements: e.g. for *fresh* the nearest neighbors are (*flavour*, 0.393), (*soup*, 0.368), (*vegetables*, 0.365), (*potato*, 0.362), (*chicken*, 0.36). Thus, the association (*fresh, lobster*) receiving a very low human value (0.01) is estimated by LSA at 0.2.

An interesting effect occurs for associations between some concepts and their salient properties, e.g. (*snow, white*) which is estimated at 0.408 by humans and at 0.09 by LSA. The nearest neighbors found by LSA for *snow* belong to the “weather forecast” domain: (*snowfalls*, 0.65), (*winds*, 0.624), (*weather*, 0.612), (*slopes*, 0.61), (*temperature*, 0.608). It is straightforward to suppose that since the feature of ”being white“ for snow is so natural for our language community, people do not talk much about it in newspapers.

Concerning word senses LSA is known to generate neighbors of the prominent meaning only and to suppress other domains (cf. Rapp, 2003; Wandmacher, 2005). This effect can lead both to over- and to underestimation in comparison with human values. For example the pair (*nurse, hospital*) gets a relatively high LSA value of 0.627 (while the human value is 0.156), because LSA has selected the nearest neighbors for *nurse* from only one (and very specific) domain: (*nurses*, 0.64), (*hospital*, 0.627), (*patient*, 0.597), (*doctors*, 0.554), (*patients*, 0.525). On the other hand, (*eve, adam*) receives only 0.024 by LSA (while the human value is 0.567), because LSA has selected another meaning for the homonym *eve*: (*christmas*, 0.657), (*festive*, 0.535), (*yuletide* 0.456), (*festivities*, 0.453), (*presents*, 0.408).

Besides the already mentioned effects we have noticed some more regularities. It is often the case (for 9 out of 22 collocations in the correlation task data sets) that LSA assigns a low value (< 0.1) to term pairs forming a collocation, e.g. (*peg, clothes*, hum.: 0.225, LSA: 0.001), (*shotgun, wedding*, hum.: 0.402, LSA: 0.06), (*core, apple*, hum.: 0.776, LSA: 0.023). The problem here is that the terms in such collocations have no other overlap in their meanings (e.g. the nearest neighbors for *shotgun* are (*gun*, 0.536), (*pistol*, 0.506), (*shooting*, 0.463), (*shotguns*, 0.447), (*firearms*, 0.445), which most of the time have nothing to do with weddings) and the given collocations are rare in the corpus.

As for the auxiliary words (like prepositions, pronouns and conjunctions), LSA produces rather unstable results. A general observation is that the association strength for such pairs is mostly underestimated because of their low specificity (cf. Wandmacher, 2005). However, there is not enough data in the considered data sets to investigate this effect.

It is worth reminding that the semantic similarity estimated by LSA is symmetric, whereas it is obviously not the case for human scores. For example

[13] 15 from 19 oppositions found in the correlation task data sets have got LSA values $>$0.22.

the association of terms *wrong* and *right* which is assigned an LSA value of 0.493, is estimated by humans at 0.717 in the direction from *wrong* to *right* and at 0.42 in the opposite direction.

6 Discussion and Conclusion

In this paper, we have described the results of three tasks[14] in order to get an understanding of the relationships between human free associations and similarity measures produced by LSA. In reply to the title's question, we have to report that no strong correlation between human associations and LSA similarity could be discovered. Likewise, our prediction results are relatively bad (as compared to those by Wettler et al. 2005). However, Wettler et al. (2005) have used a lemmatized corpus, which is not the case for our study. The effect of lemmatization on the training data should be investigated in more detail.

We did however investigate the effect of the size of the co-occurrence window, and we have found larger windows (of around ± 75 words) to provide significantly better results in all tasks than windows of smaller sizes.

Another effect that we have observed is that LSA estimates for weakly associated terms are much closer to those of humans than for strongly associated terms. Then, we have reported a regular underestimation by LSA. We have also pointed out the fact that the clear preference for one association in human responses is not established by LSA; the average distance between the first and the second LSA neighbor is much lower (section 5.1.2).

Furthermore, we have added some comments on the LSA similarity estimates for different parts-of-speech and kinds of lexical relations. Finally, we have tried to establish some qualitative regularities in the disagreements between LSA and human estimations (section 5.2).

For further investigation it will be interesting to look not only at the first words coming into the mind of a subject after being presented a cue but also at further associations. This will probably help to understand to which domains do these associations belong and to compare these domains with the domains found for the cue by LSA.

[14]The files containing our results can be found at *http://www.ikw.uos.de/∼twandmac/FA-Results-WOA.zip*.

References

S. Cederberg and D. Widdows. 2003. Using LSA and Noun Coordination Information to Improve the Precision and Recall of Automatic Hyponymy. In *Proc. of CoNNL*.

D. A. Cruse. 1986. *Lexical Semantics*. Cambridge University Press, Cambridge, G.B.

S. C. Deerwester, S. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. 1990. Indexing by Latent Semantic Analysis. *American Society of Information Science*, 41(6):391–407.

S. Dumais. 1995. Latent Semantic Indexing (LSI): TREC-3 Report. In D. Harman, editor, *Proc. of TREC-3*, volume 500-226, pages 219–230. NIST Special Puclication.

T. K. Landauer and S. Dumais. 1997. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.

T. K. Landauer, D. Laham, B. Rehder, and M. E. Schreiner. 1997. How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In M. G. Shafto and P. Langley, editors, *Proc. of the 19th annual meeting of the Cognitive Science Society*, pages 412–417, Mawhwah, NJ. Erlbaum.

T. K. Landauer, P. W. Foltz, and D. Laham. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*, 25:259–284.

K. Lund and C. Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behaviour Research Methods, Instruments and Computers*, 28(2):203–208.

R. Rapp. 2003. Word Sense Discovery Based on Sense Descriptor Dissimilarity. In *Proc. of the Machine Translation Summit IX*, New Orleans.

G. Salton and M. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, NY.

H. Schütze. 1998. Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97–123.

D. Wade-Stein and E. Kintsch. 2003. Summary street: Interactive computer support for writing. Technical report, University of Colorado.

T. Wandmacher. 2005. How semantic is Latent Semantic Analysis? In *Proc. of TALN/RECITAL'05*, Dourdan, France, 6-10 june.

M. Wettler, R. Rapp, and P. Sedlmeier. 2005. Free word associations correspond to contiguities between words in texts. *Quantitative Linguistics*, 12(2-3):111–122.