Contextual Effects on Metaphor Comprehension: Experiment and Simulation

Benoît Lemaire (Benoit.Lemaire@upmf-grenoble.fr)
L.S.E., University of Grenoble
BP 47
38040 Grenoble Cedex 9, FRANCE

Maryse Bianco (Maryse.Bianco@upmf-grenoble.fr)
L.S.E., University of Grenoble
BP 47
38040 Grenoble Cedex 9, FRANCE

Abstract

This paper presents a computational model of referential metaphor comprehension. This model is designed on top of Latent Semantic Analysis (LSA), a model of the representation of word and text meanings. Comprehending a referential metaphor consists in scanning the semantic neighbors of the metaphor in order to find words that are also semantically related to the context. The depth of that search is compared to the time it takes for humans to process a metaphor. In particular, we are interested in two independent variables : the nature of the reference (either a literal meaning or a figurative meaning) and the nature of the context (inductive or not inductive). We show that, for both humans and model, first, metaphors take longer to process than the literal meanings and second, an inductive context can shorten the processing time.

Introduction

How do humans process a metaphor embedded into a context and how can we account for that by means of a computer simulation? This paper contributes to the answer on both questions. First, we performed an experiment with humans to know more about the way the context can influence the metaphor processing. In parallel, we implemented a computational model of metaphor comprehension, ran it on the same experiment and compared both data.

The kind of metaphor we are interested in are called referential metaphors. As opposed to the largely-studied predicative metaphors like *my lawyer is a shark*, a referential metaphor refers to a concept previously mentioned in the discourse. A referential metaphor therefore follows a piece of text. For instance, A *few days after the judgment, I got a mail from my lawyer. It was the bill. I found it very expensive. The* shark *did not told me that...*

Cognitive theories say that appropriate contexts facilitate the comprehension of the metaphor. Martin (1994) asserts that the comprehension of a metaphor can be determined from the way the context predicts the future occurrence of the metaphor. Gildea and Glucksberg (1983) claim that metaphors are comprehended as easily as literal sentences. However, this claim may only apply to predicative metaphors: various experiments show a higher processing time for referential metaphors (Gibbs, 1994, Noveck et al., 2001, Onishi & Murphy, 1993) or even a worse accuracy in answering comprehension questions (Budiu & Anderson, 2002).

In this paper, we study the way a human and a computational model process texts that have the following format:

- one sentence containing a source word;
- four neutral contextual sentences;
- one sentence containing a reference to the source word;
- an additional sentence.

The following is the English translation of one of our texts. Note the word *scientist* in the 6th line, which refers to the word *child* in the first line.

Aged 10, Pierre is a surprising ***child****.*
Pierre loves books.
He has an answer for everything.
During a gymnastics training, someone said:
"Who invented the Olympics?".
The **scientist** *exclaimed: "It's the Greeks!".*
The instructor turned round.

The two independent variables we study are the *nature of the reference*, either a metaphoric reference (Rm), like in the previous example, or a synonymic reference (Rs), and the *nature of the context*, either metaphorically-inductive (Im) or synonymically-inductive (Is). A metaphorically-inductive context is intended to induce the metaphor whereas a synonymically-inductive context should not. The main idea is to check the effect of an inductive context on the comprehension of a metaphor and whether or not this inductive context is more effective for a metaphor than for a synonym.

A context is said to be inductive if its purpose is to enhance the comprehension of the reference by means of words that are associated with it. To operationalize this idea, we decided that an inductive context should contain at least three words whose semantic similarity with the reference is higher than the semantic similarity with the other kind of reference. For instance, a word W is part of an inductive context for a metaphor M, S being the corresponding synonym, if:

similarity (W,M) > similarity (W,S),

and vice versa.

We relied on Latent Semantic Analysis (Deerwester, 1990) in order to compute the various semantic similarities. Since this tool will be also used as part of our model, we will present it later.

This gives us four versions of each text:

- synonymic reference/synonymic induction (RsIs);
- synonymic reference/metaphoric induction (RsIm);
- metaphoric reference/synonymic induction (RmIs);
- metaphoric reference/metaphoric induction (RmIm).

It is worth noting that the difference between Is and Im is slight: Im is just a rephrasing of Is with 3 or 4 specific words added. The main idea remains identical. We designed 8 French texts, thus a total of 32 versions. The following is the translation of the synonymically-inductive context of the previous example (the synonym is *boy* and the metaphor is *scientist*). Inductive words are underlined.

Aged 10, Pierre is a surprising ***child****.*
He likes teasing <u>girls</u> with his <u>friends</u>.
It does not prevent him from being serious at <u>school</u>.
One day, he was playing with his <u>mother</u>.
Someone said: "Who knows the origin of the Olympics?"
The ***boy/scientist*** *exclaimed: "It's the Greeks!".*
The instructor turned round.

This synonymically-inductive context contains words that are associated with the word *child*. On the opposite, the metaphorically-inductive context contains words that are associated with *scientist*:

Aged 10, Pierre is a surprising ***child.***
He is a very <u>cultured</u> person.
His professors are surprised by so much <u>ingenuity</u>.
His <u>knowledge</u> is well developed.
One day, someone said: "Who knows the origin of the Olympics?"
The ***boy/scientist*** *exclaimed: "It's the Greeks!".*
The instructor turned round.

The dependent variable we defined is the time it takes to process the reference (metaphor or synonym). This is quite a common practice in the literature to model the comprehension difficulties. The less the time taken to read a sentence, the easiest the comprehension. We will see in the next sections how this was implemented for both humans and computer.

First, we will present the human experiment, then our computational model. Secondly, we will compare both data.

Human Experiment

Fifty-nine undergraduate students volunteered to participate to this experiment. All were French native speakers. Subjects read two texts in each of the four conditions. The order of the presentation was counterbalanced across subjects. Texts were presented one line at a time on a computer screen. Subjects had to press the space bar to display the next line. After the 7th line, subjects had to answer two comprehension questions, that we will not detail in this paper. The goal of these questions was to keep subjects concentrated on the comprehension of the text. Two practice texts were presented to subjects beforehand. In order to control both the variation of reading speed and the variation of number of syllables of the line containing the reference, the dependent variable was actually the difference between a theoretical reading time and an effective reading time of the reference line. The theoretical time is the average time it takes for subjects to read a line with the same number of syllables. This information comes from the measure of the reading speed of all the lines subjects had read. If T_{ref} is the effective reading time of the reference line and S_{ref} its number of syllables, T_{total} the total reading time and S_{total} number of syllables, the dependent variable A is the additional time it takes for a subject to process the reference :

$$A = T_{ref} - S_{ref} * T_{total} / S_{total}$$

Results are presented in Table 1. For instance, if the reference was a synonym with a synonymically-inductive context, subjects spent an average of 111 less milliseconds to process the line than for a standard line with the same number of syllables. In the opposite, they spent 120 more milliseconds than a standard line if the context was metaphorically-inductive.

Table 1 : Mean additional time for processing different references in different contexts (human data)

Additional time	Syn. induction (Is)	Met. induction (Im)
Syn. reference (Rs)	-111	120
Met. reference (Rm)	488	326

The first result is concerned with the reference variable. The experiment show that metaphors takes longer to process than synonyms (F(1,58)=22,5; p<.01). This result is coherent with the literature: as we mentioned earlier, it takes usually more time to process a referential metaphor than the literal meaning.

The second result concerns the context variable. The experiment shows that it takes shorter to process a reference when the context is inductive ($F(1,58)=4.5$; $p<.05$): synonymically-inductive contexts profits to the synonyms and metaphorically-inductive contexts profits to the metaphors. Under an inductive context, the mean gain is 231ms for synonyms and 152ms for metaphors. If we admit that the reading duration is a good indicator of the comprehension, it is interesting to note that a slight change in the context affects significantly the comprehension of a reference.

Finally, there was no interaction between the two factors.

These results were quite expected. We will now present our computational model, apply it to the same 32 versions of texts, and compare human and model data.

Simulation

The goal of this section is to present our computational model for simulating the comprehension of referential metaphors. First, we will present the knowledge representation we rely on, then the computational model per se, then how it can simulate the previous experiment.

Knowledge Representation

A computational model of metaphor comprehension needs to be based on a representation of word meaning. In the domain of metaphor or metonymy processing, several formalisms are used in the literature: lexical networks (Duvignau et al., 2002), semantic networks (Markert & Hahn, 2002; Martin, 1994) or connectionnist representations (Thomas et al., 2001). The main problem with these approaches is that they represent a limited set of word meanings. Only a few metaphors can therefore be processed. Another approach consists in representing the meanings of all words, which permits to test the model on any kind of metaphor. Latent Semantic Analysis (LSA) is such a knowledge representation tool (Landauer & Dumais, 1997).

LSA analyses the co-occurrences of words in huge corpus (several millions of words) of unprocessed texts. The idea is more elaborated than just occurrence counting which already proved its limits (see Landauer, 2002 for a nice review on that topic). The main hypothesis is that two words are semantically *similar* if they occur in semantically *similar* paragraphs. In the same way, two paragraphs are considered semantically *similar* if they contain semantically *similar* words. This mutual recursion is solved by a mathematical procedure called singular value decomposition (Deerwester et al., 1990).

Once the whole corpus has been analyzed, all words and paragraphs are represented by high-dimensional vectors (generally around 300 dimensions). This representation is clearly not as explicit as symbolic representations, like semantic networks or conceptual graphs, but we are not interested in the representation per se but rather in comparison relationships between words and/or texts.

What is interesting is that LSA offers us a way to compare two words or groups of words from a semantic point of view. Since every word and every paragraph is represented by a vector, the cosine function easily returns a measure of similarity. This measure is a number between -1 (lowest similarity) and 1 (highest similarity).

Another interesting point is that any new piece of text can be easily associated to a vector, even if this sequence of words did not appear as such in the original corpus. To do that, LSA just sums up the vectors corresponding to the words of the new piece of text.

To give an idea of how LSA works, the following are examples of semantic similarities, computed from the corpus "General_Reading_up_to_1st_year_college" on the LSA web site of the university of Colorado (http://lsa.colorado.edu):

- similarity("mice","mouse") = .79
- similarity("computer","disk") = .70
- similarity("computer","duck") = .01
- similarity("every mouse was scared by the feline", "mice were afraid of the cat") = .74

Numerous experiments showed that LSA mimics quite well the human similarity judgement (Foltz, 1998; Landauer & Dumais, 1997; Rehder et al. 1998; Wolfe et al., 1998). LSA is therefore a very good candidate for representing the meaning of words in computational simulations.

Kintsch' model

Kintsch (2000) relied on LSA to design a computational model of metaphor comprehension. He was interested in predicative metaphors, whose form is *A is a B* (i.e., *the mosquito is a vampire*). His goal was to construct a vector for the metaphor from the vector for the topic A (i.e., *mosquito*) and the vector for the vehicle B (i.e., *vampire*). Kintsch found that just adding the two vectors does not work: the new vector is far from vectors for landmarks, which are words that express the new emergent meaning of the metaphor. In the shark metaphor, such landmarks can be *vicious* or *aggressive*. Landmarks are obviously defined by hand.

Instead of just adding the topic vector and the vehicle vector, Kintsch rather linked LSA to his construction-integration model (Kintsch, 1998). The idea was to find words that could represent the meaning of the metaphor, so that a better vector could be constructed for the metaphor. Words that are associated to both the topic and the vehicle are good candidates.

One way to do that is to consider the semantic neighborhood of the vehicle. Kintsch's idea was therefore to select neighbors of the vehicle that are also associated to the topic. For instance, *sea* is associated to *shark* but not to *lawyer* whereas *vicious* is associated to both, although it is not immediate. In order to do that, a network composed of the vehicle, the topic and a fixed set of neighbors of the vehicle is constructed. These items are connected by links whose value is given by LSA cosines. The integration component is then run: an iterative spreading activation algorithm selects the nodes that are connected to both the topic and the vehicle. The *k* (usually 5) bests activated nodes are added to the topic vector and the vehicle vector to form the vector for the metaphor.

Kintsch showed that this procedure works with several metaphors, since the new vector is closer to landmarks he has selected than the other vectors. However this model has some limits:

- the fact that the set of neighbors has a predetermined size is a drawback: it should be set beforehand by a trial and error procedure which is not cognitively plausible;
- tests are not systematic: Kintsch picked some metaphors, some landmarks and claimed that the model works but this experiment reflects a lack of stringency.

Comparing the model with humans data can be performed at various levels. First, one can compare the new meaning that emerged from the association of the topic and the vehicle (for instance the idea of *vicious* from the metaphor *my lawyer is a shark*). That is what Kintsch did. In this paper, we are more interested in simulating the differences in time processing according to the different contexts.

Our model

In order to model the time it takes to process a referential metaphor, we made some changes on Kintsch's model. Firstly, since we do not work on predicative metaphors but rather on referential metaphors, we do not have a single topic but rather a whole context. We are therefore interested in selecting words that are common to both the vehicle and the context. Secondly, the search of common neighbors is not performed on a predetermined set of neighbors but rather done incrementally. The idea consists in scanning all neighbors of the vehicle, from the closest to the farthest, and keeping those that are close enough to the context. The procedure ends when 5 neighbors are found. Let's take an example, for illustrative purpose only. Consider two different contexts preceding the word *illuminate*:

- *During winter days, the city is dark very early. The local council plans to buy more street lamps. They would (illuminate)...*
- *Student complains that they cannot understand the theorem. The teacher decides to draw an example. His comments (illuminate)...*

In the first case, *illuminate* has a literal meaning whereas it has a figurative meaning in the second context. Our model will scan the neighbors of *illuminate*, starting from the closest one and progressively going away. The 5 common neighbors will be found quickly in the first case because *illuminate* is expected in this context. Words like *projector* or *light* will be gathered.

However, it may ”take time” to find them in the second context because words that are likely to be close to the context (like *clear* or *explain)* are further away in the neighborhood of *illuminate*.

This is how we model the time it takes to process a metaphor. The depth of this search (i.e., the total number of neighbors the model has to consider before finding 5 of them) will be compared with the human dependent variable (additional time) presented in the previous section.

This model has been programmed in C on top of the LSA routines. To be precise, we need to describe several parameters that are used in this program:

- *corpus* : we used a French corpus of 24 millions words, coming from all the texts published in the newspaper *Le Monde*, during year 1999. This corpus was analyzed by LSA to be represented in a 300-dimensions space. This procedure took about 30 hours, but then computing similarities between pieces of texts or getting the closest neighbors of a word takes a few seconds.
- *neighbor weight* : words for which LSA has not enough knowledge either because their frequency in the corpus is too low or because they appear in two many different contexts (like articles, pronouns, etc.) are ruled out. We rely for that on the weight LSA associates to each vector. We experimentally found that words whose weight is above .85 or below .2 should not be kept.
- *Similarity with the context* : we said that vehicle neighbors whose similarity with the context is high enough should be kept. This idea of "high enough" needs to be operationalized. Kintsch uses a threshold of two standard deviations above the mean similarity between words, which gives a value of .14. We found that a threshold of .2 gives better results on our French corpus.
- *Number of neighbors to keep* : like Kintsch, we keep 5 neighbors.

It is worth noting that other values for these parameters do not change dramatically the behavior of the model.

Simulation

Table 2 shows the results on the previous 32 texts. Except for the metaphoric reference in a synonymically-inductive context, Student tests show that all values are identical from a statistical point of view.

Table 2 : Mean search depth for processing different references in different contexts (model data)

Search depth	Syn. induction (Is)	Met. induction (Im)
Syn. reference (Rs)	385	357
Met. reference (Rm)	1432	393

The search depth is tremendously higher for metaphoric reference under a synonymically-inductive context. In a previous experiment (Lemaire et al., 2001), we showed that, contrarily to this kind of model, humans can stop processing a weird metaphor although they do not have a clear understanding of it. A similar phenomena may have occurred here. The model keeps scanning neighbors although the expected benefit is low.

Another point is that there are no differences between the two different contexts for the synonymic reference. An inductive context benefits to the figurative meaning but seems to have no effect on the processing of the literal meaning.

Conclusion

Comparison between humans and model

Overall, simulation and empirical data are in accordance. In both cases, metaphors are harder to process than synonyms and an inductive context facilitate the processing. The main difference lies in the processing of metaphor under an inductive context. The model is much more sensitive to an inductive context, to such an extent that metaphors are processed as quickly as synonyms. In the case of human data, the inductive context also reduces the time processing for metaphors, but not enough for reaching the synonym value. It is like the model works better than the humans! The reason for that could be twofold.

First, the text material might benefit mainly to the model due to the way we designed it. As we mentioned earlier, inductive texts contain words that are close to the reference and we relied on LSA to control these similarities. We are aware of this circularity: LSA is assessed by means of an experiment that rely on LSA. However, many tests have been performed in the literature which show that LSA mimics well the similarity judgment. It seemed therefore reasonable to use LSA for controlling similarities.

Second, the reason for the discrepancy between human and model could be that humans process metaphors in a way which is not accounted for by the model. The problem is that we do not know whether this difference has to do with the reference processing or with the metaphor processing. The model is designed to process metaphors, not references, and it is what it does. However, humans surely process simultaneously metaphor and reference. That might explain the extra delay needed by humans. In order to know more about that, we plan to perform another experiment in which the reference processing would be better controlled, for example by requesting subjects to perform a lexical decision task.

Relationships with cognitive theories

We will now analyze our model with respect to the cognitive theories. There is a large debate in the literature about the way a figurative meaning is cognitively processed compared to a literal meaning. The old stage theory (Searle, 1979) does not seem to be longer valid. This model considered that the literal meaning is first processed and, in case of failure, that it is reanalyzed as a metaphor. Current theories hypothesize a single mechanism for processing both meanings : metaphors would be processed by the same mechanism as literal sentences, and in parallel (Gibbs, 1994; Glucksberg, 2001). Our model fits in with these theories: there is no specific procedure to deal with metaphors.

If most researchers agree with the idea of a single mechanism, it does not mean that figurative and literal meanings are processed in equivalent time. This a second debate. It seems that there is no extra time needed to process predicative metaphors (McElree & Nordie, 1999). However, people found differences for referential metaphors as we mentioned in the introduction. We also found a difference, in both human data and computational data, which confirms these results.

LSA for computational models

LSA appears to be an interesting framework for designing computational models in the domain of language. It proposes a solution to the main drawback of most computational models, namely the lack of exhaustiveness and objectivity.

Exhaustiveness is an important issue: word meanings form a large network that it is important to consider in full. Working on just a small set of words might alter the model performances since the meaning of a word depends on all the other words. In addition, exhaustiveness permit to test models on an infinite set of texts.

Objectivity is possible with LSA since the human intervention is minimal. Hand-coded representations largely depend on the people who designed it, especially in the domain of language. If we want experiments to be replicated, we need objective procedures.

In the other hand, LSA suffers from being dependent on a large set of empirical parameters. Nobody knows the optimal number of dimensions nor the correct size of the corpus. Actually, no research teams seem to have a corpus it is happy with. In this experiment, we rely on a corpus from a newspaper, but this is far from perfect. Cognitive models need to be based on inputs that are as close as possible to the human ones. This is a problem with LSA since the inputs (all what a human being is exposed to) are huge and occur on a period of many years. Our future work will consist in gathering specific texts in order to have a better corpus.

Acknowledgments

The authors acknowledge support from the French program Cognitique. We are grateful to Coralie Gillon, Sabrina Mazot and Emmanuel Sylvestre for their help in the design of the text material, to Guy Denhière, Philippe Dessus and Walter Kintsch for reading an early draft of this paper and to the Telcordia Technologies company for kindly providing us with the LSA routines.

References

Budiu, R. & Anderson, J.R. (2002). Comprehending Anaphoric Metaphors. *Memory and Cognition*, 30, 158-165.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.

Duvignau, K., Gasquet, O., Gaume, B. & Gineste, M. (2002). Categorization of Actions By Analogy: From the Analysis of Metaphoric Utterances to a Computational Model. *Proceedings of the 15th International FLAIRS Conference*. AAAI.

Foltz, P., Kintsch, W. & Landauer, T.K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, 25, 285-307.

Gibbs, W.R. (1994). Figurative Thought and Figurative Language. In Gernsbacher, M.A (Ed.) *Handbook of Psycholinguistics*. Academy Press. 411-446.

Gildea, P. & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, 22, 57-590.

Glucksberg, S. (2001) On the fate(s) of litteral meanings. In *Proceeding of the Workshop "Towards on Experimental Pragmatics*". Lyon, France.

Kintsch, W. (1998). *Comprehension, a paradigm for cognition.* Cambridge University Press.

Kintsch, W. (2000). Metaphor comprehension: a Computational Theory. *Psychonomic Bulletin & Review*, 7(2), 257-266.

Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review* 104, 211-240.

Landauer T.K. (2002). On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The psychology of learning and motivation*, 41, 43-84.

Lemaire, B., Bianco, M., Sylvestre, E. & Noveck, I. (2001). Un modèle de compréhension de textes fondé sur l'analyse de la sémantique latente. In H. Paugam-Moisy, V. Nyckees, J. Caron-Pargue (Eds.) *La cognition entre individu et société* (Proceedings of the ARCo'2001 Conference). Hermès, 309-320.

McElree, B. & Nordlie, J (1999). Literal and Figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, 6, 486-494.

Markert, K. & Hahn, U. (2002). Understanding metonymies in discourse. *Artificial Intelligence*, 135(1-2), 145-198.

Martin, J.H. (1994). *A Corpus-Based Analysis of Context Effects on Metaphor Comprehension.* Technical report, University of Colorado. CU-CS-738-94.

Noveck, I.A., Bianco, M. & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, 16(1-2), 109-121.

Onishi, K.H. & Murphy, G.L. (1993). Metaphoric reference: when metaphors are not understood as easily as literal expressions. *Memory and Cognition*, 21(6), 763-772.

Rehder, B., Schreiner, M.E., Wolfe, M.B.W., Laham, D., Landauer, T.K. & Kintsch, W. (1998). Using Latent Semantic Analysis to Assess Knowledge: some Technical Considerations. *Discourse Processes*, 25, 337-354.

Searle, J. (1979). Metaphor. In O. Ortony (Ed.) *Metaphor and thought*, Cambridge University Press, 92-123.

Thomas, M.S.C., Mareschal, D. & Hinds, A.C. (2001). A connectionist account of the emergence of the literal-metaphorical-anomalous distinction in young children. *Proceedings of the twenty-third annual conference of the Cognitive Science Society*, 1042-1047.

Wolfe, M.B.W., Schreiner, M.E., Rehder, B., Laham, D., Foltz, P., Kintsch, W. & Landauer T.K. (1998). Learning from text: Matching Readers and Texts by Latent Semantic Analysis. *Discourse Processes*, 25, 309-336.