Title
To See and Not To See, That Is The Question

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 3(0)

Authors
Andreewsky, E.
Andreewsky, A.
Deloche, G.
et al.

Publication Date
1981

TO SEE AND NOT TO SEE, THAT IS THE QUESTION

E. Andreewsky[1], A. Andreewsky[2]
G. Deloche[1] and D. Bourcier[3].

[1]INSERM U 84, Hopital de la Salpetriere, 47, Bd de l'hopital, 75014, PARIS-FRANCE.

[2]LIMSI-CNRS, Orsay FRANCE

[3]ERA 430 CNRS-FRANCE.

Both psychological and computational theories of the lexicon usually consist of the specifications of the semantic relations between words. The present paper proposes an alternative, and simpler lexicon which does not include the particular meanings for each item in it. This lexicon consists instead of pointers to a general knowledge data-base, on the basis of which the meanings of words are to be computed.

World knowledge is indeed required to understand properly any word. This is demonstrated by examples like the following : in "a green salad" vs."a green cadaver", the color of "green" is not the same! But the requirement may be more obvious for a lexical ambiguity. For instance, to account for the meanings of the verb "study" in : "Francois studies english" vs. "Chomsky studies english", one cannot help but take into account much knowledge about how english is not the french people (as Francois refers to) mother tongue, on one hand, and, on the other hand, about Chomsky's scientific activity. The psychological meaning of words does indeed depends on world knowledge : it is more likely to result from some processing (taking this stored knowledge into account) than to be just retrieved out of any structure.

In this paper, we will, first, give a computational model for the resolution of lexical ambiguity; it is grounded on an abstract "lexicon", with pointers to the knowledge data-base. We will then present psycholinguistic data which call for such a pre-processing as the above abstract lexicon may provide.

RESOLUTION OF POLYSEMIES. AN A.I. APPROACH.

Given a (french) sentence, including a lexical ambiguity, m, we want to disambiguate m. The present resolution of this problem is based on an automatic information retrieval system, in natural language, so called "SPIRIT" (A. Andreewsky et al, 1980)which is currently use in Paris, with various data-base.

The "world knowledge" of SPIRIT is its stored documents (the french texts $D_1, D_2 \dots D_n$).

The "lexicon" of SPIRIT is very poor : it contains mainly morphological and syntactical properties of words.

SPIRIT computes the "distances" d_i, between any given french sentence s (that is, any request to the system) and its n stored documents; these distances d_i enable the system to answer to s , by means of an hierarchically arranged list of numbers :

$$A_s = (i_1, i_2, \dots \quad i_m)$$

i.e. a set of weighted numbers pointing to the documents $Di_1, Di_2, \dots Di_m$, answering the best to the request s.

SPIRIT includes a syntactical disambiguator: For instance, a word such as "can" (he can open the can) has two different entries in its lexicon, whereas a word such as "bachelor" has only one. To handle the p different meanings of such polysemies, the following steps are taken :

- Usual dictionnary definitions of words (one definition for a regular word, p for a p-polysemous word) are input, as given "request" to SPIRIT; the system's answers (one to p for each word) are sets of numbers A_i; these sets are linked to each lexical entry, and stored. The lexicon of the system include now, together with morphological and syntactical data, pointers to the knowledge data-base the A.

- If a polysemous word m has the p meanings : $m_1, m_2 \dots m_p$, its lexical entry include the sets : $A_1, A_2 \dots A_p$. Given, now, the word m in a sentence s the following operations enable its disambiguation, in a very simple way :

- A_s is the answer of SPIRIT to the sentence s.
- $A_1, A_2 \dots A_p$ are the answers of SPIRIT to the p different definitions of the meaning of m.

the greatest of the following intersections :

$A_1 \cap A_s$, $A_2 \cap A_s \dots \dots$ $A_p \cap A_s$ give which definition of m shares a maximum of related documents with the sentence s. It is the meaning of m in the given sentence s.

In order to exemplify how the system actually works, consider the disambiguation of the word : "instruction". In french, this word has three meanings :

1) instruction = "teaching or education, the taking in charge of school-age children."

Given the first definition, the answer of the system is :

A_1 = 120, 513, 519, 1829, 611, 1361, 207. 4212

2) instruction = "Proceedings which bring a case or a law-suit to trial".

to this definition, the answer of the system is :

A_2 1761, 1760, 1376, 1393, 1369, 1723, 276.

3) instruction = "Directions, instructions or informations given for indicative purposes".

answer of the system :

A_3 1802, 2144, 1761, 1367, 2720, 2490.

(all the numbers here refers to a data-base of 3000 laws, chosen for the present experiment).

Now, given the following sentence :

s "Qui se charge de l'instruction de dossiers de recouvrement d'impots" (Who is in charge of the examination of records in tax collection cases), including "instruction", the answer of the system is:

A_s = 276, 1367, 2121, 1761, 1760, 1376, 1393, 1361.

What is the meaning of "instruction" in the sentence s? The intersection of the answer of the system to s, with the answers to the three definitions of "instruction" are :

$A_1 \cap A_s$ 1361
$A_2 \cap A_s$ 276, 1761, 1760, 1376, 1393
$A_3 \cap A_s$ 1367, 1761.

Thus, the 2nd meaning of "instruction" triggers the greatest intersection, therefore it is the meaning of this word in the context at hand. An example of the use of "instruction" in the same meaning is taken out of the document N° 1761 . "le service d'assiette procède à l'instruction de la demande"(the tax-assessment service proceeds to investigate claims)

Here, there is a lexicon without any semantic information which enables, however (out of its pointers to the knowledge data-base) to handle lexical ambiguities in sentences.

* *
*

PSYCHOLINGUISTIC EXPERIMENTS.

How people process lexical ambiguities? In general, there is two contreversed models to deal with this problem (cf Carey P.W. et al, 1970). requiring a lexicon from which meanings are retrieved. If more than one meaning are retrieved, the wrong ones are then inhibited; the inhibition hypothesis is also invoked to explain subluminary experiments were people are able to implicitely use some semantic properties of words they are unable to report, or even to notice (cf D.A. Allport, 1977)

We will present two psycholinguistic experiments which require either a meaning-retrieval and inhibition mechanism, or, alternatively, some abstract lexicon such as the one above.

Experiment 1.

Written words were tachistoscopically presented in pairs, with visual masking conditions, in a speed which allowed subjects to report at most one of them.

All the words were nouns, and they include the french homograph : "fils" (which means either "son" or "threads"); this word is not an homophone, and when it means "son" it is uttered /fis/, and /fil/, for "threads". The written word "fils" was displayed together with either the word "father" or the word "needle". (E. Andreewsky et al, 1978). The subject's sole pronounciation /fis/ or /fil/ testifies in favor of the implicit resolution of the written polysemy; this resolution was found to be in accordance with the "ungrasped" co-displayed noun (father or needle)

The subjects' utterances testifies therefore in favor of implicit use of the "meaning" of words, displayed but not understood.

Experiment 2.

M.L. Albert et al (1973) propose a method so-called the "odd-word-out-test". Subjects are asked to point to the odd item in a written list, such as "hat" in ; "cat, dog, pig, hat, wolf". There are alexic subjects, who can neither read aloud nor understand (i.e. match with proper pictured items) any of the written nouns in the list, but can, nevertheless, perform the "odd-word-out-test". We have reported (G.Deloche et al, on press), in the context of the above task, the behavior of one of those alexic subjects : he gives 8 correct answers out of 10 lists of 5 written items, all of which he cannot understand. Therefore, since semantic cues are obviously required for the selection of odd items, such a patient displays, here again, the ability to make implicit use of the "meaning" of words he "sees", without being able to understand.

* *
*

The question arises then : how people can manage "to see and not to see" written items, that is make an implicit use of the meaning of written words which are not understood? This is the case in the two above experiments, such as in many others (Cf A. Marcel, on press).

Under a structural model of the human lexicon, from which meaning of words should be stored and retrieved, it is impossible to explain these experiments without a strongly ad hoc hypothesis on an inhibition mechanism, following the retrieval of a lexical meaning (explaining how one can make use of a word meaning, and not understand this word).

An alternative explanation can be carried out, grounded on such an abstract lexicon as described in the A.I. part of this paper. In the system described above, the lexicon does not include semantic informations, but only sets of pointers to the knowledge data-base. The retrieval of pointers is a preprocessing towards word understanding. It is clear that these pointers provide enough information -even not semantic- to explain how people may "see and not see" the meaning of words. For instance, in the case of the alexic patient, if such pointers would be retrieved, without further processing of the meaning of words, the patient will perfectly well be able to point to the odd word, out of an intersection between the pointers of the list, without having any idea on any written word meaning.

Therefore, our A.I. approach not only allows a very simple way to handle lexical ambiguities, but also provides a theory to explain how people can manage to see and not to see a word, that is to implicitely use semantic properties of non-understood written items.

REFERENCES.

- Albert, M.L., Yamadori, A., Gardner, H. and Howes, (1973), Comprehension in Alexia, "brain", Vol. 96, part 2, 317-328.
- Allport, D.A. (1977) On knowing the meaning of words we are unable to report : the effect of visual masking. In "Attention and Performance", Dornic (ed) Lawrence Erlbaum, New Jersey, 505-533.
Andreewsky, A., Debili, F., Fluhr, C. (1980). Apprentissage Syntaxe semantique lexicale."Revue du Palais de la Decouverte", Vol. 9, N° 83, 17-40.
Andreewsky, E., Deloche, G., and Kossanyi, P. (1979). Towards a procedural understanding of reading INS Conf., Noordwijkerhout, Holland. (in "Les processus de la Lecture", special issue, 1978).
- Carey, P.W., Mehler, J. and Bever, T.G. (1970) Judging the veracity of ambiguous sentences. "Journal of Verbal learning and Verbal Behavior", 9, 243-254.
- Deloche, G., Andreewsky, E., and Desi, M. (on press) Surface dyslexia : a case report and some theoretical implications to reading models. "Brain and Language".
- Johnson-Laird, P.N. (1980). Mental models in cognitive science. "Cognitive Science", 4, 71-115.
- Marcel, A.J. (on press). Conscious and unconscious reading : The effects of visual masking on word perception. "Cognitive Psychology"