ARTICLES AND RESOURCE CONTROL

Janusz S. Bien

Institute of Informatics, Warsaw University

P.O.B. 1210, 00-901 Warszawa, Poland

Abstract

The paper discusses how the resource control hypothesis introduced earlier by the author accounts for the rather mysterious fact that English articles are rendered in Slavonic-languages by word order and vice versa. The definite versus indefinite distinction is viewed as a manifestation of the variable depth of nominal phrase processing. The depth of processing is determined by the availability of resources, which is indirectly controlled by the speaker with sufficient precision; articles appear to be only some of several resource control devices available in natural language.

Articles

Teaching the proper use of articles to the students whose native language does not posses them is one of the most difficult tasks. Despite the effort of both the teachers of English who are and who are not native speakers, no satisfactory teaching methods have been developed.

In hope to find some hints useful for teaching English to Polish students, an experiment was performed by Smolska (1976). A short English text was deprived of all articles, and the underscore character was inserted at all the potential places of their occurence. Next, 4 educated native speakers of English willing to be the subjects of the experiments were found in Warsaw; the reader should appreciate that it was not a trivial task. The subjects were independently asked to fill in the articles when appropriate. The result was rather surprising: the subjects agreed with themselves and the author of the original text in 86 % cases.

The Smolska's experiment strongly suggest that the semantic impact of articles is negligible, but of course it does not prove the case. Unfortunately, the native speakers of English are too scarce in Poland to allow an experiment on a larger scale. Therefore the onus lies on those opponents which have no difficulty to gain cooperation of sufficient number of native speakers. For the purpose of the present paper I simply assume the claim to be true.

Let us notice now that articles are not just historical remnants, but emerged relatively lately as a new linguistic mechanism, so it is natural to assume that they serve some specific purpose. If they do not affect WHICH meaning is assigned to an utterance, they may influence HOW the meaning is assigned. In other words, it may be the case that articles control in some respect the process of computing the meaning.

Resource control hypothesis

The resource control hypothesis has been formulated in (Bien 1980) as a further development of the multiple environments model of natural language advocated e.g. in (Bien- 1975, 1976a,b). Below we discuss briefly only those aspects of the hypothesis which are relevant for the present paper.

We accept the wide-spread assumption that information in memory is clustered into frames. At every moment some frames are in the focus of attention; the saliency of other frames may be represented by an appropriately defined distance from the focus of attention. Changes in saliency are viewed as a suitable displacement; such a displacement requires always some time and other resources. For more details of the proposed DISPLACEMENT MODEL OF MEMORY see (Bien 1980).

A frame in the strict sense describes a specific object or notion, e.g. a specific copy of a specific book; more general information is stored in frame prototypes. Instantiating a prototype usually requires memory search to locate the relevant data; in consequence, it is a process with substantial demands for resources. The frame prototypes are of course interrelated, e.g. the prototype for a book contains a slot for its author, the prototype for an author contains a slot for his/her books etc. The interrelation is represented by means of frame pointers. They are just data structures allowing to locate the appropriate prototype when needed. Constructing a frame pointer requires a negligible amount of resources.

We assume that there is some limit for the human language processor capacity and that the linguistic processes must compete for resources. The lower level proceses (acoustic, phonetic and syntactic analysis) are data-limited processes, while the higher level processes (semantic and pragmatic processes, including memory search and spontaneous inferences) are resource-limited (Norman,Bobrow 1975). Data-limited processes must have higher priority, otherwise the resource-limited processes would use up all the resources available; in consequence, the semantic and pragmatic processing uses only the resources left by the lower level processes.

A nominal phrase in an utterance is always transformed at first into a frame pointer. Depending on the availability of resources, the pointer can be developed into a prototype, accomodating some additional information contained in the utterance. The prototype can be then instantiated by memory search. Finally, the frame instance can be displaced toward the focus of attention. In other words, the availability of resources determines the depth of processing for a given nominal phrase.

Controlling the depth of processing

Most devices for controlling the allocation of resources for the resource-limited processes and, in consequence, controlling the depth of processing seem to be language universalis, but in different languages they are used to different extent.

The most general way of allocating the resources consists in choosing for a phrase an appropriate distance from the beginning of the utterance; we will call it the time ordinate principle. Putting a phrase close to the beginning results in the early initiation of its processing, and gives it more chances to collect the appropriate amount of resources. We assume that the scheduling of the resource-limited processes is governed by the preference for deterministic processes (Bien 1980:4, Kowalski 1979:100). This strategy distributes the resources to the memory search, the inferences and the focusing in the proportions depending on the context of computation. If a phrase is placed at the beginning of a sentence, the global effect is determined by several factors. We will mention here only the extreme cases. The most typical one is when the phrase refers to some information known already, which is then retrieved and brought to the focus of attention; the information from memory becomes available soon enough to influence the procesing of the rest of the sentence. If the phrase introduces new information, as in (Bien 1980:15, Szwedek 1981:74)

(1)   A man was coming.

the inferencing tries to establish its relation to the known information; on the other hand the new information becomes the focus of attention.

Another case is represented by English cleft sentences of the type

(2)   It was John who did it.

where memory search is usually not needed and all the resources are used for focusing and spontaneous inferencing, which results in the effect of strong emphasis. As it can be seen now, the claim that there is a universal tendency to put the given or definite information at the beginning of the utterance (Cark Clark 1977:548) is just a rough approximation of the reality.

Another way of controlling the resources consists in using demonstratives. We take for granted that the primary use of demonstratives is to draw the hearer's attention to physical objects in his environment. The hearer's strategy to process demonstratives is therefore to assign the top priority to the task (the object pointed should be located as soon as possible, because e.g. it can be moving and the speaker's description may soon become obsolete). Succesful processing of demonstratives requires accomodation of the information perceived and not contained in the utterance, therefore the nominal plArase containing a demonstrative should be processed at least at the level of frame instances. The speaker's strategy for using demonstratives includes the requirement for the adequate description of the object, so the hearer may assume the description to be precise and not requiring any additional inferences. Now it should be noted that what we actually perceive e.g. visually is only a little spot (the data from the retinas of our eyes) in our mental image of the environment, so the hearer's strategy can be formulated simply as: locate the memory representation of the object, assuming that the description is precise, and store the retrieved information in a frame instance as quickly as possible. In consequence, both in English and Slavonic languages the nominal phrase containing demonstratives may refer to the information given earlier in the text; cf. e.g. (Hawkins 1978:149).

The next method of controlling the depth of processing exploits some properties of numerals. The primary use of numerals expresses the cardinality, but the side-effect of it is the low level of processing: it is impossible to construct 1000 frame instances for the phrase "thousand soldiers", it would be useless to construct 10 frame instances for "ten soldiers", so by analogy no separate instances are created (unless forced by other reasons) for "two soldiers" although it is feasible. We claim that the analogy holds also for the numeral meaning one, i.e. that a nominal phrase containing the numeral "one" is not processed deeper than to the level of frame pointer or a single frame prototype. It explains why the numeral is occasionally used both in English and Slavonic languages to signal new information.

The next very important method applies only to spoken language but had, at least for English, a far-reaching consequences also for written language: it is the sentence stress. When we assign the sentence stress to a phrase, we pronounce it more clearly and in a longer time span than usually (the claim is based only on my intuition, but 1 am not aware of any evidence contradicting it). The clearer pronounciation makes the task of the acoustic analysis easier, so it consumes less resources; the larger time span generally increases the availability of resources. The global result is the substantial increase of resources available for the higher level processes. As usual, the resources can be allocated to them in several ways. Typically, the inferences and focusing provide the effect of emphasis. On the other hand, unstressing a phrase or a word deprives it of resources. To the best of my knowledge, in all languages articles originated from unstressed numerals or pronouns (personal or demonstratives); articles has therefore much in common with their source. Our treatment of articles, which intends to incorporate the results of (Hawkins 1978) by interpreting them in the framework of (Bien 1980), will account for this fact.

The definite article in English originated from demonstratives; depriving demonstratives of resources by unstressing them does not allow to assign the top priority, but still the task is the frame instance depth of processing. The process of establishing the frame instance takes now longer, and in the meantime some useful results of spontaneous inferences may become available; therefore it would be unefficient to stick to the precision requirement. The frame instance may be needed to accomodate either the information retrieved from the memory (the anaphoric u e of definite article) or for the elaborate description contained in the utterance (the forward-pointing use), or for other reasons (e.g. the generic use). The indefinite article originated from the numeral; its function is simply to signal that shallow processing is sufficient. In the next section we will try to demonstrate that the Hawkins' notions of inclusiveness and exclusiveness are just the manifestation of the different depth of processing.

Examples

As it is always risky to design sophisticated examples in a

non-native language we take both the examples and their meanings from (Hawkins 1978); we preserve the original numbering of examples.

At first we will illustrate the basic uses of articles.

(3.02)   Fred was discussing an interesting book in his class. 1 went to discuss the book with him afterwards.

In the first sentence we have a typical use of indefinite article, which signals that the processing level of the frame pointer to the book prototype would be sufficient; the attribute ''interesting'' may be attached to the pointer or, if there are resources available, it may be accomodated in the instance of the book prototype located by means of the pointer. In both cases the nominal phrase gets only shallow processing. In the second sentence we have an anaphoric use of definite article; the book prototype is used to create a frame instance, then the memory is searched in an attempt to further instantiate it. The antecedent pointer or frame instance is easily found and merged with the instance which initiated the search.

(3.03)   Fred was discussing an interesting book in his class. He is friendly with the author.

In the second sentence we have the associative anaphoric use of definite article. The frame instance for the author is created and the memory is searched as before. However, this time only one of its slots ("the author of what ") will be succesfully matched against the frame instance in the memory.

(3.02")   Fred was discussing an interesting book in his class. 1 went to discuss a book with him afterwards.

In this text we have two separate pointers or frame prototypes (i.e. frame instances containing only the information stored in the prototype) for "book". Hawkins (1987:87) is wrong claiming that the phrases cannot refer to the same book, as the continuation is possible "It appeared later to be the same book'' which results in collapsing two frame instances into a single one.

(3.03")   Fred was discussing an interesting book in his class. He is friendly with an author.

In Hawkins' terms, "an author'' (and "a book" in the preceding example) is a non-located (in the shared knowledge of speaker and hearer) specific indefinite, which EXCLUDE (this is for Hawkins the essential property of indefinite reference) an infinite number of potential referents. This sophisticated mental construction decribes, in our opinion, the simple fact that a shallowly processed nominal phrase represents an object unrelated to other instances of the same prototype stored in the memory, unless such a relation is established by some additional effort; it is not worth mentioning that an unlimited number of the new instances of the same prototype may be created later. The exclusiveness of indefinite reference is more intuitive in the following example:

(4.16)   I've just deckled to inspect a house. I decided not to buy it because a roof was leaking.

Hawkins claims that "a roof'' cannot refer to the roof of the house mentioned, because it has to exclude at least one other possible referent; as there is only one roof talked about, the exclusion is impossible and results in the unacceptability of the text. In our opinion, (4.16) is fully analogous to (3.02') and (3.03"), the difference consisting only in the fact that even a

temporary assumption of the non-coreferentiality of "a r o o f" with the inferred roof of "a house" leads to the incoherence of the text under consideration. In consequence, we accept the exclusiveness rule only as an interplay of the shallow depth of processing with general coherence rules. Our position is additionally supported by the Hawkins' observation that the exclusiveness condition disappears with the verbs "to have", "to be" and the set-existential verbs:

(4.137)   1 have a head.

(4.145)   There is a roof on my house.

(4.158)   Do you remember the other day 1 was talking about a student called Smith ?

Such sentences may be called attention-shifting, because usually they do not introduce new information, but change the saliency of the already known facts. According to Hawkins, the essential property of definite reference is its 1NCLUSIVENESS (which in case of singular count nouns turns into uniqueness), e.g.

(3.175)   Fred brought the wickets in after the game of cricket,

"is understood as making a claim about all the wickets in question". We think the property originates from the fact that the deep processing of a nominal phrase fully associates it with earlier mentions of the referent.

References

(Bien 1975). J.S.Bien. Toward a Multiple Environments Model of Natural Language. Advance Papers of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi 1975, pp 379-382.

(Bien 1976a). J.S.Bien. Multiple Environments Approach to Natural Language. American Journal of Computational Linguistics microfiche 54.

(Bien 1976b). J.S.Bien. Computational Explication of Inlensionality. 1976 International Conference on Computational Linguistics preprint no 41.

(Bien 1980). J.S.Bien. A Preliminary Study on Linguistic Implications of Resource Control in Natural Language Understanding. 1SSCO Working Paper no 44, Geneva.

(Clark,Clark 1977). H.C.Clark, E.V.Clark. Psychology and Language. New York 1977.

(Hawkins 1978) J.A.Hawkins. Definiteness and Indefiniteness. A Study in Reference and Grammaticality Prediction. London: Croom Helm Ltd. 1978.

(Kowalskil979). R.A.Kowalski. Logic for Problem Solving. North Holland 1979.

(Norman,Bobrow 1975). D.A.Norman, D.G.Bobrow. On Data Limited and Resource-Limited Processes. Cognitive Psychology 7 (1975), pp 44-64.

(Smolska 1976). J.Smolska. Unpublished note.

(Szwedek 1981). A.Szwedek. Word Order, Sentence Stress and Reference in English and Polish. WSP Bydgoszcz 1981.