

ALEKSANDER SZWEDEK, PEDAGOGICAL UNIVERSITY, BYDGOSZCZ

E L E M E N T S O F T E X T G R A M M A R

1. Recent developments in linguistics have shown that sentence grammars even of the most elaborate form are insufficient for an adequate and complete description of a language. Elements have been discovered that either escaped linguistic analysis or have been included in the description as sentence elements. It must also be added that elements considered until recently as sentential have been found to be subject to textual constraints. It is important to remember that text analysis has been recognized for some time now as more natural and explanatory than sentence analysis. For example Gleason /1968/ wrote that "Discourse analysis cannot reach down to the sentence and stop. When discourse organization extends into the sentence, it must be treated as part and parcel of wider structure. Sentences will never be fully described apart from the discourse in which they occur, and by which they are moulded" /45-46/.

2. One of the elements that linguists have found difficult to account for within the frames of S-grammars is reference, and connected with it coreference, i.e., an association of two /or more/ NP's with one referent. Coreference plays a crucial role in such linguistic processes as definitization, pronominalization, reflexivization, equi-NP deletion. So far they have been described as sentence phenomena although definitization /at least

some cases of it/, pronominalization and equi-NP deletion are genuine text phenomena /cf. Krzeszowski 1975, Palek 1968, van Dijk 1972, and many others/. In all of these processes it is necessary to refer back to a possible antecedent in the previous sentence and determine whether the two nouns refer to the same referent or not. Only on this condition the rules can apply. Thus, the referential definite article can be inserted before a noun only if the latter's antecedent has been determined as referring to the same referent. For example, the second occurrence of the noun in /1/

/1/ I saw man yesterday. Man was wearing black hat.

can be definitized if it is found to refer to the same person as the man in the first sentence, thus producing /2/

/2/ I saw a man yesterday. The man was wearing a black hat.

The same condition is required for pronominalization:

/3/ I saw a man yesterday. He was wearing a black hat.

and equi-NP deletion:

/4/ John hoped to go.

from

/5/ John hoped [John will go].

One of the fundamental questions concerning these problems is when and how the identity of reference is determined. Two extreme solutions have been proposed. In the standard transformational

theory of pronouns and reflexives, NP's are changed into pronoun or reflexive forms on the basis of morphological identity and intended coreference with other NP's in the sentence. There is, however, a number of difficulties with this approach. One of the most important is the infinite recursion of deep structures in sentences such as /6//first pointed out in Bach 1970/:

/6/ The man who deserves it will get the prize he wants.

If the underlying structure of pronouns is a fully specified NP identical with its coreferent, both it and he in /6/ must have infinite deep structures:

The man [who deserves the prize [which the man [who ...] wants]] will get the prize [which the man [who deserves the prize [which ...] wants]]

Another difficulty discussed by Dougherty /1969/ is offered by sentences such as /7/:

/7/ Lips that touch liquor shall never touch mine

/= Lips that touch liquor shall never touch my lips/

since they cannot be derived if the identity condition obtains, because the sentence does not say that lips have touched, do touch, or ever will touch liquor. Jackendoff /1972/ observed that also "pronominal epithets" /the bastard, the poor guy, etc./ cannot be accounted for in such a theory. Pronominal epithets are, however, part of a much wider phenomenon mentioned by Gleason already in 1968 in the following example: "A tall, lanky high-school basketball star folded himself into the little foreign car with his short, pudgy father beside him. The son drove carefully down the busy highway, obviously restrained by

the older man's presence. He watched the other traffic intently. He signalled before each turn and came to a full stop at the blinker. Classmates stared in amazement, but the driver took no notice of them. The athlete's strange behaviour seemed totally inexplicable. The poor boy was obviously in some trouble" /55/. Here the same person is referred to as a basketball star, the son, the driver, the athlete, the poor boy and no present grammar can give an adequate account of these phenomena.

Jackendoff /1972/ writes that "The drastic way out in the transformational approach to reference is to generate only referential indices in NP position, then bring in lexical material from outside conjoined clauses" /110/ and that "this solution requires the addition of clauses to the deep structure whose precise syntactic form have no bearing on the surface structure" /110/. He then suggests an alternative approach to the problems at issue, the interpretive approach: "If the reference of pronouns is determined by a rule of semantic interpretation, the deep structure ... contains the pronouns themselves, so there is no recursion. Furthermore, in the process of semantic interpretation, a pronoun need not be replaced with a duplicate of the noun phrase with which it is coreferential /which would again bring up the problem of recursion/, but rather it may just be marked coreferential with another noun phrase" /110/.

A commonly used device for indicating coreference is the index of coreference introduced by Chomsky /1965:146/. Chomsky suggests that each noun phrase is marked with an integer /or index/, and two NP's are coreferential if they have the same

Index. Jackendoff /1972/ proposes a different formalism, a table of coreference in which coreference relations /coreferential - noncoreferential/ will be expressed explicitly. No two NP's may be left unexpressed in the table of coreference and "Three or more NP's can be understood as mutually coreferential only if they have been marked pairwise coreferential" /111/.

It does not seem to be very difficult to extend Jackendoff's table of coreference over more sentences than just one and thus to use the device in a text grammar. Since the table of coreference operates on lexical items /i.e. after lexical insertion/ the problem of different lexical items for the same referent should be easy to handle, and even some of Jackendoff's conditions /e.g. consistency condition on coreferents/ will also hold here.

It is also possible to account for cross-sentence reference in a slightly modified generative semantics model. In 1974 Krzeszowski stated that "there are both theoretical and practical reasons motivating attempts to construct grammars which would enumerate not only well-formed sentences in a language but also their well-formed combinations" /45-46/. He suggested that instead of S the initial symbol should be Σ , an input to the derivation of a sequence of n sentences where $n \geq 1$. He also remarked that "any equivalent sequence of sentences in any two languages will be derived from an identical configuration while the number of surface-structure sentences within the equivalent sequences of sentences may be unequal" /119/. He also adopted Sampson's /1969/ proposal of associating "mental entities /let's call them referents/ with a set of integers from 1 to n in such

a way that each referent is labelled with a unique integer, the cardinality of the set of referents thus formed being finite at a given moment of time for an individual language user and determined by his former linguistic career. It will be furthermore possible to associate each referent with one or more linguistic expressions by attaching the respective indices to those portions of the derivations which underlie these expressions" /Krzyszowski 1974:33-34/. This combination of extended generative semantics and Sampson's theory of reference handles nicely all the difficult cases mentioned above. Which of the two /Jackendoff's or Krzyszowski's/ theories is better suited to describe reference is still to be seen.

3. One of the more neglected aspects of relations between sentences is time /tense category. A possible reason for this neglect is the number of factors that seem to contribute to the determination of time/tense and a rather obscure relation between them.

Among such factors, often mentioned in the literature, are "relations of cause and consequence between the different clauses in a complex sentence" /van Dijk 1972:63/ and "assumptions made by the speakers concerning the relationship between the actual time of the speech-act and that of an event in the physical universe, and the perception by the speaker of the temporal gap between event and speech" /h. Lakoff 1970:848/.

It is quite obvious that, since everything must occur in time and space, the text must reflect the time of the event it describes which may also be determined by the Participant /in Gleason's 1968 sense of the word/ as in /8/ and /9//see also

J. Anderson 1976/:

/8/ Einstein taught me physics.

/9/ I have been taught physics by Einstein.

Also aspect of the verb /in the sense of Polish linguistics, i.e., perfect vs imperfect/ is determined by the Participant /whose interpretation is linked with focus, see point 4/, as in Polish examples /10/ and /11/:

/10/ Dziewczynka pisała wiersz.

/Girl was writing poem/

/11/ Wiersz pisała dziewczynka.

/Poem wrote /fem./ girl/

To a certain extent the choice of the tense, for example past or historical present, is a reflection of the speaker's attitude towards the event, but in the subsequent sentences it is constrained by "cause and consequence relations", as in van Dijk's examples:

/12/ a/ Peter will write an article for our review.

b/ It was published in a special issue.

/13/ a/ Yesterday John bought a car.

b/ The car is very expensive.

with inversion of the latter sequence impossible:

/14/ a/ The car is very expensive.

b/ Yesterday John bought a car.

It seems, however, that what makes sequence /14/ incorrect is not only the tenses. We may have the same past tense in both

a/ and b/ and /14/ will still be wrong:

- /15/ a/ The car was very expensive.
- b/ Yesterday John bought a car.

while inversion of a/ and b/

- /16/ b/ Yesterday John bought a car.
- a/ The car was very expensive.

renders a correct sequence.

Two conditions seem to come into play in the above examples: one we could call the condition on reference /a car - the car, not vice versa/ and also a general pragmatic condition that a prior event must have the same or earlier tense /as in /13/ or /16/ than later event. The first condition would exclude cases like /14/ and /15/, the other, cases such as /12/.

A condition similar to the latter holds for location of referent or event. The same referent or event cannot be situated in two places at the same time.

All in all, it seems that van Dijk is to a certain extent correct when he concludes his discussion of time/tense dimension by saying that "the temporal elements of the sentence as represented in tenses, temporal adverbs and conjunctions, establish, together with discourse referents, presupposition and topic and comment, the main ordering conditions for sentences in a text" /81/. However, further research on the above mentioned problems is necessary.

There is also a group of verbs, as Karttunen noticed in 1969, which do not establish reference and thus exclude

assertive sentences with definite NP's as sequence sentences. The group contains the modals and verbs such as want, expect, try, plan, intend, hope, etc. as in the following examples /Karttunen 1969:8-9/:

- /17/ a/ Bill can make a kite.
b/ The kite has a long string.
/18/ a/ Mary expects to have a baby.
b/ The baby's name is Susan.

In this respect they play the same role in discourse as NEG which does not establish reference either, as in /19/:

- /19/ a/ Bill does not have a car.
b/ The car is black.

Karttunen finds that the common feature of these verbs and NEG is that they "represent a yet untrue proposition at the time specified by the tense and time adverbials in the main clause" /Karttunen 1969:9/.

4.. One of the most prominent elements of a text is the distribution of new and given information /called also topic/comment, focus/presupposition, theme/rheme; since, however, these terms are also used in slightly different senses I prefer the more neutral terms new/given information/. Research on this subject can be traced back to the middle of the 19th c. /A.M. Bell 1842, H. Weil 1844/ but it was fully developed in the works of the Prague School linguistics, works initiated by V. Mathesius /1930, 1939, 1941/ and continued by F. Daneš, J. Firbas, P. Sgall and many others.

Two assumptions underlie the research:

- a/ Apart from the grammatical and semantic structure, every sentence has a thematic structure, i.e. can be analysed as consisting of two parts: one containing new information, the other given information. This is only logical since, within a text, if a sentence contains only given information then it is unnecessary, if it contains only new information then it is in no way linked to the preceding sentence /this statement is an oversimplification as will become clear in section 5/. Even in an opening sentence a part of it is at least assumed to be given /cf. Mathesius 1941/.
- b/ This "organization of the utterance" /Daneš 1964/ must be somehow expressed on the surface, i.e., new and given information must be signalled in different ways.

Every language uses different signals to mark the new/given information distribution. In Czech, as Mathesius observed, sentence initial position signals that the element is "given", sentence final position that the element is "new". In English, where word order is used to indicate grammatical relations, the new/given information distribution is signalled by the place of the sentence stress and the articles; in Polish by word order and the place of the sentence stress /for a fairly detailed account of these phenomena see Szwedek, 1976/.

Connected with the organization of the utterance is ellipsis. It is obvious that ellipsis can affect only those parts of the sentence the absence of which will not break coherence. It seems that only the absence of given information

will not break coherence. However, ellipsis is subject to certain language specific constraints. For example, in Polish the subject pronoun must be deleted if the object pronoun is deleted, but not vice versa, as the examples below demonstrate:

/20/ Janek uderzył Tomka a potem on kopnął.

/John hit Tom, and then he kicked/

/21/ Janek uderzył Tomka a potem go kopnął.

/John hit Tom and then him kicked/

/22/ Janek uderzył Tomka a potem kopnął.

/John hit Tom and then kicked/

In English omission of object pronoun is impossible, as

/23/-/25/ show:

/23/ John hit Bill and then kicked.

/24/ John hit Bill and then kicked him.

/25/ John hit Bill and then he kicked him.

More extensive research on ellipsis has been done by T. Shopen /1974/.

5. The relations between the above mentioned elements are by no means simple ones. There are interactions between reference and time/tense category, where time/tense is determined to a certain extent by referents, between reference and modality /modal verbs and want, plan, expect, etc./, reference and negation and interrogation, relations between time and the attitude of the speaker, relations of cause and consequence. All of them seem to operate at various levels of derivation and the exact nature and place of them in the system will have to be determined.

The elements sketched above are used to ensure coherence of the text. Some of them are necessary conditions /time/tense category seems to be a necessary condition/, some may not be overtly expressed but we would expect them to be present in the deep structure. For example, in /26/

/26/ John bought a book yesterday. He gave the book to Mary.

coherence is expressed in the thematic structure /pronominalization, sentence stress on new element - a book, to Mary/ and time/tense consistency / both tenses are past/. The thematic structure is not present in an example similar to Lakoff's /1971/, example 23:

/21/ Nixon was elected. The blacks will revolt.

Only time/tense condition is clearly present. Otherwise there is no necessary connection between the two sentences. It seems that sequence /21/ is an example of ellipsis. This kind of ellipsis differs from ellipsis discussed at the end of section 4. We could call ellipsis in example /24/ grammatical /syntactic/, while the ellipsis in /21/ could be called semantic /?/. The exact nature of the two /perhaps there are more/ kinds of ellipsis remains to be determined. Ellipsis as an element of the text is hinted at by Gleason /1968/ when he says that "Sentences will never be fully described apart from the discourse in which they occur, and by which they are moulded" /45-46/. Van Dijk /1972/ is more explicit on the subject: "Since presuppositions are always represented as sentences, we may consider the set of presuppositions,

followed by the sentence/s/ presupposing them, to be part of a text" /103/, and "we assume that presuppositions are preceding sentences, or rather preceding semantic representations, because some of them may be optionally deleted if implied by other preceding sentences".

It would seem then, that if presuppositions take the form of sentences, the deep structure of a text will contain a set /sequence/ of sentences each of which will represent a minimal step in the "cause and consequence" continuity. In such a case rules would have to be formulated which would tell us which sentences of the sequence have to surface and which not. Thus, for a deep structure of the text T:

$$T_{SR} = S_1 - S_2 - S_3 - S_4 - S_5 - S_6 - S_7 - S_8 - S_9$$

we could have a rule saying that if the sequence T_{SR} satisfies certain conditions /specifying which of the S's can be deleted, i.e., which of them are presuppositions of which/ then

$$S_1 - S_2 - S_3 - S_4 - S_5 - S_6 - S_7 - S_8 - S_9 \Rightarrow S_1 - S_5 - S_9$$

in which case S_2, S_3, S_4 and S_6, S_7, S_8 are presuppositions to S_5 and S_9 respectively. Before this is done we have to know what a presupposition is. As is well known the concept of presupposition has no clear and uniform definition. Various types of presuppositions have been described: lexical presupposition /Fillmore 1969, 1971/, speech act presupposition /Keenan 1971; his pragmatic presupposition seems to concern speech act only/, logical presupposition /Keenan 1971/, pragmatic presupposition /R. Lakoff 1971/. In view of the diversity of presuppositions and their importance in linguistic analysis it is imperative that the nature and types of

presuppositions should be determined.

REFERENCE

- Anderson, J. 1976. Perfect possibilities and existential constraints. SAP. VII:3-6.
- Bach, E. 1968. Nouns and Noun Phrases. In Bach and Harms, 1968, 91-124.
- Bach, E. 1970. Preblominalization. Linguistic Inquiry 1:121.
- Chomsky, N. 1965. Aspects of the theory of syntax. MIT Press: Cambridge, Massachusetts.
- Daneš, F. 1964. A three-level approach to syntax. Travaux Linguistiques de Prague 1, 225-240.
- Dijk, T. van. 1972. Some aspects of text grammars. Mouton: The Hague.
- Dougherty, R. 1969. An interpretive theory of pronominal reference. Foundations of Language 5:4, 488-519.
- Firbas, J. 1964. From comparative word-order studies. Brno Studies in English 4, 111-28.
- Gleason, H.R., Jr. 1968. Contrastive analysis in discourse structure. In Alatis 1968, 39-64.
- Jackendoff, R. 1972. Semantic interpretation in generative grammar. The MIT Press: Massachusetts.
- Karttunen, L. 1969. Discourse referents. Indiana Univ. Linguistics Club: Bloomington.
- Krzeszowski, T. 1974. Contrastive generative grammar. Uniwersytet Łódzki: Łódź.
- Lakoff, G. 1971. The role of deduction in grammar. In Fillmore and Langendoen 1971, 63-70.
- Lakoff, R. 1970. Tense and its relation to participants. Language 46, 838-849.
- Mathesius, V. 1930. K pořádku slov v hovorové češtině. NŘ 14, 117-21.

Mathesius, V. 1939. O tak zvaném aktualním členění vety. SaS 5, 171-4.

Mathesius, V. 1941. Základní funkce pořádku slov. SaS 7, 169-80.

Sampson, G. 1969. Towards a linguistic theory of reference.
An unpublished doctoral dissertation. Yale University.

Shopen, T. and W. Świeczkowski. 1974. Some remarks on ellipsis in Polish and English. PSiCL 4, 111-26.

Szwedek, A. 1976. Word order, sentence stress and reference in English and Polish. Linguistic Research, Inc.: Canada.