IS SENTENCE VIABLE?[1]

Andrej A. Kibrik (Institute of Linguistics, Russian Academy of Sciences, kibrik@comtv.ru)
Vera I. Podlesskaya (Russian State University for the Humanities, podlesskaya@ocrus.ru)

Linguists tend to think of natural language as consisting of sentences. Sentence is usually thought to be among the basic units of language, and some theories explicitly claim it *the* basic unit. Evidently, a sentence is supposed to be a sequence of words separated by final punctuation marks, such as periods. This view is often uncritically borrowed into other sciences, for example psychology, as represented in most textbooks. Even psycholinguistics, inherently predisposed to linguistic 'performance', has 'sentence processing' as one of its central research fields.

Under a close examination, sentence turns out to be a rather problematic notion. This study is based on the data of spoken language. Quite obviously, the spoken mode of language is the original and central one, while written language is derived from it. In spoken language, there are no periods that can help to identify sentences. The research questions of this paper are: is it possible to identify sentences in speech? and if so, is it a basic unit of language?

The kind of spoken language used in this study is a corpus of Russian stories, known as 'Night Dream Stories', see Kibrik and Podlesskaya in press. Stories, mostly representing the narrative discourse type, largely consist of one kind of illocutionary act, that is, of statements. So this discussion is restricted to statements and does not include other illocutionary types, such as questions and directives. If story as a whole is a statement, does it actually split into intermediate hierarchical units, comparable to what is known as 'sentence'?

Spoken discourse consists of minimal quanta or moves – elementary discourse units (EDUs). EDUs are units of physiological and cognitive processing and prototypically coincide with clauses. The discourse function of an EDU is conveyed in speech by means of a pitch accent placed on the constituent that is the EDU's informational center. In particular, some EDUs are marked with a falling pitch accent, which is often taken as a signal of the end of a sentence. Conversely, EDUs that are non-terminal within a sentence are marked with a rising accent; the direction of tone in this case is an anticipatory adaptation to the fall that is expected to come in the sentence-final EDU. In the transcription system developed for the corpus in question the sentence-terminal falling accent is marked with a period, and the rising accent with a comma. A simple example of this kind can be seen in an excerpt from a story, provided in the format of slightly simplified discourse transcription with a free English translation.

1.	...(0.6) /Мы через них ..(0.3) \про<u>е</u>-ехали.	We drove through them.
2.	....(0.8) /Потом ..(0.2) нам встретился <тут | <u>какой-то</u>> /м<u>о</u>стик,	Then we came across a little bridge,
3.	...(0.6) очень ..(0.1) ' \<u>у</u>зенький,	a very narrow one,
4.	...(0.7) <и> /мы через него еле /про<u>е</u>хали тоже,	and we hardly drove through it too,
5.	....(1.2) /потом подъехали к /пл<u>я</u>жу,	then approached a beach,
6.	...(0.5) (/Я вообще плавать не \ум<u>е</u>-ею.	(In principle I cannot swim.
7.	..(0.4) Не \ум<u>е</u>ла тогда,	I could not then,
8.	*когда мне это \сн<u>и</u>лось.*)	when I dreamed this.)
9.	....(1.1) \в<u>от</u>,	OK,
10.	и-и' ....(1.1) /я почему-то \поплыл<u>а</u>.	and I swam for some reason.
11.	/Нырн<u>у</u>ла,	I took a dive,
12.	и \поплыл<u>а</u>.	and swam.

This excerpt consists of 12 EDUs. Dotted ovals in Figure 1 indicate the crucial pitch accents in each of the EDUs, and are numbered accordingly. EDUs marked with a clear rising accent (typical comma intonation) are ##2, 4, 5, and 11. Apparent instances of the falling accent (typical period intonation) are EDUs ##1, 6, 8, 10,12. As can be seen in Figure 1, the target level of period falling is 190 to 210 Hz. However, a deeper analysis of prosody reveals that not all falling accents are the same. In the speech of each individual storyteller there is a significant difference between what can be called a definitive or terminal falling (numbers provided above),

and a quibbling of intermediate falling. The latter is found in EDUs 3 and 9: its target level is 230 to 240 Hz, approximately 3 semitones higher than the level of the terminal falling. Experts trained to identify pitch accents do not characterize this phenomenon as identical to period falling. In terms of discourse semantics it rather qualifies as a variant of a comma accent. So one has to recognize that the non-terminal accent can be realized in two physically contrasted ways: rising and intermediate falling. Furthermore, the falling comma intonation can be prosodically complicated by tone rising on post-accent syllables. In this case the falling may actually be quite low. An example of this pattern is found in EDU #7 of the example. This way or another, stories abound with 'falling commas', and they are nearly as common as 'rising commas'.
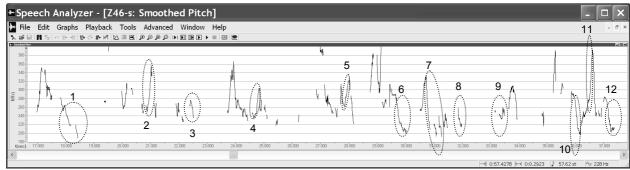

Figure 1. F0 graph of the discourse excerpt.

Sometimes the contrast between the intermediate and terminal falling is quite sharp and is easy to identify. If it were always this way, the notion of sentence in speech would be problem-free. However, in the speech of many storytellers a boundary between the 'comma falling' and the 'period falling' can be posited only through a complex analytic procedure. The acoustic reality does not unequivocally tell us: this is a falling comma intonation, this is a period intonation. So identification of sentences turns out a tricky task. These difficulties are not insurmountable. When substantial discourse material is recorded from a given speaker, a comprehensive analysis of all prosodic features (target levels of falling, post-accent rising, pausing, latching vs. reset, steepness and the interval of falling) usually provides an experienced expert with an understanding of the individual's prosodemic distinction between the two patterns.

However, formulation of such analytic techniques requires significant effort. The method developed for one discourse type may require significant tune-up when applied to other types, such as conversation. Moreover, such techniques must be developed and tested for each specific language. (There is evidence that many languages differentiate between the two kinds of falling.)

What is the function of sentence? Why would speakers use them, provided that the whole story is one statement illocutionary act? Most likely, the reason is pursuit for the ease of cognitive processing. The cognitive system favors intermediate hierarchical nodes rather than highly complex unstructured sets of EDUs. (This explanation is not incompatibe with the notion of 'superfoci of consciousness' put forward by Chafe 1994: 140 as a cognitive explanation of sentence.) Still speakers enjoy significant freedom in their sentential strategies. Some do not posit story-internal sentence boundaries; the whole story is produced as one sentence. Some other gravitate toward extra short sentences, often coinciding with EDUs.

To conclude, sentences are groups of EDUs found not only in written language but also in speech. In order to identify sentences one must employ a complex analytic procedure. Sentences are very far from being self-evident units of language, and identification of sentences can be clear and feasible to a various degree. Individual strategies of sentence formation are highly variable. Overall, sentence is a difficult, non-elementary, and elusive notion. Unlike clause and EDU, sentence should not be considered a basic unit of language.

REFERENCES

Chafe, Wallace. 1994. Discourse, consciousness, and time. Chicago: Univ. of Chicago Press.
Kibrik, Andrej A., Vera I. Podlesskaya (eds.) In press. Night dream stories: A corpus study of
    spoken Russian discourse.