UNDERSTANDING SPOKEN DISCOURSE:
THE CONTRIBUTION OF THREE INFORMATION CHANNELS[1]

Andrej A. Kibrik (Institute of Linguistics, Russian Academy of Sciences, kibrik@comtv.ru)
Ekaterina M. Èl'bert (CJSC GLITNIR Securities, ekaterina_elbert@inbox.ru)

Modern linguistics generally presumes that linguistic form is a sequence of phonological segments. Segments form words, words make up phrases, etc. Thus, according to this approach, dominating modern linguistics, language is ultimately segmental. Admittedly, there exist non-segmental aspects inherent to language and discourse. The first of these is prosody, that is non-segmental phonetics. Prosody includes a variety of components, including pausing, accents, tone, tempo and length, registers, reduction, phonation, loudness, etc. As has been demonstrated by a number of authors (for example, Kodzasov 2001), prosodic features have a significant semantic content and thus bear a significant load in linguistic communication. It is possible to learn much about the character of communication taking place behind a wall, when one can only hear prosody. Prosody, as well as the segmental component, is an important information channel in discourse.

In addition to these two vocal channels, there exists non-vocal communication via body language, including gestures, mimics, postures, proxemics, etc. Here these various visually transmitted elements are collectively called the visual channel in discourse. The visual channel affects communication severely. If one is telling a compliment with a grimace of disgust, this changes the message entirely. Gestures are now viewed by many (see e.g. McNeill 1992) as a system not separate from language but forming a single complex with it. We thus tentatively split all elements that can possibly convey information in discourse into three channels: visual, prosodic, and verbal (=segmental).

Prosody and body language are taken seriously and explored by many linguists, but are generally neglected in mainstream linguistics. In contrast, in some other disciplines dealing with human communication a very different view is taken. In particular, in applied psychology it is often stated that body language conveys more than a half of any message. Specific figures, describing the relative contribution of the three channels, circulate in the literature: body language conveys 55% of information, prosody 38%, and the verbal component the remaining 7%. These figures probably go back to Sulger 1986. So the question is: which viewpoint has more truth to it: that of linguistics or that of applied psychology?

In this paper we attempt a scientific study of relative contributions of the three channels. We have used an experimental excerpt from a Russian TV serial "Tajny sledstvija" – "Mysteries of the investigation" (3 min. 20 sec. long), preceded by a 8 minutes context (that starts from the beginning of the series). The excerpt fully consists of a conversation, to ensure that we are testing the understanding of discourse rather than of the film in general.

The three channels – verbal, prosodic, and visual – have been isolated from each other and presented in all possible combinations. The visual channel by itself is video alone (without sound), the verbal channel is subtitles running in temporal alignment with the original film, and the prosodic channel is the original audio component with a superimposed filter creating the effect of a conversation behing the wall.

Each version of the experimental excerpt was shown to a group of subjects. Altogether there were eight (=2^3) experimental groups, including the group that watched the original excerpt (all three channels together) and the control group that watched no experimental excerpt at all (that is, watched the context only). Each of the groups numbered 10 to 17 persons, from 18 to 71 years old. Every subject was instructed to watch the context and the experimental excerpt and then answer a set of questions concerned with the experimental excerpt alone. A subject was supposed to choose only one answer out of four listed variants. An example of a question and four answers, in an English translation:

What Tamara Stepanovna offers Masha before the beginning of the conversation:
  to take off her coat
  to have a cup of tea
  to have a seat
  to have a drink

The third answer is fully correct, and the first two are plausible. The fourth one is implausible provided that the story is about a schoolboy's mother visiting his teacher at school.

The questionnaire originally contained 29 questions, but six of them were subsequently discarded as they turned out either trivial or prone to guessing. So after a testing phase 23 questions were kept. The percentage of correct answers was used as a measure of understanding discourse. Table 1 shows the mean percentage of correct answers in each of the eight experimental groups.

Group number	1	2	3	4	5	6	7	8
Experimental material	original	sound	subtitles plus video	prosody plus video	subtitles	prosody	video	nothing (context alone)
Information channels	verbal prosodic visual	verbal prosodic	verbal visual	prosodic visual	verbal	prosodic	visual	[*none*]
Mean %% correct answers	87.4%	70.4%	73.9%	51.2%	72.0%	51.1%	61,7%	38,3%

Table 1. Mean percentage of correct answers in the eight experimental groups

Conclusions from these results are the following.
1. Each of the three information channels, taken in isolation, is quite informative. The percentages in groups 5 through 7 are significantly higher than the percentage in group 8.
2. The hierarchy of informativeness can be represented as follows: verbal > visual > prosodic.
3. Interestingly, combining the verbal channel with one additional channel does not increase the percentage of correct answers (compare group 5 with groups 2 and 3).
4. Adding the visual channel to the prosodic channel does not result in increase in correct answers (compare group 6 with group 4). The combination 'prosodic plus visual' (group 4) displays significantly lower result than in other pairs of channels (groups 2 and 3). Evidently, this combination is not customary for subjects, and they have trouble integrating information from prosody and video.

The third and the fourth conclusions suggest, cumulatively, that subjects use the verbal channel as the leading one and information from two other channels is primarily used through integration with the verbal channel.

In order to estimate the relative contribution of the three channels, the following simple technique can be used. Assuming, for the sake of simplicity, that all three channels are independent, one can sum the percentages in groups 5 through 7 (72+51+62=185) and then normalize these individual contributions to 100%. The resulting contributions are: 39% for the verbal channel (72:1.85≈39), 28% for the prosodic channel, and 33% for the visual channel. Therefore:
- All information channels are highly significant, and the traditional linguistic viewpoint is erroneous
- The verbal channel is the most central, and the viewpoint popular in applied psychology is erroneous

REFERENCES

Kodzasov, Sandro V. 2001. Kombinatornye metody v fonologii. Moscow: MGU.
McNeill, David. 1992. Hand and mind: what gestures reveal about thought. Chicago: University of Chicago Press.
Sulger, François. 1986. Les gestes-vérité. Paris: Editions Sand.