SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 104–111.
Los Angeles, California, 21–23 September 2011.

Local Discourse Structure of Chat Dialogues: Evidence from Keystroke Logging

Evgeny Chukharev-Hudilainen
The A. A. Hudyakov Center for Linguistic Research
Russia
evgeny@chukharev.ru

Abstract

While the global discourse structure that describes how utterances are grouped into larger units of discourse has received much attention both in oral and in computer-mediated dialogues, the local structure (i. e. the structure of individual utterances) of chat conversations has not been previously studied in a psycholinguistic perspective. In this paper we explore some evidence of cognitive processing in spontaneous electronic language production in an experimental web chat. Keystroke logging is used to detect hesitation pauses in chat, which are then mapped onto the local discourse structure as marked up in the corpus of chat dialogues by four independent coders.

1 Introduction

As a type of discourse, **dialogue** (or conversation) is distinct from another discourse type, **text**. This distinction is so crucial that Dixon and Bortolussi (2001) argued that text is not communication at all, because there is no feedback from the author at the time of reading, and therefore it always remains unclear what exactly the author had in mind when writing the text. Before the era of computer-mediated communication (CMC), these two discourse types were strongly associated with communication media: texts were mostly written, while conversations were usually oral. Now that CMC has become ubiquitous, this association has loosened considerably. Online text-based chats, for instance, are actually conversations rather than texts (cf. Beißwenger, 2003).

1.1 Levels of Discourse Structure

Both discourse types are structured on two levels, which Kibrik (2003) termed **global** and **local**. On the global level, discourses are structured into units larger than individual utterances, such as paragraphs in texts or contributions (or turns) in conversations. The local structure describes the basic units from which utterances are built. There are at least two types of such units (cf. Polanyi, 2001): elementary discourse constituent units (or predicate expressions) and extrapropositional discourse operators.

A **predicate expression** is typically defined as a linguistic sign denoting a single state of affairs (a situation or fact). Examples of predicate expressions include clauses, phrases with secondary predication, event names, etc. **Discourse operators** are non-propositional elements of utterances which do not express any states of affairs.

It has been argued that the local discourse structure reflects the workings of the mind in the course of the utterance production. According to Hudyakov's (2000) model of semiosis, the production of an utterance begins with the construction of a proposition in the speaker's mind. Said proposition is then embodied in a predicate expression in the local discourse structure. However, since the speaker's intentions usually exceed simply asserting states of affairs, extrapropositional discourse operators are further introduced into the utterance, in order to endow it with sense, in addition to the propositional semantics (or meaning) it already has. In Hudyakov's view, it is the sense, and not the meaning (semantics), which is at the core of the communication process.

The global discourse structure of dialogues (both oral and CMC) has received much attention from researchers to date, while, to our knowledge, the local structure of computer-mediated conversations has not been studied yet. In the present work, we investigate such structure in relation to hesitation pauses viewed as indirect psycholinguistic evidence of cognitive processing.

1.2 Hesitation Pauses

Hesitation pauses have been treated as a manifestation of the more general blocking of activity which occurs when organisms are confronted with situations of uncertainty, and when taking the next step requires an act of choice. According to Goldman-Eisler (1968), spontaneous speakers (and writers / typists) keep making three kinds of choices while objectifying their utterances: a) content decisions, which can be either completely non-verbal or tied to key words standing out as semantic landmarks without any syntagmatic ties; b) syntactic choices, which are crucial for any kind of coherent speech; c) lexical choices, i. e. selecting words to fit the syntactic framework in accordance with the semantic plan. It has been shown that all three types of choices made in the course of spontaneous speech must be accompanied by an arrest of the speech objectification process, i. e. by pausing (unless, of course, the speech has some degree of preparedness and some planning is done before the utterance begins).

Though hesitation phenomena have been thoroughly investigated in oral speech only, they also occur in spontaneous CMC as observed in chats and instant messengers.

Generally, CMC appears to be an easier object of linguistic research compared to oral speech because it does not require transcription of the raw material before including it in corpora for quantitative analysis. However accurate, transcription of oral speech productions inevitably fails to render every detail of intonation or capture non-verbal cues with complete precision. A log of a chat conversation, on the contrary, inherently contains all information that was actually exchanged by the interlocutors in the course of the conversation, and this information is readily available in a form suitable for corpus analysis.

On the other hand, the study of hesitation pauses in text-based CMC is challenging due to the quasi-synchronous nature of communication. Quasi-synchronous communication is similar to synchronous in that the delays in the communications channel are barely (if at all) noticeable, and the recipient gets the messages nearly instantaneously, i. e. approximately at the same time when they are objectified by the sender. The difference between quasi-synchronous and fully synchronous types of communication is that in the former case the message objectification process is hidden from the addressee: first the sender types the message in an edit box, and then it is sent to the recipient (Hård af Segerstad, 2002; Dürscheid, 2003). This implies that though the sender is likely to pause while typing the message, these pauses will remain unseen by both the recipient and the meta-observer who would study the message logs (Beißwenger 2003).

The only way to detect hesitation pauses in chat is through keystroke logging. The use of keystroke logging as a research method in linguistics is not a new field of study; however to date this area of research has primarily focused on written composition and translation studies. In our work, we aimed at extending the contributions of keystroke logging to spontaneous CMC in chat. If a keystroke log is available to the researcher, hesitation pauses may be defined as prolonged intervals between consecutive keystrokes.

Indeed, according to Rumelhart and Norman's (1982) model, a complex mechanism of motor schemata coordinating simultaneous movements of several fingers is employed to shorten inter-keystroke intervals in fluent typists. Obviously, hesitation terminates this mechanism, and when typing is resumed additional time is required to prepare and start executing a new motor program. This time together with the duration of the hesitation pause *per se* (when the speaker makes a linguistic choice) constitutes the observed hesitation interval between consecutive keystrokes.

It still remains to be decided how exactly long an interval should be classified a hesitation pause. In previous work, a cut-off value of some 1–2 seconds was chosen and delays in typing exceeding this value were treated as pauses (cf. Alves et al., 2007). In our opinion, such choice of the cut-off value is somewhat arbitrary, and a better grounded way of distinguishing hesitation from non-hesitation pauses should be established.

2 Methods

In our experiment a novel web application (Figure 1) was used to log keystrokes made by chat users in a game, in order to measure the duration of inter-keystroke intervals, and further to analyze these durations in relation to the units of the local discourse structure. The web chat was hosted at http://www.justchat.ru and made freely available to the public.

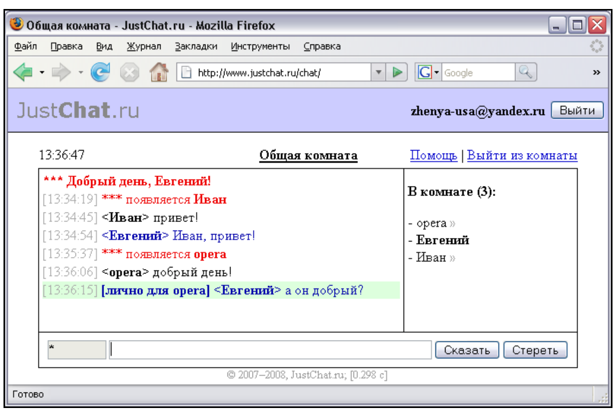

Figure 1: Web Chat Interface

When signing up for a free account at the chat website, everyone had to accept a user agreement and give their explicit permission to use any information gathered during their communication for the purposes of the present research. Therefore, all chat users of the chat served as subjects in our experiment. As we discovered later by interviewing our subjects, most of them had not actually read the agreement before clicking "I Agree" and thus ended up being unaware of the fact that their communication was logged along with keystroke timings for research purposes. In fact, it was good for the research as subjects behaved more naturally than they would do otherwise.

We used the chat to hold on-line intellectual games designed after the popular Russian game called "What? Where? When?." All our game sessions were held in the Russian language.

There are two versions of this game. The initial version is a popular Russian TV show that has been on since 1975. In the show, a team of six players are posed questions that they have to answer under a time limit. They are given one minute to discuss each question inside the team, and then the team captain announces the final answer. The show host announces the correct answer to the question. The score is kept in order to determine whether the team won or lost the game.

The sports (or competitive) version of the game was invented by the TV show fans so that more people could play the game without having to take part in the show. In the sports version several teams compete in finding answers to the questions, which are posed to all teams at the same time.

To answer the questions correctly, no special knowledge is usually required, but rather common knowledge along with logical reasoning skills. Good partnership and collaboration within a team is known to be one of the key success factors in this game. For the sake of illustration, here are two sample questions translated into English:

- Margaret Thatcher believes that no one would remember the Good Samaritan if he'd only had good intentions. What else, according to the "Iron Lady," did he have to have? *(Correct answer: The money, to give to the man in need.)*

- What color is the longest line on the map of the London Underground? *(Correct answer: Blue. It is the River Thames.)*

Our chat games were based on the sports version of "What? Where? When?," but differed in that the team players did not have any personal contact during the game, the time limit was increased from one to four minutes per question, and the number of players on a team was unlimited. Questions for each game session were randomly drawn from an online database at http://db.chgk.info. The players were unfamiliar with the questions before the game started, which guaranteed spontaneity of their communication.

The multi-room feature of the chat enabled several teams to play the game at the same time. Each of the teams occupied a separate chat room, where they could discuss the questions in private. Questions were posed to the teams through chat bots, one per room, impersonating the show host. One player on each team was chosen to be the team captain. After a team had finished discussing a question, it was the captain's responsibility to formulate the final answer and send it to the bot, who then announced both the correct answer and whether the team's answer was accepted as correct. Since the answers could be worded differently, a human operator was employed to judge the answers behind the scene in real time. Teams who succeeded in answering a question were awarded one point each, and the winning team was the one having earned the most points by the end of the game.

After the game was over, the team rooms were closed and all players were automatically transferred to a common chat room where they could discuss the game or just enjoy talking.

The web chat software was designed to keep a keystroke protocol reflecting inter-keystroke intervals with a resolution of 1 ms.

3 Results

A total of 34 games were held, in which anyone could participate. Invitations to join the games were sent out to people by e-mail and posted on the "What? Where? When?" fan forums on-line. 47 chat sessions in the team rooms and 39 sessions in the common room where all players met before or after the games were logged. The logs contained 22,501 messages (contributions) overall.

To reduce the size of the corpus while keeping it representative, the following procedure was applied:

1) Data from subjects who produced less than 10 messages each were dropped.
2) If a subject produced less than 100 messages, all sessions this subject took part in were retained in the corpus.
3) If the deleting of a session from the corpus would cause the number of remaining messages produced by at least one subject fall below 100, such session was retained in the corpus.
4) Sessions not matched by rules 2 and 3 were dropped from the corpus.

Following this procedure, the corpus shrank by 48.8%. 25 team room sessions and 18 common room session were retained, containing a total of 11,518 messages (over 68,000 tokens) produced by 36 subjects (14 women). All subjects were native Russian speakers, their average age was 23.8 ± 3.9 years (range 17–38 years), average computer experience 3.4 ± 3.7 years (range 1–18 years). According to the data provided by the subjects through the sign-up form, 10 of them were IT professionals, 12 were college students (including 5 IT students), 14 were home or office computer users. Only 8 out of 36 subjects touch typed, others were keyboard gazers. 22 subjects were using chats or instant messages on a daily basis. The subjects' typing rate averaged at 110 ± 52 keystrokes per minute.

The distribution of messages among subjects appeared very uneven. The top five subjects produced as many as 58.1% of messages, while the bottom nine produced less than 100 messages each. To make balanced judgments from the corpus data, all statistics were first computed separately for each of the subjects and then an average value was found.

First we analyzed the durations of time intervals between consecutive keystrokes in order to identify and classify hesitation pauses. Then we studied these pauses in relation to the elementary discourse constituent units.

3.1 Hesitation Pauses

Obviously, not all of delays in typing were due to linguistic hesitation. First of all, an effort was made to eliminate noise in the delays arising, for example, from a subject pausing to drink coffee some time during the chat, switching to another application on their computer, or anticipating other's responses. To do so, initial pauses (i. e. pauses before the onset of the typing of a new message) as well as those associated with the keyboard focus loss by the input field in the chat window were excluded from further analysis. Pauses

It was our aim to analyze pauses appearing while typing messages, not while editing them. So keystrokes made to append characters to the end of messages were only studied, and keystrokes used to delete characters or insert characters in the middle of the message were excluded from analysis.

Then we tried to establish a cut-off value between **motor pauses** (non-hesitation) solely attributable to the motor execution of typing, and **hesitation pauses** that were linguistically grounded.

Figure 2 displays a typical histogram showing the distribution of pauses between keystrokes made by one of our subjects (Subject #3).

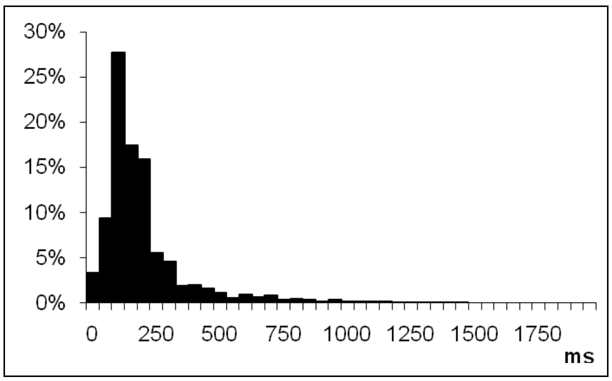

Figure 2: Pause Distribution (Subject #3)

Clearly, this distribution can be roughly split into two parts. The left-hand part of the histogram closely resembles normal distribution, while the long right-hand "tail" corresponds to pauses which are not distributed normally. Since motor pauses depend upon many random factors, they would probably be distributed normally if no hesitation at all were present. (Typos when letters are typed in the wrong order may be treated as "negative" pauses to account for the positive probability mass assigned to the "negative" pauses by the normal distribution law.) So it was natural to assume that the right-hand portion of the distribution corresponded to the actual hesitation pauses.

Suppose that we find two border values for each subject, t_m and t_h, so that if a pause t_i is shorter than t_m ($t_i < t_m$) it is most probably a motor pause. If a pause is longer than t_h ($t_i > t_h$), it is most probably a hesitation pause. Finally, if a pause is between the two border values ($t_m \le t_i \le t_h$), it could be either.

To find t_m, we chose the skewness value of

$$S(X) = \frac{E(X_j - \overline{X})^3}{\sigma^3}$$

as a rough metric of how close a sample was to the normal distribution, and then for each of our subjects we searched for such a value of t_m that the sub-sample of pauses shorter than this value $\{t_i \mid t_i < t_m\}$ would be the most symmetric, that is, the skew would be minimal ($|S| \to \min$). We did it through a brute force search among integer values from 50 to 1000 ms. Then we assessed the parameters of the distribution of motor pauses by computing the statistics (μ, σ^2) of the sub-sample $\{t_i \mid t_i < t_m\}$. Since 97.7% of normally distributed values are within three standard deviations from the mean, we defined $t_h = \mu + 3\sigma$. Not unexpectedly, the value of t_h varied across subjects and depended upon their typing rate, averaging at 386.9 ± 102.9 ms.

For Subject #3, whose data is shown in Figure 2, $t_m = 296$, $\mu = 154.7$, $\sigma = 60.2$, $t_h = 335$ (ms).

Since the minimal unit of typing is generally agreed to be a token, not an individual character (cf. Rumelhart & Norman, 1982), we proceeded to study pauses which occurred on the token level. We automatically tokenized the corpus, and for each token we found the longest pause $p_j = \max \{t_i\}$ which occurred while typing this token. We called it the **peak pause** of this token.

Then we analyzed the distributions of the peak pauses individually for each of the subjects. For the sake of illustration, a distribution obtained from Subject #3 is shown in Figure 3. All of the peak pause distributions looked very similar to the distribution of all inter-keystroke intervals described above (cf. Figure 2), though the border point between the symmetric and the asymmetric parts was obviously shifted to the right. In order to find a border value between the two parts of the distribution, we used the same statistical procedure as described above. The border value p_h computed similarly to t_h also varied across subjects, averaging at 937.9 ± 357.4 ms. We hypothesized that pauses shorter than this value marked hesitation in the production of individual tokens (**lexical hesitation**), while longer pauses occurred when hesitation was associated with a higher-level planning decision (we called those **segment hesitation pauses**, for the lack of better term).

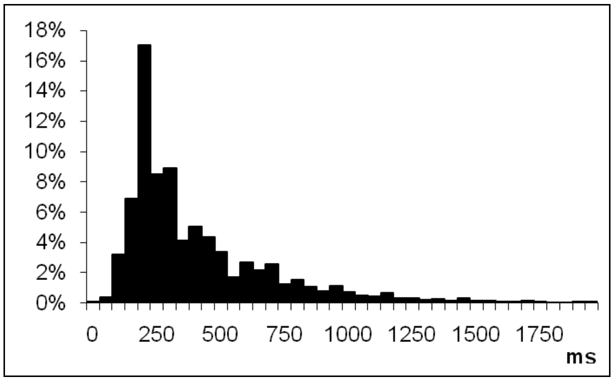

Figure 3: Peak Pause Distribution (Subject #3)

In order to verify that our findings indeed had some "physical meaning," we performed the following test. We randomly split the corpus into strings of consecutive characters. The lengths of the strings were distributed precisely as the lengths of the actual tokens in the corpus. We called these strings **pseudo-tokens**. For the pseudo-tokens, peak delays were also determined and analyzed in the aforesaid manner. The distribution function for the pseudo-token peak delays appeared to differ profoundly from that of the actual tokens. In particular, the boundary value p'_h for the pseudo-tokens averaged at 709.6 ± 270.7 ms, and for 34 out of 36 subjects (94%) it was lower than the boundary value selected for the actual tokens ($p'_h < p_h$). Thus the connection between lexical hesitation pauses and tokens was confirmed.

3.2 Segment Hesitation and Predicate Expressions

We had four independent coders mark up predicate expressions in our corpus of chat messages. A novel markup tool was used to perform this task (see Figure 4). Additionally, in each predicate expression a **vertex** was marked up, defined as the word representing the semantic predicate of the underlying (mental) proposition.

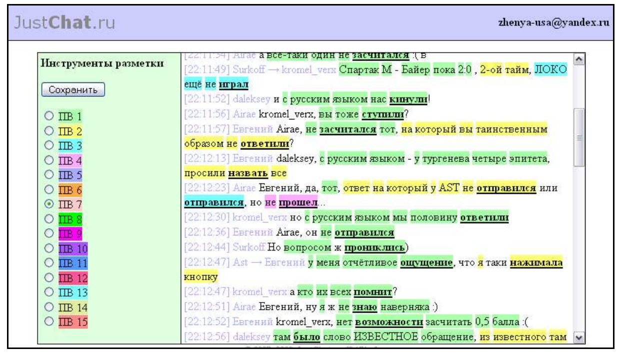

Figure 4: Corpus Markup Tool

Note that predicate expressions are semantic units that do not always correspond to the syntactic structure of the utterance. Moreover, the vertex of the predicate expression does not necessarily coincide with the linguistic predicate of the sentence. For example, the sentence

Инженеры выполняют работы по проведению
Inžen'ery vypoln'ajut raboty po prov'ed'en'iju
'Engineers perform work on carrying out

эксплуатации системы.
ekspluatatsii s'ist'emy.
of exploitation of [the] system'

denotes a single situation (the engineers operate the system), verbalizes a single proposition, and is actually a single predicate expression despite containing a finite verb and three event names. Though the verb *выполняют* 'perform' is the linguistic predicate of the sentence, it is partially delexicalized, and the word *эксплуатации* 'exploitation' actually functions as the vertex of the predicate expression.

Segments of discourse not coded as part of any predicate expression were automatically marked as extrapropositional discourse operators for further analysis.

Krippendorff's α was computed to assess intercoder reliability (see Krippendorff, 2007). The value of 0.79 was obtained for the predicate expression markup. For the vertices of predicate expressions, α = 0.84. These values are interpreted as excellent reliability according to Strijbos and Stahl (2007), which indicates the psycholinguistic relevance of predicate expressions and their vertices in the discourse.

An example of a chat message with semantic markup follows. This particular message appeared in the context of discussing whether the city of Leninabad had been renamed or not.

{теперь не ***знаю****} {как* ***называется****}, но наверное*
t'ep'er' n'e znaju kak nazyvajets'a no nav'ernoje
'now I don't know what [it] is called but perhaps

*{****переименовали****}. хотя может и {нет}*
p'er'eim'enoval'i hot'a možet i n'et
[they have] renamed [it] though maybe not'

Here braces indicate the borders of predicate expressions, and the underlined words were those marked up as the vertices. Note that there is no explicit vertex in the last predicate expression, *нет* 'not.'

For the analysis of segment hesitation pauses, data were dropped from the subjects for whom less than 30 such pauses were observed, which left us with 24 subjects (66%). For each of them, the following sets of tokens were analyzed: T – the set of all tokens produced by this subject; V – the set of vertex tokens of predicate expressions; I – the set of initial tokens of predicate expression. Within each of the sets, subsets of tokens marked with segment hesitation (i. e. where $p_j > p_h$) were found, labeled T_h, V_h, I_h, respectively.

The following inequalities were tested for each of the subjects:

$$\frac{|I_h|}{|I|} > \frac{|T_h \setminus I_h|}{|T \setminus I|} \tag{1}$$

$$\frac{|V_h|}{|V|} > \frac{|T_h \setminus V_h|}{|T \setminus V|} \tag{2}$$

$$\frac{|V_h \cup I_h|}{|V \cup I|} > \frac{|T_h \setminus (V_h \cup I_h)|}{|T \setminus (V \cup I)|} \tag{3}$$

Inequality (1) held true for 17 subjects (71%), inequality (2) held true for 16 subjects (67%), and inequality (3) held true for 19 subjects (79%). It means that in our data initial and vertex tokens of predicate expressions were more frequently marked with segment hesitation pauses than non-initial and non-vertex tokens.

These results support the claim that hesitation pauses are associated with the production of predicate expressions in the chat discourse. On one hand, hesitation while typing the initial token of the predicate expression can be attributed to the fact that the semantic and syntactic structures of the latter has not been finalized by the onset of typing. On the other hand, the vertex represents the mental (relational) predicate, i. e. the semantic center of the proposition, demands that substantial cognitive effort be applied to choose both the concept for the predicate and the most appropriate word for it.

Our data indicated no association between hesitation pauses and extrapropositional discourse operators.

There was no difference in the distribution of hesitation pauses between the task-related discussions (game sessions) and the free conversations that took place before or after the games.

4 Conclusions

In our study, we have applied keystroke logging as a method of linguistic research to spontaneous CMC in an experimental web chat. Our experiment yielded a representative corpus of chat messages logged along with keystroke timings. The communication environment was very naturalistic: our subjects were generally unaware that their conversations were logged, and they were vividly interested in the game because they enjoyed it and used our chat to practice for their real-life games which most of them played on a regular basis.

Due to the nature of the game, the players' communication primarily consisted of brainstorming, i. e. generating ideas in multiple simultaneous threads, therefore the impact of the interactional contingencies (interruptions, turn-taking, turn-yielding, holding the floor, etc.) on the timing phenomena was limited. It allowed us to focus on the hesitation pauses which occurred in the individualistic turn formulation.

It was possible to establish a statistical criterion for the detection of true hesitation pauses in chat dialogues. Furthermore, hesitation pauses were classified into two types: lexical and segment hesitation. Segment hesitation was strongly associated with the production of elementary discourse constituent units (predicate expressions).

We believe that the reported results may be applied to the detection of users' hesitation in various human-computer dialogue systems. For example, an on-line L2 learning system on which we are currently working will use hesitation patterns in typing to identify the student's fluency level at completing certain linguistic tasks.

References

Alves R. A., São Luís Castro, Liliana de Sousa and Sven Strömqvist. 2007. Influence of Typing Skill on Pause Execution Cycles in Written Composition. Writing and Cognition: Research and Applications. Elseveir, Amsterdam; Oxford. 55–65.

Beißwenger M. 2003. Sprachhandlungskoordination im Chat. Zeitschrift für Germanistische Linguistik. 31(2). 198–231.

Dixon P. and Bortolussi M. 2001. Text Is Not Communication: a Challenge to a Common Assumption. Discourse Processes. 31(1). P. 1–25.

Dürscheid Ch. 2003. Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Probleme. Zeitschrift für Angewandte Linguistik. 38. 37–56.

Goldman-Eisler F. 1968. Psycholinguistics. Experiments in Spontaneous Speech. Academic Press, London; New York.

Hård af Segerstad Y. 2002. Use and Adaptation of Written Language to the Conditions of Computer-mediated Communication. Göteborg University, Göteborg.

Hudyakov A. A. 2000. Semiozis prostogo predloženija. Pomor University, Arkhangelsk.

Kibrik A. A. 2003. Analiz diskursa v kognitivnoj perspektive. Russian Academy of Sciences, Moscow.

Krippendorff K. 2007. Computing Krippendorff's Alpha reliability. http://www.asc.upenn.edu/usr/ krippendorff/webreliability.doc

Polanyi L. 2001. The Linguistic Structure of Discourse. The Handbook of Discourse Analysis. Blackwell, Oxford. 265–281.

Rumelhart D. E. and Norman D. A. 1982. Simulating a Skilled Typist: A Study of Skilled Cognitive-motor Performance. Cognitive Science. 6(1). 1–36.

Strijbos J.-W. and Stahl G. 2007. Methodological Issues in Developing a Multi-dimensional Coding Procedure for Small-Group Chat Communication. Learning and Instruction. 17(4). 394–404.