Компьютерная лингвистика и интеллектуальные технологии

По материалам ежегодной международной конференции «Диалог» (2017)

Выпуск 16

Том 2 из 2
Компьютерная лингвистика: лингвистические исследования

Computational Linguistics and Intellectual Technologies


Issue 16

Volume 2 of 2
Computational Linguistics: Linguistic Research
Редакционная коллекция: В. П. Селегей (главный редактор), А. В. Байтеин, В. И. Беликов, И. М. Богуславский, Б. В. Добров, Д. О. Добровольский, Л. М. Захаров, Л. Л. Иомдин, И. М. Кобозева, Е. Б. Козеренко, М. А. Кронауз, Н. И. Лауфер, Н. В. Лукашевич, Д. Маккарти, П. Наков, Й. Нивре, Г. С. Осипов, А. Ч. Пиперски, В. Раскин, Э. Хови, С. А. Шаров, Т. Е. Янко


Сборник включает 71 доклад международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2017», представляющих широкий спектр теоретических и прикладных исследований в области описания естественного языка, моделирования языковых процессов, создания практически применимых компьютерных лингвистических технологий.

Для специалистов в области теоретической и прикладной лингвистики и интеллектуальных технологий.

© Редколлегия сборника «Компьютерная лингвистика и интеллектуальные технологии» (составитель), 2017
INTERACTION AND EMPATHY
AS ELEMENTS OF NARRATIVE STRATEGIES
IN THE RUSSIAN CLIPS CORPUS

Bergelson M. B. (mirabergelson@gmail.com),
Khudyakova M. V.
National research University Higher School of Economics,
Moscow, Russia

This paper focuses on evaluation of discourse abilities of speakers with
brain damage: people with dynamic aphasia (PWA(d)) and right hemisphere
damage (RHD) as compared to healthy speakers of Russian language. The
study is based on the material from the Russian CliPS corpus that contains
retellings of the Pear Film produced by PWA and RHD, as well as neurologi-
cally healthy controls.

The nature of the narratives in the corpus allows for a comparative in-
vestigation of discourse on the level of micro-structure (grammatical and
lexical phenomena) and on the macro level: narrative structure, coherence
and cohesion, interactional patterns and narrative discourse strategies.
In this paper we present results of the comparative analysis of some macro
level discourse strategies: the way interaction and empathy are realized
in the stories by PWA(d), RHD and healthy speakers.

We have found significantly higher numbers of attitude expression
markers, as well as significantly lower numbers of cognitive difficulties
markers, in healthy speakers as compared to PWA(d). These results sup-
port what is known about difficulties that PWA(d) demonstrate in discourse
production tasks. While PWA(d) use interactive markers to get a break from
keeping with the story plan, they avoid using epistemic predicates whose
subjects are the story characters.

We also present qualitative analysis of the discourse strategies of
healthy speakers.

Key words: corpus linguistics, narratives, aphasia, interaction markers,
empathy, discourse strategies

1 The paper was prepared within the framework of the Academic Fund Program at the National
Research University Higher School of Economics (HSE) in 2016–2017 (grant №16-05-0024)
and supported within the framework of a subsidy granted to the HSE by the Government
of the Russian Federation for the implementation of the Global Competitiveness Program
1. Introduction

The clinical discourse studies are aimed at studying language use by persons with acquired language disorders due to focal brain damage or persons with psychiatric and neurodegenerative diseases. The ‘discourse approach’ in clinical linguistics can be attributed to the change of perspective from assessing separate linguistic skills in various disorders to the idea that communication skills should be assessed as a whole (see discussion in Linnik et al., 2015). Also, it has been discovered that difficulties at the word and sentence level do not exclude a sensible story and a communicative success (Olness et al., 2010). The reverse may be true as well: spared grammar and lexis not necessarily result in a coherent story (Sherratt & Bryan, 2012). On the other hand, investigation of language in clinical populations provides a chance to test models of discourse comprehension and production and investigate the correlation between linguistic skills in the domain of general linguistics.

Our project is focused on evaluation of discourse abilities of speakers with brain damage: people with dynamic aphasia syndrome (PWA(d)) and right hemisphere damage (RHD) as compared to healthy speakers of Russian language. The study is based on the material from the Russian CliPS corpus that contains retellings of the Pear Film produced by people with different types of aphasia, RHD, and neurologically healthy speakers.

In this paper we present results of the comparative analysis of macro level discourse strategies. Our goal is to look at the way interaction and empathy are realized in the stories by PWA(d), RHD and healthy speakers in order to decide whether these linguistic concepts can be used as parameters for discriminating between the pathological and healthy discourse. We call the units that signal interaction and empathy interaction markers (IMs), and empathy markers (EMs), respectively.

The two clinical groups (PWA(d) and RHD) were chosen for this kind of analysis because the nature of their specific language deficit is known to affect the narration (see 3.2).

Our hypothesis predicts that certain IMs allow compensating for the linguistic deficits of the clinical populations, while some other IMs are only cognitively accessible to the healthy speakers. To prove or falsify this hypothesis we perform intergroup comparisons.
2. Interaction and empathy

The crucial difference between such narratives as report, chronicles, or traditional myths, and stories lies in the fact that the former don't 'interact' with the addressee. Stories, on the contrary, are not so much about providing facts, but about expressing opinions on and giving evaluations of the facts described using all discourse production types, or discourse passages (Plungyan, 2008; Polanyi, 1989).

We understand linguistic interaction in the specific situations of telling stories as an interpolation of the following concepts:

- narrative vs. discursive mode (see Paducheva, 2008: нарративный vs. речевой режим);
- fore- and backgrounding (Weinrich, 1964);
- the world of story vs. the world of narration (Norrick, 2000).

When telling a story the primary system of communication is the world of story and the collateral system corresponds to the world of narration (an instance of communication at certain time and place). Building on that, we look at the narrative communicative event from the perspective of the discourse strategies used by the narrator (the speaker) to interact with the recipient of the story (the addressee). Realization of these strategies does not interfere with the regular story components (Fleischman, 1991; Hoey, 1994; Labov, 1972; Labov & Waletzky, 1967; Longacre, 1983; Polanyi, 1989) as they belong to two different dimensions, or worlds, of narration and story, respectively.

Linguistic devices that earmark interaction (IMs) may be found inside all types of the discourse passages (Kibrik, 2009; Longacre, 1990)—descriptive, narrative, instructive, argumentative, and expository. Some of them mark the end or the beginning of the elementary discourse units (EDUs), utterances, or whole episodes. They are different in their specific functions, but all of them serve to signal switching from the world of story to the world of narration.

Interaction allows the narrator to have a break from the story world with its demands for informational consistency: the necessity to balance narration and description, logical and chronological sequence of events, complications, and denouement.

The world of narration is comprised as two separate communication grounds. One is the real situation of storytelling where IMs are directed onto the addressee as part of the conversation. The other is a mental space where the narrators communicate with themselves while planning next portions of discourse. IMs then, are those linguistic elements that reveal narrators’ cognitive difficulties with planning, retrieval, or physical producing discourse. For the purposes of analysis that involves clinical discourse, this latter ‘communication ground’ and IMs serving it are especially prominent.

(1) мужчина интересного вида да конец лета надо полагать собирает урожай зелёный на видимо грушу по приставной лестнице
[a funny man, yes, it is the end of the summer, I believe, (he) harvests the fruit, is climbing the, seemingly, pear tree using the ladder]
On the other hand, empathy in the storytelling stands for the ability of the narrators, so to say, immerse themselves in the world of the story, to communicate what is going there from the character’s viewpoint. Such understanding builds upon the concept of empathy in syntax and semantics (Kuno, 1987; Kuno & Kaburaki, 1977), though is closer to the stylistic phenomena in the paradigm of interactional linguistics (Kalliokoski & Verschueren, 1991)—see below our definition in 3.3.1. If the narrators tell us about feelings and decisions of the characters (the boy understood that..., he decided to..., he did not like...) it would mean they are putting themselves in the characters’ shoes, taking, as it were, their roles in the story world, inventing the ways another person could think or feel.

3. Method and Material

3.1. The Russian CliPS corpus

The Russian CliPS corpus contains retellings of the “Pear story” film (Chafe, 1980) by people with different aphasia types (efferent, dynamic, acoustic-mnestic, and sensory), RHD and healthy speakers. The stories were audio-recorded and transcribed in ELAN (Wittenburg et al., 2006) on various linguistic levels, including lexical and grammatical information, segmentation into elementary discourse units, coherence and coreference annotation, as well as error classification (see Khudyakova et al., 2016; Toldova et al., 2016 for details).

3.2. Subcorpus for the study

Texts by three groups of speakers from the corpus were selected for the study of interaction and empathy components: PWA(d), RHD and healthy speakers. These two groups are united by difficulties they experience on the macro level of discourse.

Dynamic aphasia is a language impairment, resulting from damage to the language-dominant hemisphere (usually left), characterized by non-fluent speech output. The primary deficit in dynamic aphasia is in the planning of the utterance stage, while motor components and understanding are intact (Akhutina, 2015; Luria, 1972; Luria & Hutton, 1977).

Damage to the right (language non-dominant) hemisphere does not result in aphasia, but people with RHD demonstrate speech disturbances on discourse level: lower coherence, inability to maintain the main theme of the narrative. Nevertheless, they do not have any deficits on microlinguistic level (Glosser & Deser, 1990; Marini et al., 2005).

For the study we have analyzed 8 narratives by PWA(d) (5 females; mean age—49.5; age range—41–58), 5 narratives by people with RHD (2 females; mean age—54.2; age range—41–74) and 10 narratives by healthy speakers (4 females; mean age—65; age range—42–84).
3.3. Annotation

3.3.1. Empathy

For this specific communicative task, we define empathy as the narrator reporting of inner states (feelings, opinions and decisions) of the characters in the movie. The retellings of the movie are in no way personal stories, and the narrator has not experienced these feelings, nor made these decisions in his/her previous experience or even had positive information to guess about them. To report them one has to take a certain perspective, has to empathize with the subject of these inner states. Thus, in our annotation system we tag all predicates expressing inner states (we call them quasi-events) if their subject is a story character and the corresponding stance (discourse passage) belongs to the main line of the story.

(3) и тут мальчик подумал (Emp), что груши можно унести [so the boy thought (Emp) that he might take the pears away]

3.3.2. Interaction markers

The IMs are also annotated as a separate annotation layer. Everything that pertains to the world of storytelling where the speaker interacts with the listener—directly by appealing to him/her or by demonstrating to the listener his/her cognitive difficulties—is regarded as interaction. Fillers, feedback markers, appellations to the listener, but also: word search, false starts, repetitions. Interaction markers can occur both within the clause (e.g., let’s call the boy Vovochka, where let’s is an element of interaction within a descriptive clause) or comprise a separate clause (e.g., if you say so).

Functionally, all the IMs fall into several groups. Besides those that serve to reveal cognitive difficulties (see 3.3.2), there are pace regulating markers, attention drawing markers and expressing attitudes IMs.

The list of tags for the IMs by groups:

- **Pace regulating markers:**
  - **Fill**—(often repetitive) discourse markers (DMs) of cognitive search
  - **Reg**—DMs mostly signalling the beginning or the end of EDUs, utterances or stances (passages) that serve to pace narration and mark its chunks

- **Cognitive difficulties markers:**
  - **Rp**—repetitions of words and phrases
  - **Fs/Rep**—false starts followed by repairs
  - **Rep**—repairs that happen without preceding false starts
  - **SFs**—structural false starts

---

2 Another approach to the macro level analysis of the stories is based on the genre schema and its components, for example: abstract, introduction, main line events, evaluation, coda. The latter two, being part of the world of narrator, also belong to interactive aspect of storytelling. We will be dealing with componential analysis of stories in CliPS at the next stage of our project.

3 See Fraser 2009 for a comprehensive classification of discourse markers.
Attention drawing markers:
• *Apl*—DM that target to draw the addressee’s attention
• *DM*—other DM

Attitude expression markers:
• *Est*—markers of epistemic attitude
• *Jdg*—markers of axiological evaluation

The tag *Imp* refers to clausal or clause-like markers of interaction, as opposed to one- or two-word discourse markers.

Repetitions (*Rp*) may signal either the importance of the repeated item for the story, or difficulties in the process of mental search, which is especially true for the pathological discourse. Still the healthy speakers’ stories often demonstrate abundance of the mental search problem as part of the specific *I-am-not-Committing-Myself* discourse strategy.

### 3.3.3. Annotation procedure

The annotation of IMs and EMs was performed by two raters independently for all texts. Each item in the Lexical transcript tier was interpreted as a start of a new marker (with identification of its type), a word belonging to the previously annotated marker, or as a neutral word.

For empathy annotation the percentage agreement between raters on 5943 items was 93.6%, Cohen’s Kappa = 0.356. For interaction marker annotation the percentage agreement between raters on 5943 items was 80.9%, Cohen’s Kappa = 0.501. Mismatches in annotation were resolved by the third annotator.

### 4. Results

#### 4.1. Empathy markers

A marginally significant effect of the speaker group on the amount of EMs in narratives was found with ANOVA (one-way, between subject), $F(2, 20) = 3.28$, $p = 0.06$; as shown by posthoc analysis with Tukey HSD, healthy speakers use more EMs than people with dynamic aphasia ($p_{adj.} = 0.09$), see Figure 1 for summary and distribution of EMs across speakers and groups.

#### 4.2. Interaction markers

Given the non normal distribution of IM data, we have run four separate Kruskal-Wallis tests for each group of IM, with adjusted $p$ values using Holm method, as well as Dunn posthoc analysis. We have found no significant differences between the speakers’ groups for the use of pace-regulating and attention drawing markers. PWA(d) use significantly more cognitive difficulties markers ($p_{adj.} = 0.1$)

---

4 Interaction markers *DM* and to some extent *Apl* are introduced here to cover those cases of interaction that can not be easily ascribed to other more specific classes.
and significantly less attitude expression markers ($p < 0.05$) than healthy speakers. We have found no significant differences between the use of any type of IM by people with RHD and PWA(d) or healthy speakers. See Figure 2 for summary of the distribution of IMs across speakers and groups.

![Figure 1. Distribution of empathy markers in groups](image)

![Figure 2. Distribution of IMs across groups](image)
5. Discussion

Our results demonstrate that PWA(d) use significantly more markers of cognitive difficulties than healthy speakers and people with RHD. Obviously, this is a natural effect of the non-fluent speech output. Also, PWA(d) use significantly less attitude expression markers as compared with the healthy speakers. This corresponds with what is known about difficulties that PWA(d) experience in narration. Akhutina (1989/2012, p.45; cited in Akhutina 2015) points out that in dynamic aphasia, the story schema is disintegrated, so the patient is unable to produce and maintain the plan of the utterance. Complexity of planning depends on the task. As shown in Akhutina 1975/2002, p.50 (cited in Akhutina 2015), the hardest is narration on a given topic (3 points out of 3), while retellings may vary from 1 to 3. Too many cognitive resources of PWA(d) are spent on maintaining the story schema, keeping the main line of the story, balancing narration and description, so there is not enough left for expressing attitudes towards the described events.

Our data show that discourse in RHD does not show significant differences from that of healthy speakers in the amount of IMs, or EMs, used. However, they use less EMs than healthy speakers, though this difference does not reach significance. As stated in Traxler, (2012: 531) regarding the RHD and challenges in narration, “the right hemisphere appears to play a vital role in keeping track of the main idea or theme of a story as well as using the theme to organize the various subcomponents of the story into a coherent macrostructure”. More specifically, discussing the retelling task, Traxler (ibid) notes that RHD omit more important ideas than patients with LHD. We don’t know what specific disorder (aphasia type) is meant here by the LHD (left hemisphere damage), but in our experiment PWA(d) don’t demonstrate significant differences in empathy as compared with RHD discourse. Still, speaking about RHD, we can use the same reasoning as for the PWA(d) lack of expressing attitude markers and imagine that their difficulties with keeping themes and ideas which leads to omitting some of them, will result in certain, not intended, prioritizing of what is more important. The most important thing in the story is the main line of events. This may explain the RHD lower figures for empathy.

Our results and the results we got for another group of PWA (acoustic-mnestic aphasia type—see (Bergelson et al., 2016) show that there exist correlations between the prevalence of specific types of interactional elements and type of disorder. In this and other papers (Bergelson et al., 2015, Bergelson et al., 2016) we check these data against the interactional strategies used by the healthy speakers in the same experiment, which brings into comparison additional factors, such as situational context, personal goals, social distance and others. Altogether they fall into two parameters: the parameter of the task, which is stable and in this case amounts to retelling a film in the experimental setting, and the individual discourse strategies that healthy speakers use to cope with this specific task. We call these strategies compensation for a communication setting and task that fall out of the normal regular experience of healthy speakers. We all tell personal stories about the experiences we had, but retelling a film in an experimental setting is a somewhat awkward situation. Participants feel that some of their abilities (memory? narrative skills?) are being tested and respond to this correspondingly.
This project data provides additional evidence for the existence of at least four discourse strategies employed by healthy speakers in their retellings of The Pear Stories film in attempts to protect their social face.

- Strategy 1: Expressing Uncertainty, or *I-am-not-Committing-Myself*
- Strategy 2: Fine-grained approach, or *Staying-on-a-Safe-Side*
- Strategy 3: ‘As is’ story, or *No-Worries*
- Strategy 4: Discussing and evaluating, or *Moralist-Syndrome*

The four strategies of healthy speakers correlate with four groups of IMs, with each strategy having IMs of a certain group as its champion.

Strategy 1: Expressing Uncertainty, or *I-am-not-Committing-Myself* is manifested by abundant use of fillers (marker *Fill*), disclaimers, false starts, self-corrections, and laughter, non-standard reference and irony, some cases of epistemic markers—see Figure 3 for the use of false starts and *Fill* markers. Participants HP-v02 and HP-v10 use this strategy. While HP-v10 bases this strategy on pace regulating IMs, HP-v10 sticks to false starts and repairs.

![Figure 3. Distribution of markers of I-am-not-Committing-Myself strategy](image)

Strategy 2: Fine-grained approach, or *Staying-on-a-Safe-Side* reveals itself through lots of details in descriptions of the characters. Discourse marker *ны* is associated with this strategy. Narrators try to cover everything, which leads to abundant use of structural false starts (*SFs*) and repairs (*Rep*) unmotivated by false starts. Participants HP-v04 and HP-v12 represent this strategy as shown in Figure 4.demonstrating that markers of cognitive difficulties are inherent in the healthy discourse as well.
Strategy 3: ‘As is’ story, or No-Worries is demonstrated by minimal use of any IMs. This is best demonstrated by the participant HP-v09 and to a certain degree by HP-v05—see Figure 5.

Strategy 4: Discussing and evaluating, or Moralist-Syndrome is more difficult to capture with our data, as it involves using lots of evaluations, prolonged and detailed coda with evaluations of characters’ actions. Markers expressing attitudes
(Est, Jdg) are characteristic of this strategy. In Figure 6 we see the champion for this strategy—HP-v03.

![Figure 6. Distribution of markers of Moralist-Syndrome strategy](image)

We believe that more research is needed to be able to thoroughly describe various compensating strategies and the discourse means used by narrators for this purpose. We know that these and other compensating strategies are used by people with various types of aphasia and RHD. Yet another step will be to compare them across the groups.

**References**


