Experimental Approach to Reference in Discourse: Working Memory Capacity and Language Comprehension in Russian*

Olga Fedorova, Ekaterina Delikishkina, and Anna Uspenskaya

Department of Linguistics, Lomonosov Moscow State University,
Leninskie Gory, Moscow 119991, Russia
olga.fedorova@msu.ru

Abstract. In the experimental psycholinguistic study we consider the correlation between one of activation in working memory factors (Kibrik 2010), namely a rhetorical distance to the pronoun antecedent and a successful reference resolution in Russian. Using Russian adaptation of the working memory span test from Daneman and Carpenter 1980 we demonstrated that a rhetorical distance is really a significant factor affecting referent activation as well as that working memory capacity is correlated with some referential processes. However, we did not find an interaction between rhetorical distance and working memory group. Thus, our results supported the Separate Language Interpretation Resource theory of working memory structure (Caplan and Waters 1999).

Keywords: working memory, discourse comprehension, Rhetorical Structure Theory, pronoun resolution, Russian.

1 Working Memory Capacity and Language Comprehension

The most famous model of **Working Memory** (WM) developed by A. Baddeley and his colleagues (Baddeley and Hitch 1974; Baddeley 2001) is composed now of four components: the central executive and three short-term storage systems, namely the phonological loop, the visuo-spatial sketchpad and the episodic buffer. The phonological loop is divided into a phonological store, where the auditory memory traces decay in about two seconds, and an articulatory rehearsal component, that serves for subvocal rehearsal. The visuo-spatial sketchpad is assumed to integrate visual, spatial and possibly kinesthetic information. Episodic buffer is supposed to conjoin information from the phonological loop and visuo-spatial sketchpad within a representation of a whole episode, and to link this representation to long-term memory. The central executive is assumed as a limited capacity system, which controls all buffers and information processing within subsystems.

Shah and Miyake (1996) have hypothesized that the central executive of WM can be further divided into two sub-pools: (i) visuo-spatial component and (ii) verbal component. Moreover, in Caplan and Waters 1999 the authors have further distinguished between interpretive (on-line) and post-interpretive (off-line) components of the verbal pool. The former is described as automatic, first-pass language processing ("extraction of meaning from a linguistic signal", p. 79), the latter uses the propositional content of the sentence to accomplish tasks like reasoning, planning actions, and storing information in long-term memory.

Traditionally, there are two approaches used to investigate the question of WM resources for language processing – **Individual Differences** approach and **Dual Task** approach. In this paper we relate individual differences in verbal WM to the process of **language comprehension**. To

* The work reported in this paper was supported by a grant #08-04-00165a from the RGNF and by a grant #09-07-00366a from the RFFI.

study the individual differences in verbal WM capacity Daneman and Carpenter (1980) developed a "reading span" task that has become the standard psycholinguistic method of assessing verbal WM. In this task, participants are required to read 20 blocks of sentences, which vary in length from two to six sentences. A participant starts with the first block that contains only two sentences, reads them aloud and is required to repeat the last word of each sentence in the order they were presented. By the time the participant reaches the last block, s/he will have to report the last words of six sentences in the order in which they were presented. The score is based on the largest set-size successfully recalled, with values assigned for partially correct sets varying from 2 to 6 items.

Since the study by Caplan and Waters (1999) in the psycholinguistic literature there has been a discussion of the differences between two basic approaches to investigating the possible specialization of WM. On the one hand, participants can have a set of verbal processing resources that can be devoted to all verbal tasks (Daneman and Carpenter 1980; King and Just 1991). With respect to the individual-differences approach, the **Single Resource Theory** (SR) predicts that having a low WM capacity will reduce the resources available for sentence processing and make it less efficient. On the other hand, the part of the verbal WM system is specialized for interpretive aspects of sentence comprehension (Caplan and Waters 1999). The **Separate Sentence Interpretation Resource** (SSIR), or **Separate Language Interpretation Resource** (SLIR) theory predicts that performance on general verbal WM tasks will not predict language processing efficiency.

A number of experimental results have provided evidence that syntactic processing in language comprehension requires the allocation of WM resources. Some research has shown that sentences that have more complex syntactic structures are more difficult and time consuming to understand. For instance, King and Just (1991) reported self-paced word-by-word reading times for high and low span participants for these two sentence types: (1) *The boy that the girl pushed kissed the baby* and (2) *The girl pushed the boy that kissed the baby*. King and Just (1991) found that the reading time increases at points of syntactic complexity. They concluded that in the task assigning the structure of an object-relativized sentence such as (1) is more demanding than a participant-relativized sentence such as (2).

However, in order to argue for the SR theory it is necessary to find an interaction between sentence complexity (high, low) and WM group (high, medium and low span). King and Just (1991) provided evidence that there is a main effect of span, a main effect of extraction type, and a main effect of the interaction between syntactic complexity and span. However, Caplan and Waters (1999) noted that the required statistics were not reported. Caplan and Waters (1999) replicated the experiment but did not find the required interactions. Thus, they interpreted their results as supporting the SLIR theory.

In the following sections we concentrate on the question how not the syntactic but the **discourse complexity**, namely the rhetorical structure of discourse affects the pronoun interpretations of low and high capacity participants.

2 Reference in Discourse Analysis. Experimental Discourse Analysis

The experimental discourse analysis is a relatively new psycholinguistic field, but see van Berkum in press. In the beginning of the section (2.1) we consider the theoretical framework of reference in discourse; then (2.2) we briefly describe the hypotheses of our experimental study.

2.1 Reference in Discourse. Rhetorical Structure Theory

The process of mentioning referents in discourse is traditionally called **reference**, and linguistic elements that perform a mention of a referent are called **referring expressions.** The choice of referring expression for a referent, or **referential choice** is made by the speaker on the basis of a variety of factors that are called **activation factors**. In Kibrik 2010 is adopted a multi-factorial approach allowing to describe the integration of activation factors (that is called **activation score**, AS) in each moment of discourse stream. Each factor of activation in WM has some

value; in sum, AS can vary from 0 to 1. In the model of referential choice that we used in our study there are two main components – the first one is dealing with **the activation in working memory**; the second one, called **the referential conflict filter** allows to revise a reduced referring expression (a pronoun or a zero expression) if it creates the ambiguity effect for the addressee. Kibrik (2010) examines the reference effects from a speaker's perspective. Here we adopt the approach examining the process of reference comprehension, or **reference resolution** by the addressee.

Among many different factors of referent activation in WM we choose the rhetorical distance to the pronoun antecedent that is, according to Kibrik and Krasavina 2005, one of the most important. According to the Rhetorical Structure Theory (Mann and Thompson 1988), discourse is divided into elementary discourse units that are usually asymmetrically related by a rhetorical relation, usually a nucleus and a satellite, but sometimes two satellites. There are a number of rhetorical relations between discourse units; higher order units are further connected by the same rhetorical relations, etc. In Kibrik 1996 the author proposed the term **rhetorical distance** (RhD) and the ways to compute it.

2.2 The Experimental Goals

The conclusions of the discourse studies described in Kibrik 2010 in general are made on the basis of the observation and modeling. Here we examine some basic claims from this work by means **psycholinguistic experimental methods**.

First, we test the supposition that the rhetorical distance is a real factor affecting referent activation in WM. Second, we test whether addressees' ability to recover referents correlates with their WM reading spans. Third, we try to tease apart the SLIR and SR theories of verbal component of WM.

3 The Experiment

3.1 The Stimuli

Six experimental discourse fragments were composed, each containing about 100 words and two paragraphs. In the first paragraph (the first four sentences, see figure 1) three referents were introduced, one being **the target referent** (доктор, doktor, 'the doctor') and two other being false referents, or **distractors** (медбрат, medbrat, 'the nurse boy' and ассистент, assistent, 'the assistant'). In the second paragraph the first two sentences (the 5th and the 6th) contained no mentions of the referents, and in the 7th sentence a third person pronoun *он* 'he' is used. A participant read the discourse fragment and was thereafter asked to answer three questions, including one question addressing the identity of the target referent (mentioned by the third person pronoun *он* 'he') and two questions about facts. One complete discourse example with three questions is shown in Figure 1.

1. Был конец рабочего дня. 2. Пятая бригада скорой помощи ехала на базу после ложного вызова. 3. На носилках в кабине, набегавшись за смену, прикорнул медбрат. 4. Усталый доктор, слушавший музыку в плеере, игнорировал заискивающие взгляды молодого ассистента, горящего рабочим энтузиазмом после первого дня в бригаде.

5. В наушниках звучал «Белый альбом» битлов. 6. Безупречная мелодия качала и убаюкивала.
7a) **RhD=1** Он испытывал легкие угрызения совести за свою невнимательность к коллеге, но усталость превозмогала всё.
7b) **RhD=2** Он любил слушать эту пластинку после тяжелого трудового дня.
7c) **RhD=3** Он почувствовал, что медленно проваливается в сон.

Questions:
1. Какой номер был у бригады скорой?
2. Какая запись звучала в плеере?
3.a) **RhD=1** Кому было стыдно за невнимательность к коллеге?
b) **RhD=2** Кто любил слушать пластинку после тяжелого трудового дня?
c) **RhD=3** Кто почувствовал, что засыпает?

1. Byl konec rabočego dnja (It was the end of a working day). 2. Pjataja brigada skoroj pomošči jexala na bazu posle ložnogo vyzova (The Fifth Team of the Emergency Medical Response Unit was returning to its station after responding to a false alarm). 3. Na nosilkax v kabine, nabegavšis' za smenu, prikornul medbrat (Having been on the run for the entire day, a male nurse took a nap on an ambulance stretcher). 4. Ustalyj doktor, slušavšij muzyku v pleere, ignoriroval zaiskivajuščije vzgljady molodogo assistenta, gorjaščego rabočim entuziazmom posle pervogo dnja v brigade (A tired doctor listened to music on his player and ignored the inquisitive glances of a young assistant, still full of burning enthusiasm at the end of his first working day).

5. V naušnikah zvučal "Belyj al'bom" bitlov. 6. Bezuprečnaja melodija kačala i ubajukivala (An earphone delivered the Beatles, the White Album - a flawless melody that soothed and lulled the listener to sleep).
7a) **RhD=1** On ispytyval ljogkije ugryzenija sovesti za svoju nevnimatel'nost' k kollege, no ustalost' prevozmogala vsjo (He felt slightly guilty because of his inconsiderate attitude towards a colleague; however, fatigue had taken over).
7b) **RhD=2** On ljubil slušat' etu plastinku posle tjažjologo trudovogo dnja (He loved to listen to this music after an exhausting and stressful working day).
7c) **RhD=3** On počuvstvoval, čto medlenno provalivaetsja v son (He felt he was slowly sinking into a restful sleep).

Questions:
1. Kakoj nomer byl u brigady skoroj pomošči? (What is the team number of the Emergency Medical Response Unit?)
2. Kakaja zapis' zvučala v plere? (Which artist was playing the music that was delivered by earphone?)
3.a) **RhD=1** Komu bylo stydno za nevnimatel'nost' k kollege? (Who was feeling guilty for being inconsiderate to a colleague?)
b) **RhD=2** Kto ljubil slušat' plastinku posle trudnogo trudovogo dnja? (Who listened to soothing music at the end of a stressful working day?)
c) **RhD=3** Kto počuvstvoval, čto zasypaet? (Who was slowly sinking into a restful sleep?)

Figure 1: Discourse example (original text in Russian and transliteration with English translation).

So, each text had three variations with the rhetorical distance to the antecedent (=to the target referent доктор, doktor, 'the doctor') in 1, 2, and 3 discourse units. Rhetorical trees marked for the first paragraph are depicted in Figure 2:

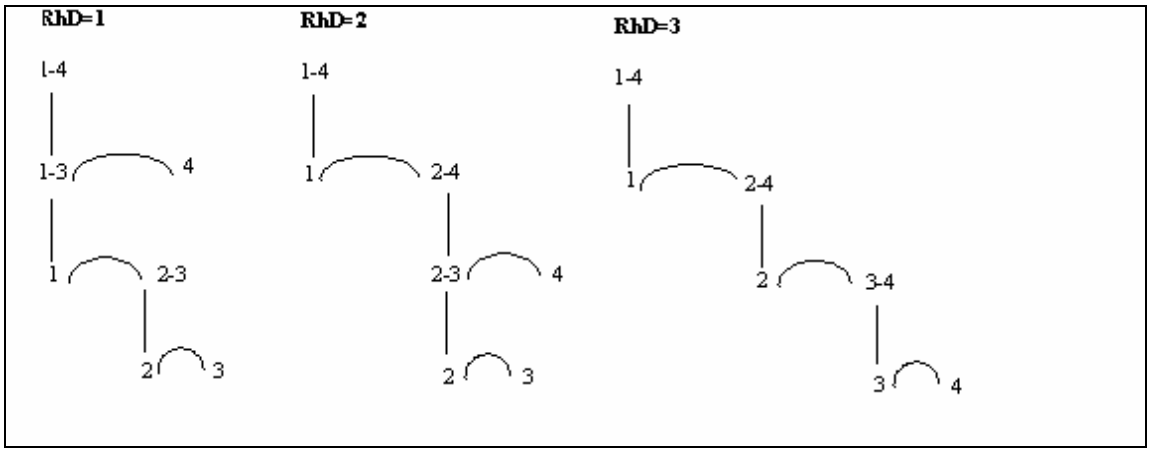

Figure 2: RST trees.

In Figure 3 we describe activation scores for three referents calculated according to Kibrik 2010; here abbreviations *LinD, animacy, introductory referent, synt role,* and *ParaD* are used for denoting some other activation factors.

AS computed without RhD:

The first distractor (medbrat, 'the nurse boy') = -0.3 (LinD=4) + 0.2 (animacy: Human) – 0.1 (introductory referent: Yes) – 0.3 (ParaD=1) = **-0.5**

The second distractor (assistent, 'the assistant') = -0.2 (LinD=3) + 0 (synt role: Other) + 0.2 (animacy: Human) – 0.3 (ParaD=1) = **-0.1** or **-0.3**

The target referent (doktor, 'the doctor') = -0.2 (LinD=3) + 0.4 (synt role: S of a main clause) + 0.2 (animacy: Human) – 0.1 (introductory referent: Yes) – 0.3 (ParaD=1) = **0**

Complete ASs for the target referent computed with RhD:

AS (RhD=1) = 0 + 0.7 = **0.7**
AS (RhD=2) = 0 + 0.5 = **0.5**
AS (RhD=3) = 0 + 0 = **0**

Figure 3: Activation scores for the referents with and without RhD (from Kibrik 2010).

3.2 Preliminary Experiments

In the first preliminary experiment we verified whether our texts have never really had a referential conflict. 24 participants read each modification of each text and answered the same questions. But in this experiment they were allowed to look the text during answering the questions. There were no errors in their answers. Thus, we concluded that there were no referential conflicts.

In the second preliminary experiment we test the naturalness (on the scale from 0 to 4) of our six texts. The results demonstrated an acceptable level of all six fragments, see table 1.

Table 1: The results of the naturalness test (96 participants).

text, №	RhD = 1	RhD = 2	RhD = 3	mean
1	2,8	2,4	2,6	2,6
2	2,5	2,4	2,5	2,4
3	3,2	3,2	2,9	3,1
4	2,3	2,5	2,3	2,4
5	2,5	2,6	2,6	2,5
6	2,9	2,5	2,9	2,8
mean	2,7	2,6	2,7	2,7

3.3 The Main Experiment

120 participants of our main experiment were given two tests: (1) a reading span test to measure the span of their WM, and (2) a reading comprehension test that asked two questions about facts and one question about pronominal reference.

In this study we used the reading span task for Russian, which we developed in 2001. For this, we adapted the Daneman and Carpenter (1980) reading span task. In our reading span task, participants read aloud sentences presented on a computer. They were instructed to press a key once they read a sentence, and a new one followed. After each block, a participant saw a blank computer screen during which s/he was required to recall the last word of each sentence in the block. All participants completed the reading portion of the experiment in less than 15 min. Based on their results we usually divided the participants into two groups: Low Span participants with reading span less than 4, and High Span participants with reading span equal to 4, 4.5 or 5. Sometimes, we regrouped our participants into three groups: High-Spans (≥ 4.5), Medium-Span (≥ 3.5) and Low-Span (≤ 3) participants. We have been collected the data from our experiments with Russian adaptation of Daneman and Carpenter (1980) test since 2001. Bringing together these data we counted that 64.5% out of more than 1000 participants had reading span < 4, i.e. that Low-Span participants constitute the majority. This percentage differs from the 'real' ratio of High-Span to Low-Span participants in analogous experiments on English material where only 50% of participants had reading span < 4.

We claim that the differences can be explained by the specifics of Russian material. Counting the number of words per sentence, we may see that Russian sentences were as long as their English analogs (12-15 words each). But counting the number of syllables we may see the difference: Russian sentences appear to be much longer since the portion of one or two syllable words in Russian is significantly less than in English. Thus, in terms of Baddeley's WM model the difference can be explained by the greater phonological loop load in Russian. A general finding that, like it has been shown for other languages, in Russian version of the test recall is better for the short last words than for the long ones supports this belief. The converging evidence is provided by the fact that in Russian the typical advantage from the auditory presentation of sentences compared with visual presentation that is also shown for other languages is greater than in English. Rich inflection, which is a characteristic of Russian, may contribute to the distribution of High and Low spans as well. This claim is supported by a substantially greater number of the word form errors made by Russian participants in comparison to English.

In the current experiment 86 from 120 participants were Low-Spans (≤ 3), 17 Medium-Spans (≥ 3.5), and 17 High-Spans (≥ 4.5).

As shown in table 2, the percentage of correct answers to the pronoun reference questions significantly declines with the increase of rhetorical distance and the corresponding decrease of AS ($cor = 0.9966159$, p-value < 0.05).

Table 2: Percentage of correct answers to the pronoun reference questions.

Participants, N	WM capacity	RhD = 1, AS = 0.7	RhD = 2, AS = 0.5	RhD = 3, AS = 0	Mean
86	Low: 2 to 3	62	38	28	43
17	Medium: 3.5 to 4	85	68	41	65
17	High: 4.5 to 5	100	94	68	73
120	Total	70	50	35	52

As shown in table 3, the percentage of correct answers to the fact questions did not declined with the increase of rhetorical distance and the corresponding decrease of AS. Thus, we can assume that there is **a separate module of WM dealing with the reference processes**.

Table 3: Percentage of correct answers to the fact questions.

Participants, N	WM capacity	RhD = 1, AS = 0.7	RhD = 2, AS = 0.5	RhD = 3, AS = 0	Mean
86	Low: 2 to 3	74	70	74	73
17	Medium: 3.5 to 4	78	74	75	76
17	High: 4.5 to 5	94	94	89	92
120	Total	78	74	76	76

As shown in both tables, we obtained the significant main effect of the WM both for the pronoun reference questions (cor = 0.5640427, p-value < 0.01), as for the fact questions (cor = 0.4241487, p-value < 0.01). However, we did not obtained the main effect of an interaction between rhetorical distance (RhD = 1, RhD = 2, RhD = 3) and WM group (high, medium and low span). Thus, we can interpret our results as **supporting the SLIR theory**.

Now if we suppose that the SLIR theory is really correct, we have doubts of the important idea from Daneman and Carpenter's study (1980) about good and poor readers because their test exploits the post-interpretive (off-line) component of the verbal pool. However, the process of reading is governed by the interpretive (on-line) component. The doubt is well agreed to the recent idea from the study of Otten and Van Berkum (2009); they concluded that "even low WM readers, who are typically believed to have less computational resources available can rapidly anticipate upcoming information in a fairly sophisticated way. In fact, at least with the materials tested here, there is no evidence in our data that they do so less, or less effectively, than their high WM counterparts." (p. 97).

4 Conclusion

In the study we demonstrated the significant correlation between a rhetorical distance to the pronoun antecedent and a successful reference resolution. We concluded that a rhetorical distance is really a significant factor affecting referent activation in WM.

We assumed also that there is a separate module of WM dealing with the referential processes as well as that there is the significant relatedness between WM and referential processes.

However, we did not find an interaction between rhetorical distance and WM group. Thus, our results supported the SLIR theory of WM structure.

References

Baddeley, Alan D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4(11), 417–23.

Baddeley, Alan D., and Graham J. Hitch (1974). Working memory. In: Gordon A. Bower (ed.), *Recent advances in learning and motivation.* New York: Academic Press, 47–90.

Caplan, David, and Gloria S. Waters (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22(1), 77–126.

Daneman, Meredyth, and Patricia A. Carpenter (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450–66.

Fox, Barbara A. (1987). *Discourse structure and anaphora in written and conversational English.* Cambridge: Cambridge University Press.

Givón, T. (1983). Topic continuity in spoken English. In: T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study.* Amsterdam: Benjamins, 345–63.

Kibrik, Andrej A. (1996). Anaphora in Russian narrative discourse: A cognitive calculative account. In: Barbara A. Fox (ed.), *Studies in anaphora.* Amsterdam: Benjamins, 255–304.

Kibrik, Andrej A., and Olga N. Krasavina (2005). A corpus study of referential choice: The role of rhetorical structure. In: Irina M. Kobozeva, Aleksandr S. Narin'jani, and Vladimir P. Selegej (eds.), *Dialogue'2005. Computational linguistics and intellectual technologies.* Nauka, 561–9.

Kibrik, Andrej A. (2010). *Reference in discourse.* Oxford: Oxford University Press.

King, Jonathan, and Marcel Adam Just (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30(5), 580–602.

Mann, William, and Sandra A. Thompson (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text* 8(3), 243–81.

Otten, Marte, and Jos J. A. van Berkum (2009). Does working memory capacity affect the ability to predict upcoming words in discourse. *Brain Research* 1291, 92–101.

Shah, Priti, and Akira Miyake (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General* 125(1), 4–27.

Tomlin, Russell S. (1987). Linguistic reflections of cognitive events. In: Russell S. Tomlin (ed.), *Coherence and grounding in discourse.* Amsterdam: Benjamins, 455–79.

van Berkum, Jos J. A. (In press). The electrophysiology of discourse and conversation. In: Michael Spivey, Marc F. Joanisse, and Ken McRae (eds.), *The Cambridge Handbook of Psycholinguistics.* Cambridge: Cambridge University Press