

## COGNITIVE INFERENCES FROM DISCOURSE OBSERVATIONS: REFERENCE AND WORKING MEMORY

ANDREJ A. KIBRIK

*(Institute of Linguistics, Russian Academy of Sciences)*

### **1. Preliminaries: Linguistics and cognitive science**

Linguistics is generally viewed as one of the main components of cognitive science. This presumably means that linguistics both is fed by and feeds cognitive science as a broader discipline. True, cognitive linguists have a fair record of taking the data provided by cognitive psychology into serious account; cf., for example, the impact of Rosch and other psychologists on the work of Lakoff (e.g. 1987), or the psychologically minded linguistic work, such as Tomlin 1994 or Dickinson & Givón 1997. On the other hand, the impact of linguistics on cognitive science has been quite modest. A typical textbook in cognitive psychology or cognitive science contains a chapter on language in which the only linguistic framework covered with any degree of detail is that of generative grammar. I believe that this situation is not tolerable and that cognitive linguists should try to explain their points to psychologists more effectively. After all, language is the main natural phenomenon that consistently and abundantly demonstrates overt traces of cognitive processing, and it cannot be ignored by the general enterprise of cognitive science.

In this paper I address some problems in the study of one important cognitive system, known as **working memory**. I will try to demonstrate that linguistic analysis can contribute to the resolution of these problems. This paper attempts to develop the tradition of cognitively oriented discourse analysis (Chafe 1994, Tomlin 1994, Givón 1995: Ch. 8, inter alia) and to establish links between linguistics and cognitive psychology.

In section 2 I introduce some concepts in the study of working memory. In the main body of this paper I look into a well-known linguistic phenomenon: **choice of referential expression** for a referent in discourse, and present my cognitive-linguistic analysis of this phenomenon. (For another recent cognitive-linguistics approach to reference see van Hoek 1997, Langacker 1996.) After that, I demonstrate that this analysis can contribute to a more general enterprise: the study of working memory in cognitive science. That is, from observing reference in natural discourse one can make certain inferences about the human ability of working memory<sup>1</sup>.

## 2. *Working memory*

Working memory (WM; otherwise called short-term memory or primary memory) is a small and quickly updated storage of information. The study of WM is one of the most active fields in modern cognitive psychology (for reviews see Baddeley 1986; Anderson 1990: Ch. 6; some recent approaches are represented in Gathercole (ed.) 1996). WM is also becoming an important issue in neuroscience, see Smith & Jonides 1997.

The range of classical questions about WM includes, inter alia, the following three:

- CAPACITY: how much information can there be in WM at one time
- CONTROL: what is the mechanism through which information enters WM
- FORGETTING: what is the mechanism through which information quits WM

A contribution to these classical questions I propose below is a side product of a linguistic study (described in Kibrik 1996 and in sections 3–4 below) that relied, in its turn, on cognitive work. At some point I discovered that the model I developed to explain and predict discourse phenomena has implications for more general cognitive issues.

Below I will briefly outline the study Kibrik 1996 on referential choice in Russian narrative discourse (section 3), report an analogous study of English reference (section 4), and then proceed with the three issues in working memory mentioned above (sections 5, 6, and 7). Conclusions are presented in section 8.

## 3. *Referential choice in discourse: A cognitive calculative approach*

A number of linguists have proposed that referential choice in discourse is dependent on the current memorial status<sup>2</sup> of the referent (Tomlin & Pu 1991, Chafe 1994, Givón 1995: 380ff.; cf. also Kibrik 1987; Ariel 1988; Gordon, Grosz & Gilliom 1993; Gundel, Hegberg & Zacharski 1993). This hypothesis amounts to the following:

- (1) If a referent is currently highly activated in the speaker's working memory, it is coded by a reduced NP (anaphoric pronoun or zero), and if the referent's activation in the WM is below a certain threshold, it is coded by a full NP<sup>3</sup>.

Hypothesis (1) presupposes that entities can pertain to WM to different degrees and, therefore the boundaries of WM are not clear-cut. For example, Chafe (1994) distinguishes three degrees of activation of referents: active, semiactive, and inactive. In the present study, activation is interpreted as gradual closeness to the center of WM. Activation is maximal when the entity

is in the center of WM, and minimal when the entity is totally out of WM; however, all intermediate degrees are possible, too.

Hypothesis (1) is compatible with the research of such psycholinguists as Gernsbacher 1990, Clifton & Ferreira 1987, Vonk, Hustinx & Simons 1992, inter alia. Below I take hypothesis (1) for granted, relying on the work of the mentioned authors, including experimental work. The hypothesis will be substantially specified below. In section 6 I present some additional evidence in favor of hypothesis (1).

In Kibrik 1996 I proposed a model of referential choice in Russian narrative prose, that is, the choice between a full NP and a reduced NP (in Russian, usually a third person pronoun *on*). My goals in that study were:

- to explain all occurrences of referential expressions in the selected discourse corpus
- to identify all factors influencing referential choice
- to design a calculative model of interaction between the relevant factors.

Below I will briefly outline that model; illustrations are postponed until the discussion of English data (section 4).

The model I proposed for Russian narrative discourse includes seven factors related to the properties of either the referent or the previous discourse, such as distance to the antecedent (three different measurements of distance are used), syntactic and semantic role of the antecedent, stable features of the referent (such as animacy).

The list of factors was identified empirically as necessary and sufficient to make predictions about referential choices (see below). All the factors contribute to referential choice, but not directly. They give rise to an integral characterization of a referent at the given moment in discourse. This integral characterization is called **activation score** (AS), that is, the status of the referent in the speaker's working memory. AS varies between 0 and 1. If AS=0, the referent is completely out of the speaker's WM; if AS=1, the referent is maximally activated in the WM. All intermediate grades of activation are possible, the minimal point on this scale accepted to be 0.1. When a particular referent has a high AS, it takes up a significant portion of the overall WM; this issue will be discussed in detail in section 5 below.

Each factor can be realized by two or more different features. For example, linear distance to the antecedent measured in discourse units ( $\approx$  clauses) can be 1, 2, 3, 4, or more than 4. To each of the features a numerical activation value corresponds, positive or negative, that can contribute to AS in particular cases. Since for each mention of a referent in each clause all factors and their features are easily and operationally identifiable, the corresponding numerical values are available too. They are

simply added to each other, thus giving rise to the AS of the particular referent at the particular point in discourse.

Numerical activation values had been specified through a cyclic heuristic procedure; they were being refined until the complete fit of the predictions made on the basis of the values (see below) to the observed data was reached.

An illustration of how activation factors are organized is found in Table 1. The structure of the most powerful activation factor (Rhetorical distance; see explanation in section 4 below), including its possible features, and the numerical activation values of each feature, are shown there.

Factor	Feature	Activation value
Rhetorical distance to the antecedent	1	0.7
	2	0.4
	3	0
	4+	-0.3

Table 1: *An illustration of an activation factor, its features and numerical activation values*

Below, in the course of the discussion of English data, I will present more comprehensive information on each of the activation factors, specifically in their application to English discourse.

Once each factor's feature is operationally identifiable at each point in discourse for each referent, one knows all the relevant numerical values. These values are summed up, and give rise to the current AS of the referent. The referential options available to the speaker in this situation are provided by so-called referential strategies, summarized in the chart in Figure 1. These strategies guide the choice between the full NP and the primary Russian reduced referential device: third person pronoun *on*. Figure 1 demonstrates the referential strategies as developed for Russian narrative discourse in Kibrik 1996.

AS:	<div style="display: flex; justify-content: space-around; width: 100%;"> <span>full NP only</span> </div>	<div style="display: flex; justify-content: space-around; width: 100%;"> <span>full NP most likely, <i>on</i> questionable</span> </div>	<div style="display: flex; justify-content: space-around; width: 100%;"> <span>either full NP or <i>on</i></span> </div>	<div style="display: flex; justify-content: space-around; width: 100%;"> <span><i>on</i> only</span> </div>
	0   0.1   0.2   0.3	0.4   0.5   0.6	0.7   0.8   0.9	1

Fig. 1: *Referential strategies in Russian narrative discourse: referential choice based on the referent's activation score*

That is, if AS is maximal, the third person pronoun only should be used. When AS is relatively high, both a pronoun and a full NP are appropriate. If

AS is in the middle of the scale, a pronoun is unlikely (probably depending on the idiolect), and if AS is low, pronouns are ruled out.

Figure 2 represents a flow chart of the process of referent mentioning in discourse. One component of this model not discussed above is “Filters”. The most important filter is referential conflict, or ambiguity, that will be discussed at some length in section 7 below; filters are not the main focus of this paper.

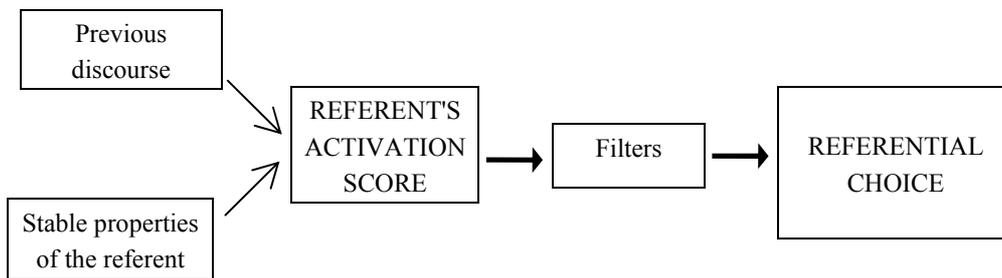


Fig. 2: *The process of referential choice*

#### 4. *An application to English data: the Margaret study*

##### 4.1 *General*

The model developed on the basis of Russian narrative prose was further tested on a corpus of English narrative discourse. This corpus was a children's story “The Maggie B.” by Irene Haas. The brief plot of the story is as follows.

A young girl called Margaret, or Maggie, daydreams of sailing her own ship. After she goes to sleep, she finds herself in a ship with her little brother James. There is a number of animals on board, and several trees. Margaret cleans the deck, cooks, feeds James, teaches him. Then a storm starts, and she fixes everything on the ship. After the dinner, Margaret plays fiddle to James, and the day is over.

There are 117 discourse units in the corpus. 76 different **referents** are mentioned in it, not counting 13 more mentioned in the quoted songs. There are 225 referent **mentions** in the discourse (not counting those in quoted text). This is a relatively small corpus, and I view it as a pilot study and am planning to test the present approach on a much more extensive corpus of English discourse.

There are 14 different referents mentioned in discourse that are **important** for this study. They are those mentioned at least once in a context

where any degree of activation can be possibly expected. Among the important referents, there are three protagonist referents: “Margaret” (72 mentions altogether, including 6 in collective references and 4 mentions by first person pronouns in quoted speech), “James” (28 mentions, including 6 in collective references), and “the ship” (12 mentions).

Any referent, including an important referent, can be mentioned in different ways, some of which (for example, first person pronouns in quoted speech) are irrelevant for this study. Those that are **relevant** for this study fall into two large formal classes: references by full NPs and references by activation-based pronouns. By “activation-based pronouns” I mean the unmarked, general type of pronoun occurrences that cannot be accounted for by means of any kind of syntactic rules, in particular, for the simple reason that they often appear in a different sentence than their antecedents. In order to explain and predict this kind of pronoun occurrences, one necessarily needs to construct a system of the type described in this paper, taking into account a variety of factors related to discourse context and referent properties. Typical examples of activation-based pronouns are given in (2) below.

- (2) 1607        Lightning split the sky  
       1608        as she ran into the cabin  
       1609        and slammed the door against the wet wind.  
       1610        Now everything was safe and secure.  
       1701        When she lit the lamps,  
       1702        the cabin was bright and warm.

There are two occurrences of the activation-based pronoun she in (2), and the second one is even used across the paragraph boundary from its antecedent.

There is a different type of pronoun occurrences that can be called syntactically based. Syntactically-based pronouns (and zeroes) are the third largest class among the mentions of important referents in the corpus. Examples of syntactically-based pronouns and anaphoric zeroes are given in (3), (4) below.

- (3) 0901        On her little stove, Margaret set a big pot of broth to bubble  
                   and boil
- (4) 1601        She took in the sail  
       1602        and ∅ tied it tight.

In principle, such occurrences of pronouns can be accounted for by the activation-based rules. The case is that syntactic anaphora is **grammaticalization** of activation-based anaphora, and factors operating in both cases are quite similar. However, pronoun occurrences like in (3), (4) can be simpler and probably more psychologically adequately treated as syntactically induced. They appear in such tight and stereotypical contexts with their antecedents that trying to explain them through the sophisticated apparatus of activation factors would be an overcomplication. The simple syntactic factor working in (3) is control of the antecedent being the subject of the clause. The simple syntactic factor in (4) is control of the antecedent over the subject zero (or the object pronoun) taking place when they occupy parallel syntactic positions in two conjoined clauses (so-called parallel function, or role inertia, described in the work of Caramazza & Gupta 1979).

There are 25 syntactic occurrences of overt pronouns in the corpus; they are not of primary importance in this study and are not included in the discussion below.

#### 4.2 *The database*

Thus the focus of this study is restricted to 39 full NP references and 40 activation-based pronominal references. As was found out in the study Kibrik 1996, within each of the referential types — full NPs and pronouns — there is a crucial difference: whether the referential form in question has **an alternative**. For example, there is a big difference between a pronoun that can be replaced by a full NP and a pronoun that is categorical, that is, allows no referential alternative. These two kinds of pronouns would correspond to different levels of activation. In (5) below an illustration of a pronoun usage is given that can vary with a full NP: in unit 1601 the full NP Margaret could well be used (especially provided that there is a paragraph boundary in front of unit 1601).

- (5) 1502 A storm was coming!  
 1503 Margaret must make the boat ready at once.  
 1601 She took in the sail  
 1602 and tied it tight.

To the contrary, there are occurrences of categorical pronouns. Consider an example which is a direct continuation of (5):

- (6) 1603 She dropped the anchor  
 1604 and stowed all the gear <...>

In 1603, it's impossible to use the full NP Margaret, only a pronoun is appropriate.

For the English data, it was found that referential forms of each type (for example, pronouns) fall into three categories: those allowing no alternative (= categorical), those allowing a questionable alternative, and those allowing a clear alternative. Thus there are six possible correspondences between the five potential types and two actual realizations, see Table 2.

Potential referential form	full NP only	full NP, ?pronoun	pronoun or full NP	pronoun, ?full NP	pronoun only
Actual referential form	\   / full NP		\   / pronoun		

Table 2. *Actual and potential referential forms*

The information about referential alternatives is crucial for establishing referential strategies. Of course, attribution of particular cases to one of the categories is not a straightforward matter. There were two sources of information on referential alternatives used in this study.

#### 4.3 *Judgments on referential alternatives*

First, a native speaker of English who was a linguist and had a full understanding of the problem and the research method was requested to supply her intuitive judgments on all thinkable referential alternatives in all relevant points of discourse. Each referential alternative was considered independently, under the assumption that the rest of the discourse is intact. Each referential alternative was subject to a four-way judgment: (i) appropriate (ii) slightly awkward (iii) questionable or significantly awkward (iv) clearly inappropriate. Those referential alternatives that were attributed to category (iv) — clearly inappropriate — were excluded from further consideration and the corresponding referential choices were considered to be “pronoun only” or “full NP only”.

Of course, it is not permissible to fully rely on the intuitions of one subject, so in order to objectivize the attribution of reference to particular referential types, an experiment of the following design was conducted. The idea of the experiment was to modify the original referential forms, present it to a subject and see whether the subject identifies the replacement as a linguistic (or “stylistic”) error. Of course, in order to keep the general well-formedness of the discourse one cannot make too many modifications at a time, because there is a threat of interference between modifications in the adjacent parts of the discourse. Seven modified discourses were made up from the original discourse. All relevant references were subject to modification in

this or that modified discourse. In each particular version of the discourse the adjacent changes never appeared closer than across a paragraph boundary, and usually had at least two paragraph boundaries between them.

Thus, there were 8 different variants of the discourse (one original and seven modified ones). They were presented to 12 students of the University of Oregon, native speakers of English, who did the job of assessing the felicity of the discourse 20 times altogether. Most of the subjects did the assessment job twice (with the time interval of two days), but some only once. Those who did the assessment twice were presented distinct variants of the discourse. No dependency of the assessment on the number of the trial (first vs. second) was discovered. Each of the 8 variants of the discourse was assessed 2.5 times on the average.

4.4 *Weighted judgments and referential strategies*

A special task was to bring all the linguist expert's and the student subjects' judgments together and build an integral judgment of each referential alternative. A system of weights was set up for this purpose. The linguist expert's judgments were attributed the following weights: (i) appropriate — 2; (ii) slightly awkward — 0; (iii) questionable or significantly awkward — -2.

The student subjects normally used only two options — they either did not notice the referential replacement or pinpointed it in the discourse and rejected it by returning to the original referential choice. The “default” acceptance of the referential alternative was attributed weight 1. The rejection of the alternative was attributed weight -2. The difference in the absolute values is due to the fact that pinpointing an “error” and rejecting it is a much more conscious and volitional act than default acceptance. I would be even inclined to attribute the weight of -3 to the rejection, except for it is not always clear to what degree the referential alternative is awkward.

All weights of the judgments (including the linguist expert and the student subjects) were summed together and averaged. The integral judgments on referential alternatives were obtained through the numerical scale shown in Figure 3.

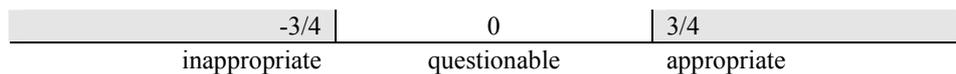


Fig. 3: *Averaged judgments of referential alternatives*

Referential alternatives with the value of  $-3/4$  or less are considered inappropriate. Alternatives falling in the range between  $-3/4$  and  $3/4$  are judged

questionable. And referential alternatives with the value of 3/4 or more are considered as appropriate as the actual referential choices in the original discourse.

Referential strategies I arrived at in this study are represented in Figure 4. Five categories of potential referential forms correspond to five different intervals on the activation scale. Specific activation factors and their numerical values giving rise to ASs at particular points of discourse are explained below in section 4.5.

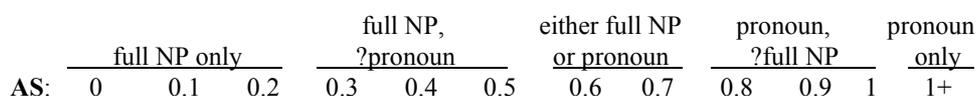


Fig. 4: *Referential strategies in English narrative discourse*

The way in which five categories of potential referential forms are represented in the corpus numerically, is represented in Table 3.

full NP only	full NP, ?pronoun	pronoun or full NP	pronoun, ?full NP	pronoun only
15	17	22 (including: 7 actual full NPs and 15 actual pronouns)	18	7
<b>Total:</b> actual full NPs — 39, actual pronouns — 40				

Table 3: *Frequencies of referential forms in the corpus*

#### 4.5 Activation factors

The system of activation factors that was developed for the Margaret discourse corpus is presented in Table 4 below.

Some comments regarding the activation factors and their structure and values are in order. I present a concise version of these comments here, and refer the reader to a more extensive discussion in Kibrik 1996. Any discussion of why certain factors are considered relevant for referential choice while others are not, is omitted here; see Kibrik 1996. Appendix to this paper contains a sample (the first three paragraphs) of the explored discourse.

The first three activation factors listed in Table 4 are related to the distance from the point in question to the antecedent. This distance can be measured in three different ways. Givón (1983), among others, proposed **linear distance** to the antecedent, measured in clauses, as an important determiner of referential choice. Fox (1987a) demonstrated that it is hierarchical rather than linear structure of discourse that is relevant for the

relationship between the point in question and the antecedent. The model of hierarchical discourse structure used in Fox 1987a and assumed in this study as well is Rhetorical Structure Theory of Mann and Thompson (see e.g. Mann, Matthiessen &

Factor	Feature	Value
Rhetorical distance (RhD)	1	0.7
	2	0.5
	3+	0
Linear distance (LinD)	1	0
	2	-0.1
	3	-0.2
	4+	-0.3
Paragraph distance (ParaD)	0	0
	1	-0.3
	2+	-0.5
Linear antecedent role	Linear distance is 4+	0
	Linear distance $\leq 3$ :	
	S of the main clause	0.4
	other active S	0.3
	DO, passive S, Pred suppressed NP	0.2
Other	-0.3	
Animacy	LinD $\leq 2$	0
	LinD $\geq 3$ :	
	inanimate	0
	animate non-human	0.1
human	0.2	
Protagonisthood	RhD+ParaD $\leq 2$	0
	RhD+ParaD $\geq 3$ :	
	-	0
+	0.2 (first in series) 0.1 (second in series)	
Super contiguity (contiguous words or same clause)	-	0
	+	0.2
Temporal / spatial shift	-	0
	+	-0.2
Weak referent	-	0
	+	-0.2
Predictability	-	0
	+	0.1

Antecedent is introductory	–	0
	+	-0.1

Table 4: *Activation factors and their values, as identified for English narrative discourse*

Thompson 1992). In Rhetorical Structure Theory, discourse is represented as a net of discourse units (roughly equaling clauses) connected by so-called rhetorical relations (such as sequence, cause, purpose, condition, concession, etc.). Each discourse unit is rhetorically connected to at least one other discourse unit, and via it, ultimately, to any other discourse unit. **Rhetorical, or hierarchical, distance** is measured as the amount of discourse units to the unit containing the antecedent; the “path” to such antecedent discourse unit is found in accordance with the rhetorical net. Discourse sample in the Appendix contains a simplified representation of the rhetorical structure demonstrating how discourse units can be hierarchically related to one another. Even though rhetorical distance is indeed a more powerful and explanatory parameter than linear distance, the latter has an important value, too. The third distance factor is **paragraph distance**; this factor was emphasized by Marslen-Wilson, Levy & Tyler 1982, Fox 1987b, Tomlin 1987, and others. Paragraph distance is measured as the number of paragraph boundaries between the point and question and the antecedent. Rhetorical distance is by far the most influential among the distance factors, and in fact among all activation factors: it can add up to 0.7 to the activation score of the referent. Linear and paragraph distance can be called penalty, factors, since they can only deduct something from AS if the distance is too high.

The next factor indicated in Table 4 is that of **grammatical role of the linear antecedent** (note that because of the different principles of identifying rhetorical distance and linear distance one referent mention can have two distinct antecedents: a rhetorical and a linear one). The logical structure of this factor is rather complex. First, it applies only when the linear distance is short enough: after about four discourse units it gets forgotten what the role of the antecedent was, only the fact of its presence may still be relevant. Second, this factor has a fairly diverse set of features. As has long been known from studies of syntactic anaphora, subject is the best candidate for the pronoun's antecedent. Different subtypes of subjects, though, have different weights, ranging from 0.4 to 0.2. Other relevant features of the factor include direct object, nominal part of the predicate, and “suppressed NP” — a non-mention of a referent that is however semantically implied in the discourse unit (though not being syntactically identifiable as, for instance, a zero subject in a coordinate structure); an example of such a suppressed mention of “the peaches” appears (or rather does not appear) in discourse unit 1707:

- (7) 1706 She sliced some peaches  
 1707 and put cinnamon and honey on top,  
 1708 and they went into the oven, too.

The antecedent role factor is the second most powerful after rhetorical distance and is an important source of activation.

The next couple of factors are related not to the previous discourse but to the relatively stable properties of the referent in question. **Animacy** specifies the permanent characterization of the referent on the scale of the “great chain of being”. **Protagonisthood** specifies whether the referent is the main character of the discourse (on some procedures of protagonist identification see Givón 1990: 907–908). Protagonisthood and animacy can be called rate-of-deactivation correction factors. They capture the observation that important discourse referents and human referents deactivate slower than those referents that are neither important nor human. In the formulation presented in Table 4, protagonisthood is connected with the rhetorical and paragraph distance: when these two together are high enough, a protagonist referent gains some extra activation; when they are not, protagonisthood does not matter; “series” is a group of clauses all containing mentions of a referent preceded by a group of at least three clauses containing no mentions of the referent. Animacy is connected here with the linear structure of discourse: under high linear distance human referents deactivate less than other referents.

The final group is second-order, or “exotic”, factors, including the following ones. **Supercontiguity** comes into play when the antecedent and the discourse point in question are in some way extraordinarily close. **Temporal or spatial shift** is similar to paragraph boundary but is a weaker episodic boundary; for example, occurrence of the clause-initial *then* frequently implies that the moments of time reported in two consecutive clauses are distinct, in some way separated from each other rather than flow from one to the other. **Weak referents** are those that are not likely to be maintained — such as “bed” in 0105 (see Appendix). **Predictability** is a relation of the current discourse unit to the preceding, such that it can be predicted that a certain referent must be mentioned at this point; this happens with the referent “Margaret” in discourse unit 1202:

- (8) 1201 After juice-and-cookie time, she gave James his counting  
 lesson,  
 1202 and this is how she did it.

Finally, **introductory antecedent** means that when a referent is first introduced into discourse it takes no less than two mentions to fully activate it.

#### 4.6 *Remarks on the system of activation factors*

Numerical activation values of each feature cited in Table 4 were obtained through a long heuristic “trial-and-error” procedure, performed in cycles until the whole array of data was explained. There is no space here to demonstrate in detail how the system of activation factors works and predicts/explains particular referential choices in accordance with the referential strategies (Figure 4). Also, some important components of the model are not mentioned in this paper for the sake of brevity. However, it should be stated that all referential facts contained in the original discourse, and obtained through experimentation with modified discourses, are indeed predicted/explained by the combination of activation factors with their numerical values, and the referential strategies.

It should be mentioned that the arithmetical approach employed allows AS to turn out somewhat higher than 1 in some cases. For example, the system of numerical values was set up in such a way that categorical pronouns received the AS of 1.1. This is interpreted as “extremely high activation” that gives the speaker no full NP option to mention the referent. The AS of 1 is then interpreted as “normal maximal” activation. Also, low AS frequently turns out to be negative. Such values are simply rounded to 0. It is definitely possible to arrange the mathematics of the present model in such way that the calculated AS never goes beyond the interval between 0 and 1; however, I prefer to use the rough and simplistic approach rather than complicate the model with sophisticated and hard-to-understand mathematics.

The system of activation factors developed for English is similar to that developed for Russian in its main traits, but has some important differences. Exploration of language typology from the viewpoint of this approach is a matter of future research.

Now we can address the questions about working memory that were posed at the beginning of this paper.

### 5. *Capacity*

The question of the capacity of WM is the following: how much information can there be in WM at one time? Of course, there are different kinds of information processed in WM at any given time. It is clear that among those kinds there is information about **specific referents** thought of or spoken of, and most likely it constitutes an important portion of WM.

Smith and Jonides (1997), relying on psychological and neurological experimentation, suggest that there are multiple working memories, devoted to different types of information: spatial, verbal, related to visual objects. Developing this line of reasoning, it seems plausible that there must be WM for specific referents, or at least a specialized section of WM devoted to specific referents. If so, the question of the maximal capacity of such WM section can be legitimately raised. In different situations, WM for specific referents is differently divided into parts for particular referents. Highly activated referents take up a large portion of the overall WM capacity, while referents of low AS take up a tiny fraction of the capacity.

The system of activation factors and their numerical values was developed in order to explain the observed and potential types of referent mentions in discourse. In the first place, only those referents that were actually mentioned in a given discourse unit by the author were considered. But this system was discovered to have one additional advantage: it operates independently of whether a particular referent is actually mentioned at the present point in discourse. That is, the system can identify any referent's activation at any point in discourse. For example, the AS of the referent "Margaret" can be identified for every discourse unit no matter whether the author chose to mention "Margaret" in that unit.

If so, one can find out activation of **all referents** at a given point in discourse. Consider discourse unit 0302 (see Appendix). Among the protagonist referents, only "Margaret" was selected by the author to be mentioned in that unit; its AS was low (0.3), so the author could only have used a full NP. For another protagonist referent, "James", that was not chosen by the author for discourse unit 0302, AS can also be easily calculated: it is 0.6. Likewise, for "the ship" it is 0.4. In addition, one more referent has a non-zero AS, too: "the rooster" has the AS of 0.9.

Summed together, ASs of all referents will produce **grand activation** — the summary activation of all referents at a given point in discourse. In 0302, grand activation is 2.2. Remember that the value of 1 on this scale is the maximal activation of a single referent.

Grand activation gives us an estimate of the capacity of the specific-referent portion of WM. Figures 5 and 6 depict the dynamics of activation processes in portions of Russian and English discourses. In each case, two major protagonists' activation is demonstrated, as well as grand activation.

Observation of the data in Figures 5 and 6 make it possible to arrive at several important generalization.

- (9) Grand activation varies normally within the range between 1 and 3, with the mean of about 2 or somewhat less, where 1 is the maximal

activation of a specific referent; we thus have an estimate of a very important portion of WM.<sup>4</sup>

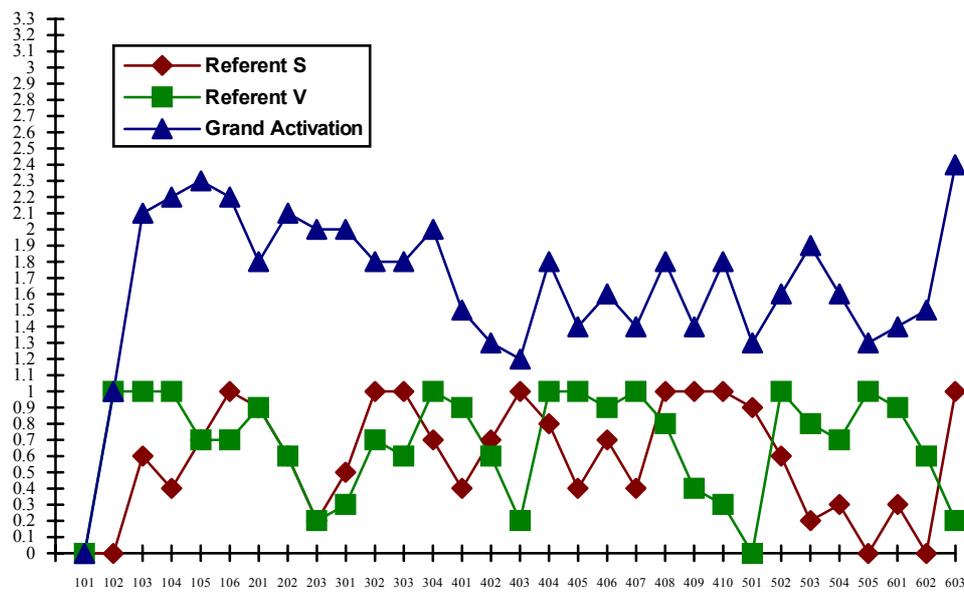


Fig. 5: Dynamics of the protagonist referents' activation and grand activation in the initial fragment of a Russian story

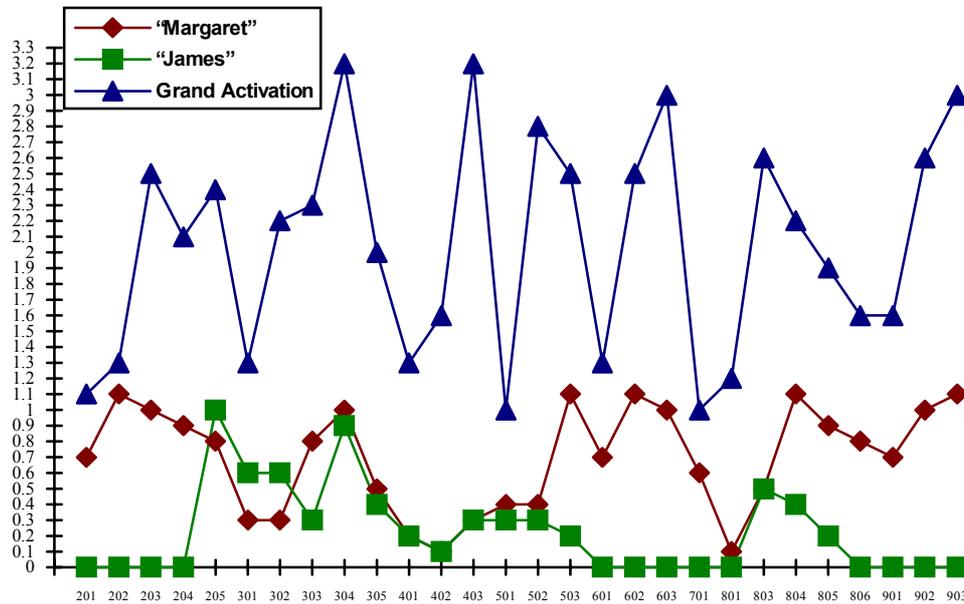


Fig. 6: Dynamics of the protagonist referents' activation and grand activation in a fragment of the English children's story

- (10) Grand activation varies much less than activation of individual referents which fluctuates between 0 and 1 all the time; the maximal grand activation value is only about 3 times higher than its minimal value.
- (11) Strongest shifts of grand activation are found at paragraph boundaries; even a visual examination of the graph in Figure 6, for instance, demonstrates that grand activation values at the beginnings of all paragraphs are local minimums; for the English excerpt the mean grand activation at the beginnings of the paragraphs is 1.2; apparently one of the cognitive functions of a paragraph is a threshold of activation update.<sup>5</sup>

There are some differences between the Russian and the English activation pattern. First, mean grand activation for the Russian extract is 1.7, while for English it is 2. Second, paragraph boundaries seem to have a more radical significance in English than in Russian. Both of these tendencies are observed not only in the illustrative excerpts of Figures 5 and 6 but throughout the discourse corpora employed in this study. It is not clear whether those differences can be attributed to a difference between languages, or discourse genres, or are significant at all. These questions call for further research.

## 6. *Control*

The question of control of WM is the question of how information comes into WM. Consider the following three statements.

(a) Current cognitive literature connects attention and WM. The mechanism controlling WM is what has long been known as attention. This view is expressed and motivated by Baddeley (1990), Cowan (1995) and, on the neurological basis, by Posner & Raichle (1994: 173). According to the latter authors, information flows from executive attention, based in the brain area known as anterior cingulate, into WM, based in the lateral frontal areas of the brain.

(b) The linguistic manifestation of attention is grammatical roles. As has been convincingly demonstrated in the experimental study of Tomlin (1994), focal attention in many languages, including English, is consistently coded by speakers as the subject of the clause.

(c) Subjecthood and reduced forms of reference are causally related: antecedent subjecthood is among the most powerful factors leading to the selection of a reduced form of reference. In both English and Russian, antecedent subjecthood can add up to 0.4 to the overall activation of a referent. In both English and Russian discourse corpora, 86% of pronouns allowing no referential alternative have subjects as their antecedent.

If considered together, these three sets of facts lead one to a remarkably coherent picture of an interplay between attention and WM, both at the linguistic and at the cognitive level:

- (9) Attention feeds WM, i.e. what is attended at moment  $t_n$  becomes activated in WM at moment  $t_{n+1}$ . Linguistic moments are discourse units. Focally attended referents are coded by subjects; at the next moment they become activated (even if they were not before) and are coded by reduced NPs.

The relationships between attention and WM, as well as between their linguistic manifestations, are represented in Table 5.

Moments of time (discourse units)	$t_n$	$t_{n+1}$
Cognitive phenomenon	focal attention	high activation
Linguistic reflection	mention in the subject position	pronominal reference
Examples	<u>Margaret</u> , <u>she</u>	<i>she</i> , <i>her</i>

Table 5: *Attention and working memory in cognition and in discourse*

## 7. *Forgetting*

How does information get forgotten from WM? There is a long debate in cognitive psychology between two competing hypotheses (for a review see, e.g., Baddeley 1986: 6–71). The first one, sometimes called “trace decay”, suggested that forgetting is a function of time. The second hypothesis, admittedly a more sophisticated one, proposed that information gets forgotten not simply because of the time factor but due to interference of or displacement by other incoming information.

The factor of time is captured in the model outlined above by means of the distance factors. As distance (in its different aspects) becomes greater, activation goes down<sup>6</sup>. That is, the model developed here apparently is in line with the trace decay hypothesis. What is linguistic evidence in favor of this view?

First, referents clearly deactivate, even in the absence of other incoming strongly competing referents. Consider an example of paragraph 3 of the English story. In discourse unit 0304, there are two pronouns and three highly activated referents: “Margaret”, “James”, and “the sun”, with ASs of 1, 0.9, and 1.1, respectively. Now compare that example with the already familiar discourse unit 0302 in which “Margaret” is highly unlikely to be mentioned by a pronoun because its AS is 0.3. Is non-pronominalizability of “Margaret” in 0302 due to interference of other referents? Candidate referents for such a role would be “James” (AS=0.6) and the newly introduced referent “the rooster” (AS=0.9). Keeping in mind what is permissible in 0304, it looks very implausible that referents with such low ASs could displace “Margaret” from WM and deprive it of high activation<sup>7</sup>. What does really deactivate “Margaret” since its previous occurrence in 0204 is distance — paragraph, rhetorical, and linear.

Second, a limitation on the number of concurrently activated referents does not necessarily require the concept of displacement or interference. It can be explained by the already stated limitation on the capacity of WM. Since grand activation rarely exceeds 3, three strongly activated referents, as in 0304, is about as much as there can be in discourse at one time. And this is due not to the displacement effect but to the balanced system of activation factors that activate and deactivate referents in accordance with the limits of the WM store.

The phenomenon of competition between referents is, however, real. Suppose that there are two highly activated referents at a certain point. Suppose the speaker needs to mention only one of them at that point, and uses a reduced form of reference. Since the addressee also knows that there are two highly activated referents, how would s/he recover the correct referent from the reduced form? This situation is called **referential conflict**, or, more traditionally, ambiguity. Every language possesses a repertoire of devices aiding to discriminate between such referents, for example, gender (as in

0304). A typology of such devices was outlined in Kibrik 1991. Referential conflict, if not eliminated by such supportive devices, can prevent the speaker from using a reduced referential form even in case of very high activation. The important point here is that referential conflict and all processes associated with it are **a separate component** of the referential system. Referential conflict is not an activation (or deactivation) factor, it is a filter coming into play **after** the activation factors computed the ASs of referents (see Figure 2 above).

If the discourse data support the trace decay hypothesis of forgetting, there seems to be a clear contradiction between them and the quite advance cognitive-psychological experimental studies proposing the other alternative: the interference/displacement hypothesis. An explanation to this contradiction is hinted by the study Hockey 1973. In that study a difference between the compulsory and the passive strategies of operating WM was emphasized. According to Hockey, under a passive strategy, when the pace of performance is chosen by the subject rather than by the experimenter, the pattern of forgetting approaches the prediction made by the trace decay hypothesis. The problem is that in many psychological experiments the cognitive system of a subject undergoes such pressure that never or rarely occurs in natural conditions. In other words, in experiments, it is not the attentional system of an individual himself but rather the will of the experimenter exploits WM and brings there too many referents at a time, and the effect of interference can indeed be observed. It is very likely that in natural conditions (under the “passive” strategy), on the other hand, the attentional system brings as many referents to WM as WM can normally accommodate and process.

## 8. *Conclusions*

The main conclusions about the functioning of WM we arrived at in this study include the following.

- capacity of WM for referents is severely limited (about 3 times maximal activation of a single referent)
- referents enter WM through the mechanism of attentional control
- referents can be forgotten from WM by the mechanism of decay.

If these conclusions are correct, that means that linguistic discourse analysis can indeed contribute to explorations of the human cognitive system.

The discourse excerpt below is broken into discourse units. Each discourse unit has a four-digit number. The first two digits designate the paragraph number, and the following two digits — the number of the discourse unit within the current paragraph.

- 0101 This is a story of a wish come true.
- 0102 Margaret Barnstable wished on a star one night —
- 0103 *“North Star, star of the sea, I wish for a ship named after me*
- 0104 *To sail for a day, alone and free, with someone nice for company.”*
- 0105 And then she went off to bed.
- 0201 When she woke up,
- 0202 she was in the cabin of her own ship.
- 0203 It was named *The Maggie B.* after her,
- 0204 and the nice company was her brother, James,
- 0205 who was a dear baby.
- 0301 A rooster crowed on deck,
- 0302 so Margaret knew the day was about to begin.
- 0303 She took James out to welcome the sun.
- 0304 It warmed them up
- 0305 and brightened the sky.

## NOTES

<sup>1</sup>The initial part of this project was conducted during my stay as a visiting scholar in University of Oregon, on the grant of the Fulbright Program (CIES). I am very grateful to Russ Tomlin for his support without which the experimental part of the project would not be fulfilled. The final stage of the study was supported by the grant #98-06-80442 of the Russian Basic Research Foundation. I would like to thank Gwen Frishkoff for her invaluable help in this project, especially in its part reported in section 4.3. I also appreciate the generous assistance of my English language consultants, especially Amy Crutchfield.

Various parts of this study have been reported at several academic seminars: at University of Oregon (January 1997), at Emory University, Atlanta (January 1997), at University of Hawai'i (February 1997), at Moscow Institute of Linguistics (October 1997), ICLC-97 (naturally), and at the conference Dialogue'98 in Tarusa (October 1998). I thank all colleagues who attended those presentations and supplied important comments, especially Russ Tomlin, Michael Tomasello, Leo Noordman, Wietske Vonk, Anatolij Baranov, Dmitriy Dobrovol'skij, and Ol'ga Fedorova. I am grateful to Leo Noordman and Karen van Hoek for their useful comments on a written draft of this paper. Also I would like to thank Michael Posner and Michael Anderson for their consultation on some psychological issues. Of course, none of the people mentioned above are in any way responsible for the ideas laid out here.

<sup>2</sup>Not all of the authors listed below actually talk in terms of memory; for example, Chafe prefers the concept of “consciousness”. However, in my understanding these authors have in mind essentially the same kind of cognitive phenomena.

<sup>3</sup>As has been noted by a number of authors (e.g., Chafe 1994), it may be more important how activated the referent is in the addressee's WM, according to the speaker's current assessment. But for the sake of simplification I talk about activation in the speaker per se.

<sup>4</sup>It is important to emphasize that grand activation does not depend on the amount of protagonists in the discourse. Even when there are more than two protagonists (e.g., four), at a given point in discourse not all of them act, and grand activation does not grow twice as high as in case of two protagonists.

<sup>5</sup>Of course, a drop in grand activation at paragraph boundaries is predetermined by the fact that paragraph boundary is a strong activation decreasing factor: each referent is deactivated after a paragraph boundary, and, therefore, the sum of particular ASs necessarily goes down. However, grand activation drop is not a mere artifact of the present approach. The deactivational effect of a paragraph boundary is an immanent fact that needs to be accounted for by any theory of reference in discourse. The observation of grand activation drop is a direct consequence of that immanent fact.

<sup>6</sup>As pointed out in section 4.5 above, rate of deactivation can be different for different referents: protagonists and humans deactivate slower than other referents. However, deactivation always happens with time, and even for protagonists and humans distance factors are the most powerful.

<sup>7</sup>One could argue that in such cases displacement might still take place, just the displacing information is not referents but, perhaps, other activated information – states or events being spoken of. However, as discussed in section 5 above, it is likely that working memory for specific referents is a relatively separate module of the cognitive system with its own capacity limitations.

## REFERENCES

- Anderson, John R. 1990. *Cognitive psychology and its implications*. 3d ed. New York: W. H. Freeman & Company.
- Ariel, Mira. 1988. “Referring and accessibility”. *Journal of Linguistics* 24. 65–87.
- Baddeley, Alan. 1986. *Working memory*. Oxford: Clarendon Press.
- Baddeley, Alan. 1990. *Human memory: Theory and practice*. Needham Heights, Mass: Allyn & Bacon.
- Caramazza, A. & Sh. Gupta. 1979. “The roles of topicalization, parallel function, and verb semantics in the interpretation of pronouns”. *Linguistics* 17. 497–518.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

- Clifton, Charles, Jr. & Fernanda Ferreira. 1987. "Discourse structure and anaphora: Some experimental results". *Attention and performance XII*, edited by M. Coltheart, 645–654. Hove: Erlbaum.
- Cowan, Nelson. 1995. *Attention and memory: An integrated framework*. New York – Oxford: Oxford University Press.
- Dickinson, Connie & T. Givón. 1997. "Memory and conversation: Toward an experimental paradigm". *Conversation: Cognitive, communicative, and social perspectives*. Ed. by T. Givón, 91–132. Amsterdam & Philadelphia: John Benjamins.
- Fox, Barbara. 1987a. *Discourse structure and anaphora*. Cambridge: Cambridge University Press.
- Fox, Barbara. 1987b. "Anaphora in popular written English narratives". *Coherence and grounding in discourse*, ed. by R. Tomlin, 157–174. Amsterdam & Philadelphia: John Benjamins.
- Gathercole, Susan E., ed. 1996. *Models of short-term memory*. Hove, East Sussex: Psychology Press.
- Gernsbacher, Morton Ann. 1990. *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Givón, T. (ed.) 1983. *Topic continuity in discourse. A quantified cross-language study*. Amsterdam & Philadelphia: John Benjamins.
- Givón, T. 1990. *Syntax: A functional-typological introduction*. Vol. 2. Amsterdam & Philadelphia: John Benjamins.
- Givón, T. 1995. *Functionalism and grammar*. Amsterdam & Philadelphia: John Benjamins.
- Gordon, Peter C., Barbara J. Grosz & Laura A. Gilliom. 1993. "Pronouns, names, and the centering of attention in discourse". *Cognitive Science* 17. 311–347.
- Gundel, Jeanette K., Nancy Hegberg & Ron Zacharski. 1993. "Cognitive status and the form of referring expressions in discourse". *Language* 69. 274–307.
- Hockey, G.R.J. 1973. "Rate of presentation in running memory and direct manipulation of input processing strategies". *Quarterly Journal of Experimental Psychology* 25. 104–111.
- Kibrik, Andrej A. 1987. "Fokusirovanie vnimanija i mestoimenno-anaforičeskaja nominacija" (Focusing of attention and pronominal anaphora). *Voprosy jazykoznanija* 1987. 3. 79–90.
- Kibrik, Andrej A. 1991. "Maintenance of reference in sentence and discourse". *Language typology*, ed. by Winfred P. Lehmann & Helen-Jo J. Hewitt, 57–84. Amsterdam & Philadelphia: John Benjamins.

- Kibrik, Andrej A. 1996. "Anaphora in Russian narrative prose: A cognitive account". *Studies in anaphora*, ed. by B. Fox, 255–304. Amsterdam & Philadelphia: John Benjamins.
- Lakoff, George. 1987. *Women, fire, and dangerous things. What categories reveal about mind*. Chicago: University of Chicago Press.
- Langacker, Ronald. 1996. "Conceptual grouping and pronominal anaphora". *Studies in anaphora*, ed. by B. Fox, 333–378. Amsterdam & Philadelphia: John Benjamins.
- Mann, William, Christian Matthiessen & Sandra Thompson. 1992. "Rhetorical structure theory and text analysis". *Discourse description. Diverse linguistic analyses of a fund-raising text*, ed. by W. Mann & S. Thompson, 39-78. Amsterdam & Philadelphia: John Benjamins.
- Marslen-Wilson, William, Elena Levy & Lorraine K.Tyler. 1982. "Producing interpretable discourse: The establishment and maintenance of reference". *Speech, place, and action. Studies in deixis and related topics*, ed. by Robert J.Jarvella and Wolfgang Klein, 339-378. Chichester: Wiley.
- Posner, Michael I. & Marcus E. Raichle. 1994. *Images of mind*. New York: Scientific American Library.
- Smith, Edward E. & John Jonides. 1997. Working memory: A view from neuroimaging. *Cognitive psychology* 33. 5–42.
- Tomlin, Russell. 1987. "Linguistic reflections of cognitive events". *Coherence and grounding in discourse*, ed. by R.Tomlin, 455-79. Amsterdam & Philadelphia: John Benjamins.
- Tomlin, Russell. 1994. "Focal attention, voice and word order: An experimental cross-linguistic study". *Word order in discourse*, ed. by Pamela Downing and Michael Noonan, 517–554. Amsterdam & Philadelphia: John Benjamins.
- Tomlin, Russell & Ming-Ming Pu. 1991. "The management of reference in Mandarin discourse". *Cognitive linguistics* 2. 65–93.
- van Hoek, Karen. 1997. *Anaphora and conceptual structure*. Chicago: University of Chicago Press.
- Vonk, Wietske, Letticia G.M.M.Hustinx & Wim H.G.Simons. 1992. "The use of referential expressions in structuring discourse". *Language and Cognitive Processes* 7. 301–333.