

STUDIES IN ANAPHORA

Edited by

BARBARA FOX
University of Colorado

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

Anaphora in Russian Narrative Prose: A Cognitive Calculative Account

Andrej A. Kibrik

Institute of Linguistics, Russian Academy of Sciences

In this paper I am going to propose a model of referential device selection in a sample of Russian written narrative prose. I will argue that the most important factor in a referent's pronominalizability is the cognitive notion of activation. In particular, I will propose an arithmetical model which calculates referents' activation at any given point and thus accounts for every instance of referential device selection in the sample discourse.

This paper is structured as follows. After the Introduction (1), I present a theoretical cognitively-oriented model of anaphora in section 2. In section 3 relevant facts about the discourse sample employed and about the Russian language will be reviewed. Section 4 contains a discussion of activation factors and their numerical values. All calculations explaining the referential choices in the sample discourse are also presented in section 4. In sections 5 and 6 two additional components of the model of anaphora are briefly discussed. Concluding remarks are found in section 7. Appendix 1 contains the sample discourse with an English translation. Appendix 2 is an illustration of the proposed calculative methodology.

1. Introduction

For the last fifteen years or so, anaphora has been the subject of a number of very interesting and productive accounts. One group of these accounts is very rich in taking into consideration a great number of different discourse factors influencing the choice of referential devices (e.g. Clancy 1980; Givón 1983;

1990; van Dijk and Kintsch 1983); but some of these studies represent what could be called a “statistical majority approach”, assuming that accounting for 90% of occurrences is fairly satisfactory. Another group of studies is remarkable in that they explain every occurrence of a referential device in the selected corpus of data. However, they concentrate mostly on one factor that influences referential choice — e.g. discourse episodic structure (Marslen-Wilson, Levy and Tyler 1982; Fox 1987b; Tomlin 1987) or rhetorical structure (Fox 1987a: Ch. 5). From another point of view, many studies of anaphora are, in my opinion, too text-oriented, as if relying on a belief that language users, when producing and understanding discourse, look back in the pre-text and trace antecedents there.

In this paper I will try to combine the following three principles:

- a multifactorial (rather than monofactorial) concept of anaphora, that is, an assumption that there are **many factors** influencing an on-line selection of a referential device by the speaker;
- an attempt to cover and explain — consistently, not in an ad-hoc manner — **every single occurrence** of anaphora in some (limited) sample of discourse rather than picking examples from different discourses, authors, and genres;
- a consistently cognitively-based approach based on the assumption that the verbal form of discourse is directly influenced by the **cognitive structures** in the speaker at the moment of speech (rather than by pre-text, context, etc.).

2. Cognitive Model

The model I am going to adopt in this paper is as follows. The main prerequisite for the speaker's using an anaphoric pronoun is **activation** of the referent in question. That is, the speaker needs to have the referent in his/her active memory by the beginning of the current discourse unit. Moreover, s/he needs to believe that the referent is in the active memory of the addressee.

A number of authors, including the present writer (Kibrik 1984, 1988), have suggested that the cognitive determiner of anaphoric pronouns is attention focus. However, now I am convinced by some linguistic (Chafe 1994; Tomlin 1994), psychological (Posner and Snyder 1975; Cowan 1988), and neurophysiological (Glezer 1993) work that attention and activation are

separate cognitive phenomena, both have certain linguistic reflections, and anaphora is more directly related to activation. Chafe (1994) investigates cognitive and linguistic processes connected with activation in much detail and demonstrates that the usage of an attenuated form of reference is related to the activation of corresponding referents. Tomlin (1994) has demonstrated that the process of attention focusing is responsible for such linguistic processes as grammatical subject and word order selection.

From a cognitive point of view, activation of an item in active memory normally is a result of focusing attention on this item at a previous moment in time. This cognitive relationship between attention and activation is iconically represented in language by the well-known fact that the antecedents of anaphoric pronouns frequently are grammatical subjects. What is attended at moment t_n and is therefore marked by subject, becomes activated at moment t_{n+1} and is therefore mentioned with an anaphoric pronoun.

A referent can enter the speaker's active memory from several alternative sources, including previous discourse, observed environment, and recall from long-term memory. (Activation from all of these sources can also involve peripheral activation by inference.) Since the addressee normally has no direct access to the speaker's long-term memory, this latter channel is of a very limited relevance for anaphora. Rather, this channel is the basis for introductory reference. Furthermore, since we will be dealing with written prose, activation through perception of the physical environment shared by the speaker and the addressee is also limited to very few items (such as reference to the author). Therefore, the only essential way for referents to enter the active memory of the speaker and addressee is through previous discourse.

But how are we to detect which referents are active and which are not? How can this intangible notion of activation help us to explain referential choices?

A number of approaches to anaphora have tried to find some textual, objective correlates of pronominalizability, be it distance to the closest antecedent, rhetorical dominance of the antecedent-containing clause over the current clause, or a paragraph boundary. However, it is unlikely that a speaker searches through previous text keeping its form in memory. I am going to suggest that the previous text plays a major role in determining the referents' pronominalizability but does that through the mediation of the cognitive activation structures rather than directly. At every moment of discourse production, factors of activation work in the speaker's mind (of course

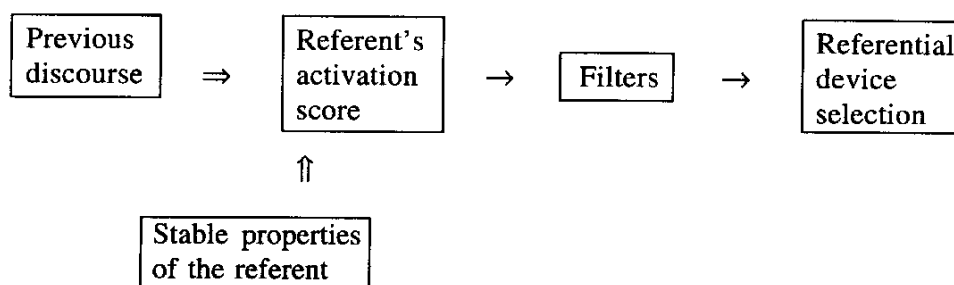
independently of his/her consciousness) and determine an activation level for each referent. If the speaker needs to mention a referent, and the activation level for this referent is high enough, an anaphoric pronoun can be used. If not, then a full NP has to be used.

Therefore, previous discourse influences the state of the active memory, and active memory in its turn determines the pronominalizability of referents. Below I will propose an extensive set of activation factors determining a referent's activation at any given point in discourse. These factors are formulated as properties of previous discourse open to public objective verification. I will propose an arithmetical model in which each factor is assigned certain numerical values. In every particular case these values, when added altogether, give rise to the current referent's **activation score** — the main predictor of referential device selection. This multifactorial arithmetical model of activation is able to explain all occurrences of referential devices in the selected sample of discourse.

I have to point out here that once the referent is in active memory due to previous discourse there is one more thing that can influence its current activation level: Stable properties of the referent, such as animacy and protagonisthood (see below).

Thus, pronominal anaphora is appropriate if a referent is highly activated. High activation is a necessary but not a sufficient condition. There are also a number of filters which block certain potential pronominal mentions. These filters will be discussed in some detail below.

Now consider the diagram in Figure 1 demonstrating how the whole process of referent mentioning works.



Legend:

⇒ arrows designating the operation of activation factors, taking place prior to the production of the current mention

→ arrows designating on-line transition from one stage to another during production

Figure 1. The process of referent mentioning

3. Discourse Sample: Preliminary Data

3.1. General

In this study I investigate a single sample of narrative prose — a short story by Russian writer Boris Zhitkov entitled “Nad vodoj” (“Over the water”). This particular sample of discourse was selected for this study for the following reasons.

(1) Narrative was selected since it is one of the basic discourse types (though not the most basic one), and I assume that any linguistic phenomenon should first be explored by the examples of its basic, prototypical manifestation, while other manifestations can be described on the basis of the prototype.

(2) Written prose was selected because it is a well-controlled mode in the sense that previous discourse is almost the only source for the referents that are anaphorically referred to, and therefore we can more easily control the ways in which referents appear and become activated.

(3) Boris Zhitkov was selected as an excellent stylist, with very simple and clear language, well-motivated lexical choices, and at the same time with a neutral, non-exotic way of writing.

(4) This specific story was selected since it is a prototypical narrative describing primarily basic events — physical events, interactions of people, people’s reflections, sentiments, and speech. The story is written in the third person, so there are no references to the author.

The discourse sample is presented in Appendix 1. The text is given there with a free English translation and, where necessary, a literal translation. In the Russian text all relevant mentions of the referents are underlined and supplied with referential indices where necessary. The text is divided into **discourse units** that are kept as close to clauses as possible. In some cases, however, a clause is broken by another clause, or contains several relative clauses, and then discourse units can happen to be smaller than clauses. All discourse units are numbered, and below I will refer to discourse units by simply mentioning their four-digit numbers. The first two digits represent the number of a paragraph, and the last two digits the number of a unit within the current paragraph.

The discourse sample comprises about 300 discourse units. There are about 500 mentions of various referents in the sample, and there are some 70

Table 1. The list of characters and their referential indices

pilot	p	aircraft	
mechanic	m	engine	h
Fedorchuk	F	carburetor	c
fat passenger	f	nut	n
lanky passenger	l	instruments	
woman	w	altimeter	
serviceman	s	door	
young passenger	y	handle	
elderly passenger	e	wing	
all passengers	a	book	
a subset of		fog	
passengers	a1, a2, etc.	snow	
all aboard	A	clouds	
		sea	

different referents appearing in the discourse. However, only a minority of them occurs more than once. There are 25 referents appearing at least once in an anaphoric context, that is, in a situation where at least a certain degree of activation can be expected. The list of these referents is given in Table 1. Mnemonic referential indices are indicated there for all human referents and some inanimate referents.

3.2. Referential devices in Russian

The major distinction in Russian referential devices is between full NPs and semantically reduced NPs. The former can appear with the demonstrative *ÉTOT* 'this'; such occurrences in principle are quite interesting cases of anaphora but they are too rare in the discourse sample so I will not discuss them separately from full NPs in general.

The reduced NPs fall into non-zero pronouns and zeroes.

3.2.1. Non-zero pronouns

The major device for anaphoric mention in Russian is the third person pronoun *ON* 'he/she/it' (masc. *on*, fem. *ona*, neut. *ono*, Pl. *oni*). It is the most neutral means, having also a deictic usage. In written discourses the two most common types of usage are what can be called activation-based (or discourse) usage and syntactic usage. Syntactic occurrences of *ON* can be predicted

solely from a syntactic context of a pronoun, in terms of superficial control of the antecedent NP, and usually are obligatory, that is they do not allow replacement by another referential device. Syntactic occurrences of ON are a minority of all ON occurrences. Activation-based occurrences cannot be explained by syntactic structure and frequently allow variation with other referential devices.

Here is a partial paradigm of ON:

	masculine	feminine	Plural
Nominative	<i>on</i>	<i>ona</i>	<i>oni</i>
Accusative	<i>ego</i>	<i>ee</i>	<i>ix</i>
Dative	<i>(n)emu</i>	<i>(n)ej</i>	<i>(n)im</i>
Instrumental	<i>(n)im</i>	<i>(n)ej</i>	<i>(n)imi</i>
Locative	<i>(n)em</i>	<i>(n)ej</i>	<i>(n)ix</i>
Genitive	<i>(n)ego</i>	<i>(n)ee</i>	<i>(n)ix</i>

Another important pronoun is the demonstrative *TOT* 'that' (masc. *tot*, fem. *ta*, Pl. *te*). *TOT* has many various usages and is basically an adjectival word but as an anaphoric pronoun it is used as a substantive. *TOT* is more frequently employed syntactically, but can also be used as an activation-based pronoun.

The main relative pronoun is *KOTORYJ*.

The reflexive pronoun is *SEBJA* (accusative), the possessive reflexive is *SVOJ*. The reciprocal pronoun is *DRUG DRUGA*. None of these will be an object of special attention in this paper.

First and second person pronouns will not be of interest for us either.

3.2.2. Referential zeroes

Zeroes are a convention used to indicate those formally unfilled argument or adjunct positions in the clauses that correspond to certain referents. I have identified the following range of zeroes in the text under investigation (they are close to covering the whole variety of zeroes found in Russian; however, inherently non-specific zeroes like the indefinite-personal zero, see #1605, and the natural force zero, see #1108, are not included here). Graphically, types of zeroes are distinguished below with the help of superscript indices (not to be confused with the subscript referential indices).

Zeroes with a syntactically identifiable position

Subject zeroes

- Ø¹ - zero subject of an independent non-coordinate clause (#2310)
- Ø² - zero subject resulting from conjunction reduction (#2302)
- Ø¹² - zero subject of a coordinate clause not resulting from obligatory conjunction reduction (#1109)
- Ø³ - zero subject of a subordinate clause (#0402)
- Ø³ - zero subject of a cosubordinate clause (#0112)
- Ø⁴ - zero subject with an infinitive (#1408)
- Ø³⁴ - zero subject of a subordinate infinitival clause (#0109)
- Ø⁵ - zero subject in a converbal (converb = adverbial participle, see Haspelmath and Koenig 1995) clause (#1204)

Non-subject zeroes

- Ø^d - zero Actor of the “state predicates”, normally requiring the dative case (#2901)
- Ø^o - zero object (#2902)
- Ø^p - zero possessor (#0308); it is marked and counted below only in the clauses where no other mention of the referent is found

Other zeroes

- Øⁿ - zero nominal with a non-zero adjectival modifier (#0105)
- Øⁱ - zero subject of an imperative clause (#2502)
- @ - zero mention of the speaker within the implicit introduction of quoted speech/thought (#2603)

Below I am not going to discuss purely syntactic zeroes (Ø², Ø⁴, Ø³⁴, Ø⁵). Irrelevant for this paper are the imperative subject Øⁱ and the zero Øⁿ occurring mostly with quantifiers. The same applies to the implicit introduction of quoted speech/thought, marked @, and the possessive zero Ø^p. Among the remaining zeroes, only Ø¹, Ø^d, and Ø^o will be briefly discussed; other zeroes are too rare in the sample for making any generalizations.

3.3. Referential device distribution in the sample discourse

In Table 2 below a classification of relevant referential devices is presented, along with their frequencies in the corpus. The figures designating frequencies (in the right hand part of the table) are aligned in accordance with the hierarchical classification of the referential devices (in the left hand part of the table).

Table 2. Frequencies of referential devices in the corpus

Reduced NPs	59			
<i>ON</i> -pronouns		33		
syntactic <i>ON</i> -pronouns			6	
de-deictic <i>ON</i> -pronouns			2	
activation-based <i>ON</i> -pronouns			25	*
without a full NP alternative				12
with a full NP alternative				13
<i>TOT</i> -pronouns		4		
\emptyset^1		4		
\emptyset^{12}		2		
\emptyset^3		2		
\emptyset^{23}		1		
\emptyset^d		7		
\emptyset^o		3		
\emptyset^p		3		
Full NPs	94			
with a reduced NP alternative		23		
with an <i>ON</i> -pronominal alternative			8	*
with a questionable <i>ON</i> -pronominal alternative			8	*
with a <i>TOT</i> -alternative			3	
with a questionable <i>TOT</i> -pronominal alternative			2	
with another alternative			2	
without a reduced NP alternative		71		*

* An asterisk marks those types of devices that will be of particular relevance to the discussion of activation below.

As is common in studies of referential device selection, the major division considered below is that between a full NP and a reduced NP. As the central type of reduced NP I will consider below the *ON*-pronouns, both because they are the most unmarked anaphoric device in Russian, and because they are most amply represented in the corpus.

A difference between a referential device allowing an alternative and the same device not allowing one is very important, as will become clear later. Roughly speaking, these two devices can be expected to have different activation levels. When making judgments as to which occurrences of referential devices allow an alternative and which do not, I relied initially on my own intuition as a native speaker. To check my intuition I ran a series of experiments aimed at collecting other speaker's opinions on what is an

appropriate device and what is not. A detailed report of these experiments is beyond the scope of this paper, but basically they were conducted as follows.

I offered the discourse sample to a subject and asked him/her to read it through. The next day I offered the subject a modified text where certain original mentions of referents were replaced by alternative ones. This replacement was done both when it was consistent with my intuition and when it was not. I asked the subject to indicate points in the discourse that seemed improperly worded and that called for rewording. In the course of these experiments I found out that my intuitions mostly coincide with those of other native speakers of Russian. In all cases of divergence my intuitions were more conservative than those of the subjects: some modifications which I expected to be on the verge of acceptability were in fact recognized by the subjects as quite appropriate. From this I infer that my intuitive estimates, with a certain amount of caution, can be taken as working hypotheses for referential device alternation.

4. Activation

In this section I identify those parameters of referents and discourse context that significantly correlate with referential device selection (4.1). Then I present the final list of activation factors with assigned numerical values (4.2) and make the calculations of activation scores for full NPs and *ON*-pronouns (4.3). Finally, I discuss other referential devices found in the corpus (4.4).

4.1. Correlations between candidate activation factors and referential choices

Several textual factors have been suggested in the literature as directly determining the choice of referential device. Best known is the suggestion by Givón (1983, 1990) that linear distance from an anaphor to the antecedent, measured in clauses, is at least one of the major predictors of referential choice.

Fox (1987a: Ch.5) has convincingly demonstrated that it is the rhetorical, hierarchical structure of discourse rather than plain linear structure that affects selection of referential devices. (The notions of rhetorical structure and rhetorical distance will be explained below in section 4.2.1.) Although

rhetorical distance measurement is indeed a much more powerful tool for modelling reference than linear (“referential”) distance, linear distance also plays a modest role.

In a number of works it has been suggested that a crucial factor of referential choice is episodic structure, especially in narratives. Marslen-Wilson, Levy and Tyler (1982), Tomlin (1987), and Fox (1987b) all demonstrate, using very different methodologies, that an episode/paragraph boundary is a border after which speakers tend to use full NPs even if the referent was recently mentioned.

One more factor is emphasized in Grimes (1978) — centrality of a referent in discourse, which I call protagonisthood below.

Several other factors have been suggested in the literature, including animacy, syntactic and semantic roles played by the NP/referent and by the antecedent, and syntactic distance to the antecedent measured in full sentences. For a recent study bringing together a rich set of factors determining the referential choice see Payne (1993).

In this section I am going to check which of the listed parameters correlate significantly with the choice of *ON*-pronoun vs. full NP. In Table 3 below I calculate the correlation between a certain feature of a parameter and the choice of referential device. These calculations will serve as preliminary data for singling out those parameters that really prove to be factors of referential device selection.

The quantitative results below were obtained through the following methodology. To analyze the discourse sample I used the Data Management Program SHOEBOS (developed by the Summer Institute of Linguistics). Several databases were compiled, according to the types of referential devices (full NPs, *ON*-pronouns, etc.). A separate record of the appropriate database corresponded to each relevant mention of a referent. A database record included as many parameters that potentially could influence referential choice as possible. Then I formulated hypotheses regarding possible correlations between a certain discourse/referent feature and referential device selection, and separated out the records that conformed vs. did not conform to this hypothesis (through a special SHOEBOS filtering procedure).

The counts below use referential device as a starting point, that is, e.g. they show how frequently *ON* has an animate referent. A count from the side of the features of referents/discourse, e.g. how frequently animate referents get pronominalized, would give other, also quite interesting results.

Table 3. Referent/discourse features' correlation with referential device selection

Parameters and their features	Percentage of correlation		
	ON	Full NP	
animacy			*
human	78	48	
inanimate	22	52	
protagonisthood			*
yes	66	46	
no	34	54	
syntactic role of the NP			
subject	53	56	
non-subject	47	44	
semantic role of the referent			
Actor	54	66	
non-Actor	46	34	
linear distance			*
1	78	10	
2	13	21	
3	3	13	
4	-	9	
5	-	10	
>	-	37	
no antecedent	6		
rhetorical distance			*
1	91	21	
2	3	18	
3	-	11	
4	-	11	
5	-	11	
>	-	28	
no antecedent	6		
syntactic distance			*
0	44	6	
1	50	27	
2	-	21	
3	-	17	
>	-	29	
no antecedent	6		
paragraph distance			*
0	91	44	
1	3	32	
2	-	10	
3	-	4	
>	-	10	
no antecedent	6		

antecedent type			
full NP	53	62	
relative pronoun	3	2	
<i>ON</i>	10	13	
<i>TOT</i>	-	-	
Zero	[34]	[20]	
\emptyset^1	-	1	
\emptyset^2	19	11	
\emptyset^3	3	-	
\emptyset^4	6	-	
\emptyset^d	3	2	
\emptyset^o	-	1	
\emptyset^p	-	2	
\emptyset^n	3	3	
Other	-	3	
syntactic role of the antecedent			*
subject	78	50	
non-subject	22	50	
semantic role of the antecedent			*
Actor	81	63	
non-Actor	19	37	

A correlation between a feature of a parameter and the usage of *ON* is considered significant if it is observed in at least two thirds of the cases (66%). The following parameters (marked with an asterisk in Table 3) and their features significantly correlate with *ON*-pronominalization: animacy (feature "human"), protagonist hood ("yes"), referential distance ("1"), rhetorical distance ("1"), paragraph distance ("0"), syntactic ("subject") and semantic ("Actor") roles of the antecedent. (The term "Actor" is an abstract semantic role; it designates the semantically central participant of a clause, with more-than-one-place verbs it is usually the agent or experiencer; see e.g. Van Valin 1993: 43ff.) One can also argue that low syntactic distance (0 to 1) also correlates with *ON*-pronominalization. There is no significant correlation between the syntactic/semantic role of the NP/referent in the current clause, and its pronominalizability. Surprisingly, also, in the discourse sample explored, recent pronominalization of a referent does not correlate with its current pronominalizability.

The decision to list the above-mentioned parameters as potential factors of activation is corroborated by the fact that these parameters display very

different patterns vis-a-vis full NPs: the percentages of such features as “subject vs. non-subject” and “Actor vs. non-Actor” do not significantly vary with full NPs, compared to the sharp contrast in the *ON* column; the distribution of animacy and protagonisthood is the mirror-image of that found in *ON*-pronouns; the percentages of different features of all distance parameters are scattered across the scale, unlike the one-polar pattern found in *ON*-pronouns.

An important question is whether we are to include syntactic distance as a significant parameter. I propose that an effect resulting from long syntactic distances is already accounted for by the referential distance and rhetorical distance measures.

Thus, we will further explore only those parameters that appear to be real factors in referential device selection. There is a whole set of such factors. Rather than trying to reduce them all to just one, I am going to suggest that many independent factors at a time affect the activation level of a referent.

4.2. Activation factors and their numerical values

In this section I assign certain numerical values to the features of the factors affecting activation. I use a very simple arithmetic model assuming that activation can vary from 0 to 1. Each value of activation factors is measured in tenths of 1. All values are summed; I propose that the resulting activation score (abbreviation: AS) motivates the selection of referential device.

Table 4 below includes seven activation factors, with their numerical values, listed partly in the order of their importance but mostly in an order of convenience. This set of seven factors proves to be necessary and sufficient for the explanation of referents' activation in the corpus. Note that the set of factors appearing in the table below includes one additional factor compared to what we arrived at in the previous section. This factor (“sloppy identity”) will be explained below. Formulation of some factors are more complex than they appeared before, and this again will be motivated below.

Table 4. Activation factors and their numerical values

Factor	Feature	Activation value
Rhetorical distance to the antecedent	1	0.7
	2	0.4
	3	0
	>	-0.3
Syntactic and semantic role of the antecedent The closest linear antecedent is:	S and A	
	Rhetorical antecedent is S and A	0.4
	Rhetorical antecedent is not S and A	0.3
	either S or A	0.2
	non-S and non-A	0
	Protagonisthood	yes
	first mention in a series	0.3
	second mention in a series	0.1
	more than second mention in a series	0
	no	0
Animacy	Human	
	Rhetorical distance is:	
	3 or more	0.2
	2	0.1
	1	0
	Inanimate	0
Linear distance	1	0
	2	-0.1
	3	-0.2
	4	-0.3
	>	-0.5
Paragraph distance	0	0
	1	-0.2
	>	-0.4
Sloppy identity	no	0
	yes	-0.2

4.2.1. Explanation of activation factors

Rhetorical distance to the antecedent. This factor determines distance measured in discourse units from the current unit back to the rhetorically closest one containing an antecedent. Rhetorical distance is measured on the basis of a rhetorical structure constructed for discourse in accordance with the Rhetorical Structure Theory, as developed by Mann, Matthiessen, and Thompson (1992). Rhetorical Structure Theory claims that every discourse unit is connected to at least one other unit by a certain “rhetorical relation”, such as sequence, cause, result, concession, etc. As a discourse unit is produced, the most important question is to which other unit it adds new content, and which unit it linearly follows is less important. Fox (1987a) has demonstrated that referential device selection is particularly sensitive to the rhetorical structure: most pronominalizable are those referents that were mentioned in the rhetorically preceding unit. In this study Rhetorical Structure Theory is applied with minor adaptation. I give an example of how rhetorical structure is constructed in the first paragraph of the sample discourse (see Appendix 1). In this example a simplified rhetorical structure is shown; specific rhetorical relations are not indicated, but it is made clear for each discourse unit, to which other unit it is directly rhetorically related.

A four-fold distinction appears to be relevant in the factor of rhetorical distance: rhetorical distance can be 1, 2, 3, and above three. As rhetorical antecedents I counted all kinds of mentions, even the most implicit ones, as e.g. unexpressed author of quoted speech/thought (indicated by “@” in the text). For instance, for the referent “mechanic” the rhetorical distance is 1 in both #2308 and #2310, even though in the latter case the antecedent is found not in the linearly preceding unit. This is because #2310 is directly rhetorically related to #2308, not to #2309. In other cases rhetorical distance can be more than 1 while there is an antecedent in the linearly preceding clause. I distinguish between the linear antecedent (linearly closest) and the rhetorical antecedent (closest according to the rhetorical structure). The two antecedents do not have to always coincide. If they do, I merely say “antecedent”.

Syntactic and semantic role of the antecedent. This factor captures the fact that those referents that were last mentioned as subjects and/or Actors in their clauses are more pronominalizable. Most commonly the properties of subject and Actor coincide, but the combinations subject/non-Actor (e.g. in passive clauses) and Actor/ non-subject (e.g. with state predicates requiring the dative

case in Russian) are also quite usual. I found it useful to distinguish between three situations: when the antecedent is both subject and Actor, when it is either of these two, and when it is neither. In the first case, there is also a subtle distinction between a situation when the rhetorical antecedent is also subject/Actor, and a case when it is not. Thus in some cases both the linear and the rhetorical antecedents separately affect the activation level.

For example, for the referent “Fedorchuk” the antecedent is subject/Actor in #0403, non-subject/Actor in #1213a, and non-subject/non-Actor in #2505.

Protagonisthood is a factor which sometimes affects activation score but not always. It is an important factor in the situation of referent reactivation (cf. Grimes 1978; Givón 1990: 907-908). Protagonist referents are easier to reactivate than referents that are peripheral to the narrative. When a referent is already active anyway, there is not much difference whether it is a protagonist or a peripheral referent. To capture this observation, I have applied the following technique. Protagonisthood counts only in the cases of reactivation: at the beginning of what I call a *series*. A series is a sequence of at least three consecutive units, such that: (1) All of these units mention the referent in question; (2) this sequence is preceded by a gap of at least three consecutive units not mentioning the referent in question. At the beginning of a series, that is, in the situation of reactivation, protagonisthood helps a referent to regain activation. In this particular sample of discourse I treated as protagonists the referents “pilot”, “mechanic”, “Fedorchuk”, and “passengers” (as a set). For a discussion of how to measure a referent’s centrality see Givón (1990: 907-909).

For example, protagonisthood is at play for the referent “Fedorchuk” in #2503, and is irrelevant in #1214.

Animacy is represented in the present corpus by two features: human and inanimate. Humanness, an inherent property of a referent, can increase a referent’s activation, but not always. Much like protagonisthood, the influence of animacy is dependent on rhetorical distance. With longer distances, humanness demonstrably helps to keep activation higher, and with shorter distances inanimate referents gain and keep activation to the same degree as human referents. I found out that with rhetorical distance of 3 or more the influence of animacy is relatively high, with 2 it is slight, and with 1 it is none.

For example, humanness contributes to activation of the referent “me-

chanic” in #1404; but in the case of rhetorical adjacency, there is no difference between a human (#1408) and an inanimate (#2309) referent.

Linear distance might seem to be an unnecessary parameter, given that we assume rhetorical distance as the most powerful factor. However, one can see that a short rhetorical distance with a short linear distance is not the same as a short rhetorical distance with a long linear distance. Even with rhetorical distance of 1, long linear distance can considerably decrease the activation score. I found it sufficient to distinguish between five possible referential distances: 1, 2, 3, 4, and above 4. While rhetorical antecedents include all kinds of antecedents, even the most implicit ones, from linear antecedents I exclude the syntactically unidentifiable ones, particularly, unexpressed authors of quoted speech (as in #3002).

An example of the influence of linear distance can be seen in #3504: the rhetorical distance for “Fedorchuk” is 1 (rhetorical antecedent in #3501), and the linear distance is 2 (linear antecedent in #3502).

Paragraph distance is a factor reflecting the importance of episodic structure in discourse. Normally, within a paragraph activation is preserved relatively well, while a paragraph’s boundary is reflected cognitively as an activation distribution update. I distinguish between the zero paragraph distance (antecedent within the current paragraph), 1, and above 1.

A clear example of the influence of paragraph boundary is found in #2801 (“Fedorchuk”). If there were no paragraph boundary here, a full NP would be hardly appropriate.

“**Sloppy identity**” is the last factor in this list. It was not mentioned before because of its limited relevance. “Sloppy identity” is used in some works as a term implying incomplete referential identity of an anaphor with its antecedent.

One can identify several types of sloppy identity. For example, in #1106-1107 the antecedent is a specific referent, and the anaphor is generic; the pronoun in #0603 does not have a literally coreferential antecedent; in #1303 and #1305 the two referents are clearly different.

Why is the set of factors suggested here exactly the necessary and sufficient set properly accounting for all evidence? There are at least two possible ways to justify the selection of factors. One is to isolate every single factor, manipulate it independently of all other factors, and demonstrate that it is exactly the changes in this factor that are responsible for activation score changes and resulting referential device selection. Another way is to empiri-

cally assign certain numerical values to all features of all factors, and to set up a system accounting for all the observed data.

Though both approaches to justification are feasible, here we concentrate on the second approach. Below I provide the activation counts for the whole set of data, and demonstrate that they really do explain all observed evidence.

4.2.2. Explanation of the factors' numerical values

The procedure for calculating the activation scores can be described as a method of prizes and penalties. Points are added when a feature increasing activation appears, and they are subtracted, when a decreasing feature occurs. Some factors, as one can see from Table 4, are increasing-only — e.g. animacy, protagonist-hood, and antecedent role: Some features of these factors add something to the activation score, and the other simply have no influence. Another kind of factor, namely rhetorical distance, works both ways: Short distances increase activation, long distances decrease it. Finally, the factors of linear distance, paragraph distance, and sloppy identity are decreasing-only: They have an unmarked feature that does not affect activation, while other features strongly decrease activation.

The specific numerical values for each feature were found empirically, in several successive adjusting trials to account for the occurrences of the *ON*-pronouns. When this was finally achieved, it turned out that all other occurrences of referential devices are properly explained by this set of numerical values without any further adjustment. I interpret this fact as evidence supporting the correctness of the developed system.

4.3. Calculations of activation scores for major referential devices

The whole set of referent mentions that are of interest for us here includes about 130 items. They fall into several classes, as described in Table 2 above. For each referent mention I have calculated the current activation score by simply adding the numerical values of activation factors, both positive and negative. For example, the calculation of activation score for the pronoun *nego*₁ in #0905 goes as follows:

$$0.7 \text{ (Rhet. distance = 1)} + 0.4 \text{ (Antec. role: S and A)} = 1.1$$

Numerical values for all other factors are simply zero. A more complicated example is presented by the full NP *Fedorčuk*_F in #1403:

-0.3 (Rhet. distance = 4) + 0.4 (Antec. role: S and A) + 0.3 (protagonist-hood: yes, first mention in series) + 0.2 (Animacy: human, Rhet. distance > 3) - 0.3 (Lin. distance = 4) - 0.2 (para. distance = 1) = **0.1**

For one more extensive example of activation score calculations in a sequence of 15 discourse units see Appendix 2.

Following the outlined methodology, an activation score was calculated for each referent mention. The system of numerical values is constructed so that actual activation scores can sometimes be less than 0 or more than 1. I keep these numbers in all examples below to make the calculation process more transparent. However, the cognitive interpretation of the activation score presupposes that it can vary between 0 and 1. Therefore, all negative scores should be read as rounded to 0, and all scores above 1 as rounded to 1.

The **basic hypothesis** on how activation score determines the choice between *ON*-pronouns and full NPs can be formulated as follows:

- (1) referents with an activation score of 0.3 or less are coded by full NPs and cannot be mentioned by *ON*-pronouns
- (2) referents with an activation score of 0.4 through 0.6 are most likely to be coded by full NPs, but to a certain degree/for certain native speakers a mention by an *ON*-pronoun seems also appropriate
- (3) referents with an activation score of 0.7 through 0.9 can be equally successfully mentioned by both *ON*-pronouns and full NPs.
- (4) referents with an activation score of 1.0 (or more) are coded with *ON*-pronouns and cannot be mentioned by full NPs.

Strategies 1 and 4 are absolute, while strategy 3 and partly strategy 2 leave a degree of freedom for the speaker to use either referential device. To use an analogy from syntactic work, in case (3) both options can be considered well-formed, in cases (1) and (4) only one can be considered well-formed, the other being starred, and in case (2) one form is appropriate and the other would have a question mark.

All four strategies that constitute the basic hypothesis, and the corresponding types of *ON* and full NP occurrences are summarized in Figure 2. Other referential devices in my corpus are not numerous enough to make definite generalizations, but I will consider some aspects of them somewhat later.

	(1) full NP only				(2) full NP most likely, ON questionable			(3) either full NP or ON			(4) ON only
AS:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Figure 2. Major referential device selection, as determined by the referent's activation score

Below I present five tables containing activation score counts for all four types of *ON* and full NP occurrences. Two tables (6 and 7) naturally correspond to type (3): Table 6 contains counts for the registered occurrences of *ON*, and Table 7 for the registered occurrences of full NPs. Each table is supplied with a prediction of the potential activation score range which was made according to the basic hypothesis.

Table 5. Activation scores for *ON*-pronouns allowing no full NP referential alternative (prediction: 1.0 or higher)

Unit #	AS	Comment	Unit #	AS	Comment
0110	1.1		1603	1.2	
0403	1.1		1904	1.1	
0905	1.1		2203	1.1	
1104	1.1		2308	1.1	
1214	1.0		2503	1.1	
1216	1.2		3502	1.2	

Obviously, the counts in Table 5 demonstrate 100% agreement with the prediction made by the basic hypothesis.

Table 6. Activation scores for *ON*-pronouns allowing a full NP referential alternative (prediction: between 0.7 and 0.9)

Unit #	AS	Comment	Unit #	AS	Comment
0102	0.7		2304	1.1	*
0104	1.0	*	2309	0.7	
0306	0.9		2320	0.7	
0603	0.7		2504	1.1	*
1107	0.9		2602	0.7	
1404	0.8		2803	0.7	
2204	0.9				

In the set of data represented in Table 6, there are three cases (marked with an asterisk) when the calculated activation score is higher than what is predicted by the basic hypothesis. However, the only thing these data contradict is my intuition on whether a full NP would be appropriate in these contexts. I have certain suggestions as to why this contradiction might have arisen. (In all cases refer to Appendix 1 for relevant contexts.)

As for #0104, the linear antecedent, relative pronoun *kotorye_a* found in #0103, might be cognitively interpreted not as a separate mention, and the whole unit #0103, a relative clause, might be treated as a part of the preceding unit. Then the score would be 0.8, thus conforming to the hypothesis' prediction. I think that this kind of flexibility of a system calculating activation scores is an asset rather than a shortcoming.

As for #2304 and #2504, I think that the cognitive effect taking place here is what might be called "early activation score update". In both cases another referent appears in the position of subject/Actor. Supposedly this fact should influence the activation scores of all referents no earlier than in the next unit. But given the left-to-right orientation of reading, especially slow analytical reading, it might impose other referents' activation lowering even within the same unit. This must be the reason why speakers (in this case myself) admit the plausibility of a full NP alternative here.

Table 7. Activation scores for full NPs allowing an ON-pronominal referential alternative (prediction: between 0.7 and 0.9)

Unit #	AS	Comment	Unit #	AS	Comment
0501	0.9		2801	0.8	
1213a	0.8		3101	0.7	
1701	0.9		3504	0.8	
2316	0.8		3507	0.7	

Again, in Table 7 there is full agreement with what was expected.

Table 8. Activation scores for full NPs questionably allowing an ON-pronominal referential alternative (prediction: between 0.4 and 0.6)

Unit #	AS	Comment	Unit #	AS	Comment
0503	0.4		2702	0.6	
1211	0.4		3001	0.4	
2318	1.1	Ref. conflict [0.9]	3204	0.6	
2503	0.4		3302	0.6	

In Table 8, the only difference from the prediction made by the basic hypothesis is the full NP *mexanika_m* in #2318 that received a score of 1.1 which is higher than expected. However, an additional mechanism comes into play here influencing referential device selection: referential conflict, discussed below in section 6. The comment for #2318 in Table 8 indicates the activation score of the competing referent “Fedorchuk” [0.9]. Note that the difference in the activation scores of the two referents is small; this might be the reason why the referent “mechanic” still seems to be marginally pronominalizable (according to the native speaker’s intuition).

Table 9. Activation scores for full NPs allowing no ON-pronominal alternative (prediction: 0.3 or lower)

Unit #	AS	Comment	Unit #	AS	Comment
0303	0.3	*	1906	0.2	
0303	0.3	*	2211	–	
0312	0.8	World boundary	2403	–	
0401	0.6	Ref. conflict [1.0]	2407	0.4	Ref. conflict [1.1]
0706	0.3		2504	0.1	World boundary
0902	0		2505	0.7	Ref. conflict [1.1]
1005	0.2	World boundary	2602	0.2	
1101	0.3	World boundary	2702	0.4	World boundary
1303	0.1		3505	0.1	
1304	–		3506	0.3	
1403	0.1		3604	–	
1407	0.2				

In the discourse sample there are 71 occurrences of no-alternative full NPs. However, in table 9 I included only 23 of them: those that had high numerical values at least for rhetorical distance or antecedent role, that is those that had at least a slight chance of not having a very low activation score. For all scores lower than 0 the score “–” is indicated. In this set of data we see three cases when a relatively high score, in principle allowing a potential ON-pronoun, is overcome by a referential conflict that rules a pronoun out. There are also several cases of “World boundary” block — see section 5 below. Other occurrences that deserve a special note are two mentions of the referents “mechanic” and “Fedorchuk” in #0303 (marked with asterisks in Table 9). They have antecedents quite close and could have a high activation score if these antecedents were not mentioned in a coordinate noun phrase (#0301) along with a third referent: *pilot, mexanik i učenic* ‘the pilot, the mechanic,

and the trainee'. One could postulate a separate "coordinate antecedent constraint" or ascribe non-pronominalizability to referential conflict but I am inclined to believe that the overall activation score of a triple referent in #0301 is divided by three ($1.0 : 3 = 0.3$). When there are two coordinate referents, further pronominalization of each one of them is not excluded if no referential conflict is there; I think this is because a relatively high score is divided by two, and the quotient is still sufficiently high.

4.4. Minor referential devices

Now brief notes are due about the minor referential devices used in the sample discourse.

4.4.1. Syntactic *ON*-pronouns

As was already mentioned above, syntactic anaphoric pronouns, unlike the activation-based anaphoric pronouns just treated in 4.3, can be most efficiently described in terms of syntactic control of the antecedent over the anaphor. Of course, syntactic anaphora is derivative from activation-based anaphora, and the latter can be formulated more broadly in order to include the former. However, for relatively simple syntactic *ON*-pronouns, the refined technique we employed for activation-based *ON*-pronouns is excessively complex.

There are six occurrences of syntactic *ON*-pronouns in the sample discourse. They are found in units #0108, #0111, #0114, #1702, #1802, and #2207. All of these occurrences belong to two major contexts of syntactic *ON*-anaphora found in Russian:

1. In complement clause constructions, the subject of the main clause triggers *ON*-pronominalization of coreferential arguments in the complement clause. Syntactic *ON*-pronouns in #0108, #0111, #1702, and #2207 are controlled by their subject antecedents in the respective main clauses.
2. In coordinate clauses, an argument of the linearly first clause triggers *ON*-pronominalization of a non-subject argument of the linearly second clause. Syntactic pronouns in #0114 and #1802 are controlled by their antecedents in the respective previous coordinate clauses.

4.4.2. “De-deictic” *ON*-pronouns

Both occurrences of what I called “de-deictic” *ON*-pronouns (#0803 and #1219) are found within reported speech/thought of characters. If this reported material were rendered as quoted speech/thought, first person pronouns *my* ‘we’ and *ja* ‘I’, respectively, would be used there. Apparently we face here a way of getting *ON*-pronouns which is alternative to the one involving the mechanism of discourse-based activation factors. These *ON*-pronouns are directly derived from deictic pronouns when quoted speech is turned into reported speech in a complement construction or otherwise. In some languages (especially African) there are special pronouns used in this function, so-called logophoric. This phenomenon in Russian requires further investigation.

4.4.3. *TOT*-pronouns

TOT-pronouns are a very interesting means in the Russian system of reference-maintenance designed specifically to refer to a referent second in activation (see Kibrik 1987). However, the activation-based pattern is not as frequent in discourse for *TOT* as it is for *ON*. To the contrary, *TOT* is more often used syntactically. Both activation-based and syntactic usages of *TOT* serve to remove an otherwise inevitable referential conflict.

All four occurrences of *TOT*-pronouns (#0203, #1408, #2504, and #3602) are syntactic ones. Basically, syntactic *TOT* refers to a non-subject argument of the linearly preceding clause. The activation score of the four occurrences of *TOT* varies between 0.7 and 0.8 while there is always a competing referent with the score of 1.0 through 1.2. The Russian *TOT*-pronoun is somewhat analogous to the so-called obviative in certain American Indian languages. For additional details on *TOT* see Kibrik (1987, 1988).

4.4.4. \emptyset^1

Zero subjects in independent clauses are a relatively rare phenomenon in the discourse under analysis. There are four occurrences of it in the sample (#2310, #2420, #2703, and #3205), and one instance when an *ON*-pronoun could be easily replaced by such zero (#2308). Among the four occurrences found in the text, two are found in the units containing an unexpressed subject of quoted speech (#2703, #3205). These zero mentions reproduce the deictic zeroes used in spoken speech. There is too little evidence to discuss them further.

Two other occurrences (#2310, #2405) are more related to the issue of discourse-based activation. The analysis of these two occurrences, and the negative material of non-zero *ON*-mentions in the discourse allow us to formulate some generalizations about this latter kind of \emptyset^1 . It might be expected, according to Givón (1983, 1990) that a zero marking appears in the situation of higher activation (topic continuity). However, the corresponding referents have activation scores quite normal for *ON*-pronouns, and indeed *ON*-pronouns can happily replace zeroes in these units. There are, however, some very strict limitations on the usage of this kind of \emptyset^1 as compared to *ON*. Rhetorical distance must be 1, the antecedent must be a subject, and paragraph distance must be 0. Furthermore, the current unit and the antecedent unit must be connected by a rhetorical link of simple temporal sequencing, maybe even belong to the narrative mainline. For some further observations on the differences between \emptyset^1 and *ON* see Nichols 1985.

4.4.5. Other

Other referential means employed in the sample discourse are too rare to allow us to make any specific conclusions. One could say that the object zero \emptyset^o (#2311, #2903, and #3206) tends to be used with a middle-high activation score (between 0.5 and 1.0). The dative zero \emptyset^d accompanying stative predicates (#0502, #0705, #0809, #2901, #3102, #3203, #3206) has very diverse activation scores ranging from -0.5 to 1.1 and varies greatly referentially (from specific to indefinite reference). The remaining zeroes (\emptyset^{12} , \emptyset^3 , etc.) are so rare that hardly anything can be said about them.

5. World Boundary Filter

“World boundary filter” is a separate component in the process of referential device selection. A world boundary is a block prohibiting any reduced form of reference (anaphoric pronoun or zero), even if it is possible from the point of view of activation score. The influence of “world shifts” on referential choices was extensively discussed by Clancy (1980: 146ff.).

The only type of world shift found in the sample discourse is a shift between the physical world described by the author, and the world of a character’s thoughts and images. The most eloquent example is found in #0312 where activation score of the referent “clouds” is very high (1.1), and

therefore according to the standard referential device selection procedure an anaphoric pronoun must be chosen. However, this would be absolutely impossible here because the previous mention of the referent exists in a different world from the one of the current unit. Though the referent is apparently highly activated, transition to a world different from the one where it was activated rules out its pronominalization. One and the same referent gets activated in different worlds independently.

6. Referential Conflict Filter

The referential conflict filter is another component potentially blocking the usage of reduced referential devices. It is much more important and general in nature than the world boundary block.

Referential conflict is what is commonly spoken of as referential ambiguity. When speaking, speakers tend to preclude the usage of such referential devices that might be assigned a wrong referent by the addressee. At any time a speaker knows the activation scores for each referent, and if there is more than one significantly active referent, the speaker should choose the lexico-grammatical means distinguishing the necessary referent from the competing referent. There is a great typological variety of linguistic means used in various languages for this purpose (see Kibrik 1991). In Russian, the most common lexico-grammatical means eliminating a potential referential conflict include: number, gender, type of anaphoric pronoun (*ON/TOT*), converbs (adverbial participles), and infinitives. Moreover, as probably in any language, less conventionalized ways to eliminate a potential referential conflict exist — those connected with the context of a particular unit (clause). Two major factors within this type are: (1) The factor of engagement (the competing referent has been already engaged in the current unit, and therefore cannot conflict with the referent in question); (2) the factor of fitting the propositional context (the competing referent, unlike the referent in question, is obviously incompatible with the propositional content of the present unit). If none of these lexico-grammatical or semantic means helps to distinguish the referents, then reduced forms of reference are blocked, and a full NP must be used.

A straightforward example of how the referential conflict filter works is found in #2505. The activation score of the referent “Fedorchuk” is 0.7 but a

full NP is used here, and no pronoun seems plausible in this place. This is because a competing referent, “mechanic”, has a higher activation score here (1.1), and none of the standard referential conflict elimination means helps to readily distinguish between the two referents. It can easily be demonstrated that in all cases where a pronoun or a zero is used referential conflict is removed by some readily available means.

I am aware of two ways of arguing against the referential conflict filter as a separate component in referential device selection. On one hand, Chafe suggested that the referential ambiguity problem is imposed by “exocultural” linguists on the language and real speakers “for whom familiarity and context are likely to remove most problems of keeping third-person referents straight” (1990: 315). I believe that not all problems resulting from referential conflict are removed by context, as we saw in the example just analyzed: in #2505 the reduced referential device is excluded exactly because of these problems. Furthermore, the cases when the context removes referential conflict must be subject to linguistic study.

On the other hand, some authors, e.g. Clancy (1980), seem to list referential conflict together with the factors that I interpret here as determining activation score. Can one suggest that introduction of a referent X per se plays a role in deactivating referent Y that was mentioned before that? I think not: The number of simultaneously fully active referents is of course limited, but two and even three different referents can easily be active at a time. In natural discourse it is not a rarity to see a clause with three arguments, all marked by pronominal or zero forms. So I believe that the referential conflict filter must be given the status of a separate component in the referential selection process.

7. Conclusion

In this paper I have tried to propose a cognitively oriented model of anaphora. The key concept underlying the phenomenon of anaphoric pronominalizability is that of referent activation, that is a referent’s presence in the active memory of the speaker as well as the addressee (according to the speaker’s belief). Referents whose activation is higher than a certain threshold value can be pronominalized, and those that are below this threshold cannot. In the genre of discourse selected in this study — third person narrative written

prose — all activation can come only from previous discourse and/or stable features of referents, so the processes of activation and deactivation can be fairly effectively controlled. I have proposed a set of activation factors that can either increase or decrease activation of a particular referent. Each feature of every factor has a certain numerical value, so for each referent an activation score — a cognitive and at the same time a numerical equivalent of pronominalizability — can be calculated. The activation score of every referent at any given time emerges naturally from the previous discourse and the properties of the referent itself, and is always readily available without any special effort.

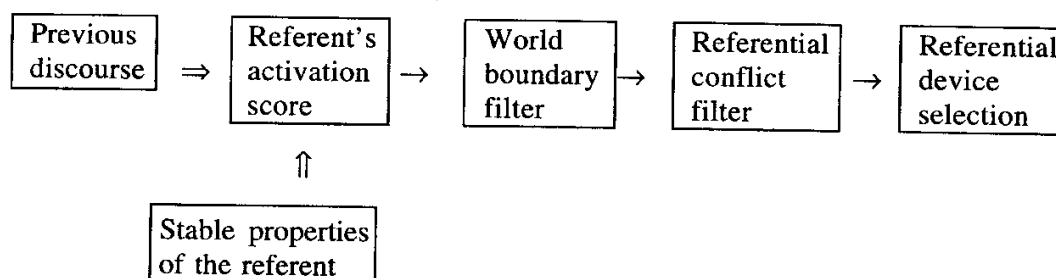
Among the activation factors that appeared crucial for the sample of Russian narrative prose are: rhetorical distance to the antecedent, syntactic and semantic role played by the antecedent, protagonist-hood, animacy, linear distance to the antecedent, paragraph distance, and sloppy identity of referents. Of course, the set and internal structure of factors have to be different for different languages, different registers and genres within one language, and possibly even for different speakers. Among the factors that are very likely to be relevant but did not prove useful for our sample is the factor of continued activation (resulting in persisting pronominalization; see Kibrik 1984, 1988; Givón 1990: 916). In principle, current activation of a referent can result from two possible causes: the fact that the referent was attended to at the previous moment of time (which is linguistically reflected as the antecedent being a subject), or the fact that the referent was already active at the previous moment of time (which is linguistically represented as the antecedent in a pronominal form). In the analyzed discourse I have found substantial evidence only for the first channel of activation.

I assume that speakers, when they are in the process of selecting a referential device, must rely on some single and cognitively simple mechanism. At the same time, it seems clear that no single parameter of context can fully explain all referential choices. The proposed model happily combines the holistic nature of activation (a single activation score at any given time) with its multifactorial origin.

The arithmetical technique employed to calculate activation score is supposed to model the cognitive interplay of activation factors. Of course, I realize that the cognitive processes cannot be that simplistic; they probably do not consist of the arithmetic operation of addition alone. I view this technique as a first approximation to a truly explanatory model of activational processes

that may become possible some time in the future and would employ a more sophisticated mathematical apparatus.

High activation is a necessary but not a sufficient condition for referent pronominalization. Each referent is further tested for two filters. The world boundary filter prohibits pronominalization of a referent activated in an alternative "world". The referential conflict filter blocks such pronouns that can be ascribed a wrong referent by the addressee, primarily because this wrong referent has a comparable activation score. Now the diagram representing the cognitive process of referential device selection would look as follows:



Legend:

- ⇒ arrows designating the operation of activation factors, taking place prior to the production of the current mention
- arrows designating the on-line transition from one stage to another during production

Figure 3. The revised process of referent mentioning

Acknowledgements

I would like to sincerely thank Barbara Fox, Thomas Payne, and Alan Cienki who took the effort to carefully read through this paper and suggest many valuable comments. I am also grateful to the participants of the Symposium on Anaphora (Allens Park, Colorado, May 1994) who took part in the discussion of the proposed ideas, particularly Russ Tomlin, Jean Newman, and Immanuel Barshi. I cordially thank Amy Crutchfield for her invaluable help with the translation of the discourse sample into English on a Russian airplane in May 1994. Of course, all of these people are in no way responsible for the contents of this paper.

The study underlying this article was partly supported by a grant from the Russian Basic Research Foundation (project 93-06-10940), and by a grant from the International Science Foundation, Moscow (project ZZ 5000/324).

REFERENCES

- Chafe, Wallace (ed.). 1980. *The pear stories: cognitive, cultural and linguistic aspects of narrative production*. Norwood: Ablex.
- Chafe, Wallace. 1990. Introduction. *International Journal of American Linguistics* 56:3, *Special issue on third-person reference in discourse*, 313-316.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: Chicago University Press.
- Clancy, Patricia. 1980. "Referential choice in English and Japanese narrative discourse." In Chafe (ed.), 1980:127-202.
- Cowan, Nelson. 1988. "Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information-processing system." *Psychological Bulletin* 104:2, 163-191.
- van Dijk, Teun, and Walter Kintsch. 1983. *Strategies of discourse comprehension*. New York: Academic Press.
- Fox, Barbara. 1987a. *Discourse structure and anaphora*. Cambridge: Cambridge University Press.
- Fox, Barbara. 1987b. "Anaphora in popular written English narratives." In Tomlin (ed.) 1987:157-174.
- Givón, Talmy (ed.) 1983. *Topic continuity in discourse: Quantified cross-language studies*. Amsterdam: Benjamins.
- Givón, Talmy. 1990. *Syntax: A functional-typological introduction. Vol. 2*. Amsterdam: Benjamins.
- Glezer, V.D. 1993. *Zrenie i myshlenie*. (Vision and cognition) St Petersburg: Nauka.
- Grimes, Joseph (ed.). 1978. *Papers in discourse*. Arlington: SIL.
- Haspelmath, Martin and Ekkehard Koenig (eds). 1995. *Converbs in cross-linguistic perspective*. Berlin: Mouton de Gruyter.
- Kibrik, Andrej. 1984. *Povtornaja nominacija: tipy i mexanizmy*. (Repeated mention: types and mechanisms). M.A.thesis. Moscow: MGU.
- Kibrik, Andrej. 1987. "Mexanizmy ustraneniya referencial'nogo konflikta v russkom jazyke (Mechanisms of referential conflict elimination in Russian)." In *Modelirovanie jazykovej dejatel'nosti v intellektual'nyx sistemax*, A.E.Kibrik and A.S.Narin'jani (eds), [128-145]. Moscow: Nauka.
- Kibrik, Andrej. 1988. *Tipologija sredstv oformlenija anaforicheskix svjazej*. (A typology of anaphoric means). Ph.D. diss. Moscow: Institute of Linguistics.
- Kibrik, Andrej. 1991. "Maintenance of reference in sentence and discourse." In *Language typology*, Winfred P. Lehmann and Helen-Jo J.Hewitt (eds), [57-84]. Amsterdam: Benjamins, 1991.
- Mann, William, Christian Matthiessen and Sandra Thompson. 1992. "Rhetorical structure theory and text analysis." In *Discourse description. Diverse linguistic analyses of a fund-raising text*, W. Mann and S. Thompson (eds), [39-78]. Amsterdam: Benjamins.
- Marslen-Wilson, William, Elena Levy, and Lorraine K. Tyler. 1982. "Producing interpretable discourse: The establishment and maintenance of reference." In *Speech, place, and action. Studies in deixis and related topics*, Robert J. Jarvella and Wolfgang Klein (eds), [339-378]. Chichester: Wiley.

- Nichols, Johanna. 1985. "The grammatical marking of theme in literary Russian." In *Issues in Russian morphosyntax*, Michael S. Flier and Richard D. Brecht (eds), [170-186]. Columbus, Ohio: Slavica.
- Payne, Thomas E. 1993. *The Twins Stories: Participant coding in Yagua narrative*. Berkeley: University of California Press.
- Posner, Michael and C. R. Snyder. 1975. "Facilitation and inhibition in the processing of signals." In *Attention and performance V*, P. M. Rabbit (ed.), [669-681]. New York: Academic Press.
- Tomlin, Russell (ed.). 1987. *Coherence and grounding in discourse*. Amsterdam: Benjamins.
- Tomlin, Russell. 1987. "Linguistic reflections of cognitive events." In R. Tomlin (ed.), 1987:455-79.
- Tomlin, Russell. 1994. "Focal attention, voice and word order: An experimental cross-linguistic study." In *Word order in discourse*, Pamela Downing and Michael Noonan (eds), [517-554]. Amsterdam: Benjamins.
- Van Valin, Robert. 1993. "A synopsis of Role and Reference Grammar." In *Advances in Role and Reference Grammar*, Robert D. Van Valin Jr. (ed.), [1-166]. Amsterdam: Benjamins.

Appendix 1: The Discourse Sample

Over the water, Boris Zhitkov (written around 1923)

The following extract comprises about 80% of the whole story. The beginning 20% of the text was excluded from consideration for certain technical reasons. The author describes there the passengers entering an aircraft (including a fat man, a lanky man, a woman, a military serviceman, and a young man), and the crew, including a pilot, a mechanic, and a trainee called Fedorchuk. Two more episodes are mentioned there: A boy filling the tank with gas takes out the filter secretly at some point and pours in dirty gasoline; the mechanic does not want to take the trainee to the flight, but the pilot insists.

I use the free English translation (printed in boldface) as the only glosses of the Russian units, when the syntactic structures of the two languages are isomorphic enough to allow this. I follow this principle even in the cases when nothing corresponds in the Russian original to English (e.g. an article). When no word-by-word free translation is possible, I use a separate line of more literal glosses, printed in regular face. Sometimes a free translation corresponds to two units, each having its own literal glosses. In the word-by-word glosses I render the meanings of the Russian cases with the help of hyphenated English prepositions.

0101 *Dva motora verteli dva vinta.*

Two engines rolled two propellers.

0102 *Za ix treskom trudno bylo slušat' drug druga passažiram,*
by their cracking hard was listen each other for-passengers

0103 *kotorye_a sideli v kajute apparata.*
which sat in cabin of-craft

The cracking sound they produced made it hard for the passengers who were sitting in the cabin of the craft to hear each other.

0104 *Oni_a perepisyvalis' na kločkax bumagi.*

They corresponded on sheets of paper

0105 *Nekotorye \emptyset _{a1} ne otryvajas' gljadeli v okna,*

Some were constantly watching through the windows

0106 *drugie \emptyset _{a2}, naoborot, staralis' smotret' v pol,*
others, to the contrary, were trying to look at the floor

0107 *čtoby kak-nibud' \emptyset ³⁴_{a2} ne uvidat',*
for somehow not see

in order not to see occasionally

0108 *na kakoj oni_{a2} vysote,*
at which they altitude

how high they were

0109 *i \emptyset ³⁴_{a2} ne ispugat'sja,*
and not get frightened

and get frightened,

- 0110 *no oni_{a2} čuvstvovali,*
but they felt
- 0111 *čtó pod nimi_{a2},*
what was beneath them anyway,
- 0112 *i ot ètogo Q²³_{a2} ne mogli bol'she ni o čem dumat'.*
and from this not could more not of anything think
and this made it impossible for them to think of anything else.
- 0113 *Dama_d dostala knižku*
The woman took out a book
- 0114 *i Q²_d ne otryvajas' v nee smotrela,*
and not tearing herself away in it looked
and constantly looked at it
- 0115 *no ničego Q²_d ne ponimala.*
but nothing not understood
but did not understand anything.
- 0201 "A *my vse podnimaemsja*", —
"We are still rising", —
- 0202 *napisal na bumažke veselyj tolstyj passažir_f, smotrevšij v okno,*
wrote on paper lively fat passenger looking at window
the lively fat passenger who was looking through the window wrote on a sheet
of paper
svoemu obaldevšemu sosedu₁.
to his distracted neighbor.
- 0203 *Tot₁ pročel,*
that read
The latter read,
- 0204 *Q²₁ maxnul razdraženno rukoj,*
waved irritatedly with-hand
waved his hand irritatedly,
- 0205 *Q²₁ natjanul ešče globže svoju šljapu*
pulled even deeper his hat
pulled his hat even lower
- 0206 *i Q²₁ niže naklonilsja k polu.*
and lower bent to floor
and bent further to the floor.
- 0207 *Tolstyj passažir_f dostal iz sakvojaža buterbrody*
fat passenger took out from suitcase sandwiches
The fat passenger took sandwiches out of his suitcase
- 0208 *i Q²_f prinjalsja spokojno est'.*
and started calmly eat
and calmly started eating.

- 0301 *A vperedi, u upravljenja, sideli pilot_p.*
In the front, at the control unit, sat the pilot,
mexanik_m i učenik_f.
the mechanic, and the trainee.
- 0302 *Vse \emptyset^n _{pmf} byli teplo odety, v kožanyx šlemax.*
All were warmly dressed, in leather helmets.
- 0303 *Mexanik_m znakami pokazyval učeniku_f na pribory:*
The mechanic used hand signs to show the trainee the indicators:
- 0304 *na al'timetr, kotoryj pokazyval vysotu,*
an altimeter that showed altitude,
- 0305 *na manometry, pokazyvavšie davlenie masla i benzina.*
on manometers showing pressure of-oil and of-gas
and manometers indicating oil and gas pressure.
- 0306 *Učenik_f sledil za ego_m žestami*
The trainee followed his signs
- 0307 *i \emptyset^2 _F pisal u sebja v knižečke voprosy korjavymi bukvami —*
and wrote at self in notebook questions with-uneven letters —
and wrote questions in his notebook with uneven letters —
- 0308 *\emptyset^p _F ruki byli v ogromnyx teplyx perčatkax.*
his hands were in huge warm gloves.
- 0309 *Al'timetr pokazyval 800 metrov*
The altimeter showed 800 meters
- 0310 *i \emptyset^2 šel vverx.*
and went up
and was rising.
- 0311 *Uže blizko oblaka.*
already close clouds
Clouds were already close.
- 0312 *“A kak v oblakax?” —*
and how in clouds
“What’s it like in the clouds?” —
- 0313 *pisal Fedorčuk_f*
wrote Fedorchuk.
- 0314 *“Čepuxa, uvidiš’”, —*
“Nothing special, you’ll see”, —
- 0315 *otvetil mexanik_m.*
answered the mechanic.
- 0401 *Učenik_f ne spešil bojat’sja,*
The trainee was not quickly frightened,
- 0402 *xot’ \emptyset^3 _F nikogda v oblakax ne byl.*
though never in clouds not was
though he had never been in clouds before.

- 0403 *Grešnym delom, on_F vse-taki podumyval,*
Somehow, he nonetheless started thinking
- 0404 *čto nepremenko dolžno vyjti čto-nibud' vrede stolknovenija.*
 that necessary must happen something like crash
of the inevitability of something like a crash.
- 0405 *Vperedi bylo sovsem tumanno,*
Ahead of them it was completely foggy
- 0406 *no čerez minutu aparat popal v polosu snega,*
but in a minute the craft got into a cloud of snow
- 0407 *kotoryj, kazalos', letel ne sverxu, a prjamo navstreču.*
 that seemed to fly not from above but directly at them.
- 0501 *Sneg zalepil okno vperedi pilota_p —*
The snow covered the window in front of the pilot —
- 0502 *vnizu \emptyset^d ničego ne bylo vidno.*
 below nothing not was visible
nothing could be seen below.
- 0503 *Pilot_p pravil po kompasu,*
The pilot steered using a compass,
- 0504 *no \emptyset^2 _p vse tak že zabiral vyše i vyše.*
 but still was getting higher and higher
bringing the craft higher and higher.
- 0505 *Stalo temnee.*
It grew darker.
- 0601 *Mexanik_m napisal Fedorčuku_F.*
The mechanic wrote to Fedorchuk:
- 0602 *"My v oblakax".*
"We are in the clouds".
- 0603 *Vokrug nix_A byl gustoj tuman,*
Around them was a dense fog,
- 0604 *i stalo temno, kak v sumerki.*
and it grew dark, as twilight.
- 0605 *Da i pozdno bylo —*
 and and late was
And it was late —
- 0606 *ostavalos' polčasa do zakata.*
 remained half an hour until sunset
only half an hour remained until sunset.
- 0701 *No vot stalo svetlee, ešče i ešče,*
 but here grew lighter more and more
But then it grew increasingly light,

- 0702 *i jarkoe solnce sovsem na gorizonte veselo zasverkalo*
and the bright sun just at the horizon joyfully played
na zaleplennyx snegom steklax.
 on covered with-snow glasses
on the windows covered with snow.
- 0703 *Daže passažiry_{a2}, čto smotrel i v pol,*
Even the passengers, who were looking at the floor
priobodrilis' i ožili.
 became lively and animated.
- 0704 *Sil'nyj veter ot xoda apparata*
The strong wind, resulting from the flight of the craft,
sdul nalipšij na stekla sneg,
 blew away stuck to glasses snow
blew away the snow stuck to the windows
- 0705 *i \emptyset^d_A stalo vidno jarkuju pelenu vnizu, do samogo gorizonta,*
 and became visible bright surface below until very horizon
and a bright surface below them, reaching the very horizon, became visible
- 0706 *kak budto nad beskonečnoj snežnoj ravninoj nessja apparat.*
 as if over endless snowy plain flew craft
as if the craft were flying over an endless snowy plain.
- 0801 *Pilot_p smotrel po časam*
The pilot looked at the watch
- 0802 *i \emptyset^2_p vysčityval v ume,*
and counted in his mind
- 0803 *gde oni_A sejčas dolžny byli byt'.*
 where they now must were be
where they were supposed to be at that time.
- 0804 *Solnce zašlo.*
The sun set.
- 0805 *Mexanik_m vključil svet,*
The mechanic turned on the light
- 0806 *i ottogo v kajute u passažirov_a stalo ujutnej.*
 and hence in cabin at passengers grew cosier
and it grew cosier in the passengers' cabin.
- 0807 *Vse \emptyset^n_a privykli k ravnomernomu revu motorov i*
Everyone got used to the monotonous roaring of the engines and
svistu vetra.
the whistling of wind.
- 0808 *V kajute bylo teplo,*
 in cabin was warm
It was warm in the cabin,

- 0809 *i* \emptyset^d_a *možno bylo zabyt'*,
and one could forget
- 0810 *čto pod apparatom poltory versty*
that under the craft there was a kilometer and a half
pustogo prostranstva,
of empty space
- 0811 *čto esli \emptyset^{34} upast'*,
that if one fell,
- 0812 *to voron kostej ne soberet,*
 then crow bones not will collect
not a trace would remain,
- 0813 *čto žizn' vse \emptyset^n_A - v iskusstve pilota*
 that life of-all in skill of-pilot
that everyone's life depended on the pilot's skill
i ispravnoj rabote motorov.
and the proper functioning of the engines.
- 0814 *Mnogie \emptyset^n_{a3} sovsem razveselilis'*,
Many passengers began to enjoy themselves
- 0815 *a tolstyj passažir, posylal vsem \emptyset^n_a smešnye zapiski.*
and the fat passenger was sending everyone funny notes.
- 0901 *Vdrug v rev motorov_{hh} vorvalis' kakie-to pereboi.*
 Suddenly in roaring of-engines entered some interruptions
Suddenly, there were interruptions in the roaring of the engines.
- 0902 *Passažiry_a bespokojno peregljanulis'.*
 passengers worriedly looked at each other
The passengers looked at each other worriedly.
- 0903 *Dolgovjazyj_l poblednel*
The lanky man grew pale
- 0904 *i \emptyset^2_l v pervyj raz vzgljanul v okno:*
 and in first time glanced in window
and glanced at the window for the first time:
- 0905 *ottuda na nego_l gljanula pustaja temnota,*
 from there at him looked empty darkness
empty darkness looked at him from there,
- 0906 *tol'ko otaženie lampočki trjaslos' v stekle.*
 only reflection of-bulb shook in glass
a lone bulb's reflection shook in the glass.
- 1001 *No pereboi prekratilis'*,
But the interruptions stopped,
- 1002 *i opjat' po-prežnemu rovnym voem reveli motory_{hh}.*
 and again further with-even roaring roared engines
and again the engines roared evenly.

- 1003 “*Ne pugajtes’*, —
“Don’t be frightened. —
- 1004 *pisal tolstjak_f*, —
wrote the fat man, —
- 1005 *esli i stanut motory_{hh}*
 if even stop engines
even if the engines stop,
- 1006 *my splaniruem*”.
we’ll glide”.
- 1007 “*V more*”, —
“Into the sea”, —
- 1008 *pripisal dolgovjazvi_i*
added the lanky man
- 1009 *i \emptyset^2_1 peredal zapisku obratno*.
and passed the note back.
- 1101 *Dejstvitel’no, apparat letel teper’ nad morem*.
Indeed, the craft was flying now over the sea.
- 1102 *Mexanik_m naprjazhenno slušal rev motorov_{hh}*
 mechanic tensely listened roaring of-engines
The mechanic listened tensely to the engines’ roaring,
- 1103 *kak doktor slušaet serdce bol’nogo*.
as a doctor listens to the heart of a patient.
- 1104 *On_m ponjal*,
He understood
- 1105 *čto byl propusk*,
 that was skip
that the engine skipped,
- 1106 *čto, verojatno, zasorilsja karbjurator_c* —
 that probably got clogged carburetor
- 1107 *čerez nego_c popadaet benzin v motor_h*,
 through it gets gas into engine
that probably a carburetor — through which gas gets into an engine — had gotten clogged,
- 1108 *a čto teper’ proneslo*;
and that this time the emergency had passed;
- 1109 *no \emptyset^{12}_m uže znal*,
 but already knew
but he knew already
- 1110 *čto benzin nečist*,
that the gas was not clean,
- 1111 *i \emptyset^2_m bojalsja*,
and was afraid

- 1112 *čto zasoritsja karbjurator_c —*
that will get clogged carburetor
that the carburetor would become clogged
- 1113 *i stanet motor_h.*
and will stop engine
and the engine would stop.
- 1201 *Fedorčuk_f sprosila,*
Fedorchuk asked
- 1202 *v čem delo.*
what the problem was.
- 1203 *No mexanik_m otmaxnulsja*
But the mechanic waved him away
- 1204 *i, Ø_m⁵ ne otvečaja,*
and, without answering,
- 1205 *Ø_m² prodolžal naprjaženno prislušivat'sja.*
kept tensely listen
kept listening intently.
- 1206 *Učenik_f staralsja sam dogadat'sja,*
trainee tried self guess
The trainee was trying to guess himself
- 1207 *otčego èto poperxnulsja motor.*
why that coughed engine
why it was that the engine had coughed.
- 1208 *Tysjača pričín: magneto, sveči, klapany —*
There could be a thousand reasons: magneto, spark plugs, valves —
- 1209 *i kakoj motor, pravij ili levij?*
and which engine, right or left?
- 1210 *V každom motore opjat' že po dva karbjuratora.*
In each engine there are, again, two carburetors.
- 1211 *Fedorčuku_f tože prixodilo v golovu,*
to-Fedorchuk also was coming to head
It also occurred to Fedorchuk
- 1212 *ne zasorilos' li.*
not clogged whether
that there might have been some obstruction.
- 1213 *"Nu, —*
"Well, —
- 1213a *podumal Fedorčuk_f, —*
thought Fedorchuk, —
- 1213b *Ø' budem planirovat' i činit'sja v vozduxe".*
we will just glide and make repairs in the air".
- 1214 *No emu_f bylo udivitel'no,*
But he was surprised

- 1215 *počemu tak perezpugalsja ètot znajuščij _____ mexanik_m.*
 why so got scared this knowledgeable mechanic
why this knowledgeable mechanic was getting so scared.
- 1216 *Takoj on_m trus*
 such he coward
Was he such a coward
- 1217 *ili v samom dele čto-nibud' ser'eznoe,*
 or was there really something seriously wrong
- 1218 *čego v polete Ø³⁴ ne ispravit',*
 that in flight not fix
not fixable during the flight
- 1219 *a on_F, novičok, ne ponimaet?*
 that he, a newcomer, did not understand?
- 1301 *No tut rev motora stal vdvoe slabee.*
 At that point the roaring of the engines decreased by half.
- 1302 *Pilot_p povernul rul'*
 The pilot turned the steering wheel
- 1303 *i Ø²_p vyključil levyj motor_{hl}.*
 and turned off the left engine.
- 1304 *Fedorčuk_F ponjal,*
 Fedorchuk understood
- 1305 *čto pravyj Øⁿ_{h2} stal sam.*
 that the right one had stopped by itself.
- 1401 *Mexanik_m poblednel*
 The mechanic grew pale
- 1402 *i Ø²_m stal kačat' ručnoj pompoj vozdux v benzinnyj bak.*
 and started pump with-hand pump air into gas tank
and started pumping air into the gas tank with a hand pump.
- 1403 *Fedorčuk_F soobrazil,*
 Fedorchuk guessed
- 1404 *čto on_m xočet naporom benzina pročistit'*
 that he wants with-pressure of-gas clean out
zasorivšijsja karbjurator,
 clogged carburetor
that he intended to clean out the dirty carburetor with the pressure of the gas
- 1405 *no Ø¹²_F znal uže,*
 but he knew already
- 1406 *čto èto ni k čemu.*
 that it was futile.
- 1407 *Pilot_p kričal na uxu mexaniku_m,*
 The pilot yelled into the ear of the mechanic

- 1408 *čtoby* tot_m *šel* *na* *krylo* Ø_m *naladit'*
that he should have gone to the wing to fix
ostanovivšijsja *motor.*
the stopped engine.
- 1501 *Al'timetr pokazyval 1200 metrov.*
The altimeter showed 1200 meters.
- 1601 *A v kajute vstrevožennye passažiry_a gljadeli*
In the cabin, the worried passengers looked
drug drugu v ispugannye lica,
at each other's frightened faces,
- 1602 *i daže tolstjak_f pisal ne sovsem četko:*
and even the fat man was writing not quite distinctly:
- 1603 *ruka ego_f trjaslas' nemnogo.*
the hand of him trembled a bit.
- 1604 *"My planiruem,*
"We are gliding,
- 1605 *sejčas ispravjat motor,*
now they'll fix the engine,
- 1606 *i my poletim".*
and we'll fly".
- 1607 *No myslenno vse Øⁿ_a pribavljali:*
But mentally everyone added:
- 1608 *"vniz golovoj v more".*
"upside down into the sea".
- 1701 *Passažiry_a ne znali,*
The passengers had no idea
- 1702 *na kakoj oni_a vysote.*
at which they altitude
how high they were flying.
- 1801 *Vse Ø^x_a bojalis' morja vnizu ,*
Everyone was as afraid of the sea below
- 1802 *i v to že vremja ix_a pugala vysota.*
and at the same time them scared altitude
as they were of the altitude.
- 1901 *Dolgovjazyj passažir_l vdrug sorvalsja s mesta*
The lanky passenger suddenly jumped from his seat
- 1902 *i zØ²_l brosilsja k dverjam kajuty;*
and rushed to the door of the cabin;
- 1904 *on_l dergal ručku,*
he jerked on the handle

- 1905 *kak budto* \emptyset^3_1 *xotel* *vyrvat'sja iz gorjaščego doma.*
 as if he were trying to escape a burning building.
- 1906 *No dver' byla zaperta snaruži.*
 But the door was locked from the other side.
- 1907 *Dama_d vypustila iz ruk knižku,*
 The woman dropped her book,
- 1908 \emptyset^2_d *diko, pronzitel'no zakričala.*
 and let out a wild, piercing cry.
- 1909 *Vse \emptyset^n_a vzdrognuli,*
 Everyone was jolted,
- 1910 \emptyset^2_a *vskočili s mest*
 jumped from his seat
- 1911 *i \emptyset^2_a stali bescel'no metat'sja.*
 and started aimlessly rushing about.
- 2001 *Tolstjak_f povtorjal,*
 The fat man repeated
- 2002 \emptyset^5_f *ne ponimaja svoix slov:*
 without understanding his own words:
- 2003 — *Ja skažu,*
 — I'll tell them
- 2004 *čtoby leteli,*
 to fly,
- 2005 \emptyset^1 *sejčas skažu!..*
 now will tell
 I'll tell now!..
- 2101 *Dama_d povernulas' k oknu*
 The woman turned toward the window
- 2102 *i \emptyset^2_d vdrug melko i slabo zabarabanila*
 and suddenly started frequently and weakly drumming
kulačkami po steklu,
 with her fists at the glass
- 2103 *no \emptyset^2_d sejčas že upala bez čuvstv poperek kajuty.*
 but right away fell in a faint across the cabin.
- 2201 *Voennyj_s blednyj kak polotno, stojal*
 The serviceman, white as a sheet, was standing
- 2202 *i \emptyset^2_s gljadel v černoje okno ostanovivšimisja glazami.*
 and staring out the black window with frozen eyes.
- 2203 *Koleni ego_s trjaslis',*
 knees his trembled
 His knees were trembling,
- 2204 *on_s ele stojal na nogax,*
 he hardly stood on legs
 he could hardly stand,

- 2205 *no* \emptyset^2_s *ne mog otvesti glaz.*
but he could not move his gaze.
- 2206 *Molodoj čelovek,* *v sinej kepke zakryl lico*
The young man in the blue cap covered his face
rukami,
with his hands,
- 2207 *kak budto u nego, boleli zuby.*
as if at him hurt teeth
as if his teeth hurt.
- 2208 *V prednem uglu požiloij passažir_e motal*
In the front corner an elderly passenger shook
bolezneno golovoj
painfully with-head
his head in pain
- 2209 *i* \emptyset^2_e *vskrikival: "Ga-ga-ga".*
and shouted: "Ha-ha-ha".
- 2210 *V takt ètomu kriku vse sil'nee dergalas' ručka dveri*
in rhythm to-this cry more stronger was-jerked handle of-door
In rhythm with this cry the door's handle was being jerked harder and harder,
- 2211 *i bol'se raskačivalsja molodoj čelovek_y,*
and more swayed young man
and the young man was swaying more and more.
- 2212 *"Ga-ga-ga" perešlo v isstuplennyj rev,*
"Ha-ha-ha" turned into a frantic roar,
- 2213 *i vdrug vse passažir_a zavyli,*
and suddenly all passengers started howling,
- 2214 \emptyset^2_a *zastonali razdirajuščim xorom.*
moaning like a heart-rending choir.
- 2301 *A mexanik_m vse vozilsja,*
And the mechanic was still fiddling with his stuff,
- 2302 \emptyset^2_m *vse podkačival pompu,*
still pumping the pump,
- 2303 \emptyset^2_m *stukal pal'cem po steklu manometra.*
hit with-finger at glass of-manometer
hitting the manometer's glass with his finger.
- 2304 *Pilot_p tolknul ego_m loktem*
The pilot hit him with his elbow
- 2305 *i* \emptyset^2_p *strogo kivnul golovoj v storonu vyxoda na krylo.*
and severely jerked his head toward the exit onto the wing.
- 2306 *Mexanik_m sunulsja,*
The mechanic started,
- 2307 *no* \emptyset^2_m *sejčas že vernulsja —*
but immediately returned —

- 2308 *on_m stal ryt'sja v jaščike s instrumentami,*
he began to dig in the box of instruments;
- 2309 *a oni ležali v svoix gnezdax, v strogom porjadke.*
they were lying in their places, in full order.
- 2310 *Ø¹_m Xvatal odin ključ,*
He pulled out one wrench,
- 2311 *Ø²_m brosal Ø^o,*
dropped it,
- 2312 *Ø²_m motal golovoj,*
shook his head,
- 2313 *Ø²_m čto-to šeptal*
 something whispered
whispered something
- 2314 *i Ø²_m snova rylsja.*
 and again dug
and reached in again.
- 2315 *Fedorčuk_f teper' jasno videl,*
Fedorchuk now clearly saw
- 2316 *čto mexanik_m strusil*
that the mechanic was a coward
- 2317 *i Ø²_m ni za čto už ne vyjdet na krylo.*
and would never go out to the wing.
- 2318 *Pilot_p razdraženno tolknul mexanika_m kulakom v šlem*
 pilot angrily pushed mechanic with-fist at helmet
The pilot angrily hit the mechanic's helmet with his fist
- 2319 *i Ø²_p tknul pal'cem na al'timetr:*
and poked his finger at the altimeter:
- 2320 *on pokazyval 150.*
it showed 150.
- 2321 *Sto pjat'desjat metrov do morja.*
One hundred fifty meters from the sea.
- 2401 *Mexanik_m utverditel'no zakival golovoj*
 mechanic affirmatively nodded head
The mechanic nodded his head affirmatively
- 2402 *i Ø²_m ešče bystree stal Ø⁴_m perebirat' instrumenty.*
 and more quicker started sort instruments
and started pawing through the instruments even more quickly.
- 2403 *Pilot_p kriknul:*
The pilot shouted:
- 2404 — *Øⁱ_m Voz'mi rul'!*
 — **Hold the steering wheel!**
- 2405 *Ø¹_p Xotel Ø⁴_p vstat'*
He wanted to get up

- 2406 *i* \varnothing^4_p *sam pojti k motoru,*
and self go to engine
and go to the engine himself
- 2407 *no mexanik_m ispuganno zamaxal rukami*
but mechanic fearfully waved with-hands
but the mechanic waved his hands fearfully
- 2408 *i* \varnothing^2_m *otkinulsja na spinku siden'ja.*
and leaned back in his seat.
- 2501 *Fedorčuk_F vskočil.*
Fedorchuk jumped up.
- 2502 — \varnothing^i_m *Davaj ključ!* —
— **Give me the wrench!** —
- 2503 *kriknul on_F mexaniku_m.*
cried he to-mechanic
he cried to the mechanic.
- 2504 *Tot_m drožašče_j rukoj sunul*
The latter, with a trembling hand, pressed
emu_F v ruki malen'kij gaečnyj ključik.
him in hands small wrench
a small wrench into his hand.
- 2505 *Fedorčuk_F vyšel na krylo.*
Fedorchuk went out onto the wing.
- 2601 *Rezkij, pronizyvajuščij veter nes xolodnyj tuman;*
A sharp, penetrating wind carried a cold fog;
- 2602 *on skol'zkoj koroj namerzal*
it with-slick crust froze
the fog froze, in a slick crust,
na kryl'jax, na stojkax, na provoločnyx tjaxax.
on the wings, struts, wires.
- 2603 $@_F$: — *K motoru!*
— **Toward the engine!**
- 2701 \varnothing^5_F *Riskuja každyju sekundu sletet' vniz,*
risking at-any second fly down
Risking a fall at any second,
- 2702 *dobral'sja Fedorčuk_F do motora_h.*
reached Fedorchuk to engine
Fedorchuk reached the engine.
- 2703 $@_F$: \varnothing^1_h *Teplij ešče.*
warm still
"It's still warm."

- 2801 *Fedorčuk_F slyšal voj iz passažirskoj kajuty*
Fedorchuk heard the howling from the passengers' cabin
- 2802 *i Ø²_F naščupyval na karbjuratore nužnuju gajku_n.*
 and groped on carburetor needed nut
and groped for the needed nut on the carburetor.
- 2803 *@_F: Vot ona_n!*
"Here it is!"
- 2901 *Ø^d_F Skol'zko stojat',*
It's too slippery to stand,
- 2902 *veter revet*
the wind roars
- 2903 *i Ø² tolkaet Ø^o_F s kryla.*
and pushes him off from the wing.
- 3001 *Vot gajka_n podalas'.*
 Now the nut yielded.
- 3002 *@_F: Idet delo!*
 goes work
"The work is progressing!"
- 3101 *Spešit Fedorčuk_F,*
 hurries Fedorchuk
Fedorchuk is in a rush,
- 3102 *i Ø^d_F už slyšno,*
 and already audible
and he can hear already
- 3103 *kak revet vnizu more.*
 how roars below sea
how the sea roars below.
- 3104 *Ešče minuta, drugaja —*
 more minute another
In one or two minutes
- 3105 *i aparat so vsemi ljud'mi_A potonet v merzloj vode.*
the craft with all aboard would sink in the freezing water.
- 3201 *@_F: — Gotovo!*
— Ready!
- 3202 *@_F: Teper' gajku na mesto!*
So now the nut goes back to its place!
- 3203 *Ø^d_F Zamerzli pal'cy,*
 froze fingers
His fingers froze,
- 3204 *ne popadaet na rez'bu prokljataja gajka_n.*
 not fits on thread damned nut
the damned nut would not go onto the thread.

- 3205 @_F: *Sejčas, sejčas* $\text{\textcircled{1}}_n$ *na meste,*
Now, now, on its place,
- 3206 *teper'* $\text{\textcircled{d}}_F$ *nemnogo ešče pritjanut'* $\text{\textcircled{o}}_n$.
 now a bit more tighten
now make it a bit tighter.
- 3301 — *Est'!* —
 — **Done!** —
- 3302 *zaoral* *Fedorčuk*_F *vo vsju silu svoix legkix.*
shouted Fedorchuk at the top of his lungs.
- 3401 *Vključili* *električeskij pusk,*
 they-turned on electric starter
The electric starter was turned on
- 3402 *i zareveli motory.*
 and roared engines
and the engines began to roar.
- 3403 *V kajute vse* $\text{\textcircled{a}}$ *srazu stixli*
In the cabin the passengers immediately became quiet
- 3404 *i* $\text{\textcircled{2}}_a$ *opustilis' gde kto byl: na pol,*
and lay down wherever they were: on the floor,
na divan, drug na druga
on the bench, on each other.
- 3405 *Tolstjak*_f *pervyj prišel v sebja*
 fat first came to self
The fat man gathered his wits first
- 3406 *i* $\text{\textcircled{2}}_f$ *stal podymat' besčuvstvennuju damu.*_d
and began to prop up the unconscious woman.
- 3501 *A* *Fedorčuk*_F *smelo lez po krylu nazad k upravljeniju.*
And Fedorchuk bravely crawled along the wing back to the control unit.
- 3502 *U nego*_F *bylo veselo na serdce.*
 at him was happy at heart
He felt happy.
- 3503 *Poryvy štormovogo vetra brosalı aparat.*
The blows of the stormy wind threw the craft about.
- 3504 *Fedorčuk*_F *vzjalsja za ručku dvercy,*
Fedorchuk held the handle of the door,
- 3505 *no soskol'znula* $\text{\textcircled{p}}_F$ *noga s obledenelogo kryla,*
 but slid foot from icy wing
but his foot slid from the icy wing,
- 3506 *ručka* *vyskol'znula iz* $\text{\textcircled{p}}_F$ *ruk,*
the handle slid out of his hand
- 3507 *i* *Fedorčuk*_F *sorvalsja v temnuju pustotu.*
and Fedorchuk fell into dark emptiness.

- 3601 *Čerez minutu pilot_p zlobno vzgljanul na mexanika_m.*
 in minute pilot viciously glanced at mechanic
After a short time the pilot glanced viciously back at the mechanic.
- 3602 *Tot_m, blednyj, vse ešče perebiral*
The latter, pale, was still pawing through
 instrumenty v jaščike.
the instruments in the box.
- 3603 *Oba Øⁿ_{pm} ponimali,*
Both understood
- 3604 *počemu net Fedorčuka_F.*
 why no Fedorchuk
why Fedorchuk was not there.

Appendix 2: An Example of Activation Score Calculations

Unit #	2306	2308	2309	2310	2315	2316	2318	2318	2320
Form	mexanik	on	oni	Ø	Fedorčuk	mexanik	pilot	mexanika	on
Referent	m	m	i	m	F	m	p	m	a
Referential device	full NP	ON	ON	Ø	full NP	full NP	full NP	full NP	ON
Alternative device	?ON, TOT	Ø	full NP	ON	-	ON	-	?ON	full NP
Predicted AS	0.4- 0.6	1.0	0.7- 0.9	1.0	0.3 or less	0.7- 0.9	0.3 or less	0.4- 0.6	0.7- 0.9

Factors and numerical values (in boldface):

Linear distance	2 -0.1	1 0	1 0	2 -0.1	>4 -0.5	2 -0.1	>4 -0.5	1 0	1 0
Rhetorical distance	2 0.4	1 0.7	1 0.7	1 0.7	>3 -0.3	2 0.4	>3 -0.3	1 0.7	1 0.7
Paragraph distance	0 0	0 0	0 0	0 0	>1 -0.4	0 0	0 0	0 0	0 0
Syntactic and semantic role of the antecedent	DO U 0	S A 0.4	Obl NC 0	S A 0.4	S A 0.4	S A 0.4	S A 0.4	S A 0.4	IO NC 0
Protagonisthood	+ 0	+ 0	- 0	+ 0	+ 0.3	+ 0	+ 0.3	+ 0	- 0
Animacy	+ 0.1	+ 0	- 0	+ 0	+ 0.2	+ 0.1	+ 0.2	+ 0	- 0
Actual AS	0.4	1.1	0.7	1.0	-0.3	0.8	0.1	1.1*	0.7

* Referential conflict filter is involved in this case, see discussion in section 4.3 next to Table 8.