

A Cognitive Calculative Approach towards Discourse Anaphora

Andrej A. Kibrik

(Institute of Linguistics of the Russian Academy of Sciences and Max Planck Institute for Evolutionary Anthropology)
Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103, Leipzig, Germany
kibrik@eva.mpg.de

Abstract

In this paper I report two studies of discourse anaphora (in Russian and English narrative discourse). I claim that referential device selection is primarily influenced by the speaker's current cognitive state, namely the referent's activation in the speaker's working memory. Activation is assumed to be a gradual parameter varying between the minimal and the maximal values. A methodology of numerical calculation of referent activation is proposed. The level of a specific referent's activation at a particular point of discourse depends on multiple factors. Each of these factors contributes (in a positive or negative numerical way) to the ultimate activation score of the referent. Activation factors are rooted in either the discourse context (several measurements of distance to the antecedent; the antecedent's syntactic role in its clause...) or in the referent's properties (animacy, protagonisthood...) The numerical model employed to calculate activation score is supposed to imitate the cognitive interplay of activation factors during discourse production. There exist soft and hard thresholds of activation determining the speaker's referential choice. For example, above the hard threshold a reduced NP (that is, an anaphoric pronoun or zero) is necessary, while above the soft threshold it is possible. The proposed model aims at predicting/explaining all referential occurrences in the sample discourse. This model has substantial consequences for the general understanding of the cognitive domain of working memory. Specific claims about the capacity and control of working memory are proposed.

1. General Assumptions Underlying this Study

In this paper, I approach discourse anaphora from the perspective of a broader process that I term referential device selection or, more simply, **referential choice**. This term differs from "discourse anaphora" in the following respects.

- 1) The notion of "referential choice" is more overtly production oriented. It is not a phenomenon found "somewhere" in discourse but a process directly performed by the speaker.
- 2) Unlike "discourse anaphora", "referential choice" does not exclude introductory mentions of referents and other mentions that are not based on already-high activation of the referent.
- 3) The notion of referential choice permits one to avoid the well-known dispute on whether "anaphora" is restricted to specialized formal devices (such as pronouns, articles, etc.) or has a purely functional definition.

Otherwise the two notions are essentially identical.

Below I formulate the list of general assumptions underlying the model of referential choice that I propose in this paper. Argument for these assumptions is beyond the scope of this article but I believe that they have a certain degree of general appeal.

- 1) An adequate model of referential choice should be able to account for every occurrence of a referential device (at least in the sample discourse), rather than content with 90% of hits.
- 2) A psychologically adequate model of referential choice cannot be based on antecedent search in the pretext (look-back), but rather must look for explanations of referential choices in the cognitive structures of the speaker at the moment of speech.
- 3) An adequate model of referential choice must explore the possibility that the choice depends on multiple factors rather than one single factor.

- 4) If the multiplicity of factors is recognized, each factor must be monitored in each case, rather than in an ad hoc manner.
- 5) If multiple factors are involved, the issue of interaction between various relevant factors must be addressed.

In accordance with these assumptions, the model proposed below is:

- sample-based
- general and predictive (accounts for all data in sample)
- explanatory and cognitively based
- speaker-oriented
- multi-factorial
- testable
- calculative.

2. The Cognitive Model

Now, a set of more specific assumptions on how referential choice works at the cognitive level is in order. Recently a number of studies have appeared suggesting that referential choice is directly related to the more general cognitive domain of working memory and the process of **activation in working memory** (Chafe, 1994; Tomlin and Pu, 1991; Givón, 1995; Kibrik, 1996, 1999). Some of these authors use slightly different terminology, for example Chafe speaks about consciousness rather than memory, but the cognitive domain he has in mind is essentially the same.) For cognitive psychological and neurophysiological accounts of working memory see Baddeley (1986, 1990), Anderson (1990), Cowan (1995), Posner and Raichle (1994), Smith and Jonides (1997). The claim that referential choice is governed by memorial processes is compatible with psycholinguistic frameworks of such authors as Gernsbacher (1990), Clifton and Ferreira (1987), Vonk, Hustinx and Simons (1992), with the cognitive approaches of Gordon, Grosz and Gilliom (1993), Gundel, Hedberg and Zacharski (1993), and van Hoek (1997), and with some computational models covered in Botley and McEnery (eds., 2000).

Thus the first element of the cognitive model can be formulated as follows:

- (1) The primary cognitive determiner of referential choice is **activation** of the referent in question in the speaker's working memory (henceforth: WM).

Activation is a matter of degree. Some chunks of information are more central in WM while some other are more peripheral. I use the term **activation score** (AS) to refer to the current referent's level of centrality in the working memory. AS can vary within a certain range – from a minimal to a maximal value. This range is not monotonous in the sense that there are certain important thresholds in it. When referents have a high AS, semantically **reduced** referential devices, such as pronouns and zeroes, are used. On the other hand, when the AS is low, semantically full devices such as full NPs are used. Thus the second basic idea of the cognitive model proposed here is the following.

- (2) If AS is above a certain **threshold**, then a semantically reduced (pronoun or zero) reference is possible, and if not, a full NP is used.

This idea will be further elaborated below.

Thus at any given moment in discourse any given referent has a certain AS. How is this AS formed? I claim that it depends on a whole gamut of various factors that can essentially be grouped in two main classes:

- properties of the referent (such as the referent's animacy and centrality)
- properties of the previous discourse (distance to the antecedent, the antecedent's syntactic and semantic status, paragraph boundaries, etc.)

These factors are also specified below in sections 3 and 4.

Now the third basic point of the model can be formulated:

- (3) At any given point of discourse **all relevant factors interact** with each other, and give rise to the integral characterization of the given referent (AS) in respect to its current position in the speaker's WM.

In other words, such oft-cited factors as referential choice as distance to the antecedent, referent centrality, etc., affect the referential choice not directly but through the mediation of the speaker's cognitive system, specifically, his/her WM. Therefore these factors can be called **activation factors**.

The actual cognitive on-line process of referential choice is a bit more complex than is suggested by the three postulates formulated above. Much work on referential choice, including by the present writer (see e.g. Kibrik, 1991), has been devoted to the issue of ambiguity of reduced referential devices. In the process of referential choice, a normal speaker filters out those referential options that can create ambiguity, or **referential conflict**. Thus it is possible that even in case of high activation of a referent a reduced referential device is still ruled out. However, this component is outside of the focus of this paper.

The cognitive model outlined above is proposed here not only in a declarative way; there is also a mathematical, or at least quantitative, or **calculative component** in it. Each activation factor is postulated to have a certain numeric value that reflects its relative contribution to the integral AS value. The general model of referential choice outlined above is assumed to be universal but the set of

activation factors, especially their relative numerical weights and thresholds in the AS range are language-specific. In this article I report two studies that have been conducted for Russian and English written narrative discourse. The quantitative component of the model is therefore explained in sections 3 and 4 below addressing these two different languages. The Russian study is described first (section 3), in quite a sketchy way but with a certain discussion of research heuristics. The English study (section 4) is described in more detail and more formally. All linguistic examples are delayed till section 4.

3. The Russian Study

In that study (for details see Kibrik, 1996) I investigated a single sample of narrative prose – a short story by the Russian writer Boris Zhitkov “Nad vodoj” (“Over the water”). This particular sample discourse was selected for this study for the following reasons.

(a) Narrative prose 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.

(b) Written prose was selected because it is a well-controlled mode in the sense that previous discourse is the only source for the recurring referents; an advantage of written language is that we can fully control the processes of activation by fairly objective discourse data.

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

(d) This specific story was selected because it is again 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 not numerous references to the narrator.

The sample discourse comprised about 300 discourse units. There are about 500 mentions of various referents in the sample, and there are some 70 different referents appearing in the discourse. However, only a minority of them occur 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 fundamental opposition in Russian referential choice is between full NPs and the third person pronoun *on*. Discourse-conditioned referential zeroes are also important, but they are rarer than *on* (for further details see Kibrik, 1996).

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 is at least one of the major predictors of referential choice. Givón measured linear distance in terms of clauses, and that principle turned out to be very productive and viable. In many later studies, including this one, discourse microstructure is viewed as a network of discourse units essentially coinciding with clauses.

There are certain reservations regarding this coincidence, but they are irrelevant for this paper.

Fox (1987a: Ch. 5) argued that it is a rhetorical, hierarchical structure of discourse rather than plain linear structure that affects selection of referential devices. Fox counted rhetorical distance to the antecedent on the basis of a rhetorical structure constructed for a text in accordance with the Rhetorical Structure Theory (RST), as developed by Mann and Thompson (see Mann and Thompson, 1992; in application to anaphora see Fox, 1987a). RST is probably the best current theory of discourse embracing both macro- and microstructure. According to RST, each discourse unit (normally a clause) is connected to at least one other discourse unit by means of a rhetorical relation, and via it, ultimately, to any other discourse unit. There exists a limited but open inventory of rhetorical relations, such as joint, sequence, cause, elaboration, etc. In terms of RST, each text can be represented as a fully interconnected graph consisting of nodes (discourse units) and connections (rhetorical relations). Rhetorical distance between nodes A and B is then the number of steps one needs to make to reach B from A along the graph. One example of a rhetorical graph is given in Appendix 2. Fox was correct in suggesting that rhetorical distance measurement is a much more powerful tool for modelling reference than linear distance. However, linear distance also plays its role, though a more modest one.

In a number of works it was 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) have all demonstrated, though using very different methodologies, that an episode/paragraph boundary is a borderline after which speakers tend to use full NPs even if the referent was recently mentioned.

One more factor was emphasized in Grimes (1978) – the centrality of a referent in discourse, which I call *protagonisthood* below. For a discussion of how to measure a referent’s centrality see Givón (1990: 907-909).

Several other factors have been suggested in the literature, including animacy, syntactic and semantic roles played by the NP/referent and by the antecedent, syntactic distance to the antecedent measured in full sentences, and the referential status of the antecedent (full/reduced NP).

From the maximal list of potentially significant activation factors I picked a subset of those that prove actually **significant** for Russian narrative prose. The criterion I used is as follows. Each factor can be realized in a number of features, for example a distance factor may have features 1, 2, etc. Each potentially significant factor has a “privileged” feature that presumably correlates with the more reduced form of reference. For example, for the linear distance to the antecedent it is the feature of “1”, while for the factor of the antecedent’s syntactic role it is “subject”. Only those potential factors whose privileged feature demonstrated a high co-occurrence (in at least 2/3 of all cases) with the reduced form of reference have been considered significant activation factors. Also, significant activation factors display different patterns of co-occurrence vis-à-vis full NPs. Some quantitative data demonstrating the patterns of co-occurrence between the significant activation factors and referential choices are shown in Table 1 below.

It is easy to see that the factor of rhetorical distance patterns vis-à-vis pronouns and full NPs in a nearly mirror-image way. There is a high co-occurrence of the feature of 1 with pronominal reference, and a high co-occurrence of the opposite end of the scale with full NP reference. Other factors such as animacy and syntactic role of the antecedent pattern differently for pronouns and full NPs but not so contrastively: their privileged feature (e.g. “human” for animacy) co-occurs highly with the pronominal choice but there is an even distribution as concerns full NP mentions.

<i>Activation factors</i>	<i>The features of activation factors</i>	<i>Percentage of cooccurrence</i>	
		with the <i>on</i> -pronoun	with full NPs
Animacy	human	78	48
	inanimate	22	52
Rhetorical distance to the antecedent	1	91	21
	2	3	18
	3		11
	4+		50
Syntactic role of the antecedent	subject	78	50
	non-subject	22	50
Referential type of the antecedent	full NP	53	62
	<i>on</i>	10	13
	other pronoun	3	2
	zero	34	22

Table 1: Examples of co-occurrence (in %) between the potential activation factors and referential choice. (The names of significant activation factors are printed in boldface in the left column.)

Still other potential factors did not display any significant co-occurrence with referential choice. In particular, the parameter of the syntactic/semantic role of the NP/referent in the current clause and the referential type of the antecedent do not correlate at all with the referent’s current pronominalisability. The latter parameter, for the sake of comparison, is included in Table 1. Obviously, it does not demonstrate any interesting co-occurrence with referential choice: 53% of *on*-pronouns have full NPs as their antecedents and 47% have reduced NPs as their antecedents.

Seven significant activation factors have been detected. Here is their list with the indication [in brackets] of the privileged feature co-occurring with pronominal reference: animacy [human], protagonisthood [yes], linear distance [1], rhetorical distance [1], paragraph distance [0], syntactic [subject] and semantic [Actor¹] roles of the antecedent, and sloppy identity².

¹ The term “Actor” is an abstract semantic macrorole; it designates the semantically central participant of a clause, with more-than-one-place verbs usually agent or experiencer; see e.g. Van Valin (1993:43ff).

² The factor of sloppy identity is what can be called a secondary or “weak” factor. It is relevant in far fewer cases than other factors. Sloppy identity slightly reduces activation when the

After the set of significant activation factors has been identified, certain numerical values have been assigned to their features. I postulated that the referent's AS can vary from 0 to 1. Each value of the activation factors' features is measured in tenths of 1. In each particular case all values of all involved factors can be summed and the resulting activation score is supposed to predict referential choice.

Table 2 below lists a selection of activation factors, each factor with the features it can accept and the corresponding numerical values. This set of seven factors proves necessary and sufficient for the prediction of referential choices in the sample discourse.

<i>Activation factor</i>	<i>Feature</i>	<i>Numerical activation value</i>
Rhetorical distance to the antecedent	1	0.7
	2	0.4
	3	0
	4+	-0.3
Paragraph distance to the antecedent	0	0
	1	-0.2
	2+	-0.4
Protagonisthood	Yes, and the current mention is: the 1st mention in a series	0.3
	the 2nd mention in a series	0.1
	otherwise	0
	No	0

Table 2: Examples of activation factors, their features, and numerical values

Table 2 contains three examples of activation factors and illustrates their main logical types. Some factors are sources of activation. The strongest among these is the factor of rhetorical distance to the antecedent. When the rhetorically previous mention is closer, the activation is higher. In the model of Kibrik (1996) this factor also works in the other direction: when RhD is 4 or more, activation is deducted.

The factor of paragraph distance is never a source of activation; vice versa, it is, so to speak, a penalizing factor. In the default situation, when the antecedent is in the same paragraph (paragraph distance = 0), this factor does not contribute to AS at all. When the antecedent is separated from the current point in discourse by one or more paragraph boundaries, the activation is lowered.

The third factor illustrated in Table 2, that of protagonisthood, has a still different logical structure. It can be called a compensating factor. It can only add activation, but does that in very special situations. When a referent is not a protagonist, this factor does not affect activation. When a referent is a protagonist, this factor can help to regain activation if it has lowered. The notion of "series" used in the definition means a sequence of consecutive discourse units, such that: 1) all of them mention the referent in question, and 2) this sequence is preceded by at least three consecutive discourse units not

referent in question and the antecedent's referent are not exactly identical.

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. When the activation is high anyway, this factor does not matter.

How were the numerical values such as those in Table 2 obtained? This was done through a heuristic procedure of trials and errors. I am not aware of any way to compute such values deductively. After a number of successive adjusting trials the numerical system turned out to predict a subset of referential choices correctly: reduced referential forms were getting ASs close to 1, and full NPs were getting ASs much closer to 0. When this was finally achieved, it fortunately turned out that all other occurrences of referential devices are properly predicted by this set of numerical values without any further adjustment. I interpret this fact as an evidence supporting the correctness of the developed system.

One more crucial point needs to be made about this model. When one observes actual referential choices in actual discourse, one can only see the ready results of referential device selection by the author – full NPs, pronouns, or zeroes. However, the real variety of devices is somewhat greater. It is important to distinguish between the **categorical** and **potentially alternating** referential choices. For example, the pronoun *on* in a certain context may be the only available option, while in another context it could be well replaced by an equally good referential option, say a full NP. These are two different classes of situations, and they correspond to two different levels of referent activation. The referential strategies formulated in Kibrik (1996) for Russian narrative discourse are based on this observation. Those referential strategies shown in Table 3 below represent the mapping of different AS levels onto possible referential choices.

Referential device:	Full NP only	Full NP most likely, <i>on/zero</i> unlikely	Either full NP or <i>on/zero</i>	<i>On/zero</i> only
AS:	0–0.3	0.4–0.6	0.7–0.9	1

Table 3: Referential strategies in Russian narrative discourse

As is clear from Table 3, there are different kinds of thresholds in this system. When the AS exceeds 0.9 or does not exceed 0.3 the referential choice is categorical. These are two hard thresholds. The third threshold, that of 0.6, is soft: both full NPs and reduced NPs are in principle possible below it and above it. Only the probability of their appearance is different.

A question may arise at this point: what governs the speaker's referential choice when the AS is within the interval of the activation scale that allows variable referential devices (especially 0.7 through 0.9)? I do not have a definitive answer to that question at this time. The choice may depend on idiolect, on discourse type and genre, or perhaps even be random. On the other hand, there may be some additional, extra-weak, factors that come into play in such situations.

The proposed model of referential choice can be summarized as follows. At each point of discourse for each referent the features of all relevant activation factors are identifiable for the speaker. If so, the numerical values

of those features are also available. Then the speaker can compute the integral AS of the referent at this point, and on the basis of the AS and the referential strategies select a referential device.

4. The English Study

The model developed for Russian narrative discourse was subsequently applied to a sample of English narrative discourse, which required a fair amount of modification. That study was described in Kibrik (1999), and here I report its main results and provide some additional details. The sample (or small corpus) was the children’s story “The Maggie B.” by Irene Haas. There are 117 discourse units in it. 76 different **referents** are mentioned in it, not counting 13 more mentioned in the quoted songs. There are 225 referent **mentions** in the discourse (not counting those in quoted text). There are 14 different referents mentioned in discourse that are **important** for this study. They are those mentioned at least once in a context where any degree of activation can be possibly expected. Among the important referents, there are three protagonist referents: “Margaret” (72 mentions altogether), “James” (28 mentions), and “the ship” (12 mentions). An excerpt from the sample discourse, namely lines 1401–2104, is given in Appendix 1 below.

Any referent, including an important referent, can be mentioned in different ways, some of which (for example, first person pronouns in quoted speech) are irrelevant for this study. Those that are **relevant** for this study fall into two large formal classes: references by full NPs and references by activation-based pronouns. By “activation-based pronouns” I mean the unmarked, general type of pronoun occurrences that cannot be accounted for by means of any kind of syntactic rules, in particular, for the simple reason that they often appear in a different sentence than their antecedents. In order to explain and predict this kind of pronoun occurrence, it is necessary to construct a system of the type described in section 3, taking into account a variety of factors related to discourse context and referent properties. Typical examples of activation-based pronouns are given in (4) below. (In the examples, as well as in Appendix 1, each line represents one discourse unit. In line numbers the first two digits refer to the paragraph number in the story, and the last two digits to the number of the discourse unit within the current paragraph.)

- (4) 1607 Lightning split the sky
 1608 as *she* ran into the cabin
 1609 and slammed the door against the wet wind.
 1610 Now everything was safe and secure.
 1601 When *she* lit the lamps,
 1602 the cabin was bright and warm.

There are two occurrences of the activation-based pronoun *she* in (4), and the second one is even used across the paragraph boundary from its antecedent.

The focus of this study is restricted to 39 full NP references and 40 activation-based pronominal references. As was pointed out in section 3 above, within each of the referential types – full NPs and pronouns – there is a crucial difference: whether the referential form in question has **an alternative**. In (5) below an illustration of a pronoun usage is given that can vary with a full NP: in

unit 1601 the full NP *Margaret* could well be used (especially provided that there is a paragraph boundary in front of unit 1601).

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

Contrariwise, there are instances of categorical pronouns. Consider (6), which is a direct continuation of (5):

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

In 1603, it would be impossible to use the full NP *Margaret*; only a pronoun is appropriate.

For the English data, it was found that referential forms of each type (for example, pronouns) fall into three categories: those allowing no alternative (= categorical), those allowing a questionable alternative, and those allowing a clear alternative. Thus there are six possible correspondences between the five potential types and two actual realizations; see Table 4.

Potential referential form	Full NP only	Full NP, ?pronoun	Full NP or pronoun	Pronoun, ?full NP	Pronoun only
Actual referential form	full NP			pronoun	

Table 4: Actual and potential referential forms

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

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

Second, referential choices were tested with a group of students, native speakers of English, through the following experimental study. The idea of the experiment was to modify the original referential forms, present it to a subject and see whether the subject identifies the replacement as a linguistic (or “stylistic”) error. Seven modified discourses were made up from the original discourse. All relevant references were subject to modification in this or that modified discourse. In each particular version of the discourse the adjacent changes never appeared closer than across a paragraph boundary,

and usually had at least two paragraph boundaries between them. Thus there were 8 different variants of the discourse (one original and seven modified ones). They were presented to 12 students who did the job of assessing the felicity of the discourse 20 times altogether.

In order to bring all these judgments together and build an integral judgment for each referential alternative, a system of weights was developed. Its details are reported in Kibrik (1999). At the end all weights of the judgments (including the expert's and the students') were summed together and averaged, and integral judgments on referential alternatives were obtained. Thus all referential alternatives were classified as either appropriate, questionable, or inappropriate.

The referential strategies I arrived at in this study are represented in Table 5. As in section 3, the referential strategies indicate the mappings of different intervals on the AS scale onto possible referential devices.

Referential device:	Full NP only	Full NP, ?pronoun	Either full NP or pronoun	Pronoun, ?full NP	Pronoun only
AS:	0–0.2	0.3–0.5	0.6–0.7	0.8–1.0	1.1+

Table 5: Referential strategies in English narrative discourse

The quantitative system in this study was designed so that AS can sometimes exceed 1 and reach the value of 1.1 or even 1.2. This is interpreted as “extremely high activation” (it gives the speaker no full NP option to mention the referent, see the value in the rightmost column of Table 5 and below). The AS of 1 is then interpreted as “normal maximal” activation. Also, a low AS frequently turns out to be negative. Such values are simply rounded to 0.

According to the referential strategies represented in Table 5, the five categories of potential referential forms correspond to five different intervals on the activation scale. There are four different thresholds in this system. The thresholds of 0.2 and 1.0 are hard: when the AS is 0.2 or less a pronoun cannot possibly be used, and when the AS is over 1.0 a full NP cannot be used. There are also two soft thresholds. When the AS is 0.5 or less a pronoun is unlikely, and when the AS is over 0.7 a full NP is unlikely.

The way in which five categories of potential referential forms are represented in the sample discourse in terms of frequency, is shown in Table 6.

Full NP only	Full NP, ?pronoun	Full NP or pronoun	Pronoun, ?full NP	Pronoun only
15	17	22 (including: 7 actual full NPs and 15 actual pronouns)	18	7
Total: actual full NPs – 39, actual pronouns – 40				

Table 6: Frequencies of potential referential forms in the sample discourse

The system of activation factors that was developed for the analysis of the sample discourse is presented in Table 7.

The first three activation factors listed in Table 7 relate to the distance from the point in question to the antecedent. The distance can be measured in three different ways that have been already explained in section 3 above. By far the most influential among the distance factors, and in fact among all activation factors, is the factor of **rhetorical distance**: it can add up to 0.7 to the activation score of a referent. **Linear and paragraph distances** can only penalize a referent for activation; this happens if the distance to the antecedent is too high.

To see how rhetorical (hierarchical) structure of discourse can be distinct from its linear structure, consider the rhetorical graph in Appendix 2. It depicts the rhetorical structure corresponding to lines 1801–2104 of the excerpt provided in Appendix 1. Rhetorical distance is counted as the number of horizontal steps required in order to reach the antecedent's discourse unit from the current discourse unit. For example, the pronoun *him* in discourse unit 1802 has its antecedent *James* in discourse unit 1801. Hence the rhetorical distance is 1. In narratives, the fundamental rhetorical relation is that of sequence. Three paragraphs of the four depicted in Appendix 2 (#18, #20, and #21) are connected by that relation, and within each of these paragraphs there are sequenced discourse units, too. If there were no other rhetorical relations in narrative besides sequence, rhetorical distance would always equal linear distance. However, this is not the case. In the example analysed, one paragraph, namely #19, is off the main narrative line. It provides the background scene against which the mainline events take place. Likewise, discourse unit 1904 reports a result of what is reported in 1903. The difference between the linear and the rhetorical distance can best be shown by the example of discourse unit 2001. For the referents “Margaret” and “James”, mentioned therein, the nearest antecedents are found in discourse unit 1802. It is easy to see that the linear distance from 2001 to 1802 is 6 (which is a very high distance) while the rhetorical distance is just 2 (first step: from 2001 to 1803, second step from 1803 to 1802).

Perhaps the most convincing examples of the power of rhetorical distance as a factor in referential choice are the cases of long quotations. Consider the following extract from the sample discourse.

- (7) 1201 After juice-and-cookie time, *she* gave James his counting lesson,
 1202 and this is how *she* did it.
 1203 *One, two, three, four, five, once I caught a fish alive,*
 1204 *six, seven, eight, nine, ten, but I let him go again.*
 1205 *Why did you let him go?*
 1206 *because he bit my finger so.*
 1207 *Which finger did he bite?*
 1208 *This little one upon the right.*
 1209 And *she* gave James' little finger a nibble <...>

Of interest here is the pronoun *she* in discourse unit 1209. If linear distance were the only kind of distance factor, most certainly under the distance of 7 back to the antecedent a pronominal reference in 1209 would not be possible. However, it is perfectly OK, and this is because the rhetorical distance to the antecedent in 1202 is just 1: the entire quotation in 1202–1208 is a satellite to 1201,

and 1209 is also directly connected to 1201 by the sequence relation.

<i>Activation factor</i>	<i>Feature</i>	<i>Numerical activation value</i>
Rhetorical distance to the antecedent (RhD)	1	0.7
	2	0.5
	3+	0
Linear distance to the antecedent (LinD)	1	0
	2	-0.1
	3	-0.2
	4+	-0.3
Paragraph distance to the antecedent (ParaD)	0	0
	1	-0.3
	2+	-0.5
Syntactic role of the linear antecedent	LinD ≥ 4	0
	LinD ≤ 3:	
	S of a main clause	0.4
	other active S	0.3
	DO, passive S, Pred	0.2
Other	0	
Animacy	Inanimate	0
	Animate, and	
	LinD ≤ 2	0
	LinD ≥ 3	
	animal	0.1
	human	0.2
Protagonisthood	No	0
	Yes, and	
	RhD+ParaD ≤ 2	0
	RhD+ParaD ≥ 3	0.2
Supercontiguity	No	0
	Yes	0.2
Temporal / spatial shift	No	0
	Yes	-0.2
Weak referent	No	0
	Yes	-0.2
Predictability	No	0
	Yes	0.1
Introductory antecedent	No	0
	Yes	-0.1

Table 7: Activation factors and their values, as identified for English narrative discourse

The following factor, indicated in Table 7, and the second most powerful source of activation, is the factor of **syntactic role of the linear antecedent** (note that one referent mention often has two distinct closest antecedents: a rhetorical and a linear one). The logical structure of this factor is quite complex. First, it applies only when the linear distance is short enough: after about four discourse units it gets forgotten what the role of the antecedent was; only the fact of its presence may still be relevant. Second, this factor has a fairly diverse set of features. As has long been known from studies of syntactic anaphora, subject is the best candidate for the pronoun's antecedent. Different subtypes of subjects, though, make different contributions, ranging from 0.4 to 0.2. Other relevant features of the factor include the direct object and the nominal part of the predicate. It is very

typical for pronouns, especially for categorical pronouns (allowing no full NP alternative) to have subjects as their antecedents. For example, consider three pronouns in paragraph #16 (see Appendix 16): *she* (discourse unit 1603), *her* (1606), and *she* (1608). According to the results of the experimental study reported above, the first and the second pronouns are categorical (that is, *Margaret* and *Margaret's* could not be used instead) and they have subject antecedents. But the third one has a non-subject antecedent, and it immediately becomes a potentially alternating pronoun (*Margaret* would be perfectly appropriate here)³.

The following two factors are related not to the previous discourse but to the relatively stable properties of the referent in question. **Animacy** specifies the permanent characterization of the referent on the scale "human – animal – inanimate". **Protagonisthood** specifies whether the referent is the main character of the discourse. Protagonisthood and animacy are rate-of-deactivation compensating factors (see discussion in section 3). They capture the observation that important discourse referents and human referents deactivate slower than those referents that are neither important nor human. In the formulation presented in Table 7, protagonisthood is connected with the rhetorical and paragraph distance: when these two together are high enough, a protagonist referent gains some extra activation; when they are not, protagonisthood does not matter. Animacy is connected here with the linear structure of discourse: under high linear distance human referents deactivate less than other referents.

The final group is second-order, or "exotic", factors, including the following ones. **Supercontiguity** comes into play when the antecedent and the discourse point in question are in some way extraordinarily close (e.g. being contiguous words or being in one clause). **Temporal or spatial shift** is similar to paragraph boundary but is a weaker episodic boundary; for example, occurrence of the clause-initial *then* frequently implies that the moments of time reported in two consecutive clauses are distinct, in some way separated from each other rather than flowing one from the other. **Weak referents** are those that are not likely to be maintained, they are mentioned only occasionally. Such referents often appear without articles (cf. NPs *rain*, *cinnamon and honey*, *supper* in the text excerpt given in Appendix 1) or are parts of stable collocations designating stereotypical activities (*slam the door*, *light the lamps*, *give a bath*). **Predictability** is a relation of the current discourse unit to the preceding, such that it can be predicted that a certain referent must be mentioned at this point. Finally, **introductory antecedent** means that when a referent is first introduced into discourse it takes no less than two mentions to fully activate it.

As in case of the Russian study, the numerical activation values of each feature cited in Table 7 were obtained through a long heuristic "trial-and-error" procedure, performed in cycles until the whole array of the data was explained. It should be emphasized that all

³ This demonstration of one factor operating in isolation is not intended to be conclusive, since the essence of the present approach is the idea that all factors operate *in conjunction*. It does, however, serve to illustrate the point.

referential facts contained in the original discourse and obtained through experimentation with modified discourses, are indeed predicted/explained by the combination of activation factors with their numerical values, and the referential strategies.

To demonstrate how predictively the calculative system of activation factors works, I present below several examples of actual calculations. All examples are taken from the text excerpt given in Appendix 1. Examples are different in that they pertain to different referential options possible on the AS scale (see Table 5 above). There is one example for each of the following referential options: (a) full NP, ?pronoun; (b) either full NP or pronoun; (c) pronoun, ?full NP; (d) pronoun only. The calculations are summarized in Table 8 below. The

upper portion of Table 8 contains a characterisation of each example: its location in the text, the actual referential form used by the author, the referent, the type of referential device and possible alternative devices, as obtained through the experimental study described above. Also, the AS interval corresponding to the referential option in question is indicated, in accordance with the referential strategies given in Table 5 above. The middle portion of Table 8 demonstrates the full procedure of calculating the ASs, in accordance with the numerical values given in Table 7 above. The last line of Table 8 indicates whether the calculated AS fits within the range predicted by the referential strategies.

Referential option	(a) Full NP, ?pronoun	(b) Full NP or pronoun	(c) Pronoun, ?full NP	(d) Pronoun only
Line number	1802	1701	1802	1603
Referential form	<u>Margaret</u>	<u>She</u>	<u>him</u>	<u>she</u>
Referent	“Margaret”	“Margaret”	“James”	“Margaret”
Chosen referential device	full NP	pronoun	pronoun	pronoun
Alternative referential device	?pronoun	full NP	?full NP	—
Corresponding AS interval	0.3–0.5	0.6–0.7	0.8–1.0	1+
Relevant activation factors				
RhD	FEATURE: 3 NUM. VALUE: 0	2 0.5	1 0.7	1 0.7
LinD	FEATURE: 3 NUM. VALUE: -0.2	2 -0.1	1 0	1 0
ParaD	FEATURE: 1 NUM. VALUE: -0.3	1 -0.3	0 0	0 0
Lin. antecedent role	FEATURE: S NUM. VALUE: 0.4	S 0.4	passive S 0.2	S 0.4
Animacy	FEATURE: Human, LinD ≥ 3 NUM. VALUE: 0.2	Human, LinD ≤ 2 0	Human, LinD ≤ 2 0	Human, LinD ≤ 2 0
Protagonisthood	FEATURE: Yes, RhD+ParaD ≥ 3 NUM. VALUE: 0.2	Yes, RhD+ParaD ≥ 3 0.2	Yes, RhD+ParaD ≤ 2 0	Yes, RhD+ParaD ≤ 2 0
Calculated AS	0.3	0.7	0.9	1.1
Fit within the predicted AS interval	Yes	Yes	Yes	Yes

Table 8: Examples of calculating the referents’ ASs in comparison with the predictions of the referential strategies

5. Consequences for Working Memory

The studies outlined in sections 3 and 4 above rely on work in cognitive psychology, but they are still purely linguistic studies aiming at explanation of phenomena observed in natural discourse. However, it turns out that the results of those studies are significant for a broader field of cognitive science, specifically for research in working memory (WM).

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

- CAPACITY: how much information can there be in WM at one time?

- CONTROL: what is the mechanism through which information enters WM?
- FORGETTING: what is the mechanism through which information quits WM?

The linguistic study of referential choice sheds light on these issues, at least in respect to the portion of WM dealing with specific referents. Here I will only mention some results related to the issues of capacity and control. For more details refer to Kibrik (1999).

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

author chose to mention “Margaret” in that unit or not. If so, one can calculate the activation of **all referents** at a given point in discourse. Consider discourse unit 1608 (see Appendix 1). Only two referents are mentioned there: “Margaret” and “the cabin”. However, the following other referents have an AS greater than 0 at this point: “the anchor”, “the gear”, “rain”, “the deck”, “thunder”, “lightning”, and “the sky”. The sum of ASs of all relevant referents gives rise to **grand activation** – the summary

activation of all referents at the given point in discourse. Grand activation gives us an estimate of the **capacity** of the specific-referents portion of WM. Figure 1 below depicts the dynamics of activation processes in a portion of the English discourse (lines 1401 through 2104, see Appendix 1). There are three curves in Figure 1: two pertaining to the activation of the protagonists “Margaret” and “James”, and the third representing the changes in grand activation.

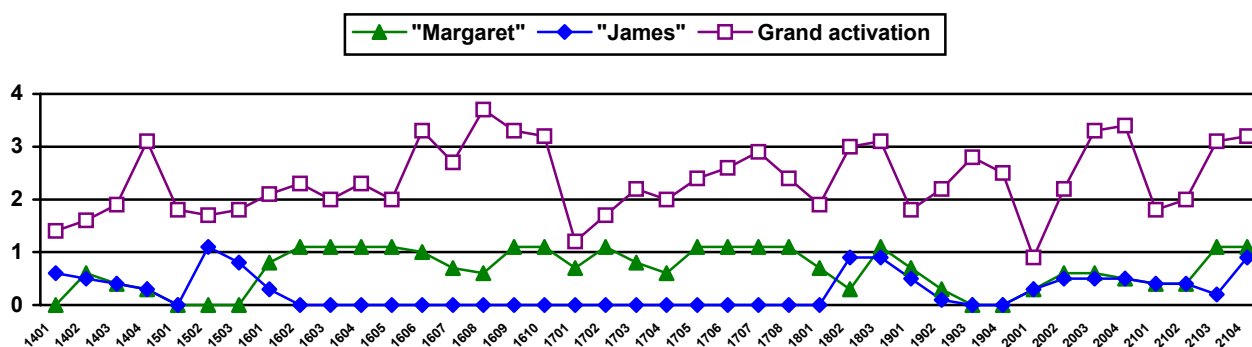


Figure 1: The dynamics of two protagonist referents’ activation and of grand activation in an excerpt of English narrative (given in Appendix 1)

Observations of the data in Figure 1 make it possible to arrive at several important generalisations. Grand activation varies normally within the range between 1 and 3, only rarely going beyond this range. Thus the variation of grand activation is very moderate, the ratio between its maximal and minimal values being only about 3. (It is of course infinite for individual referents.) Given that the numerical setup of the calculative system described above is such that the maximal activation of an individual referent can reach 1.1 or even 1.2, then the maximal grand activation exceeds this number about three times. This gives us an estimate of the maximal capacity of WM related to specific referents in discourse: three fully activated referents. (This maximum is rarely reached though.) Furthermore, there are strong shifts of grand activation at paragraph boundaries; even a visual examination of the graph in Figure 1 demonstrates that grand activation values at the beginnings of all paragraphs are local minimums; almost all of them are below 2. On the other hand, in the middle or at the end of paragraphs grand activation usually has local maximums. Apparently one of the cognitive functions of a paragraph is a threshold of activation update⁴.

The question of **control** of WM is the question of how information comes into WM. The current cognitive literature connects attention and WM. The mechanism

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

At the same time, as has been convincingly demonstrated in the experimental study by Tomlin (1994), attention has a linguistic manifestation, namely grammatical roles. Focal attention in many languages, including English, is consistently coded by speakers as the subject of the clause. As has been demonstrated in the present paper, subjecthood and reduced forms of reference are causally related: antecedent subjecthood is among the most powerful factors leading to the selection of a reduced form of reference. In both English and Russian, antecedent subjecthood can add up to 0.4 to the overall activation of a referent. In both English and Russian sample discourses, 86% of pronouns allowing no referential alternative have subjects as their antecedent.

Considered together, these several sets of facts from cognitive psychology and linguistics lead one to a remarkably coherent picture of an interplay between attention and WM, both at the linguistic and at the cognitive level. Attention feeds WM, i.e. what is attended at moment t_n becomes activated in WM at moment t_{n+1} . Linguistic moments are discourse units. Focally attended referents are coded by subjects; at the next moment they become activated (even if they were not before) and are coded by reduced NPs. The relationships between attention and WM, and between their linguistic manifestations, are represented in Table 9.

Moments of time (discourse units)	t_n	t_{n+1}
-----------------------------------	-------	-----------

⁴ Of course, a drop in grand activation at paragraph boundaries is predetermined by the fact that paragraph boundary is a strong activation decreasing factor: each referent is deactivated after a paragraph boundary, and, therefore, the sum of particular ASs necessarily goes down. However, grand activation drop is not a mere artefact of the present approach. The deactivational effect of a paragraph boundary is an inherent fact that needs to be accounted for by any theory of reference in discourse. The observation of grand activation drop is a direct consequence of that inherent fact.

Cognitive phenomenon	focal attention	high activation
Linguistic reflection	mention in the subject position	reduced NP reference
Examples	<u>M</u> argaret, <u>s</u> he	<i>she, her</i>

Table 9: Attention and working memory in cognition and in discourse

6. Conclusion

In this paper I hope to have demonstrated that a predictive and explanatory model of referential choice in discourse is possible. The approach outlined above aims to predict and explain all referential occurrences in the sample discourse. This is done through a rigorous calculative methodology allowing for no exceptions. For each referent at any point in discourse, the numerical values of all involved activation factors can be objectively and publicly verified. The objective fluidity of the process of referential choice is addressed through the distinction between the categorical and potentially alternating referential devices.

The linguistic study of referential choice in discourse was based on cognitive-psychological research, and it proved, in its turn, relevant for the study of cognitive phenomena in a more general perspective.

7. References

- Anderson, J.R., 1990. *Cognitive Psychology and its Implications*. 3d ed. New York: W. H. Freeman and Company.
- Baddeley, A., 1986. *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A., 1990. *Human Memory: Theory and Practice*. Needham Heights, Mass: Allyn and Bacon.
- Botley, S., and A.M. McEnery (eds.), 2000. *Corpus-based and Computational Approaches to Discourse Anaphora*. Amsterdam and Philadelphia: John Benjamins.
- Chafe, W., 1994. *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Clifton, C. Jr. and F. Ferreira, 1987. Discourse structure and anaphora: Some experimental results. In M. Coltheart (ed.), *Attention and Performance XII*. Hove: Erlbaum.
- Cowan, N., 1995. *Attention and Memory: An Integrated Framework*. New York – Oxford: Oxford University Press.
- Fox, B., 1987a. *Discourse Structure and Anaphora*. Cambridge: Cambridge University Press.
- Fox, B., 1987b. Anaphora in popular written English narratives. In R. Tomlin (ed.), *Coherence and Grounding in Discourse*. Amsterdam and Philadelphia: John Benjamins.
- Gathercole, S.E. (ed.), 1996. *Models of Short-term Memory*. Hove, East Sussex: Psychology Press.
- Gernsbacher, M. A., 1990. *Language Comprehension as Structure Building*. Hillsdale, NJ: Erlbaum.
- Givón, T. (ed.), 1983. *Topic Continuity in Discourse. A Quantified Cross-language Study*. Amsterdam and Philadelphia: John Benjamins.
- Givón, T., 1990. *Syntax: A Functional-typological Introduction. Vol. 2*. Amsterdam and Philadelphia: John Benjamins.
- Givón, T., 1995. *Functionalism and Grammar*. Amsterdam and Philadelphia: John Benjamins.
- Gordon, P.C., B.J. Grosz and L.A. Gilliom, 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17: 311–347.
- Grimes, J., 1978. *Papers in Discourse*. Arlington: SIL.
- Gundel, J.K., N. Hedberg and R. Zacharski, 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69: 274–307.
- Kibrik, A.A., 1991. Maintenance of reference in sentence and discourse. In W.P. Lehmann and H.J. Hewitt (eds.), *Language Typology*. Amsterdam and Philadelphia: John Benjamins.
- Kibrik, A.A., 1996. Anaphora in Russian narrative prose: A cognitive account. In B. Fox (ed.), *Studies in Anaphora*. Amsterdam and Philadelphia: John Benjamins.
- Kibrik, A.A., 1999. Reference and working memory: Cognitive inferences from discourse observation. In K. van Hoek, A.A. Kibrik, and L. Noordman (eds.), *Discourse Studies in Cognitive Linguistics*. Amsterdam and Philadelphia: John Benjamins.
- Mann, W., C. Matthiessen, and S.A. Thompson, 1992. Rhetorical structure theory and text analysis. In W. Mann and S. Thompson (eds.), *Discourse Description. Diverse Linguistic Analyses of a Fund-raising Text*. Amsterdam and Philadelphia: John Benjamins.
- Marslen-Wilson, W., E. Levy and L.K. Tyler, 1982. Producing interpretable discourse: The establishment and maintenance of reference. In R.J. Jarvella and W. Klein (eds.), *Speech, Place, and Action. Studies in Deixis and Related Topics*. Chichester: Wiley.
- Posner, M.I. and M.E. Raichle, 1994. *Images of Mind*. New York: Scientific American Library.
- Smith, E.E. and J. Jonides, 1997. Working memory: A view from neuroimaging. *Cognitive Psychology* 33: 5–42.
- Tomlin, R., 1987. Linguistic reflections of cognitive events. In R. Tomlin (ed.), *Coherence and Grounding in Discourse*. Amsterdam and Philadelphia: John Benjamins.
- Tomlin, R., 1994. Focal attention, voice and word order: An experimental cross-linguistic study. In P. Downing and M. Noonan (eds.) *Word Order in Discourse*. Amsterdam and Philadelphia: John Benjamins.
- Tomlin, R. and M. Pu, 1991. The management of reference in Mandarin discourse. *Cognitive Linguistics* 2: 65–93.
- van Hoek, K., 1997. *Anaphora and Conceptual Structure*. Chicago: University of Chicago Press.
- Van Valin, R.D. Jr., 1993 A synopsis of Role and Reference grammar. In R.D. Van Valin, Jr. (ed.) *Advances in Role and Reference Grammar*. Amsterdam: Benjamins.
- Vonk, W., L.G.M.M. Hustinx and W.H.G Simons, 1992. The use of referential expressions in structuring discourse. *Language and Cognitive Processes* 7: 301–333.

Acknowledgements

I express my gratitude to the Max Planck Institute for Evolutionary Anthropology and Bernard Comrie, the director of its Linguistics department, for their generous help that facilitated the writing of this paper and made my participation in DAARC 2000 possible. The research into the rhetorical structure of discourse in its connection with referential choice was supported by grant 2055/459/1999 from the Research Support Scheme (Soros Foundation). I would like to thank all colleagues who assisted me in the studies reported here at various stages, especially Russ Tomlin and Gwen Frishkoff.

Appendix 1: An Excerpt from an English Narrative ("The Maggie B." by Irene Haas)

1401 Margaret and James were cold.
 1402 The sky grew darker.
 1403 The goat and chickens fled into their little shelter,
 1404 the toucan flew screeching into the cabin.
 1501 James started to cry.
 1502 A storm was coming!
 1503 Margaret must make the boat ready at once.
 1601 She took in the sail
 1602 and tied it tight.
 1603 She dropped the anchor
 1604 and stowed all the gear,
 1605 while rain drummed on the deck
 1606 and thunder rumbled above her.
 1607 Lightning split the sky
 1608 as she ran into the cabin

1609 and slammed the door against the wet wind.
 1610 Now everything was safe and secure.
 1701 When she lit the lamps,
 1702 the cabin was bright and warm.
 1703 It was nearly suppertime
 1704 so Margaret mixed up a batch of muffins
 1705 and slid them into the oven.
 1706 She sliced some peaches
 1707 and put cinnamon and honey on top,
 1708 and they went into the oven, too.
 1801 James was given a splashy bath in the sink.
 1802 Margaret dried him in a big, warm towel,
 1803 and then supper was ready.
 1901 Outside, the wind howled like a pack of hungry wolves.
 1902 Rain lashed the windowpanes.
 1903 But the sturdy little Maggie B. kept her balance
 1904 and only rocked the nicest little bit.
 2001 Margaret and James ate the beautiful sea stew
 2002 and dunked their muffins in the broth,
 2003 which tasted of all the good things that had
 cooked in it.
 2004 For dessert they had the peaches with cinnamon
 and honey, and glasses of warm goat's milk.
 2101 When supper was over,
 2102 Margaret played old tunes on her fiddle.
 2103 Then she rocked James in his cradle
 2104 and sang him his favorite song.

Appendix 2: An Example of a Rhetorical Graph (lines 1801–2104 of the excerpt given in Appendix 1)

