Andrej Bekeš
Ljubljana

MEASURES OF TOPIC CONTINUITY AND THE WA-TOPIC IN JAPANESE

In this paper I examine three statistical measures of topic continuity, i.e., Topic Quotient (TQ), Referential Distance (RD) and Topic Persistence (TP), using the text of a short novel, Rashomon by Ryunosuke Akutagawa. It turns out that these measures are very unreliable as predictors of the WA-topic in Japanese. Even worse, in the case of TP, and for different referents, contradictory results were obtained. At closer inspection it turns out that this is due to the differences in status which referents possess within some segment of a text. What matters is not the numerical frequency of a referent, but its status, i.e., whether it referrs to a topic entity, or, from the expression point of view, to a topic chain of referential forms within the text, or not.

Key words*: topic, topic continuity, Topic Quotient, Referential Distance, Topic Persistence, coreference*

0. Introduction

Strict definition of functional notions in linguistics is a never-ending uphill battle with the menace of circularity, topic not being an exception. The following quotation (Lambrecht 1994:131) shows, how intuitively clear the notion of topic is:

> A referent is interpreted as the topic of a proposition if in a given situation the proposition is constructed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent.

The problem is that this and other definitions of topic do not provide sufficient means to identify particular topics in actual propositions. The reason is that topic is a functional notion and definitions of functional notions tend to be circular: form and function are difficult if not impossible to separate.

Topic (in the Prague tradition 'theme') has been studied from two different perspectives. The first is the treatment of topic on the level of a single sentence. This involves discovering syntactical properties of sentences with a topic (in the case of Japanese, Mikami 1953, 1960; Kuno 1973; Kuroda 1972). The second approach tries to catch the regularities involving topics from the wider perspective of text and context. This approach lays stress on the functional aspects of topic as seen from the point of view of text and context and is concerned with text-pragmatic issues and conditions for topicalization. Pioneer work within this approach has been done by Yamada (1908), Matsushita (1928, reprinted 1978) for Japanese and by Danep 1974, Chafe 1976, Givón 1979, 1983, 1987 from a more general linguistic perspective.

The first perspective, which I will call the local approach, usually postulates the topic as a sentence element accompanied by a topic marker, for example in Japanese,

a noun phrase marked with the so called topic particle WA (see for example Noda 1994).

The second, the global approach, on the other hand is preoccupied with identifying topic elements in their contexts and with marking of such elements.

To provide an independent criterion for "topicness" of elements, a number of empirical parameters have been proposed, such as *topic quotient (TQ)*, *referential distance (RD)*, and *topic persistence (TP)* etc (Givón 1983, 1989, Myhill, J. 1992).

The purpose of this study is to 1) examine the validity of the aforementioned statistical measures, proposed to overcome the inherent circularity of the definition of a topic and 2) to find a possible reason for their inherent insufficiencies by relating the notion of topic to the way the narrative is structured around different referents.

An earlier version of this study was presented at the JLAO workshop, EHESS, Paris, May 15-16, 2000.

1. Topic in Japanese

In this section, a short sketch of topic in Japanese will be given. The most common topic marker is particle *wa*, though other markers, such as *nanka*, *nara*, *toieba*, etc, are used in particular contexts as well. I will limit my short introduction to topics marked with wa. Syntactically, a sentence with a topic is seen as having the following structure (Shibatani, 1978):

[TOPIC] [PROPOSITION]

There was a long discussion concerning the question whether there are cases with topic being syntactically incorporated into proposition, but pragmatic considerations point towards the above structure as being generally valid for such sentences (Shibatani, ibid.).

(1) *Basu   ga   ki-   ta.*   [sentence without a topic (*mudai bun*)]
    bus   NOM.   come-PAST.
    The bus came

(2)a *Watasi WA   Tanaka   desu*   [sentence with a topic (*yuudai bun*)]
    I   TOPIC T.(name) copula-PRESENT-FORMAL.
    I am Tanaka.

  b [context] *Atarasii gakka ga dekita* (A new department has been opened.)
    Gakkatyoo WA   Tanaka   sensei   da.
    Head   TOPIC T. (name) professor.   copula-PRESENT-INFORMAL.
    The head is Prof. Tanaka.

  c *Kuzira WA   honyurui   da.*
    whale TOPIC mammal   copula-PRESENT-INFORMAL.
    Whales are mammals

Only elements accessible in their context can become topics, e.g.:

a) those accessible directly in the immediate context of communication (ex. 2a);
b) those accessible from the context of communication on the basis of our general knowledge of the world (ex. 2b);
c) generic assertions, not depending on any particular situation (ex. 2c);

2. Methodological remarks

2.1 Choice of data

Empirical analysis is based on the short novel *Rashomon* (R. Akutagawa, computer readable version from "*Aozora bunko*" (http://www.voyager.co.jp/aozora/, approx. 6750 characters). The choice was motivated with the rich narrative structure of this work, with many animate (human) and inanimate referents appearing throughout the text. The human referents are: Servant, Old woman, Corpse(s) (referred to when they were still alive), Woman, Crow(s), and Author. The most frequently mentioned inanimate referents are: the gate Rashomon, Rain, Hair (of the dead people), Fire, Kyoto, Twilight, Furuncle (on servants cheek), and Cricket. The ease of accessing the computer-readable version can stimulate further analysis of this work.

2.2 Parsing the text into clauses

Statistical measures examined in the next section are based on how coreferential forms appear in subsequent clauses of a text. It has been shown that clause in spoken language is a primary phenomenon, related to human cognitive capabilities (Chafe 1980, 1987 etc.). As argued in Bekes (1987, 1994), clause can also be validly considered in the same way in written language in spite of the differences in its production.

The question of what to consider as a clause and how to handle discontinuous topics in Japanese, with its rich system of modal suffixing on the predicate, posed some problems. I defined clause boundaries, basing my decisions on Minami (1974) as described in Bekep (1994).

I treated discontinuous topics or topics, which are, shared by several clauses as separate units. An example of a discontinuous topic is given in (3).

(3)a *Yuuzin wa*
    Friend TOPIC.

  b *"denwa no koe wa toku ni kawatta yoosu ga nakatta"*
    "there was nothing weird about his voice on the phone"

  c *to iu.*
    [he] says.

Coreferential noun phrases, both modified and unmodified, often appeared in shorter forms in the coreferential chain. Since the main goal of this study is to shed light on

the topicalization of coreferential noun phrases and not on the referential form itself, I considered any noun phrase of this sort as an instance of the same NP.

In the text of "Rashomon", a total of 495 clauses were found.

3. Statistical measures of "topicness"

3.0 Introduction

To circumvent the built-in circularity of functional notions, various statistical measures have been proposed for measuring the "topicness" of the referents. The best overall presentation of methodologies involving such measures is given in Myhill (1992).

In the following subsections I will examine the three most common statistical measures of "topicness", i.e., topic quotient (TQ), referential distance (RD) and topic persistence (TP) and their relation to various referential forms.

The main purpose of this study is to examine the predictive power of the above statistical measures for the appearance of the canonical topic in Japanese, i.e. noun phrase + topic particle *wa* (N+WA). Other important referential forms are zero anaphora (Ø) and full noun phrase accompanied by a case particle (N+P). Thus, the referential forms considered in this study are:

 a) noun phrases, including full reference to the referent + case particle (N+P)
 b) noun phrases, including full reference to the referent + topic particle *wa* (N+WA)
 c) zero anaphora (Ø)
 d) other forms, involving other particles, such as *mo* (also) etc.

The three statistical measures will be examined as both necessary and sufficient predictors for a "noun phrase+*wa*" and other referential forms. Here, "necessary" means that whenever a certain referential form is attested, there is a high probability that a given statistical measure is within a certain range of values, either high or low. On the other hand, "sufficient" means that a given statistical measure being within a certain range of values, either high or low, implies with high probability a certain choice of referential form.

3.1 Topic quotient (TO)

TQ is supposed to measure "topicness" of a referent over the whole text. It is measured as the proportion of clauses in a given text, referring to the given referent, i.e., the number of clauses referring to the given referent (frequency) divided by the total number of clauses in a text (Myhill 1992). The relation between TO and referential forms of various referents is shown in TABLE 1.

It can be seen from the table that different referents appear in the text with different frequencies. TQ does not seem to predict anything. The only correlation easily observed is the correlation between +animate (human) referents and high frequency of

344

zero anaphora on one hand, and inanimate referents and high frequency of nontopical full referential forms on the other. This simply reflects the fact that the narration in "Rashomon" is built around the actions of human referents.

TABLE 1: TOPIC QUOTIENT (R. Akutagawa: "Rashomon")

Referent	Frequency	TQ (F/495)	-------------------- REFERENTIAL FORMS --------------------			
			N+P	N+WA	Ø	Other
+animate (human)						
Servant	211	0.43	0.12 (26/211)	0.19 (42/211)	0.66 (140/211)	0.03 (3/211)
Old woman	115	0.23	0.29 (33/115)	0.10 (12/115)	0.57 (65/115)	0.03 (2/115)
Corpse	39	0.08	0.44 (17/39)	0.05 (2/39)	0.46 (18/39)	0.05 (2/39)
Woman	17	0.03	0.17 (3/17)	0.12 (2/17)	0.71 (12/17)	0
Crow	12	0.02	0.33 (4/12)	0.09 (1/12)	0.58 (7/12)	0
Author	7	0.01	0	0.14 (1/7)	0.86 (6/7)	0
-animate						
Rashomon	55	0.11	0.75 (41/55)	0.07 (4/55)	0.16 (9/55)	0.02 (1/55)
Rain	17	0.03	0.59 (10/17)	0.12 (2/17)	0.29 (5/17)	0
Hair	14	0.03	0.72 (10/14)	0.14 (2/14)	0.14 (2/14)	0
Fire	13	0.03	0.61 (8/13)	0.08 (1/13)	0.31 (4/13)	0
Kyoto	11	0.02	0.36 (4/11)	0.36 (4/11)	0.28 (3/11)	0
Twilight	7	0.01	0.86 (6/7)	0.14 (1/7)	0	0
Furuncle	4	0.01	1.00 (4/4)	0	0	0
Cricket	2	0.004	0.50 (1/2)	0	0	0.50 (1/2)

Only the highest TQ (the case of referent Servant) seems to predict necessarily and sufficiently slightly higher relative frequency of *wa* topics.

TQ is basically a rough measure of referent anaphoricity and thus its accessibility. TQ might somehow work in analyses of shorter segments of text, centered around just a few referents (see Bekes 1995). Yet, as we can see from the above table, it may safely be concluded that in any longer text, such as "Rashomon", TQ is a very unreliable predictor of wa topics or any other referential form.

3.2 Referential distance RD

I calculated referential distance by finding the most recent previous mention of the referent of the NP and then counting how many clauses back it occurred (Myhill 1992: 34). Contrary to standard practice, I included relative clauses in the count of intervening clauses because they seemed to be as demanding for processing as other subordinate clauses.

There is a problem of how to treat RD in the case of discontinuous topics and topics extending over several clauses. Discontinuous topics are common in reported speech, where the reported part can be of any length, as in example (3) in section 2. Here, *Yuuzin* (friend) in (3)a is the topic of the clause consisting of the "[reported part] + *to iu*". In counting the intervening clauses belonging to the reported part I treated the

topic separately. Thus I analyzed (3) as three units, counting the topic element *yuuzin wa* as a separate unit. However, since *yuuzin wa* is not a clause, I did not include it in the count of intervening clauses. This made it easier to count the RD between the topic element and the coreferent in preceding clauses as well as the RD between a referent in the last clause with a discontinuous topic and coreferents in the clauses that follow. Thus, I counted the RD between (3)c and (3)a as 1, while I counted RD=0 (co-occurrence in the same clause) for the case where there were no intervening clauses conveying the content of the reported speech, such as *Taroo wa iu* (Taro says).

I used a similar reasoning in my treatment of topics extending over several clauses, such as shown in (4) below. The only difference was that the intervening clauses share the topic element.

(4)a *Yukisan wa*
     Y.   TOP.

  b *[ø] tooka gogo sitizi han koro ni zitaku ni denwa o ireta no o saigo ni*
     telephoning home for the last time on the tenth at about half pastseven

  c *[ø] syoosoku o tatta*
     stopped informing [about herself]

Again, I did not include the topic unit [i.e., (4)a] in the count of intervening clauses. In cases such as (4), the RD from the ellipted topic in (4)b to the topic (4)a was counted as 0 and in the clause(s) following the clause, adjacent to the topic element [i.e., (4)c], RD was 1.

Referential distance is a rough measure of referent anaphoricity and thus of its accessibility. The largest meaningful distinction of RD is up to 20 clauses. Occurrences of a referent beyond 20 clauses are treated as having the referent as inactive as a reference 20 clauses away (Givón 1983, 1989).

To get meaningful data including a variety of RD only the most frequently appearing referents were chosen, i.e. Servant and Old woman among human and "Rashomon", "Hair" and "Rain" among inanimate referents.

TABLE 2: REFERENTIAL DISTANCE

• SERVANT				• OLD WOMAN				• INANIMATE*			
• RD	N+P	N+WA	Ø	• RD	N+P	N+WA	Ø	• RD	N+P	N+WA	Ø
1-4	16	29	131**	1-4	23	10	61	1-4	28	5	13
4-20	8	12	6	5-20	15	4	1	5-20	36	6	2

* = "Rashomon", "Hair" and "Rain"
** numbers of cases exceeding half of the total in each category are underlined.

Table 2 shows RD data for each referent or group of referents split into two groups. Shorter RDs, from 1 to 4, being one group, and RDs, equal to or longer than 5, being the other group [1]. The property of RD as a measure of accessibility is reflected in the correlation of short RDs with zero anaphora in the case of both human referents. In the

case of inanimate referents, even short RDs co-occur most often with full nontopical noun phrases. With longer RDs, less frequent referents, such as Old woman and inanimate referents, also correlate with full nontopical noun phrases. Either way, there seems to be no pattern correlating topical noun phrases and RD. In the case of Servant, topical noun phrases even become the relatively most frequent referential form for long RDs.

3.3 Topic persistence (TP)

Topic persistence is a measure of importance of a referent in the context that is following some particular occurrence. It is defined as the number of clauses referring to the referent within 10 clauses following the particular occurrence (Givón 1983, 1989).

TP is similar to TQ in that it is supposed to roughly reflect the referent's "topicness" in its context. The more topic-like the referent is, the more likely it is that it will continue to be referred to in the context, following some particular occurrence.

I have examined the distribution of referential forms in relation to topic persistence for the two main human referents in the story, the Servant (references in 211 clauses) and the Old woman (references in 115 clauses). The number of references being quite high, much higher than what we find in short conversations and newspaper articles, it is reasonable to expect that the trends, supposedly predicted by TP, would emerge quite clearly.

TABLE 3a: TOPIC PERSISTENCE (SERVANT - 211 references)

TP	N+WA	Ø	N+P	OTHER	TOTAL
0-3	0.15 (8)	0.70 (36)	0.15 (8)	0.00 (0)	52
4-6	0.24 (15)	0.61 (38)	0.15 (9)	0.00 (0)	62
7-10	0.20 (19)	0.68 (66)	0.09 (9)	0.03 (3)	97

TABLE 3b: TOPIC PERSISTENCE (OLD WOMAN - 115 references)

TP	N+WA	Ø	N+P	OTHER	TOTAL
0-3	0.06 (2)	0.63 (22)	0.31 (11)	0.00 (0)	35
4-6	0.13 (5)	0.57 (23)	0.25 (10)	0.05 (2)	40
7-10	0.13 (5)	0.50 (20)	0.37 (15)	0.00 (0)	40

Let us have a look at Table 3 (a, b) above. This table shows the referential form as a function of topic persistence in the case of the coreferential chains of the Servant and the Old woman. In the case of Servant (Table 3a), N+WA (noun phrase + particle WA), the form that is supposed to indicate the topic is exhibiting a rather modest relative frequency. The relative frequency, though increasing toward mid range of TP, is then slightly decreasing again. On the other hand, in the case of Old woman (Table 3b), the relative frequencies of N+WA are even lower, though increasing with higher TP.

Ellipsis shows a more or less permanent relative frequency for both referents, in the range between 0.50 and 0.70. The reason why ellipsis is so frequent is that, most often,

after the referent is introduced in full form, either as N+WA or N+NP, it is then repeated in ellipted form. Within one sentence, this is more or less obligatory, but it happens also across sentence boundaries. The behavior of N+P is even more interesting. In the case of Servant its relative frequencies are decreasing with higher TP values, as TP is supposed to predict. On the other hand, in the case of the Old woman, the relative frequencies are **increasing** as the TP gets to the highest range. The absolute frequency of such cases increases, too. This is exactly the contrary of what the TP is supposed to predict.

Since the topics have higher TP, it is natural to suppose that a form such as N+WA, if being marked for "topicness", should systematically exhibit stronger affinity with higher TP. Let us have a look at Table 4 (a, b), with TP expressed as a function of referential form.

TABLE 4a: TOPIC PERSISTENCE (SERVANT - 211 references)

	TP					
	0-3	4-6	7-10	TOTAL	Av. TP	MD
N+WA	0.19 (8)	0.36 (15)	0.45 (19)	1.00 (42)	5.93	6.0
Ø	0.26 (36)	0.27 (38)	0.47 (66)	1.00 (140)	5.56	6.0
N+P	0.30 (8)	0.35 (9)	0.35 (9)	1.00 (26)	4.65	5.0
OTHER	0	0	1.00 (3)	1.00 (3)	9.00	

TABLE 4b: TOPIC PERSISTENCE (OLD WOMAN - 115 references)

	TP					
	0-3	4-6	7-10	TOTAL	Av. TP	MD
N+WA	0.16 (2)	0.42 (5)	0.42 (5)	1.00 (12)	5.17	5.0
Ø	0.34 (22)	0.35 (23)	0.31 (20)	1.00 (65)	4.65	5.0
N+P	0.30 (11)	0.28 (10)	0.42 (15)	1.00 (36)	5.17	5.5
OTHER	0	1.00 (2)	0	1.00 (2)	6.00	

In Table 4 (a, b) the observations from Table 3 (a,b) become even clearer. In the case of Servant, the relative frequencies of higher range TP, associated with both N+Wa and ellipsis, tend to increase. On the other hand, TPs associated with N+P tend to have the same relative frequency regardless of their range. This is just what the TP is supposed to predict. But then again, in the case of the Old woman the picture is the reverse of what we would expect TP to predict. We not only have an increase of relative frequencies for mid and high range TP values, associated with N+WA, but also an **increase** and not decrease for TP values associated with N+P. Here, the relative frequencies of TP associated with ellipsis are the same regardless of the TP range.

From the above empirical observation it can be concluded, that TP as such is not a reliable predictor of the referential form, in particular of the form supposed to signal a topic.

3.4 Topic as a chain of coreferential noun phrases

The contradictory behavior of the two coreferential chains that we have observed in the previous section indicates that there might be coreferential chains that manifest a topic-like behavior (i.e. the Servant), and chains that do not (i.e. the Old woman). Indeed, a closer inspection of the text confirms this prediction.

For example (5), a segment of Rashomon (S80-S83), reveals two parallel coreferential chains. One is the Servant's, around which the narration is centered, and which manifests a topic-like behavior. The other is the Old woman's. In this narration in general, and in this segment in particular, the Old woman is subsidiary to the Servant.

So it can be said that what was called topic until now is just the local **manifestation** at the sentence level of a global property, i.e. the speaker's (narrator's) choice of the topic entity, i.e. a referent, around which some particular segment of discourse is built.

(5) TOPIC COREFERENTIAL CHAIN AND NONTOPIC COREFERENTIAL CHAIN

Servant chain ref. form	Old woman chain ref. form	Sent. No.	TEXT (each line a clause, a conjunction or a topic)
Nwa		80:	*sore hodo, kono otoko no aku o nikumu kokoro WA*
	Np		*- rooba no yuka ni sasita*
			- matu no kigire no yooni,
Ø			*- ikioi yoku moeagaridaasite ita no de aru.*
Nwa		81:	*Genin ni WA,*
	Nga		*- motiron, naze rooba ga sinin no kami no ke o nuku ka*
Ø			*- wakaranakatta.*
		82:	*Sitagatte,*
	Np		*- gooriteki ni wa, sore o zen-aku no izure ni katadukete yoi ka*
Ø			*- siranakatta.*
Nwa		83:	*Sikasi, genin ni totte WA*
	Ø		*- kono ame no yoru ni, kono Rasyoomon no ue de,* *sinin no kami no ke o nuku*
Ø			*- toiu koto ga, sore dake de sude ni yurusu bekarazaru*
Ø			*- aku de atta.*

TP(old woman/R)   =   3   (explicit ref. =2)   / nontopic coreferential chain
TP(servant/G)   =   5-6   (explicit ref. =2)   / topic coreferential chain
Sent. No.   =   number of sentence in the text
frq   =   frequency of pinting paragraph boundaries

This is the entity which, in the simplest of cases, can be manifested in some segment of the text as a coreferential chain of noun phrases, referring to the topic entity in the textual world. It is this choice and not merely numerical parameters that are responsible for the local topic marking [2].

4. Conclusion

From the above analysis we can conclude that the parameters that we have examined are, at least as far as the Japanese language is concerned, very unreliable predictors of "topicness" and referential forms that signal it. In the light of their mechanical definition, this should not surprise us.

In section 3.3 in particular, it has been enlightning to see how just the mechanical adherence to one such parameter, TP, does not reveal important and so far unnoticed facts.

As stated in section 3.4, the most important point is what to consider as a topic. As we have seen, sheer mechanical repetition, related to cognitive accessibility, obviously is not enough. Topic is a textual phenomenon, and has to be analyzed and explained from the point of view of text. Its manifestation in some particular sentence is but a local aspect of the complex issue we can call "topic". As can be seen from the example of the two parallel topic chains, the Servant's and the Old woman's, the "topicness" itself is a question of speakers choice, how to structure the narration and what entities to choose for this purpose. Though starting from a different starting point, Maynard (1987) also arrives at a similar conclusion. It is this creative aspect that lays behinds the elusive nature of the numerical parameters examined here.

To stress again, coreferential chain is just the simplest case of text coherence, of structuring a text, involving the topic entity and descriptions of actions and states centered on it. More complex ways of global structuring of a text do exist as well.

Also, to further clarify the choice of some particular referential form, marked for topic, it is necessary to examine the coreferential chain in relation to the way the text is structured as a narration. The primary candidate for this examination is the content paragraph structure of the text.

NOTES

[1] The line could be drawn between shorter RDs (3 or 4) without changing the picture in any substantial way.

[2] It is not a coincidence, that at approximately the same time Y. Sunakawa (in print) has arrived at a similar conclusion, discovering topical chains and non topical chains in expository prose texts. Applying the same general methodology, centered on empirical examination of whole texts, such a view was the natural outcome of the adopted methodology.

LITERATURE

Bekes, A. （ベケシュ・A.）(1995)「文脈から見た主題化と「ハ」、野田尚史他編『日本語の主題と取り立て』、くろしお出版、東京。

--- (1994)「原文残存認定単位の定義と分類基準」、佐久間まゆみ編『要約文の表現類型－－日本語教育と国語教育のために－－』第 3 章pp 28-47、ひつじ書房、東京。

--- (1992)「文の形成と節の内容的つながり」、文化言語学編集委員会編『文化言語学　その提言と建設』、三省堂、東京。

--- (1987)『テクストとシンタクス』、くろしお出版、東京。

Chafe, W. (1987) Cognitive constraints on information flow, in R. Tomlin (ed.).

--- (1980) The deployment of consciousness, in W. Chafe (ed.).

--- (ed.) (1980) The pear stories. NJ: Ablex, Norwood.

--- (1976) Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (ed.) Subject and topic. NY: Academic Press.

Comrie, B. (1981) Language universals and linguistic theory. Oxford: Basil Blackwell.

Danep, F. (1974) 'Functional sentence perspective and the organization of the text', in Danep (ed.).

Danep, F. (ed.) (1974) Papers in functional sentence perspective, Mouton, The Hague and Academia, Prague.

Firbas, J. (1974) 'Some aspects of the Czecoslovak approach to problems of functional sentence perspective', in Danes (ed.).

Givón, T. (1989) Mind, code and context, Ch.6, Hillsdale NJ: Lawrence Erlbaum.

--- (1983) Topic continuity in discourse: an introduction, in Givón (ed.) (1983).

--- (ed.) (1983) Topic continuity in discourse, Amsterdam: John Benjamins.

--- (1979). On understanding grammar. NY: Academic Press.

Halliday, MAK & R. Hasan (1976) Cohesion in English, Longman, London.

Hamada, M. (1983) Referential choices in theme, subject, and ellipsis in written narrative discourse: a case study of Japanese folktales, unpublished MA thesis, Cornell U.

Hinds, J. (1987) Thematization, assumed familiarity, staging, and syntactic binding in Japanese. In J. Hinds et al. (eds.)

---(1983) Topic continuity in Japanese, in T. Givón (ed) (1983).

Hinds, J. et al. (eds.) (1987) Perspectives on topicalization: the case of Japanese WA. Amsterdam: John Benjamins.

Kuno, S. (1973) 『日本文法研究』、大修館書店、東京。

Kuroda, S.Y. (1972) The categorical and thetic judgement: Evidence from Japanese. Foundations of Language 9:153-185.

Lambrecht, Knud (1994) Information structure and sentence form: A theory of topic, focus, and the mental representation of discourse referents. Cambridge: Cambdridge University Press.

Matsushita, D. (1928) 『改撰標準日本文法』中文館書店。

Minami, F. (1974) 『現代日本語の構造』、大修館書店、東京。

Maynard, S.K. (1987) Thematization as a stageing device in Japanese narrative. In J. Hinds et al. (eds.).

Mikami, A. (三上章) (1953) 『現代語法序説』、刀江書院。

--- (1960) 『像は鼻が長い』、くろしお、東京。

Myhill, J. (1992) Typological discourse analysis: quantitative approaches to the study of linguistic function. Oxford: Blackwell.

Nagano, S. (1986) 『文章論総説』、朝倉書店、東京。

Noda, H. (永野賢) (1994) 日本語とスペイン語の主題化、『言語研究』105:32-53。

Prince, E. (1981) Toward a taxonomy of given/new information. In P. Cole (ed.) Radical Praqgmatics. NY: Academic Press.

Shibatani. (1978) 『日本語の分析』、大修館書店、東京。

Sunakawa, Y. (in print).

Thompson, C. (1990) On treatment of topical objects in chepang: Passive or inverse? Studies in Language 14:405-27.

Tomlin, R. (ed.) (1987) Coherence and grounding in discourse. John Benjamins, Amsterdam.

Yamada, Y. (山田孝雄) (1908) 『日本文法論』、宝文館

CORPUS

1. R. Akutagawa "Rashomon"
2. 4 short articles from Asahi Shinbun

Povzetek

MERE TEMSKE ZVEZNOSTI IN TEMA S ČLENKOM *WA* V JAPONSKEM JEZIKU

V članku obravnavam tkzv. teme s členkom *WA* v japonskem jeziku, in sicer v luči treh besedilno statističnih parametrov, tematski kvocient (TQ), referenčno razdaljo (RD) in tematsko vztrajnost (TP), ki naj bi odražali tematskost dane samostalniške fraze v besedilu. V uvodu na kratko podam historiat proučevanja tematizacije in v razdelku 1. orišem temo v japonščini. V razdelku 2. opišem metodo določanja teh parametrov na osnovi manifestiranja koreferenčnih samostalniških fraz v besedilu, segmentiranem v stavke (clause). Za gradivo je izbrana novela R. Akutagawa, "Rashomon", in sicer zaradi bogate narativne strukturiranosti ter zaradi dolžine, ki omogoča statistično obravnavanje koreferenčnih samostalniških fraz. V razdelku 3. obravnavam posamične parametre. Izkaže se, da če parametri napovedujejo karkoli, je to kvečjemu elipsa samostalniških fraz pri visokih vrednostih ter morda pojav netematiziranih polnih samostalniških fraz pri nizkih vrednostih. Posebne pozornosti je deležen parameter TP, kjer pri sicer primerljivo pogostih pojavljanjih dveh referentov v besedilu pride do povsem nasprotnega obnašanja. Tam, kjer naj bi višje vrednosti TP napovedovale pojav tematiziranih samostalniških fraz v sobesedilu, en referent to napoved izpolnjuje, drugi pa se obnaša popolnoma nasprotno. Natančnejši pregled besedila pokaže, da je pojav tematiziranih samostalniških fraz samo lokalna manifestacija pripovednikovega izbora nekega referenta kot izhodišča, okoli katerega gradi pripoved. Temskost teme je torej v izboru referenta in v besedilu lahko govorimo o temskih koreferenčnih verigah in ne-temskih koreferenčnih verigah. Izbor tega, okrog katerih referentov bo pripoved gradil, je odvisen od govorca (pripovednika) in ga ni moč mehansko napovedati samo iz statističnih posebnosti konteksta. Tako vsi trije parametri v najboljšem primeru izražajo samo potrebne pogoje za tematiziranje, zadostnih pa zaradi kreativne narave govornega dejanja ne morejo.