

Gwen Eva Janda, Axel Wisiolek & Stefanie Eckmann (Munich)

## Reference tracking mechanisms and automatic annotation based on Ob-Ugric information structure

The following paper is concerned with information structure in the Ob-Ugric languages and its manifestation in reference tracking and its mechanisms. We will show how both knowledge on information structure and on reference tracking mechanisms can be used to develop a system for a (semi-)automatic annotation of syntactic, semantic and pragmatic functions. We assume that the principles of information structure, i.e., the balancing of the content of an utterance, are indicated by the use of anaphoric devices to mark participants in an on-going discourse. This process in which participants are encoded by the speaker and decoded by the hearer is called reference tracking. Our model distinguishes four important factors that play a role in reference tracking: inherent (linguistic) features of a referent, information structure, referential devices and referential strategies. The interaction between these factors we call reference tracking mechanisms. Here, the passive voice and the dative shift are used to exemplify this complex interaction system. Drawing conclusions from this, rules are developed to annotate both syntactic, semantic and pragmatic roles of discourse participants (semi-)automatically.

### 1. Introduction

This paper deals with information structure in Ob-Ugric, its effects on reference tracking and the application of both in the development of (semi-)automatic annotation tools. It covers three main topics: (i) an analysis of linguistic devices used for reference tracking in Ob-Ugric, (ii) a description of information structure and its impacts on reference tracking, and, concluding from these, (iii) a discussion of how these can be used to design and set rules for a (semi-)automatic annotation of syntactic, semantic and pragmatic functions. After a short introduction to the Ob-Ugric languages and their specific typological features (section 1), we will define the notions of information structure and reference tracking (section 2). Then we will describe the complex system of reference tracking mechanisms exemplified by the passive voice and the so-called dative shift (section 3). The last section shows how the regularities of reference tracking mechanisms can be used to develop a (semi-)automatic annotation tool for Ob-Ugric text samples.

The Ob-Ugric languages form a branch of the Uralic language family and are spoken in Western Siberia on the river Ob and its tributaries. The two languages Khanty and Mansi are further divided into dialectal groups, which in turn consist of several sub-dialects. Like many of the other minority languages spoken in the Russian Federation, both Khanty and Mansi are highly endangered (Moseley 2010). The material of our investigation is taken from the corpus of the EUROCORES project “Ob-Ugric languages: conceptual structures, lexicon, constructions, categories”

and its proceeding DFG/FWG project “Ob-Ugric database: analysed text corpora and dictionaries for less described Ob-Ugric dialects”.<sup>1</sup> The corpus offers glossed and translated texts from several Khanty and Mansi varieties, such as Northern and Western Mansi as well as Surgut Khanty and its subdialect Yugan Khanty.

From a typological point of view, Khanty and Mansi are, like most Uralic languages, mainly agglutinative languages. As known from Mordvin, the Ugric and Samoyedic branch, there is subject as well as object agreement on the verb, resulting in two sets of verbal paradigms. The first set of endings agrees only with the subject of the sentence (traditionally called *subjective conjugation*), while the second set of endings agrees with both the subject and the direct object of the sentence (traditionally called *objective conjugation*). Nouns are marked for person (traditionally referred to as *possessive inflection*), number and case. In addition to a large inventory of adverbial cases, Khanty and Mansi dialects also use postpositions, mainly to mark spatial and other adverbial relations.

However, regardless of their large inventory of suffixes, the Ob-Ugric languages tend to not mark syntactic core roles by case, but by word order (SOV) and verbal inflection. In Northern Mansi, for instance, as a result of the loss of the Proto-Ob-Ugric accusative *\*-m(V)*, both the subject and direct object role are unmarked on nouns. In these dialects, however, the accusative is still found with pronominals. The same applies for Surgut and Yugan Khanty.

In the Western Mansi dialects, another strategy to mark direct objects has developed: in Pelym Mansi, e.g., some nouns in direct object role are marked with the dative-lative, but not all of them. In other dialects of Mansi, e.g., in the eastern group or the Middle Lozva dialect of the Western group, the accusative suffix can still be found, which was still used at the time of the extinction of the dialect. For now, we will mainly focus on those dialects with unmarked nominal direct objects.

Furthermore, the Ob-Ugric languages are pro-drop languages, i.e., the subject does not need to be expressed overtly in the sentence. The same applies for certain types of direct objects. The fundamental premise is the aforementionedness of subject and direct object in the discourse, i.e., they represent previously given information.

## 2. Basic notions of information structure and reference tracking

### 2.1. Information structure

In our understanding, a discourse can be defined as the comprising of grammar and its communicative value. The latter is based on (a) the text coherence and (b) the balanced content of given and new information. This balanced content is what is called information structure and is determined by the speaker’s intention and his assumption about the hearer’s knowledge. A topic is what the discourse is about and is chosen according to what the speaker intends to talk about, i.e., to share information about. The comment provides

---

1. <<http://www.oudb.gwi.uni-muenchen.de/>>

new information on the topic (cf. Loos 2004). The comment depends on the speaker's intention as well. The speaker also decides how to convey this intention to the hearer. The speaker must signal what he speaks about (his topic) and, at the same time, he must structure the information he wants to share in a way so that it can be recognized and processed correctly by the hearer. If there is too much new information which the hearer cannot connect to anything known to him, communication fails. Therefore, the speaker must establish a common ground for the hearer to connect to any new information. If, on the contrary, there is too much given information, there is no communicative value in a discourse either, since the hearer does not learn anything new (cf. Krifka 2008, Kern 2010).

Linguistically speaking, these principles result in and are indicated by the use of various anaphoric devices to refer to participants in an on-going discourse. This process is called reference tracking, which in Ob-Ugric is based on co-referentiality (cf. Comrie 1988).

## 2.2. Reference tracking

Reference tracking can be defined as the monitoring of a participant in an on-going discourse (Nagaya 2006: 3). The notion of reference tracking thus describes the encoding (speaker) and decoding (hearer) of referents in discourse. In our model, we distinguish based on Comrie (1988, 1989) four important factors that have an effect on reference tracking and its mechanisms.

Firstly, in every language there are certain inherent features of a referent that are marked linguistically (Comrie 1988). In the Ob-Ugric languages, these are number (Example 1) and – to a limited extent – animacy: there is a distinction between animate comitatives (a postpositional phrase is used instead of an adverbial case suffix, Example 2) and inanimate instrumentals (marked with instrument case suffix, Example 3). Additionally, the repeated occurrence of the noun 'knife' instead of the use of a pronominal in Example 3 indicates (apart from other factors) the distinction between animate and inanimate referents.

- (1) OUIDB Surgut Khanty Corpus. Text ID 735, Nr. 1

*kə:t i:miyən βaʎəyən.*

kə:t i:mi                -γən βaʎ -ʎ -əγən  
two old\_woman -DU live -PRS -3DU

'Once upon a time there lived two women.'

- (2) OUIDB Northern Mansi Corpus. Text ID 750, Nr. 81

*ne:matər sir (...) ma:n jotuw at we:riti*

ne:matər sir ma:n jot -uw at we:rit -i  
no\_one 1PL with -1PL NEG trust -PRS[3SG]

'No manner of water shrine, no manner of land shrine can stand against us.'

- (3) OUIDB Northern Mansi Corpus. Text ID 1238, Nr. 59

*janəy e:k<sup>w</sup>a kasaj wi -s piy -e kasaj -il*

elder woman knife take -PST[3SG] boy -SG<3SG knife -INST

ta puwtm -as -te

EMPH1 push -PST -SG<3SG

‘The old woman took a knife and stabbed her son with it.’

Secondly, there are (anaphoric) devices that can be used to refer to a referent. In Ob-Ugric, these are noun (phrases), pronouns, personal endings and the zero morpheme. Noun (phrases) are mostly used to introduce a new referent into the discourse or to point out its non-subject-status. Pronouns and personal endings do not bear any reference of their own like nouns; instead, they are substitutes for the respective noun. Whilst pronouns are overtly realized on the surface of the sentence, referents encoded in personal endings are marked on the word they are attached to, e.g., on the verb (Skrribnik 2001a). No additional overt marker is needed (Example 4).<sup>2</sup>

- (4) OUIDB Northern Mansi Corpus. Text ID 1234, Nr. 155

*matər met woweyən.*

ø matər met wow -ey -ən

ø what fee demand -PRS -2SG

‘(You will get) what kind of fee you demand.’

3SG pro-forms, however, have a limited competence regarding disambiguation and expression of co-referentiality (Kovgan 2001: 152; Kovgan 2005: 558; Kibrik 2008: 1130). Zero morphemes exhibit the least overt realization, i.e., none (Example 5).<sup>3</sup> Additionally, the inherent features are mirrored in the referential devices.

- (5) OUIDB Pelym Mansi Corpus. Text ID 1278, Nr. 33

*o:s i o:li.*

ø o:s i o:l -i

ø and PTCL live -PRS[3SG]

‘And he lives on.’

Thirdly, the underlying information structure needs to be considered. The referent’s status of accessibility and givenness is in functional relation with the morpho-phonological size of the referential device: the less accessible, i.e., new information, the

2. Zero anaphoras are marked with the wildcard ø in the glossing.

3. Information encoded with zero morphemes is indicated with square brackets in the glossing.

more encoding material is needed, i.e., noun phrases, nouns or pronouns which are overtly realized. The more accessible, i.e., given information, the less encoding material is needed, i.e., zero anaphora.

Finally, information structure decides which referential strategy is to be employed. Referential strategies are the application of a certain syntactic role, agreement, ellipsis (i.e., pro-drop), word order and/or diathesis, which will be explained in more detail in the following section. Closing the circle, the referential strategies, of course, make use of the referential devices. The combination and use of all of these factors is what we call reference tracking mechanisms. They are illustrated in the following model (Figure 1):

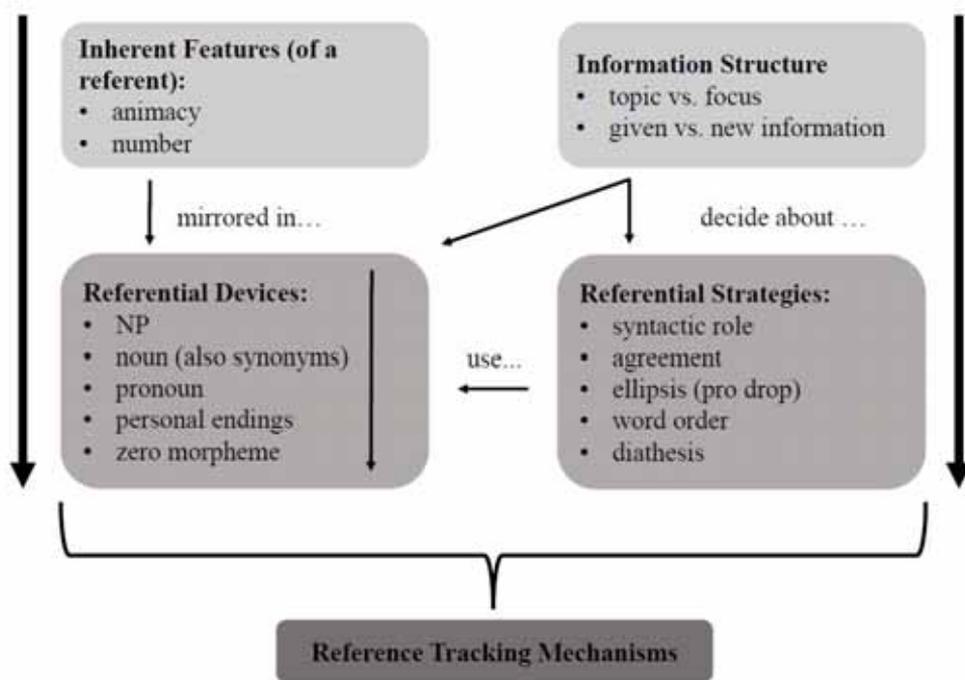


Figure 1. Model of the reference tracking mechanisms in Ob-Ugric languages.

### 3. Ob-Ugric reference tracking mechanisms

Concluding from the former section, we can state that reference tracking can be viewed as the visualization of information structure in a text. Consequently, the referent serving as the topic is the one which is monitored by reference tracking mechanisms. We understand the notion of topic as a referent's pragmatic role. Note, that in this conception we assume that there is a hierarchy of pragmatic roles (pragmatic

hierarchy). These have to be distinguished according to the referent's level of involvement in the discourse. Our text corpus mainly consists of narratives and tales. Texts of this type are well qualified to demonstrate reference tracking throughout a coherent text since in Ob-Ugric narratives there is usually only one main hero. This main hero is typically established in the beginning of the story and continuously mentioned in the plot. He therefore serves as the (role of the) *primary topic* or *discourse topic* (cf. Nikolaeva 2001). Other referents, mainly those with a relation to the main hero expressed by the verbal action, may also hold a certain topical status and serve as the *secondary topic* (cf. Nikolaeva 2001). They are not necessarily part of the whole story but may occur only in certain paragraphs and can alternatively be called *paragraph topics*. While usually there is only one discourse topic, there can be several paragraph topics since the main hero may interact with several participants or the speaker's attention may shift from one participant to another. The lowest role in the pragmatic hierarchy is represented by the *sentence topic* (cf. Reinhart 1982), which is – as inferable from its denomination – only found in a few consecutive sentences. Reference tracking mechanisms are used to track both kinds of referents – primary topics as well as secondary topics. Hereafter, the devices used to refer to the primary and secondary topics differ.

The principles of information structure in general apply to other text genres as well, but reference tracking mechanisms may differ slightly or may not be as obvious. Owing to a lack of comparable data, our analysis focuses on narrative texts.

### 3.1. Passive

In general, the passive voice is used to indicate that the semantic role of the subject is not the agent but the “patient or recipient of the action denoted by the verb” (Loos 2004). Passive voice is very frequent in Ob-Ugric and has been described in detail by, e.g., Kulonen (1989). Therefore, we will only focus on the most characteristic features of the passive voice and its relevance as a reference tracking mechanism.

Examples of a patient or a recipient as subject of a passive sentence can also be found in Ob-Ugric (Example 6). The passive is also known to demote an agent as it is not necessarily obligatory in the sentence. If it occurs, the agent is represented in a non-core syntactic role. In Mansi, for instance, the agent is marked with the dative-lative case (DLAT) (Example 6); in Khanty with the Locative case.

#### (6) OUIDB Pelym Mansi Corpus. Text ID 1277, Nr. 27

*oꝛmpnə purx itwəs.*

ø	oꝛmp	-nə	pur	-x	it	-w	-əs
ø	dog	-DLAT	bite	-INF	want	-PASS	-PST[3SG]
S/PAT	ADV/AG						

‘A dog wanted to bite him.’

Additionally, other semantic roles in subject position can also be found in passive sentences. There are, e.g., passive sentences in Ob-Ugric with a so-called locative subject (cf. Kulonen 1989). A locative subject takes the semantic role of goal in sentences with verbs of motion:

- (7) OUIDB Pelym Mansi Corpus. Text ID 1335, Nr. 11

*ta ke:m wuləmnə joxtows.*

ø	ta ke:m	wuləm	-nə	joxt	-ow	-s
ø	to an extent	sleep	-DLAT	come	-PASS	-PST[3SG]
S/LOC		ADV/AG				

‘She was completely overcome by sleep.’

Another instance where the passive is used in Ob-Ugric is when the subject takes the role of addressee with *verba dicendi*:

- (8) OUIDB Surgut Khanty Corpus. Text ID 734, Nr. 20

*muβəlinə təγə mu:nʲiʷ?*

ø	muβəli	-nə	təγə	mu:nʲiʷ	-ø	-o
ø	what	-LOC	here	tell tales	-PST	-PASS.2SG
S/	ADV/AG					

‘What has told you (to come) here?’

To conclude, the Ob-Ugric passive does not necessarily promote either the patient or the recipient to the syntactic role of subject. The passive is not restricted to promote the patient/recipient or to demote the agent. Much rather, the passive is considered as a means to decrease verb valency – since the agent is represented with a non-core syntactic role, there is only one syntactic core role left: the subject, performed by the patient. This is why the passive is usually associated with transitive verbs. The sample of verbs of motion (Example 8), however, proves that in Ob-Ugric passivization is not limited to transitives, nor is transitivity required for passivization. Instead, the passive voice is used to maintain the correlation between subject and primary topic: the assignment of a certain syntactic role is a referential strategy determined by information structure. Consequently, the primary topic is assigned the subject of a sentence in Ob-Ugric. If the semantic role is not the agent, there is a change in diathesis, i.e., passivization. The passive in Ob-Ugric therefore is a reference tracking mechanism and is not limited by verb valency or the semantic role of the referent promoted to the role of the subject.

### 3.2. Dative Shift

In the previous section, we saw that there is a strong correlation between the pragmatic role of primary topic and the syntactic role of subject. The correlation between pragmatic and syntactic roles, however, goes beyond subject and primary topic: there is also a correlation between the secondary topic and the direct object. The correlation between direct object and secondary topic is maintained with a reference tracking mechanism comparable to the passive: the so-called dative shift. (Skribnik 2001a; Givón 1984.) Consequently, there is no restriction of the patient role to the direct object, either. In a ditransitive sentence, the subject generally takes the semantic role agent, the direct object is assigned the semantic role of patient and the indirect object takes the semantic role of recipient:

(9) OUDB Surgut Khanty Corpus. Text ID 1083, Nr. 22

*nʉŋ mɛ:ntem məje v:j n'ɛ:βreməle.*

nʉŋ	mɛ:ntem	məj	-e	v:j	n'ɛ:βrem	-əle
2SG	1SG.DAT	give	-IMP.SG<2SG	small	child	-DIM.MEL
S/AG	IO/REC			DO/PAT		

‘You give me your little child.’

However, if the referent in indirect object position is the secondary topic, dative shift occurs in Ob-Ugric: the primary topic is still in subject position taking the semantic role of agent, but the secondary topic – being the recipient of the action – changes from indirect object to direct object role (Skribnik 2001: 227–229). The referent semantically serving as patient is then syntactically in indirect object position:

(10) OUDB Surgut Khanty Corpus. Text ID 1083, Nr. 23

*tʉ:t mɛ: nʉŋət tʉβətət mələm.*

tʉ:t	mɛ:	nʉŋət	tʉβət	-ət	mə	-ɪ	-əm
then	1SG	2SG.ACC	fire	-INSC	give	-PRS	-1SG
S/AG	DO/REC	ADV/PAT					

‘Then I will give you fire.’

The formal marking of the dative shift differs between Khanty and Mansi. In Surgut Khanty, the dative shift is marked in this way: the recipient as direct object (DO) is marked by accusative (if pronominal) or remains unmarked. Patients then are adverbials and marked with the instructive case (INSC) (Example 10).

In Northern Mansi, the direct object is unmarked, while the adverbial patient is inflected in the instrumental case (Example 11).

## (11) OUIDB Northern Mansi Corpus. Text ID 1229, Nr. 6

*n'a:ləl wa:rilum.*

ø	ø	n'a:l	-əl	wa:r	-i	-lum
ø	ø	arrow	-INST	make	-PRS	-SG<1SG
S/AG	DO/REC	ADV/PAT				

‘I make you an arrow.’

As previously mentioned, the choice and the morpho-phonological size of the referential device as well as the referential strategy are determined by information structure (Givón 1983). Therefore, not only can the subject as primary topic be dropped, but also the direct object as secondary topic. Furthermore, the topical status of both referents (subject and direct object) triggers the use of the objective conjugation (subject and direct object agreement on the verb) which is obligatory in this case.

In summary, the secondary topic referent triggers (i) dative shift if its semantic role is not the patient and – in Northern Mansi – and (ii) objective conjugation because of its direct object role owing to the dative shift. Thus, in these uses, the objective conjugation is not only a referential strategy by itself but can be regarded as the formal indication of dative shift. In a nutshell, the dative shift is a reference tracking mechanism that is used to assign the secondary topic to the syntactic role of direct object, and it is not limited by the semantic role of the referent promoted into the role of the direct object.

#### 4. (Semi-)automatic annotation rules for Ob-Ugric texts

The previous chapters/sections have shown that there are certain regularities regarding the use of referential devices, referential strategies and referential tracking mechanisms, all of which are mainly based on information structure. With the help of these regularities, it is possible to formulate a set of (semi-)automatic annotation rules. Such an annotation tool has recently been developed for the corpus of the EUROCORES project “Ob-Ugric languages: conceptual structures, lexicon, constructions, categories” and its proceeding DFG/FWG project “Ob-Ugric database: analysed text corpora and dictionaries for less described Ob-Ugric dialects” (see Wisiorek & Schön 2017: 389f). In this annotation system, these regularities were used as heuristic rules to tag the functional, semantic and pragmatic values of each referent throughout the whole text, followed by a review and – if necessary, a correction – by an annotator. This section will deal with the most significant regularities and their transformation into annotation rules.

One referential strategy which has not been mentioned in detail yet is word order: the unmarked word order in Ob-Ugric is SOV. We can use this regularity as a heuristic device to determine clause boundaries in complex sentences as well as to differentiate the core syntactic roles of zero marked nominal phrases: a subject-tag

is set when there is an unmarked nominal phrase and the next unmarked nominal phrase is tagged as direct object. Any nominal phrase with a case suffix is tagged as adverbial.

Yet, there are more referential devices which replace recurring nouns. Personal pronouns, which in Ob-Ugric differentiate subject and direct object by case, are tagged according to their case form. The most common references to subject as well as direct object in pro-drop languages, however, are: ellipsis as a referential strategy or zero morphemes as referential devices. Therefore, since subject and direct object do not necessarily occur on the surface of the sentence in Ob-Ugric, their position has to be visualized first in order to be able to tag the majority of syntactic arguments. Another annotation rule is thus to add a subject zero position in any clause which does not feature any unmarked nominals or pronominal phrases. If the verb is marked with objective conjugation (visible by glossing), two zero positions are added (resp. one, if there is one unmarked nominal or pronominal phrase). Referring back to word order, the position is preverbal and the sentence-initial zero is tagged as subject, the following as direct object.

With regard to semantic roles, it is suggested that subjects be tagged as agents, direct objects as patients. If there is a passive marker on the verb, however, it is suggested that the subject is tagged as patient. If there is an adverbial in the passive sentence, its suggested tag is agent. In certain cases (e.g. the combination of dative shift and passive), the suggestion will have to be corrected by the annotator.

Pragmatic roles are difficult to tag automatically. Since the size of the referential device correlates with the referent's accessibility, any zero is tagged as topic role, whereas personal pronouns are tagged as contrast. The comment part of the sentence most often corresponds with the focus part. If a sentence consists only of the predicate, this component must be new information and focus (since subject and probably direct object are given and thus dropped). This coincides with the assumption of a preverbal focus and a topic-initial word order in Ob-Ugric and is the reason why zeros are placed not only in front of the verb but in sentence-initial position. Any components (mostly adverbials) directly preceding the verb are also tagged as focus. On the other hand, there must be a reason why sometimes topics are realized as nouns within the text even though they have been dropped before. This is considered in the evaluation of the tagging results by computing the value of the nominal coded referent on the accessibility scale based on the data of the referent-tagging (s. below); introduced referents are tagged as 'RESUME', resp. 'REPEAT', if they have been mentioned immediately before.

Each referent's frequency of occurrence can be determined by numbering all participants in the text. This is where the automatic tagging reaches its limits and has to be done manually since present anaphora resolution techniques are not sufficient to recognize and disambiguate all referents correctly. Instead, the tagging relies on suggestion lists consisting of the afore-identified referents in the form of their first having been mentioned, the use of which significantly speeds up the annotation procedure.

Since proofreading of the results is always needed, too, the term (semi-)automatic annotation is used.

## 5. Conclusions and further research

The aim of this paper was to examine the Ob-Ugric reference tracking mechanisms as a set of rules based on the principles of information structure. From the interaction between reference tracking strategies and mechanisms and our knowledge of the information structure system, we can detect certain correlations: those between syntactic and pragmatic role as well as those between referential device and pragmatic role. These principles were then used to set up a tool for (semi-)automatic annotation: in combination with syntactic rules (e.g. standard word-order), it is possible to tag functional parameters. In the tagging process during the project OUIDB (which was conducted in connection with a syntactic parsing), these regularities were successfully applied to give suggestions for syntactic, semantic and pragmatic values (semi-automatic tagging). The annotated data can be used to draw further conclusions on the complexity of Ob-Ugric sentence structure and also serve as a basis for further analyses of, e.g., frequency of reference in general, the introduction of reference, referential chains, disruption in reference and re-introduction of referents in particular, as well as the choice of the actual referential device for introduction, sustainability and re-introduction.

We are just at the beginning of this kind of analysis, therefore the insights are not limited to those examples mentioned. A next task could be the inclusion of data from those (Mansi) dialects which do have direct object marking on the noun. With Middle Lozva Mansi data, the task should be easily manageable: we already tag pronominals with the accusative marker as direct objects, this should be applicable to nouns, as well. In Pelym Mansi, however, the dative-lative case is occasionally used to mark direct objects. The dative-lative gloss triggers tagging the noun phrase as adverbial and thus a more thorough research is required before we manage a correct annotation of the direct objects in question. Another valuable next step could also be an analysis of further variants and dialects (e.g. Northern and Eastern Khanty, Eastern Mansi) and the comparison of the obtained data with other recent studies in this field of research (e.g. Virtanen 2015, Filchenko 2012).

## References

- Comrie, Bernard 1988: Coreference and conjunction reduction in grammar and discourse. – John Hawkins (ed.), *Explaining language universals*. Oxford: Blackwell. 186–208.
- Comrie, Bernard 1989: Some general properties of reference-tracking systems. – Doug Arnold, Martin Atkinson, Jacques Durand, Claire Grover & Louisa Sadler (eds.), *Essays on Grammatical Theory and Universal Grammar*. Oxford: Clarendon Press. 37–51.

- Fichenko, Andrey 2012: Continuity of information structuring strategies in Eastern Khanty. – Pirkko Suihkonen, Bernard Comrie & Valery Solovyev (eds.), *Argument structure and grammatical relations: A crosslinguistic typology*. Amsterdam: John Benjamins. 115–132.
- Givón, Talmy 1983: Topic continuity in discourse: an introduction. – Talmy Givón (ed.), *Topic continuity in discourse. A quantitative cross-language study*. Typological Studies in Language 3. Amsterdam: John Benjamins. 1–41.
- Givón, Talmy 1992: The grammar of referential coherence as mental processing instructions. – *Linguistics* 30: 5–55.
- Kern, Beate 2010: *Metonymie und Diskurskontinuität im Französischen*. Linguistische Arbeiten 531. Berlin – New York: De Gruyter.
- Kibrik, Andrej 2008: Reference maintenance in discourse. – Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals 2*. Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 20/2. 1123–1141.
- Kovgan, Elena 2001: Reference-tracking in Khanty. – Tõnu Seilenthal (ed.), *Congressus Nonus Internationalis Fenno-Ugristarum*. 7.–13.8.2000. Tartu: Auctores. 145–152.
- Kovgan, Elena 2005: The textual structure of Khanty: co-reference and anaphora. – M. M. Jocelyne Fernandez-Vest (ed.), *The Uralic languages today. A linguistic and cognitive approach*. Paris: Honoré Champion. 547–560.
- Krifka, Manfred 2008: Basic notions of information structure. – *Acta Linguistica Hungarica* 55 (3–4): 243–276.
- Kulonen, Ulla-Maija 1989: *The passive in Ob-Ugrian*. Mémoires de la Société Finno-Ougrienne 203. Helsinki: Société Finno-Ougrienne.
- Loos, Eugene E. et al. (ed.) 2004: Topic. – *Glossary of linguistic terms*. <<http://www.glossary.sil.org/term/topic>> 26th July 2017
- Moseley, Christopher (ed.) 2010: *Atlas of the world's languages in danger*. Paris: UNESCO Publishing. Online version: <<http://www.unesco.org/culture/en/endangeredlanguages/atlas>>.
- Nagaya, Naonori 2006: *Topicality and reference-tracking in Tagalog*. Tokyo: University of Tokyo, Department of Linguistics. <[http://www.ruf.rice.edu/~nn1/download/Nagaya2006Topicality\\_and\\_reference-tracking\\_in\\_Tagalog.pdf](http://www.ruf.rice.edu/~nn1/download/Nagaya2006Topicality_and_reference-tracking_in_Tagalog.pdf)>.
- Nikolaeva, Irina 2001: Syntaktische Analyse der objektiven Konjugation im Ob-Ugrischen. – Seilenthal, Tõnu (ed.), *Congressus Nonus Internationalis Fenno-Ugristarum*. 7.–13.8.2000. Tartu: Auctores. 145–152.
- Reinhart, Tanya 1982: Pragmatics and linguistics: An analysis of sentence topics. – *Philosophica* 27: 53–94.
- Skribnik, Jelena 2001: Pragmatic Structuring in Northern Mansi. – Seilenthal, Tõnu (ed.), *Congressus Nonus Internationalis Fenno-Ugristarum*. 7.–13.8.2000. Tartu: Auctores. 222–239.
- Virtanen, Susanna 2015: *Transitivity in Eastern Mansi: an information structural approach*. Dissertation. University of Helsinki. <<http://urn.fi/URN:ISBN:978-951-51-0548-6>>.
- Wisiołek, Axel & Zsófia Schön 2017: Ob-Ugric database: Corpus and lexicon databases of Khanty and Mansi dialects. – *Acta Linguistica Academica* 64 (3): 383–396.