Государственный институт русского языка им. А. С. Пушкина

ТИПОЛОГИЯ
МОРФОСИНТАКСИЧЕСКИХ
ПАРАМЕТРОВ

ТОМ 3
ВЫПУСК 2

2020

Типология морфосинтаксических параметров
том 3, выпуск 2

TYPOLOGY
OF MORPHOSYNTACTIC
PARAMETERS

Typology of Morphosyntactic Parameters
volume 3, issue 2

Редакционная коллегия

Главный редактор

Екатерина Анатольевна Лютикова —

доктор филологических наук, доцент; профессор кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М. В. Ломоносова

ORCID: 0000-0003-4439-0613
Личная страница в системе ИСТИНА МГУ
Личная страница на Academia.edu

Заместитель главного редактора

Антон Владимирович Циммерлинг —

доктор филологических наук; профессор кафедры общего языкознания и русского языка Государственного института русского языка имени А. С. Пушкина; профессор кафедры контрастивной лингвистики Московского педагогического государственного университета; ведущий научный сотрудник сектора типологии Института языкознания РАН

ORCID: 0000-0002-5996-2648
Личная страница на сайте МПГУ
Личная страница на сайте ИЯ РАН
Личная страница на Academia.edu

Ответственный секретарь

Ксения Павловна Семёнова —

инженер кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М. В. Ломоносова;

Личная страница в системе ИСТИНА МГУ
Личная страница на Academia.edu

Редколлегия

Джон Фредерик Бейлин —

Ph.D., профессор университета Стоуни Брук, Нью-Йорк, США

https://linguistics.stonybrook.edu/people/_bios/_linguistics-faculty/john-bailyn.php

Олег Игоревич Беляев —

кандидат филологических наук, старший преподаватель кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М. В. Ломоносова, Москва, Россия

https://istina.msu.ru/profile/belyaev/

Яцек Виткощ —

Ph.D., профессор университета г. Познань, Польша

http://wa.amu.edu.pl/wa/Witkos_Jacek

Анастасия Алексеевна Герасимова —

аспирантка кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М. В. Ломоносова, Москва, Россия

https://istina.msu.ru/profile/Gerasimova/

Атле Грённ —

Ph.D., профессор университета г. Осло, Норвегия

https://folk.uio.no/atleg/

Нерея Мадарьяга —

Ph.D., доцент университета Страны Басков, Витория, Испания

https://ehu.academia.edu/NereaMadariaga

Мария Полинская —

Ph.D., профессор университета Мэриленда и Гарвардского университета, США

http://www.mariapolinsky.com/

Андрей Владимирович Сидельцев —

доктор филологических наук, заместитель директора Института языкознания РАН

http://iling-ran.ru/main/scholars/sidelcev

Editorial board

Editor-in-chief

Ekaterina A. Lyutikova —
Dr. Phil. Hab.; professor at the Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University

ORCID: 0000-0003-4439-0613
Personal page on Istina.msu.ru
Personal page on Academia.edu

Deputy chief editor

Anton V. Zimmerling —
Dr. Phil. Hab.; professor at the Department of General Linguistics and Russian Language, Pushkin State Russian Language Institute; principal research fellow at the Institute of Linguistics, Russian Academy of Sciences; professor at the Department of Contrastive Linguistics, Moscow Pedagogical State University, Moscow, Russia

ORCID: 0000-0002-5996-2648
Personal page on the Moscow Pedagogical State University Website
Personal page on The Institute of Linguistics RAS Website
Personal page on Academia.edu

Executive secretary

Xenia P. Semionova —
data engineer at the Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University, Moscow, Russia

Personal page on Istina.msu.ru
Personal page on Academia.edu

Editorial staff

John Frederick Bailyn —

Ph.D., professor at the Stony Brook University, New York, USA

https://linguistics.stonybrook.edu/people/_bios/_linguistics-faculty/john-bailyn.php

Oleg I. Belyaev —

Ph.D., senior lecturer at the Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University, Moscow, Russia

https://istina.msu.ru/profile/belyaev/

Jacek Witkoś —

Ph.D., professor at the Poznań University, Poland

http://wa.amu.edu.pl/wa/Witkos_Jacek

Anastasia A. Gerasimova —

Ph.D. student at the Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University, Moscow, Russia

https://istina.msu.ru/profile/Gerasimova/

Atle Grønn —

Ph.D., professor at the Oslo University, Norway

https://folk.uio.no/atleg/

Nerea Madariaga —

Ph.D., professor at the University of the Basque Country, Vitoria, Spain

https://ehu.academia.edu/NereaMadariaga

Maria Polinsky —

Ph.D., professor at the University of Maryland, professor at the Harvard University, USA

http://www.mariapolinsky.com/

Andrei V. Sideltsev —

Dr. Phil. Hab., deputy director of the Institute of Linguistics (RAS), Moscow, Russia

http://iling-ran.ru/main/scholars/sidelcev

Содержание

CONTENTS

ВЫДВИЖЕНИЕ ПРЯМОГО ДОПОЛНЕНИЯ ИЗ КОСВЕННЫХ ВОПРОСОВ С *ЛИ*

А. Ю. Вознесенская
МГУ имени М. В. Ломоносова

В данной статье исследуются островные эффекты в русских косвенных вопросах с *ли*. Такие вопросы имеют две возможные конфигурации: энклитика (*ли*) следует либо за предикатом, либо за вынесенной фокусной составляющей. В этом исследовании был проведен эксперимент, чтобы увидеть (i) как конфигурация влияет на приемлемость выдвижения прямого дополнения из косвенных вопросов с *ли*; (ii) как тип A′-передвижения влияет на выдвижение прямого дополнения из косвенных вопросов с *ли*. Результаты показывают, что оба фактора значимы.

Ключевые слова: слабые острова, косвенный вопрос с *ли*, вопросительный остров.

THE EXTRACTION OF DIRECT OBJECTS OUT OF INDIRECT YES/NO QUESTIONS IN RUSSIAN

Anastasiia Voznesenskaia
Lomonosov Moscow State University

This paper examines island effects in Russian indirect yes/no questions formed with the *li* particle. Such questions have two possible configurations: the enclitic (*li*) follows either the clause predicate or a preposed focused element. In this study an experiment was conducted to see (i.) how the configuration affects the acceptability of moving DOs out of indirect questions; (ii.) how the type of A′-movement affects the acceptability of moving DOs out of indirect questions. The results show that both factors are significant.

Keywords: weak islands, *li* questions, yes/no questions, *wh*-island, experimental syntax.

1. Introduction

An indirect question is a structure in which the embedded clause is interrogative. Such structures exhibit island effects, in this way indirect questions in English exhibit the so-called weak island effect: not all elements can move out of these structures. In the literature on the topic, such structures are called *wh*-islands (e.g. [Boeckx 2012]). The acceptability of movement out of *wh*-islands depends on several factors. One of the most discussed factors is the opposition between arguments and adjuncts [Huang 1982; Chomsky 1986]. In (1a) it is possible to move the argument out of the indirect question, while in (1b) the movement of the adjunct is unacceptable.

(1)   a. *Which girl do you wonder* [*whether Bill kissed ____*]*?*

       b. **How do you wonder* [*whether Bill kissed Sue ____*]*?*

    In Russian, indirect *wh*-questions also exhibit island effects. In (2a) *wh*-movement of a direct object is possible, while *wh*-movement of an adjunct in (2b) is ruled out.

(2)   a. *chto   ty   ne   znaesh   [gde   Petya   kupil ____*]*?*
          what   you   NEG   know.2SG   where   Petya   bought
          Lit.: 'What don't you know where Petya bought?'

       b. **gde   ty   ne   znaesh   [chto   Petya   kupil ____*]*?*
          where   you   NEG   know.2SG   what   Petya   bought
          Lit.: 'Where don't you know what Petya bought?'

    Yes/no indirect questions in Russian are distinguished by the fact that they are formed using the clitic *li* and come in two configurations: *li* refers to the clause predicate (3a) or to some XP (3b).

(3)   a. *ya   ne   znaju   [pomyl-a li Masha posud-u].*
          I   NEG   know.2SG   washed-F   LI Masha   dishes-ACC
          'I don't know whether Masha washed the dishes.'

       b. *ya   ne   znaju   [Masha li pomyl-a   posud-u].*
          I   NEG   know.2SG   Masha   LI washed-F   dishes-ACC
          'I don't know whether it was Masha who washed the dishes.'

As will be discussed in more detail in Section 2, these two configurations differ structurally: according to some analyses (e.g. [King 1994]) in (3b) the embedded CP specifier is occupied by the XP *Masha* (in English indirect questions the specifier is occupied by the *wh*-word *whether*), while in (3a) the specifier is left empty.

This study aims to find out whether the difference in configuration (3a) vs. (3b) affects the acceptability of movement out of yes/no indirect questions in Russian. Another question under study is if the acceptability is affected by the type of A′-movement. Apart from *wh*-movement, movement in relative clauses and scrambling have been considered. In order to answer these questions, an experimental study has been conducted.

2. Background

2.1. Weak islands

The seminal work [Ross 1967] was the first to describe structures that disallow moving constituents out of them — islands. Later, when new structures started adding to Ross's structures, the notions of strong and weak islands emerged. It is impossible to move any constituents out of a strong island (e.g. Complex Noun Phrase island (4a), Adjunct island (4b)).

(4) a. **What did you make* [*the claim that John bought* ___]*?*

    b. **What do you worry* [*if John buys* ___]*?*

Some elements can move out of weak islands. According to [Szabolcsi 2006], the first explanations of the weak island effect are syntactic and focus on the contrast in (5): it is believed that moving out of weak islands is possible only for arguments (5a), but impossible for adjuncts (5b).

(5) a. *What$_i$ do you wonder* [*how$_j$ to fix t$_i$ t$_j$*]*?*

    b. **How$_i$ do you wonder* [*what$_j$ to fix t$_j$ t$_i$*]*?*

In [Chomsky 1986] this contrast is explained by the Empty Category Principle.

(6) Empty Category Principle (ECP) [Chomsky 1981, 1986]:
    [$_α$ e] must be properly governed[1].

---

[1] A properly governs B iff A theta-governs B or A antecedent-governs B. A theta-governs B iff A governs B and A theta-marks B. A antecedent-governs B iff A governs B and A is coindexed with B.

In other words, a trace must be governed either by the head assigning it its theta role or by its antecedent. So, in sentence (5a), in which the argument is moved to the left periphery, t_i is theta-governed by the verb *to fix*, which assigns its theta role, and t_j is governed by the antecedent *how_j*, that is, both traces are properly governed. However, in sentence (5b), in which an adjunct moves from the island structure, t_i is not properly governed: it is not theta-governed, since there is no theta role assignment, and it is not antecedent-governed, since there is a barrier (CP) between *How_i* and t_i. For this reason, this sentence is ruled out.

However, it was soon noticed that it is not enough to be an argument to move out of a weak island [Rizzi 1990].

(7)   *What did John wonder [whether these pearls cost ____]?*

In (7) *these pearls cost* is unacceptable, that is, the extraposed *wh*-pronoun is subject to subcategorization, therefore, it can be assumed that this is an argument. Then in (7), an argument is moved out of a *wh*-island. However, this sentence turns out to be unacceptable.

Another approach to the analysis of island effects is based on the concept of Relativized minimality [Rizzi 1990]. Rizzi notes that movement cannot cross an element that has features similar to the element being moved. Thus, an A′-movement cannot cross another A′-element, A-movement cannot cross another A-element and head movement cannot cross another head.

However, in addition to the contrast between arguments and adjuncts predicted by the ECP, it is also necessary to explain the unacceptability of examples like (7). Rizzi's theory suggests that referentiality is behind the contrast. The unacceptability of (7) is explained by the fact that, although such quantitative phrases can be arguments, they do not have theta roles of event participants (referential theta roles). [Rizzi 1990] suggests that referentiality "saves" movement because referential phrases receive a referential index that allows the trace to be semantically bound. In this case, this movement is not subject to locality restrictions. Thus, the traditional contrast (5a) vs. (5b) is explained by the fact that *what,* unlike *how,* has a referential index.

It is worth noting that although referentiality is central to Rizzi's theory of weak islands, the meaning of this term is not clearly defined. It is also not entirely clear what exactly these theories mean by referential index possession in case of *wh*-expressions.

Apart from syntactic theories on weak islands, there's also a number of semantic approaches [Kroch 1989; Szabolcsi, Zwarts 1993; Honcoop 1998]. The analysis in [Abrusán 2014] is also based on weak island semantics.

Considering *wh*-islands, M. Abrusán divides matrix predicates into *know*-class and *wonder*-class: the former (8b) cause stronger island effects than the latter (8a). The author considers island effects in indirect questions with *whether* on the material of degree questions (*how many / much*-questions) such as (8).

(8) a. ?*How much wine are scientists investigating whether it is useful to drink in order to stay healthy?*

b. **How much wine did scientists discover whether it is useful to drink in order to stay healthy?*

M. Abrusán's analysis is based on the maximal informativity principle following [Dayal 1996].

(9) The Maximal Informativity Principle
     Any question presupposes that it has a maximally informative answer (i.e. a true answer which logically entails all the other true answers)

Considering the semantics of such questions, Abrusán comes to the conclusion that *wh*-islands with predicates of the *know*-class cannot get the maximally informative answer, which leads to a contradiction. The *wonder*-class predicates have the maximally informative answer only in very specific and often unnatural contexts, which makes them pragmatically odd.

In addition to the asymmetries associated with the extracted constituent, the acceptability of movement out of an indirect question may also depend on the type of movement itself. So, as mentioned in Section 1, in Russian, an indirect question turns out to be an island for the movement of *wh*-pronouns, but not relative ones [Lyutikova 2009].

(10) a. ???*tot,   kto ya   ne   znal,   chto   darit.*
        that.one   who I   NEG knew   what   gives
        'The one about which I didn't know what he is giving as a gift.'

b. *slovo   kotoroe   ya   ne   znayu   ne   tol'ko   kak vyglyadit,*
   word   which   I   NEG know   NEG only   how looks

> *no   i   chto   oznachaet.*
> but   and   what   means
> 'The word which I not only don't know what it looks like but also what it means.' [Lyutikova 2009: 467]

According to [Bailyn 2020], in Russian, there are also differences between *wh*-movement and scrambling out of an indirect question, along with perception verb complements (see, hear, etc.) and indicative clauses with the complementizer *chto* 'that'. In (11) (examples from [Bailyn 2020: 9]), *wh*-movement out of an indirect question turns out to be ruled out for adjuncts but scrambling of the same constituent turns out to be acceptable. At the same time, *wh*-movement and scrambling behave the same way with respect to the complex NP constraint, the coordinate structures constraint, and the Condition on Extraction Domains.

(11) a. $^{??}$*ty   molotk-om   sprosil kogda Mitya chinil   mashin-u.*
    you   hammer-INS   asked   when   Mitya fixed   car-ACC
    Lit.: 'With a hammer you asked when Mitya was fixing the car.'

 b. *ty   chem   sprosil kogda Mitya chinil   mashin-u.*
    you   what.INS   asked   when   Mitya fixed   car-ACC
    Lit.: 'With what did you ask when Mitya was fixing the car?'

J.F. Bailyn bases his analysis of this asymmetry on a modification of the theory of Relativized minimality proposed in [Rizzi 2004]. Within the framework of this "featural" Relativized minimality, the A′-interveners are further subdivided into groups with respect to feature classes. Relevant for A′-movement is the division of features into quantificational [+Q], which include the feature [wh] and non-quantificational [–Q]. To explain the differences between *wh*-movement and scrambling, Bailyn introduces the [Σ] feature for scrambling, classifying it as non-quantificational. Scrambling constructions, according to [Bailyn 2020], are derived as follows: some functional head F on the left periphery attracts components with the [+Σ] feature to the specifier. Thus, *wh*-pronouns in Spec, CP of indirect questions, as well as complementizers *chto* 'what', *kogda* 'when', *kak* 'how' turn out to be [+Q] interveners for *wh*-movement. However, there are no interveners for the [+Σ] components, so scrambling in the considered constructions is possible. Bailyn explains the possibility of relative movement from the *wh*-island by the fact that the relative pronoun, being a [+Q] element, behaves like a [–Q] element due to the fact that it be-

longs to the subclass of quantificational elements, which are also modification elements. According to [Abels 2012], in this case, only an element of the same subclass of the class [+Q] can be an intervener for a relative pronoun.

To sum up, the analysis based on the Empty Category Principle successfully predicts the differences between the extraction of the direct object from *wh*-islands on the one hand and adjuncts on the other hand. However, within the framework of this analysis, it turns out that (7) is unexpectedly unacceptable: if we consider *what* as an argument, respectively, moving from a theta position, the analysis suggests the possibility of such an extraction. The theory in [Rizzi 1990] accounts for the unacceptability of (7), but the explanation for the asymmetry of the constituent types is based on the rather fuzzy notion of a referential index that allows *wh*-phrases possessing it to overcome the locality constraints. The analysis in [Bailyn 2020], based on Relativized minimality [Rizzi 2004], predicts differences between different types of movement out of *wh*-islands. The semantic theory in [Abrusán 2014] assumes a contrast in the type of the extracted constituents, however, it is composed based on the semantics of questions, and, in this form, does not make predictions about other types of movement.

2.2. Indirect yes/no questions in Russian

A yes/no indirect question in Russian is structured as follows: it is formed using the particle *li*, which follows the first component of the embedded clause. In the works [Schwabe 2004; Dyakonova 2009], adhering to the cartographic system of the left periphery of the clause (see [Rizzi 1997]), *li* is analyzed as the realization of the Force head — a functional head responsible for the illocutionary force of the sentence. According to Dyakonova's work, the component preceding *li* is focused and gets to its site by focus movement. According to [Schwabe 2004], *li* is an enclitic, so it is attached to the nearest phonological word, which is usually a verb in T. This analysis predicts the word order when *li* follows the verb. In the case when it follows a certain constituent (not the predicate), it is proposed to consider that the advanced constituent is in Spec, FocP. For (12), the word order *Ivan li* is also a consequence of *li*'s status as an enclitic.

(12) a. *Ivan-a   li Petr vstretit.*
      Ivana-ACC   li Petr will.meet
      'Is it Ivan that Petr will meet?'

   b. [$_{\text{ForceP}}$ *li* [$_{\text{FocP}}$ [*Ivana*]$_i$ Foc [$_{\text{TP}}$ *Petr vstretit t$_i$*]]]

Among the works that do not rely on a cartographic approach, one can consider the analysis in [King 1994]. According to this paper, sentences in (13) have structures in (14): *li* is in C, in (13a) the focused constituent *knig-u* 'book-ACC' is in Spec, CP, while in (13b) the verb *prochitala* 'read' moves from V to C.

(13) a. *knig-u   li prochital-a   Anna.*
        book-ACC   LI read-F   Anna
        'Is it a book that Anna read?' [King 1994: 92]

    b. *prochital-a   li Anna   knig-u.*
       read-F   LI Anna   book-ACC
       'Did Anna read the book?' [King 1994: 92]

(14) [King 1994: 93]
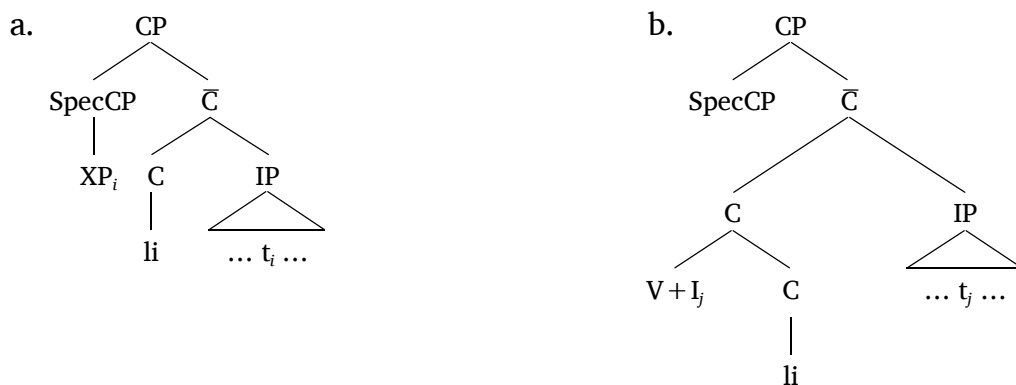
A similar analysis, but for the Bulgarian and Macedonian languages, is presented in [Rudin et al 1999]: *li* is in C° and can check the [+focus] feature. If there is no constituent before the *li*, then the verb V is raised to check the interrogative feature.

Thus, according to both groups of approaches, the XP *li* configuration differs in that XP is in the specifier, whereas with the V *li* configuration, the specifier position is not filled. In this case, the ECP predicts that in both configurations it will be possible to extract the direct object since its trace is theta-governed either way. From the point of view of Relativized minimality, for the XP *li* configuration, it should be impossible to perform A′-movement, since XP is in the A′-position and is an intervener for A′-movement. At the same time, the configuration V *li* should not present problems for A′-movement. In terms of the attribute relative minimality [Rizzi 2004], XP has the quantification attribute [+Foc], which, according to J.F. Bailyn, will make XP an intervener for *wh*-movement, but not for relative movement and scrambling.

In this study two questions were posed:

(15) i. Does the choice of configuration (XP *li* or V *li*) affect the acceptability of extraction out of yes/no indirect questions in Russian?

    ii. Does the type of A′-movement affect the acceptability of extraction out of yes/no indirect questions in Russian?

To answer these questions, an experimental study was carried out. Due to the fact that the native speakers' judgments regarding the acceptability of movement out of this structure (and weak islands in general) are very different, it was decided to use an experimental methodology for this study, which will allow to generalize the scattered data, as well as analyze the statistical significance of the expected contrasts.

3. Experimental study

3.1. Design

In order to investigate how the configuration of the indirect question and the type of A′-movement affect the acceptability of movement of the constituents, an experiment was conducted with a factor design. According to questions i. and ii. formulated in (15), two factors have been investigated: (i) type of the constituent before the particle *li* (V or XP); (ii) type of movement (*wh*-movement, relative clause movement, or scrambling). Thus, the experiment had a 2×3 design (6 experimental conditions). Following the Latin square method, 6 experimental lists were compiled, in which each block of experimental conditions ($2 \times 3 = 6$ conditions) was represented by 4 variants of lexical content. Thus, 24 experimental sentences were presented on each sheet. In addition, 24 filler sentences were drawn up, which were the same for all experimental lists. In the experiment, acceptability judgments were collected: the respondents were asked to rate the acceptability of sentences on a scale from 1 to 7.

3.2. Materials

To maintain maximum uniformity of the lexical material, the experimental sentences were composed in accordance with the following principles. Firstly, in all sentences, the direct object underwent movement since according to the above theories of weak islands, the DO is most likely to be able to move out of indirect questions. Subjects of the matrix clause were 2nd person singular pronouns in all the sentences. In sentences with a relative clause, *kotoryj* 'which'

was always chosen in inanimate, masculine form, since for this form the nominative and accusative are homonymous, which excludes the influence of the case factor on the results of the experiment. At the end of each sentence, there were adverbial prepositional groups (for example, for the evening in (16)). The constituent before *li* in the XP *li* configuration clauses was the subject of the embedded clause. As far as predicates are concerned, the decision was made to use the matrix predicates of the *wonder*-class, since, as discussed above, the predicates of this class have weaker island effects. The predicates *sprashivat'* 'ask' and *proveryat'* 'check' as the most frequently used. In each experimental list, each of these two predicates is combined with each embedded clause predicate from the list: *napisat'* 'write', *slomat'* 'break', *poteryat'* 'lose', *prochitat'* 'read', *posmotret'* 'see', *sjest'* 'eat', *narisovat'* 'draw', *zametit'* 'notice', *vybrosit'* 'throw out', *spet'* 'sing', *vyuchit'* 'learn', *otmyt'* 'wash off'. This list was compiled on the basis of frequency, and also (and to a greater extent) on the basis of the compatibility of verbs with matrix predicates. Since the investigated constructions are not easily perceived in written speech: such constructions, if used, then in oral speech and in general can be difficult to interpret without context (especially sentences with scrambling, where we are interested in a certain intonation). Therefore, before the presentation of the experimental sentence, the respondent was shown the context for this sentence. The block of experimental sentences looked like this:

(16) a. context: {I believe you tried to find out how Tanya was doing with some assignment that was due tomorrow.}

 chto ty sprashival napisal-a li Tanya za vecher?
 what you asked wrote-F LI Tanya in evening
 Lit.: 'What did you ask whether Tanya wrote in one evening?'

 b. context: {I believe you asked: "Was it Tanya who wrote such a paper in one evening?"}

 chto ty sprashival Tanya li napisal-a za vecher?
 what you asked Tanya LI wrote-F in evening
 Lit.: 'What did you ask whether it was Tanya who wrote in one evening?'

 c. context: {The teacher suspected that Tanya spent too little time working on one of her reports.}

 uchitel' nemnogo porugal tot doklad kotoryj on sprashival
 teacher a.little scolded that report which he asked

napisal-a   li   Tanya   za   vecher?
wrote-F   LI   Tanya   in   evening
Lit.: 'The teacher scolded a little the report, which he asked whether Tanya wrote in an evening.'

d. context: {The teacher suspected that Tanya copied her report from someone else's one.}

uchitel' nemnogo   porugal tot   doklad kotoryj on   sprashival
teacher a.little   scolded that report which he   asked

Tanya   li napisal-a   za   vecher?
Tanya   LI wrote-F   in   evening
Lit.: 'The teacher scolded a little the report, which he asked whether it was Tanya who wrote in an evening.'

e. context: {The teacher gave Tanya an assignment to write an essay, a report and an abstract.}

doklad   on   sprashival   napisal-a li   Tanya   za   vecher?
report   he   asked   wrote-F   LI   Tanya   in   evening
Lit.: 'The report he asked whether Tanya wrote in an evening.'

f. context: {The students handed a poster and a report to the teacher. They spent the evening working on them.}

doklad   on   sprashival   Tanya li napisal-a za   vecher?
report   he   asked   Tanya LI wrote-F   in   evening
Lit.: 'The report he asked whether it was Tanya wrote in an evening.'

These sentences correspond to the levels of the studied factors as indicated in the table 1:

Table 1. The distribution of stimuli by factor levels

	wh-movement	relative pronoun movement	scrambling
V *li*	16a	16c	16e
XP *li*	16b	16d	16f

Fillers were also divided into two main groups: acceptable and unacceptable ones. This approach allows fillers to be used not only to distract the respondent, but also as a control group of examples to check the respondent's judgments. Among the acceptable ones were sentences with a yes/no indirect question without any movement (17a) and sentences with an indirect *wh*-question

without ant movement (17b). Among the unacceptable fillers were sentences with *wh*-movement out of a relative clause (18a), as well as scrambling out of a relative clause (18b). Judgments about the inacceptability of these structures were obtained previously from several native speakers.

(17) a. context: {Vasya called the hospital to find out who treated him.}

  on   utochnil   rabotal li doctor Nosov vchera   utrom.
  he   clarified   worked LI doctor Nosov yesterday morning
  'He found out whether Dr. Nosov was at work yesterday morning.'

  b. context: {Anton was in such a difficult situation for the first time.}

  on ne mog   reshit' chto on   dolzhen   delat'
  he NEG could   decide what he   must   do
  'He couldn't decide what he should do.'

(18) a. context: {I think Vasya went to see this movie because some famous critic liked it.}

  *komu   on posmotrel film kotoryj ponravilsya.
  who.DAT he watched film which   was.liked
  Intended: 'Who is the person that Vasya watched the film because this person liked it?'

  b. context: {Vasya has two female friends: Masha and Olya. All his friends like Olya.}

  *Mash-u on ne znaet chelovek-a kotoryj   nenavidit.
  Masha-ACC he NEG knows person-ACC which   hates
  Intended: 'He doesn't know a person who hates Masha.'

3.3. Procedure

The experiment involved 42 respondents: 27 women and 15 men, 11 of whom had a linguistic education. The experiment was performed online on the Ibex Farm web platform [Drummond 2013]. First, the context was presented to the respondent on the screen. After reading the context, the respondent went to the experimental sentence, which needed to be rated on a scale from 1 to 7. Before going through the experiment, the respondent was presented with instructions with examples of ratings for acceptable and unacceptable sentences, followed by a training part of the experiment, consisting of one sentence with context. After that, the experiment itself began. Each answer was limited by 14 seconds (not counting the time to read the context). Time had to be limited so that the

respondent did not dwell on one example for a long time (the rating of such an example would be unreliable). The time limit of 14 seconds was determined based on the average reading and rating time of some of the stimulus sentences by the study author.

3.4. Results

Before analyzing the results of the experiment, a procedure for removing outlier responses was performed. First, the responses given in less than 3 seconds were deleted. Second, outlier respondents were identified and completely excluded from consideration according to the following procedure described in [Sprouse 2018]: (a) the sum of squares of errors (deviations from the expected filler rating) was calculated for each respondent; (b) the mean and standard deviation of the values obtained were calculated; (c) the respondents for whom the calculated mean of the squared errors deviated from the general mean by more than 2 standard deviations were removed from consideration. As a result of this procedure, all the answers of 2 respondents were deleted.

Standardization was carried out for the respondents' ratings (z-scores were calculated for each respondent), and filler subgroups were combined. The overall scores for factor conditions and fillers are plotted graphically: standardized scores in Figure 1, mean scores in Figure 2.

Analysis of variance (ANOVA) was used for the statistical analysis of the data. Both factors and their interaction turned out to be statistically significant: the factor of the type of movement turned out to be significant ($F = 31.22$, $p \ll 0.001$), the configuration factor V *li* vs XP *li* turned out to be significant ($F = 36.41$, $p \ll 0.001$), the same is true for the interaction of factors ($F = 5.05$, $p = 0.00663$). Pairwise comparisons were also performed using Tukey's test. As a result of these comparisons, it was found that both the differences between relative and interrogative movement ($p = 0.0001621$) and the differences between scrambling and interrogative movement ($p = 0.0005633$) and between scrambling and relative movement ($p < 0.01$) were statistically significant.

Also, the procedure of pairwise comparisons made it possible to establish that, in *wh*-movement constructions, the differences between the configurations V *li* and XP *li* are insignificant ($p = 0.2898540$), in scrambling constructions as well ($p = 0.5282024$), but the configuration has an effect with relative pronoun movement ($p < 0.01$); the differences between relative and *wh*-movement ($p = 0.8833179$) and between *wh*-movement and scrambling ($p = 0.1611467$) were also found to be insignificant with the XP *li* configuration. With the V *li* configuration, there were no significant differences between *wh*-movement and scrambling ($p = 0.0599514$).
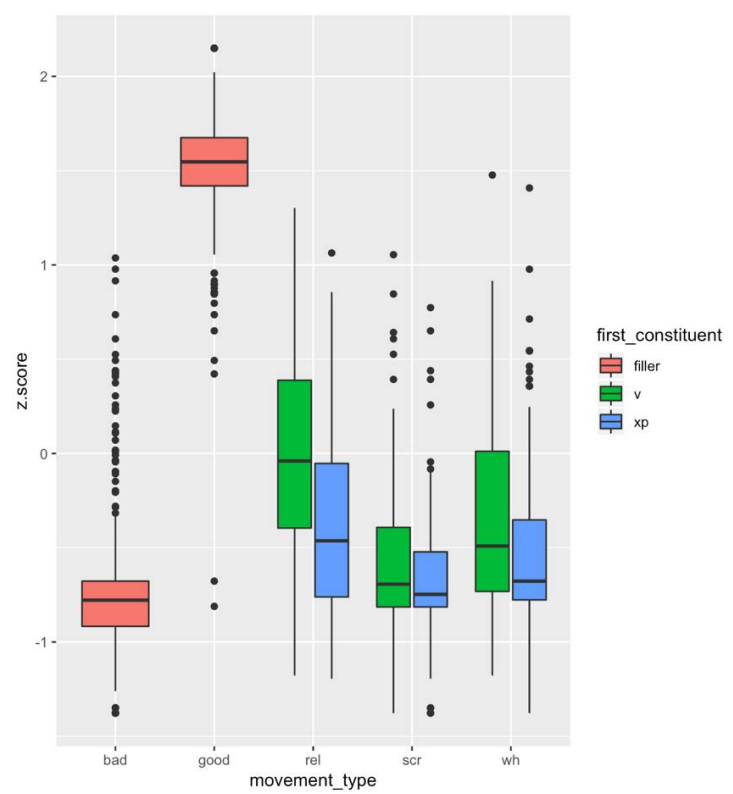
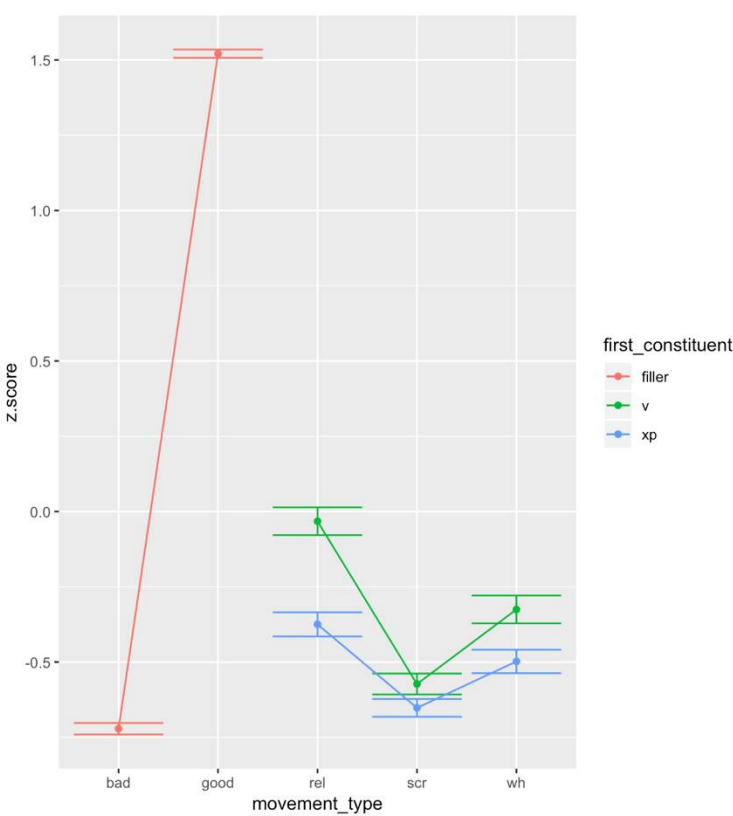Figure 1. A box-plot for standardized scores

Figure 2. Means of standardized scores

Thus, sentences with the V *li* configuration receive higher ratings than those with the XP *li* configuration. Also, sentences with relative movement were the most acceptable, followed by sentences with *wh*-movement, and sentences with scrambling having been rated the lowest.

4. Conclusion

The experiment confirmed the hypothesis that the choice of configuration V *li* vs. XP *li* does affect the acceptability of a movement out of such an indirect question. Movement out of indirect questions with V *li* configuration is rated higher than out of indirect questions with XP *li* configuration.

Out of the analyses discussed in Section 2, such a result is most directly predicted by the approach based on Relativized minimality. According to it, movement cannot cross an element that has features similar to the element being moved. Thus, A′-movement cannot cross an A′-element.

Indeed, if the A′-position of the specifier in the embedded clause in V *li* configurations is not occupied by any intervening A′-elements, the DO can move to the matrix clause. In XP *li* configurations, the XP, being an A′-element, blocks A′-movement. Thus, the contrast (19a) vs. (19b) and higher scores for the former are predicted by Relativized minimality.

(19) a. *chto ty   sprashival* [*napisal-a li   Tanya chto za   vecher*]*?*
　　　 what you   asked   wrote-F   LI   Tanya   in   evening
　　　 Lit.: 'What did you ask whether Tanya wrote in one evening?'

　　 b. *chto ty   sprashival* [*Tanya   li napisal-a chto za   vecher*]*?*
　　　　 what you   asked   Tanya   LI wrote-F   in   evening
　　　　 Lit.: 'What did you ask whether it was Tanya who wrote in one evening?'

Both experiments showed that the type of movement affects the acceptability of moving an element (direct object) out of yes/no indirect questions. At the same time, the highest scores are observed for sentences with relative movement, lower scores for sentences with *wh*-movement, and the lowest scores are for sentences with scrambling.

Pairwise comparison of the experiment results also showed that sentences with relative movement are rated significantly higher than sentences with *wh*-movement. The fact that relative pronouns are more freely moved out of indirect questions than *wh*-pronouns was also noticed in [Lyutikova 2009]. This result is predicted by the analysis in [Bailyn 2020], but this approach predicts

the same high scores for scrambling, which is not supported by the results of the experiment. In addition, the predictions of this analysis do not match the result of pairwise comparisons: with XP *li* configurations, the differences between *wh*-movement and scrambling turns out to be not significant.

Low scores for sentences with scrambling can be explained by the fact that such extractions must be communicatively motivated, which means that the context is extremely important for the respondents, which could have been chosen inaccurately or poorly perceived. It is also possible that low scores for sentences with scrambling are due to the fact that, out of the movements considered in the experiment, only scrambling is optional. It can then be assumed that the lower scores for this type of extraction are due to economic considerations.

Of the approaches discussed in Section 2, an analysis based on the empty category principle does not predict the contrast between the two configurations of yes/no indirect questions (V *li* vs. XP *li*) when extracting a direct object and the influence of the type of movement on its acceptability. Rizzi's theory outlined above predicts the discovered contrast between the configurations of the indirect question. J.F. Bailyn's approach based on "Featural" Relativized minimality expects differences between the types of movement, however, scrambling, contrary to our results, should receive the highest marks within this approach. The semantic approach of M. Abrusán considers *wh*-movement and is based on the semantics of questions. That is, this theory does not seem to make predictions about other types of movement. Although almost all of the results obtained (with the exception of the low scores for scrambling) were expected, none of the considered approaches to analyzing the weak island effects predict all the results simultaneously. Thus, a detailed analysis of the results remains for further research. It should be noted that all predictions were considered, taking the approaches to the analysis of the structure of the general indirect question presented in Section 2.2. Another subject of further research should be the consideration of predictions of theories of weak islands in other analyzes for the indirect question of whether (e.g. [Rudnitskaya 2000]).

Abbreviations

ACC — accusative; DAT — dative; F — female; INS — instrumental; LI — particle *li*; NEG — negation; SG — singular.

References

Lyutikova 2009 — Lyutikova E.A. Otnositel'nye predlozheniya s soyuznym slovom *kotoryi*: obshchaya kharakteristika i svoistva peredvizheniya. [Relative clauses with the conjunction word *kotoryj* 'which': General characteristics and movement properties]. Kiseleva K.L., Plungian V.A., Rakhilina E.V., Tatevosov S.G. (eds.). Korpusnye issledovaniya po russkoi grammatike [Corpus research in Russian grammar]. Moscow: Probel. P. 436–511.

Abels 2012 — Abels K. The Italian left periphery: A view from locality. Linguistic Inquiry. 2012. Vol. 43. No. 1. P. 229–254.

Abrusán 2007 — Abrusán M. Contradiction and grammar: The case of weak islands. Ph.D. thesis. Massachusetts Institute of Technology, 2007.

Bailyn 2020 — Bailyn J.F. The scrambling paradox. Linguistic Inquiry. 2020. Vol. 51. No. 4. P. 635–669.

Boeckx 2012 — Boeckx C. Syntactic islands. Cambridge: Cambridge University Press, 2012.

Chomsky 1981 — Chomsky N. Lectures on government and binding. Dordrecht: Foris Publications, 1981.

Chomsky 1986 — Chomsky N. Barriers. Cambridge, MA: MIT Press, 1986.

Dayal 1996 —Dayal V. Locality in WH Quantification. Dordrecht: Kluwer, 1996.

Drummond 2013 — Drummond A. Ibex Farm. 2013. http://spellout.net/ibexfarm/.

Dyakonova 2009 — Dyakonova M. A phase-based approach to Russian free word order. LOT Dissertation Series. Netherlands Graduate School of Linguistics, 2009.

Honcoop 1998 — Honcoop M. Dynamic excursions on weak islands. The Hague: Holland Academic Graphics, 1998.

Huang 1982 — Huang C.-T.J. Logical relations in Chinese and the theory of grammar. Ph.D. thesis. Massachusetts University of Technology, 1982.

King 1994 — King T.H. Focus in Russian yes-no questions. Journal of Slavic linguistics. 1994. P. 92–120.

Kroch 1989 — Kroch A. Amount quantification, referentiality and long *wh*-movement. Ph.D. thesis. University of Pennsylvania, 1989.

Rizzi 1990 — Rizzi L. Relativized minimality. Cambridge: MIT Press, 1990

Rizzi 1997 — Rizzi L. The fine structure of the left periphery. Elements of grammar. Dordrecht: Springer, 1997. P. 281–337.

Rizzi 2004 — Rizzi L. Locality and left periphery. Structures and beyond: The cartography of syntactic structures. Vol. 3. Belletti A. (ed.). Oxford: Oxford University Press, 2004. P. 223–251.

Ross 1967 — Ross J.R. Constraints on variables in syntax. Ph.D. thesis. Massachusetts Institute of Technology, 1967.

Rudin et al. 1999 — Rudin C., Kramer C., Billings L., Baerman M. Macedonian and Bulgarian li questions: Beyond syntax. Natural Language & Linguistic Theory. 1999. Vol. 17. No. 3. P. 541–586.

Rudnitskaya 2000 — Rudnitskaya E. Yes-no li questions in Russian: Interaction of syntax and phonology? Papers from the Poster Session of the 18th Annual West Coast Conference on Formal Linguistics (WCCFL 18). 2000. P. 89–98.

Schwabe 2004 — Schwabe K. The particle li and the left periphery of Slavic yes/no interrogatives. The Syntax and Semantics of the Left Periphery. Lohnstein H., Trissler S. (eds.). Berlin: Mouton de Gruyter, 2004. P. 385–430.

Sprouse 2018 — Sprouse J. Experimental syntax: Design, analysis and application. Lecture notes, 2018.

Szabolcsi 2006 — Szabolcsi A. 2006. Strong vs. weak islands. The Blackwell Companion to Syntax. Vol. 4. Everaert M., van Riemsdijk H. (eds.). Oxford: Blackwell, 2006. P. 479–532.

Szabolcsi, Zwarts 1993 — Szabolcsi A., Zwarts F. Weak islands and an algebraic semantics for scope taking. Natural Language Semantics. 1993. Vol. 1. No. 3. P. 235–284.

Анастасия Юрьевна Вознесенская

МГУ имени М. В. Ломоносова

Anastasiia Yu. Voznesenskaia

Lomonosov Moscow State University

anastvozn@gmail.com

Прототип метаграмматического описания трёхаргументных конструкций с морфологическим каузативом[*]

В. А. Генералова, С. Петижан
Университет им. Генриха Гейне в Дюссельдорфе

Возможность создания инструмента, позволяющего эффективно анализировать предложения на нескольких языках представляет интерес как для теоретической, так и для прикладной лингвистики. Одним из возможных подходов к этой задаче является создание формальных описаний языковых структур. В настоящей работе предлагается решение, позволяющее описывать синтаксическое строение трёхаргументных конструкций с морфологическим каузативом и его связь с их семантическим представлением. Анализ строится на формализованной версии грамматики Role and Reference Grammar. В качестве формального языка выбран инструмент eXtensible MetaGrammar (XMG), позволяющий описывать синтаксические деревья и семантические фреймы.

Ключевые слова: морфологический каузатив, актантная деривация, аргументная структура, формализованное описание, компьютерное представление.

A PROTOTYPE OF A METAGRAMMAR DESCRIBING THREE-ARGUMENT CONSTRUCTIONS WITH A MORPHOLOGICAL CAUSATIVE[*]

Valeria Generalova, Simon Petitjean
Heinrich Heine University of Dusseldorf

The idea of a tool for parsing several languages easily and in parallel is one of the goals for both theoretical linguistics and engineering. One approach would be to create formalized descriptions of linguistic structures. This paper suggests a solution for describing the syntactic organization and its linking to the semantic structure of three-argument constructions with morphological causatives. Our analysis grounds on formalized Role and Reference Grammar. As the description tool, we use eXtensible MetaGrammar (XMG), encoding syntactic structures in the form of trees and semantic structures as decompositional frames.

Keywords: morphological causative, verbal derivation, argument structure, formalized description, computational implementation.

1. Introduction

The research area of grammar engineering has grown popular with the overall digitalization of knowledge. There are various projects seeking to create tools for parsing and generating sentences in different languages. Although projects of formalizing individual languages' grammars are more numerous, those including cross-linguistic comparisons are receiving more attention in recent years.

Multilingual linguistic descriptions bridge theoretical knowledge with practical concerns. On the one hand, they help validating previously made observations. It is difficult to formalize a research area lacking consensus upon essential questions. Besides, the scrupulous attention to every detail imposed by the implementation requirements permits to inspect some rarely discussed parts of the theory and perhaps even expand on them. On the other hand, describing several languages in a single project helps to compare data. Moreover, these descriptions can be utilized as a source for other databases and tools.

A grammar engineering project requires several modules. First of all, it has to ground on a linguistic theory, preferably formalized enough to be used for implementation purposes. Secondly, it has to comprise descriptions of languages and/or specific constructions that will produce linguistic analyses. Thirdly, a compiler must be able to read the descriptions and create a desirable output.

Further applications of the generated analyses may be numerous. In an academic context, they can be used as illustrations for some theoretical advancements. One might find it useful to create specific databases of these analyses with an option to compare languages. In this case, the fact that the analyses have been produced within the same framework by the same tools would allow linguistic information to stand out clearly without being blurred by possible methodological differences. The generated analyses can be used in other projects such as corpora or treebanks to create syntactic and semantic annotations automatically. Originating from formal grammars designed by humans, they could surpass the accuracy of machine-learning algorithms. The restriction is the time spent on the development of the linguistic description and the limited coverage of most of the solutions of this kind.

The present paper is concerned with a single step of this pipeline, which is the creation of formalized descriptions from typological and language-specific data. This step is essential as it formalizes the existing information and presents

it in a machine-readable form. To ease its development and extension, the resource we develop is not a formal grammar but a more abstract description called metagrammar (see Section 2.4). The place the metagrammar occupies between theory and implementation determines the importance of its architecture. Therefore, we suggest a prototype of a novel architecture designed to minimize the efforts of grammar developers and protect the code from accidental errors.

The remainder of the paper is organized as follows. Section 2 presents the materials and methods used in our study. Primarily, it introduces the language data covered by our prototype. Besides, it gives an overview of Role and Reference Grammar, the theoretical foundation of this research, and of XMG, the language of our metagrammar. Section 3 reviews some existing projects providing multilingual grammatical descriptions. Section 4 demonstrates our solution, presenting various structures and the way they are related. The final Section 5 concludes the paper summarizing the key claims.

2. Materials and methods

In this section, we provide all the background information in order to make the presentation of our solution clear.

2.1. Necessary definitions

In the present paper, we investigate only constructions derived from (syntactically) transitive verbs by means of a morphological device, i. e., affix or clitic. We have narrowed the scope down to this set of constructions for several reasons.

Primarily, we are interested in modeling verbal derivations. To observe the derivation clearly, one would prefer three-argument constructions to two-argument ones. In most languages, there are quite a small number of underived three-argument constructions (e. g., ditransitives with verbs like *give*). Causative predicates based on transitive verbs appear to be three-argument predicates, suiting our goal well. (Once the valency increase mechanisms are well formalized and described, the solution could extend to encompass causatives of intransitives as well.)

However, sometimes causative constructions are bi-clausal, i. e., comprise two verbs, each with its own argument structure. This situation does not belong to the domain of valence-increasing devices, which we are interested in, so we do not cover this type of constructions (*periphrastic,* according to [Song 2013b]). On the other hand, we also eliminate lexical causatives as it is not always clear

whether a verbal derivation has taken place or not. In other words, it is not always clear from a verb whether it is a lexical causative, or a transitive demonstrating a causative alternation, or an underived ditransitive.

Saving these complicated issues for further research, we ground the solution we present in this paper on morphological causatives only. We use the definition of morphological causatives suggested by [Nedyalkov, Sil'nitskiy 1969]. They postulate an opposition V_i:V_j, where V_i denotes a situation, and V_j denotes the same situation but having been caused. The term morphological causative is used to denote the V_j, which is morphologically more complex than the V_i. This approach differs from what is suggested by [Song 2013a], where anticausatives (verbs of type V_i that are morphologically more complex than V_j) also count as a subtype of morphological causatives. We follow [Song 2013a], counting as morphological devices suffixes, prefixes and vowel change, as well as clitics.

Lastly, a terminological note about the participants of the three-argument causative construction has to be made. The added argument is most often called **causer**, whereas the former subject is usually called the **causee**. As for the former direct object, there is no consistent tradition of referring to it, as to our knowledge. In this paper, we have chosen the term **theme**, which is in line with the naming convention used for ditransitive constructions.

2.2. Language data

This paper does not present any quantitative typological observations; therefore, there is no need to operate on a typologically representative language sample. However, we aim to demonstrate different strategies of argument marking in three-argument causative constructions. We selected languages accordingly.

2.2.1. Bashkir

In Bashkir, there are two constructions with a morphological causative [Perekhval'skaya 2017], see (1) and (2). In both of them, the causer is nominative and (usually) clause-initial. The theme is accusative (which can sometimes have no overt marking) and usually precedes the verb. The marking of the cause is different: in (2) dative case is used, while in (1) — the ablative.

(1)   *Ataj   bala-lar-**ðan**   jeläk   jǝj-ðǝr-a*
      father   child-PL-**ABL**   berry   pick.up-CAUS-IPFV
      'Children are picking up berries because of the orders of the father.' Lit.: 'The father makes the children pick up berries.' [Perekhval'skaya 2017: 245]

(2) *Bala   besäj-**gä**   höt-tö   es-er-ä*
     child   cat-DAT   milk-ACC   drink-CAUS-IPFV
     'The child feeds the cat with the milk.' [Perekhval'skaya 2017: 241]

The dative case is used in Bashkir for marking recipients in ditransitive clauses (which is not unexpected from a case labeled dative). Ablative is used in a variety of contexts (mainly in complements, though), of which we would like to report one. In passive clauses with an explicit mention of the agent, it can be marked with ablative. Although the preferred way is not a pure ablative, but a prepositional phrase, the preposition *tarafənan* (which nearly corresponds to the English *by*) requires an ablative noun after it.

So, in our sample, Bashkir is a language with multiple causative constructions, one of which can be related to a ditransitive clause, and in the other one, the cause has the same marking as the demoted agent in a passive construction.

2.2.2. Kalmyk

Kalmyk is genetically close to Bashkir, and causative constructions in these languages are very much alike. Namely, Kalmyk also has two options, one being similar to ditransitive clauses (4), and the other relatable to the passive (3) clause. However, there is a different morphological case used in Kalmyk for both these options: instrumental, see (5).

(3) *ezə-n   ködəlməšč-**är**   xö   al-ulə-v*
     master-EXT   servant-INS   ram   kill-CAUS-PST
     'The master made the servant kill the ram.' [Say 2009: 387]

(4) *ekə   ürə-n-**d**-än   xašə   id-ül-ǯä-nä*
     mother   child-EXT-DAT-POSS.REFL   porridge   eat-CAUS-PROG-PRS
     'The mother feeds the child with porridge.' [Say 2009: 406]

(5) *üüidə-n   Badm-ar   sekə-gdə-v*
     door-EXT   Badma-INS   open-PASS-PST
     'The door was opened by Badma.' [Vydrina 2009: 348]

This slight difference between Bashkir and Kalmyk serves as an illustration of some of our decisions upon the architecture of the metagrammar.

In addition, Kalmyk has a strategy where both non-causer arguments are in the accusative case, see (6). It is important that there has to be an overt marker on the causee and no explicit marking on the theme.

(6) *bagšə*   *madn-**igə***   *škol-də*   ***kögǯmə***   *soŋs-ul-na*
     teacher   we-ACC   school-DAT   **music**   listen-CAUS-PRS
     'At school, the teacher made us listen to music.' [Say 2009: 411]

Without going deep into the discussion of this construction (compare [Say 2009] on Kalmyk with [Letuchiy 2006] on Khakas and [Kulikov 1998] on Tuvinian), we account for this syntactic pattern in our metagrammar as to one of the possible strategies of argument marking in the constructions in question. Causative construction with double accusative marking (or doubling of other morphological devices that are generally used to encode the direct object of a transitive clause) also occurs in other languages. The most prominent examples of them would be Bantu languages (e. g. [Baker et al. 2012]).

2.2.3. Lubukusu

Lubukusu (also known as Bukusu) is a representative of the Bantu family, well known for its valency-modifying constructions. It demonstrates a completely identical marking of both non-causer constituents in constructions with a morphological causative. Such are, apparently, most of the Bantu languages; consider examples (7)–(9) from [Baker et al. 2012: 54].

(7) *Wafula*   *a-nyw-esy-a*   *Wekesa*   *ka-ma-lwa*
     Wafula   SBJ.C1.TNS-drink-CAUS-FV   Wekesa   C6-C6-beer
     'Wafula made Wekesa drink beer.'

(8) *Wekesa*   *a-nyw-esy-ebw-a*   *ka-ma-lwa*
     Wekesa   SBJ.C1.TNS-drink-CAUS-PASS-FV   C6-C6-beer
     'Wekesa was made to drink beer.'

(9) *Ka-ma-lwa*   *ka-nyw-esy-ebw-a*   *Wekesa*
     C6-C6-beer   SBJ.C6.TNS-drink-CAUS-PASS-FV   Wekesa
     'Beer was made to be drunk by Wekesa.'

In (7), a causative construction is demonstrated. There is no overt case marking on either of the nouns, yet the subject is marked on the verb. Moreover, both non-subject participants of (7), namely the causee Wekesa (class 1) and the theme the beer (class 6), can become subjects of passive constructions, which is indicated by the class marker on the verb after the sbj morpheme in (8) and (9). So, in Lubukusu, the two constituents not only look but also behave in other syntactic contexts identically.

2.2.4. Kabardian

The causative construction in Kabardian somehow mirrors the Bantu construction with double accusative marking. In (10), both the causer and the causee receive ergative marking, while the theme is absolutive.

(10) *ł'əźə-m   ś'āla-m   pχa-r   yə-r-yə-ġa-qʷətaś*
     old.man-ERG   boy-ERG   tree-ABS   3SG-3SG-3SG-CAUS-cut-PRET-AFFIRM
     'The old man made the boy cut the tree.' [Matasović 2010: 50]

    This pattern seems to be rather rare ([Dixon 2000] also cites Trumai as having it; to some extent, Qiang can also count as one, see [LaPolla, Huang 2008]). However, there is no reason to discard it when describing typologically varied languages.

2.2.5. Nivkh

According to [Dixon 2000], Nivkh has a very rare type of causative constructions that involve a marker claimed to be unique to this construction only. In other words, the marker *-aχ-* is claimed to be used only in constructions with causatives derived from transitive bases, see (11). [Nedjalkov et al. 1969] also call this morpheme a special marker. However, [Gruzdeva 1998] demonstrates that the same morpheme appears on subjects of converbial clauses, see (12) and (13).

(11) *n'-nanx-∅*   *n'-aχ*   *pxi-roχ*   *vi-gu-d*
     1SG-elder.sister-NOM   1SG-DAT/ACC   forest-DAT/ADD   go-CAUS-FIN
     'My elder sister let me go to the forest.' [Gruzdeva 1998: 19]

(12) *hoĝat   n'yṇ   čyṇ   ṇyṇ-d'-ra*   *čyṇ-aχ*
     then   we   you   look.for-FIN-PART:PRED   you-DAT/ACC

     p'-ro-guin
     REFL-help-CVB:PURP
     'We were looking for you in order that you help [us].' [Gruzdeva 1998: 52]

(13) *if   imṇ-aχ   als   p'e-ny-vur   it-t'*
     he   they-DAT/ACC   berry   pick-FUT-CVB:RTL   say-FIN
     'He said [that] they would pick berries.' [Gruzdeva 1998: 57]

    Our solution suggests two ways of treating this strategy, either with relation to converbial clauses, or a totally independent marking pattern.

2.3. Role and Reference Grammar representations

Grammar engineering projects have to ground on some linguistic theories. We use RRG [Van Valin, LaPolla 1997], [Van Valin 2005] as it has been initially developed for typological purposes. Moreover, syntactic and semantic representations of RRG have been formalized by [Osswald, Kallmeyer 2018], which makes the implementation easier.

Syntactic representations in RRG are realized as trees; see Fig. 1. These types of structures are called the "layered structure of the clause" [Van Valin 2005: 3–4]. The upmost level of the tree is SENTENCE, which can comprise one or several CLAUSES. These notions are used in RRG in the same way as in other theories. Lower levels of the tree are specific to RRG. The CORE encompasses the predicate (which is labeled NUCLEUS), and its arguments are usually realized as RPs. Non-arguments (adjuncts) are placed outside of the CORE, in one or several PERIPHERY nodes. In this paper, we look only at monoclausal core structures, so the representations lack some levels for the sake of brevity.

Semantic representations in classical RRG are realized as predicate logic structures [Van Valin 2005: 42]. In the formalized version, decompositional frames are used [Lichte, Petitjean 2015] as they contain the maximum of the necessary information in a concise form. Besides, deriving frames from one another is easy, which is crucial for implementation purposes. Some sample frames are shown in Fig. 1: each of the (sub) frames starts with the type of the predicate and contains its role structure. Arguments of the semantic structure of the predicate are referenced as labels. The same labels are used in the syntactic structure, allowing comprehensive syntax-semantics linking.

2.4. XMG language

A metagrammar [Candito 1996] is an abstract and compact description of a grammar. The key idea of a metagrammar is to express linguistic generalizations in order to factorize the redundant parts of the grammar (e. g. see discussion in [Clément, Kinyon 2003]). Therefore, developing a metagrammar comprises two main phases: describing smaller units (*rule fragments*) and assembling them to form the grammar rules. Once compiled, a metagrammar produces the underlying grammar, which can be used for parsing or language generation.

The framework eXtensible MetaGrammar (XMG, [Crabbé et al. 2013], and XMG-2, [Petitjean et al. 2016] offers a description language that allows gener-

ating tree grammars based on any linguistic theory. This metagrammatical language is inspired by constraint programming, for the definition of the rule fragments, and logic programming, for assembling the fragments into rules.

A rule fragment, encapsulated into a unit called *class*, can be composed of a syntactic structure (a partial tree), a semantic frame fragment, or both. The description of a tree fragment consists of the enumeration of dominance and precedence relations, which have to hold between the nodes composing it. The semantic frames are described using typed feature structures [Lichte, Petitjean 2015]. Classes are organized in an inheritance hierarchy, which determines how the fragments should be assembled to form the grammar rules. The creation of the class hierarchy is made possible by three mechanisms: inheritance, conjunction and disjunction. When inheriting a class, its entire content is added to the description of the current class. The operation of conjunction is used to inherit the content of several classes cumulatively, whereas disjunction allows expressing alternatives between different classes to be imported.

Conjunction and disjunction can also be used within classes to express combinations and alternatives at a more fine-grained level (between partial syntactic descriptions, for instance). As disjunction introduces non-determinism, XMG uses backtracking to generate the rules resulting from all possible alternatives expressed in the classes. In our study, we use this mechanism to account for several typological varieties of a given construction (see Section 4.1).

In addition to this minimal logic language and backtracking, XMG borrows from logic programming tools (such as Prolog) the extensive use of unification. All structures described in classes (syntactic nodes, feature structures) can be associated with Prolog-like variables. Unifying these variables triggers the unification of the structures, allowing fragments to complete the linguistic information missing in others. This process is key to combining classes of several kinds in our architecture; see Section 4.3.

3. Review

Previously, there have been attempts to create formalized descriptions valid for multiple languages to allow further implementations. In this section, we would like to broadly review some of the key papers having a significant influence on the conception of our solution.

Several approaches to creating multilingual metagrammars have been suggested by Alexandra Kinyon and her collaborators. Namely, [Clément, Kinyon 2003]

suggest generating LFG (Lexical Functional Grammars) and TAG (Tree-Adjoining Grammars) for French and English. This paper also has a methodological discussion concerning the advantages of the metagrammatical approach over redundancies brought by LFG rules. In a later paper, [Kinyon et al. 2006] demonstrate a broader coverage of languages, including five Germanic languages and Kashmiri, yet the discussion in the paper is limited to German and Yiddish only. The paper is technically driven and demonstrates a thorough implementation of early XMG principles. The feature structures and the mechanisms of their unification are developed in much detail.

A totally different in terms of theoretical foundations and exploited tools is the LinGO Grammar Matrix project, running since 2002 at the University of Washington (see [Bender et al. 2002]). The descriptions are inspired by HPSG structures, which drives the whole analysis more towards the semantic side, letting syntactic representations look secondary. Nevertheless, the coverage of the studies related to this project is very impressive.

Among others, [Curtis 2018] is of major importance for the present research. This paper describes a library (a freely addable module of the Grammar Matrix) covering valence-modifying constructions. The central phenomenon that is modeled is the application of lexical rules. The results are illustrated with data from 8 typologically varied languages. As already mentioned, the syntactic side of the analysis is somehow shaded by the semantic one, which we consider a drawback in comparison to our approach. Also, the typological review suggested in [Curtis 2018: 13–22] seems to capture most general trends overlooking some minor varieties.

Although the realization of grammatical phenomena in our solution can hardly be compared to the above-mentioned due to differences in theoretical approaches, the very general architecture is similar. In the Grammar Matrix, there is a questionnaire eliciting all the necessary data about the grammar of a particular language. This module corresponds to the Language plugin notion (see Section 4.2). We find it advantageous to our plugins that they require less specific information than [Curtis 2018: 40] and thus are accessible to a broader audience of grammar developers.

Nevertheless, this implementation is a high-standard solution, especially powerful as a part of a larger successful project. Taking the best from it and from theories presented in Section 2, our prototype tends at least to show an equally great potential along with some keen attention to previously overlooked details.

4. Solution

This paper aims to present a solution for describing three-argument constructions with morphological causatives on the metagrammatical level. This solution has to cover a wide range of constructions encountered in languages of the world (see Section 2.2) and be expandable onto more languages once needed.

In our solution, we would like to keep cross-linguistic generalizations severed from language-specific information. This general architecture is critical for creating a solution that can be expanded in both dimensions and cover more constructions and more languages. In this section, we first present the design of metagrammatical classes describing constructions, called **construction classes**. The same classes are used for describing all languages in the sample and can be reused when other language phenomena are covered. Afterward, we present metagrammatical classes where language-specific information is stored, the so-called **language plugins**. All of them are designed in a similar way to ensure compatibility with each other and with any construction class. Finally, we couch in detail the mechanism of bringing classes of both types together to parse each particular sentence in each particular language correctly.

For the sake of clarity, we present as illustrations only prototypical situations. Namely, we talk about morphological cases, but in reality, other devices, such as clitics or vowel changes, are also covered. The full version of the metagrammar described in this paper is made available online[1].

4.1. Construction classes

Our metagrammar accounts for one single construction, a three-argument causative. Therefore, there is only one construction class. However, it is not composed from scratch but built of other classes describing construction fragments.

First, we would like to consider the structure of the causative predicate. It is made of at least two morphemes: the lexical root (or stem) and the causative affix. Each of these morphemes has its own structure. Each lexical root has its valency and role structure. These are used in non-derived verbal constructions (e. g. in a simple active non-causative clause). The causative affix always has the same frame structure, as it adds the cause subframe to the causation frame. It has no own syntactic structure, but it increases the valency of the whole sen-

---

[1] Please visit the GitHub repositrory for the actual version of the metagrammar: https://github.com/fleurdhelice/TMP2020_demo.

tence. In terms of RRG tree grammar, it creates a new RP node below the core. The addition of the causative affix is viewed as a regular grammatical operation of valency increase and applies to any base verb. Any constraints on this procedure encountered in individual languages will be brought in by Language plugins (see Section 4.2).

In Fig. 1, trees and frames are shown for sample intransitive, transitive and ditransitive verbs. Note that the syntactic structure of the predicate is always the same, but the semantic representations vary.
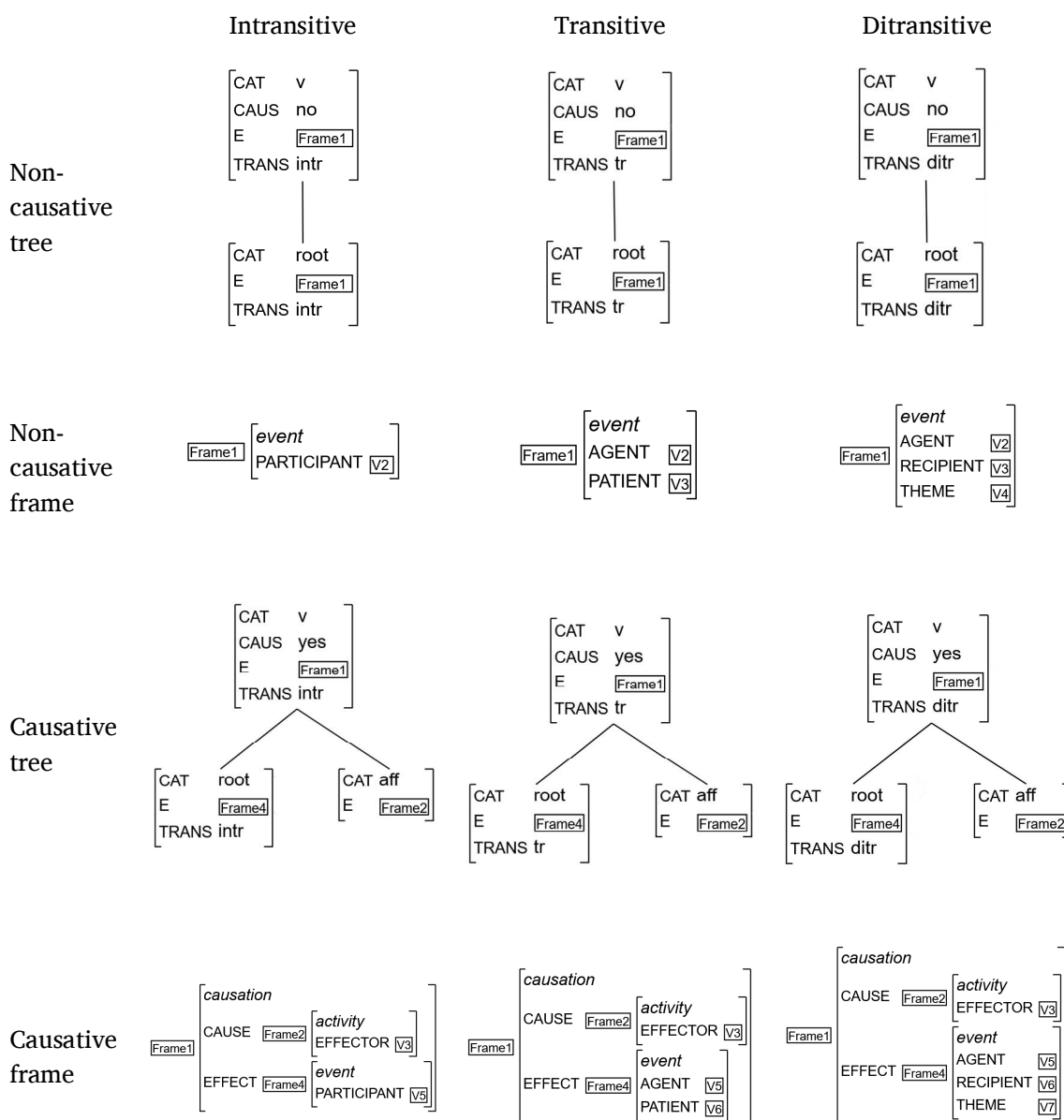

Figure 1. Syntactic and semantic structures generated by the class *Verb*

Now, let us consider the derivation of the syntactic structure. In Fig. 2, the syntactic representation of a transitive clause is shown. The two RP nodes are the subject and the direct object. Each of them has an identifier (the *i* feature). These labels are used to reference specific nodes within frames and other classes. Another feature defined for an RP is the morphological case. In Fig. 2, assigned cases are not visible. However, in the code, the label *V7* always corresponds to the case used for the syntactic subject (in RRG terms, the PSA, privileged syntactic argument) and the label *V8* — for the direct object. So, the morphological case assigned to each constituent in the non-causative transitive clause is strictly defined in the construction class.
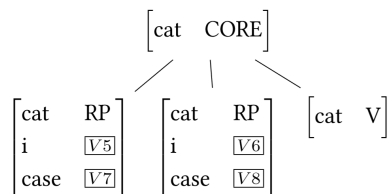
Figure 2. The syntactic tree
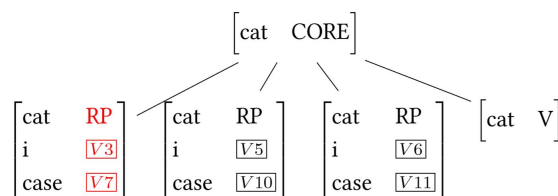for a non-causative transitive clause

Figure 3. The syntactic tree
for a causative of a transitive clause

The situation with causative constructions is not that straightforward, as there are numerous strategies for marking the arguments. Fig. 3 shows a causative construction derived from Fig. 2. There is one RP added (see new identifier). This RP corresponds to the causer. It has been generally attested that in all languages, the causer receives the case that is used for a PSA in a non-causative transitive clause. This case is assigned to the new RP as soon as it appears in the clause (consider label *V7* on the new constituent). As for the other two RP nodes, it is not clear yet which cases they are assigned. It depends on individual language properties and thus cannot be fixed within the construction class. For this reason, free variables are assigned as a value of the case feature to these constituents. At a later stage, they unify with case values imposed by other constraints.

To this point, we have shown universal structures. There are, however, some characteristics of the constructions in question that hold for only some groups of languages. On the one hand, we cannot have them fixed as, for example, the morphological case of the causer, because there are several ways how each of these characteristics can be realized. On the other hand, there are salient linguistic generalization, grouping languages according to the value of these characteristics. Given that, our solution is demanded to account for some variation dependent on a particular characteristic.

To handle this, we suggest using a disjunction inside the universal construction class. For each alternative listed in the disjunction, XMG creates a new possible syntactic structure for the grammar, resulting from the combination of the existing description and the one expressed in the alternative. Whether the resulting structure will be part of the grammar or not depends on further feature unification. All the information that is used universally (e. g., node labels or information about morphological cases) is put in the first part of the class description. Afterward, there is a set of disjoint claims. Each component of disjunction can comprise one or more conjoint claims. In the first claim, the feature across which the constructions vary. In the next claims of each of the disjoint components, the specific properties of each variety are described.

A clear example of this kind of characteristics is the word order. Indeed, there is no universal word order, but the number of groups a language may fall to is fixed and small. In the sample listed in Section 2.2, there are only two word orders: SVO (Lubukusu) and SOV (all the others). So, we have to account for two disjoint options in the metagrammar.

In the description of any tree, two main types of constraints determine its shape: dominance and precedence. In all causative constructions examined in the present paper, the three RP nodes are dominated by the core node. So, the dominance constraints are formulated in the common part of the class. As for precedence, the constraints are different. In Fig. 4, one can see two components that are disjoint and the syntactic structures that result in each case (the operator "≫" indicates linear precedence).
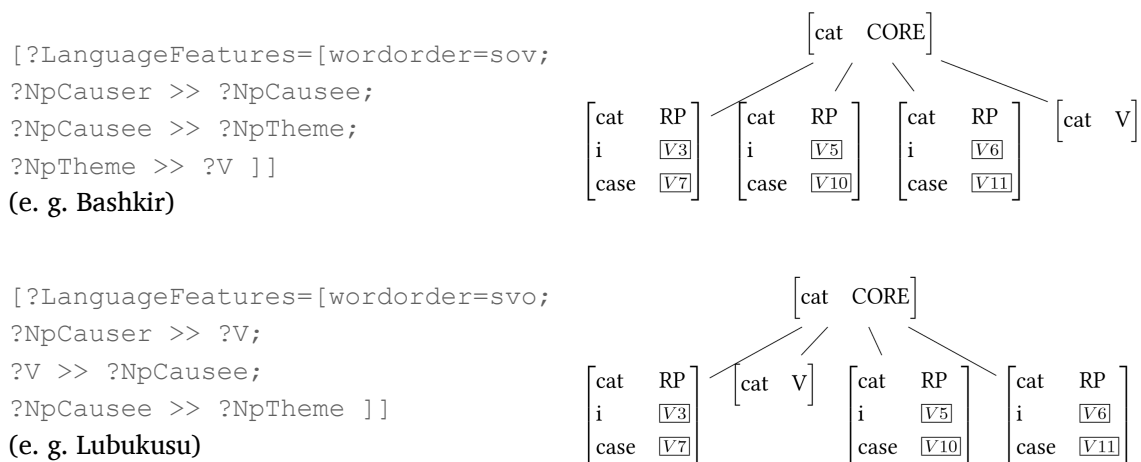

Figure 4. The example of disjoint structures

In each language plugin, this feature is assigned one single value. Once the construction class is intersected with the language plugin (see Section 4.2), the single option is chosen, and the sole description of the construction is applied to the sentence.

4.2. Language plugins

Apart from Construction classes, our metagrammar comprises Language plugins. In these XMG classes, there are no syntactic or semantic descriptions. Instead, there is one variable *LanguageFeatures* with a wide range of features. The set of features is supposed to be fixed and identical for any language. This would allow us to intersect any Language plugin with any Construction class. Having language properties as features is advantageous for one more reason. All the feature values are type-checked, which minimizes input errors.

4.2.1. Inventory features

Language plugins have two main objectives. Firstly, they have to list constructions available in each language. As shown in Section 2.2.1, there can be several causative constructions within the same language. Typological studies (e. g. [Dixon 2000]) show that there is a closed list of constructions with different marking patterns. So, they can be formulated as boolean features, each receiving positive or negative value.

Importantly, we would like to relate marking in causative constructions to marking in other constructions. This is necessary to keep patterns of constructions apart from the morphological peculiarities of each particular language. For example, both Bashkir and Kalmyk have a construction, where the causee receives the same morphological case as the demoted agent in a passive construction. Therefore we postulate a single boolean feature that stands for this kind of construction and set its value to positive for both Bashkir and Kalmyk. In Fig 5., one can see variable names and their values for each language.

In the Construction class, there is a disjunction encoding the causee marking depending on the value of the above-described boolean features, see Fig. 6. When the RP for the causee is declared in the construction class, it is assigned a free variable `?VarCauseeCase`, which serves as a label before further information is given. In the disjunction part, each conjoint claim consists of the postulation of the positive value of one boolean feature and the reassignment of

the value of the variable `?VarCauseeCase`. This is done by pulling the respective feature from the variable `?LanguageFeatures`, which is declared in language plugins. The dot operator is used to access the value of a given attribute in a feature structure introduced by a variable. No overt assignment of morphological cases is done at this point because they differ across languages. However, the marking of the causee is related to features encoding the marking of other syntactic positions in each given language.

Feature	Description of the marking pattern	Bashkir	Kalmyk	Lubukusu	Kabardian	Nivkh
CauseeLikeRecip	causee is marked identically to the recipient in a ditransitive clause	+	+			
CauseeLikeDemAg	causee is marked identically to the demoted agent in a passive clause	+	+			
CauseeLikeDo	causee is marked identically to the direct object in a transitive non-causative clause		+	+		
CauseeLikePsa	causee is marked identically to the PSA in a transitive non-causative clause				+	
CauseeSpecial	causee receives a special marker					+

Figure 5. Boolean features for varieties of causative constructions

```
class CausativeConstruction
(...)
{<syn>{(...);
    node NpCausee [case=?VarCauseeCase]; (...);

    {?LanguageFeatures.ConstrCauseeLikeRecip = yes;
    ?VarCauseeCase=?LanguageFeatures.recipCase}

    | {?LanguageFeatures.ConstrCauseeLikeDemAg = yes;
    ?VarCauseeCase=?LanguageFeatures.demAgCase}

    | {?LanguageFeatures.ConstrCauseeSpecial = yes;
    ?VarCauseeCase=?LanguageFeatures.causeeCase};
    (...) }}
```

Figure 6. A piece of the code demonstrating the disjunction between boolean features

One might note the last disjunction option listing the feature `causeeCase`, which is not related to any of the other constructions. Indeed, this is the fallback strategy for languages lacking data about some constructions or having a really particular marker for the causee. In our sample, Nivkh is an example of such a language. If we follow [Nedjalkov et al. 1969], we have to postulate the exceptional causee marker, which is not used elsewhere, and thus activate this fallback strategy. However, if we follow [Gruzdeva 1997], we could add a variety "causee marked in the same way as the subject of the converbial clause" and a feature `cvSubCase` to account for the morpheme -*aχ*-. Once this strategy is encountered in another language of the world, our metagramar is already able to handle it. Nevertheless, given the lack of data, we do not eliminate the fallback option from the list of disjunctions.

4.2.2. Morphological features

Encoding morphological properties of a language is the second main objective of a language plugin. The formalization work in this area consists of converting traditional grammar descriptions into typed features. Namely, for the purposes of this study, one needs to list all morphological cases (or other devices used for argument marking) and relate them to their most common usage contexts.

Here is a sample list of these features:

`psaCase` — the case used for the privileged syntactic argument (syntactic subject) in an active transitive clause;

`doCase` — the case used for the direct object of a transitive clause;

`recipCase` — the case used for marking Recipient in a ditransitive clause;

`demAgCase` — the case for encoding the demoted agent in a passive transitive clause, etc.

Each feature of this kind takes one morphological case from the list available in this language as its value.

Identical names of the features allow accessing them from any construction class and make syntactic structures independent from morphological input. For example, values of the feature `demAgCase` are different in Bashkir and Kalmyk, but both of them are invoked within the same construction variety, namely, the `ConstrCauseeLikeDemAg` (see Fig. 5 and Fig. 6).

4.2.3. Summary

The Language plugins in our architecture comprise one single structure (accessible via a variable) with a great number of type-checked features. Boolean features affirm or negate the presence of a given variety of the causative construction in a language. In the expanded grammar, one would add similar lists for other types of constructions covered. Categorial features encode the morphology of a language. Determining the inventory of necessary attributes and their relation to each other requires typological knowledge. Once the templates for language plugins are set, values to these features can be assigned by language specialists as long as this operation does not require advanced linguistic knowledge. This accessibility makes the barrier of entry to the metagrammar relatively low, and the potential of the metagrammar to expand on a sound number of languages relatively high.

4.3. Intersection

The final step leading to creating descriptions applicable to real sentences is the intersection of a construction class with a language plugin. At this stage, all feature values unify, and all slots left empty receive a determined value. As for disjunctions, only those parts are left that comprise statements about the features conforming to the values of these features in the Language plugin.

Technically, this is realized as *inheritance* between classes. Under this term, we understand the idea of acquiring the whole structure of one class to add something on top of it within another class. For that to be possible, structural parts of a class (variables, nodes etc.) have to be named and exported. Two entities with the same name automatically count as a single one, making it possible to put additional constraints in a new class on a node imported from elsewhere. As for the features, their values unify in that respect that a determined value (the one listed in the preamble of the metagrammar as acceptable) percolates through structures and substitutes all the variables that have been referencing it. In other words, all general and placeholder values are substituted by a determined value if it is assigned elsewhere in the inherited or inheriting classes. In case of a conflict of two determined values, no unification happens.

The intersection of a construction class and a language plugin is an extreme case of inheritance. The resulting class provides no additional descriptions, apart from importing the two classes. One might consider this step as pure unification performed apart from defining the structures. The intersected classes can be created automatically, leaving this process free from possible human errors.

5. Conclusion

We have suggested a solution that presents linguistic analyses of sentences in two dimensions, creating syntactic and semantic representations. It requires only a sentence in a language as an input to generate complete analyses based on theoretical and typological reasoning. It captures cross-linguistic generalizations due to the architecture of Construction classes, which is designed to account for many varieties of structures from the beginning. The flexibility and the modularity of the architecture are provided by the XMG language and ensure easy further development.

The current prototype of the solution has no power in terms of quantitative typology, although it might turn out to be a useful tool once more languages are covered. To add a new language in the metagrammar, one needs to know what causative constructions are allowed and how morphological cases are used.

The overall idea of formalizing a well-studied field of typological knowledge has brought its insights, too. Namely, increasing the valence of the sentence has required different solutions for the syntactic and the semantic dimensions. This underlines the importance to distinguish between syntactic and semantic valency in non-computational descriptions of languages as well. Besides, the atomic approach to causative constructions has revealed some overlooked areas. To our knowledge, little has been done in the direction of relating marking in causatives to other (except for ditransitive) constructions in a language. We find that a thorough look into these relations could tell more about what axes of similarity (cf. comparative concepts by [Haspelmath 2010]) are there in typologically varied languages.

Abbreviations

3 — 3rd person; ABL — ablative; ABS — absolutive; ACC — accusative; ADD — additive; AFF — affix; AFFIRM — affirmative; C1 — noun class 1; C6 — noun class 6; CAUS — causative; CVB — converb; DAT — dative; DITR — ditransitive; ERG — ergative; EXT — stem extension; FIN — finite; FUT — future; FV — final vowel; INS — instrumental; INTR — intransitive; IPFV — imperfective; NOM — nominative; PART — particle; PASS — passive; PL— plural; POSS — possessive; PRED — predicative; PRET — preterite; PROG — progressive; PRS — present; PST — past; PURP — purpose; RP — reference phrase; RTL — re-telling; SBJ — subject; SG — singular; TNS — tense; TR — transitive.

References

Baker et al. 2012 — Baker M, Safir K, Sikuku J. Sources of (a)symmetry in Bantu double object constructions. Proceedings of the 30th West Coast Conference on Formal Linguistics, 2012. P. 54–64.

Bender et al. 2002 — Bender E.M., Flickinger D., Oepen S. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics. Carroll J, Oostdijk N, Sutcliffe R. (eds.) Taipei, Taiwan, 2002. P. 8–14.

Candito 1996 — Candito M. A principle-based hierarchical representation of LTAGs. Proceedings of the 16th International Conference on Computational Linguistics, 1996. P. 194–199.

Clément, Kinyon: 2003 — Clément L, Kinyon A. Generating parallel multilingual LFG-TAG grammars from a metagrammar. Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, 2003. P. 184–91.

Crabbé et al. 2013 — Crabbé B., Duchier D., Gardent C., Roux J.L., Parmentier Y. XMG: extensible metagrammar. Computational Linguistics, vol. 39(3), 2013. P. 591–629.

Curtis 2018 — Curtis C.M. A parametric implementation of valence-changing morphology in the LinGO Grammar Matrix. Master's thesis; University of Washington, 2018.

Dixon 2000 — Dixon R.M.W. A typology of causatives: form, syntax and meaning. Changing valency: Case studies in transitivity. Dixon R.M.W., Aikhenvald A.Y. (eds.). Cambridge, Cambridge University Press; 2000. P. 30–83.

Gruzdeva 1998 — Gruzdeva E. Nivkh. Languages of the World/Materials, vol. 111. Lincom Europa, München and Newcastle, 1998.

Kinyon et al.: 2006 — Kinyon A., Rambow O., Scheffler T., Yoon S.W., Joshi A. The metagrammar goes multilingual: A cross-linguistic look at the V2-phenomenon. Proceedings of the Eighth International Workshop on Tree Adjoining Grammar and Related Formalisms, 2006. P. 17–24.

Kulikov 1998 — Kulikov L. Causative constructions in Tuvinian: Towards a typology of transitivity. The Mainz Meeting: Proceedings of the Seventh International Conference on Turkic Linguistics. Johanson L, et al. (eds.). 1998. P. 258–64.

LaPolla 2008 — LaPolla R.J., Huang C. A grammar of Qiang: With annotated texts and glossary. Mouton de Gruyter, 2008.

Letuchiy 2006 — Letuchiy A. Case marking, possession and syntactic hierarchies in Khakas causative constructions in comparison with other Turkic languages. Case, Valency and Transitivity. Kulikov L., Malchukov A., de Swart P. (eds.), vol. 77, 2006. P. 417–441.

Lichte, Petitjean 2015 — Lichte T, Petitjean S. Implementing semantic frames as typed feature structures with XMG. Journal of Language Modelling, vol. 3(1), 2015. P. 185–228.

Matasović 2010 — Matasović R. Kabardian causatives, reflexives, and case marking domains. Suvremena lingvistika, vol. 36(69), 2010. P. 45–64.

Nedjalkov et al. 1969 — Nedyalkov V.P., Otaina G.A., Xolodovič A.A. Morfologičeskiy i leksičeskiy kauzativy v nivxskom jazyke [Morphological and lexical causatives in Nivkh]. Typology of causative constructions: Morphological causative. Xolodovič A.A. (ed.). Leningrad, Nauka, 1969. P. 179–199.

Nedyalkov, Sil'nitskiy 1969 — Nedyalkov V.P., Sil'nitskiy G.G. Tipologija morfologičeskogo i leksičeskogo kauzativov [Typoogy of morphological and lexical causatives]. Typology of causative constructions: Morphological causative. Xolodovič A.A. (ed.). Leningrad, Nauka, 1969. P. 20–50.

Osswald, Kallmeyer 2018 — Osswald R, Kallmeyer L. Towards a formalization of role and reference grammar. Applying and expanding Role and Reference Grammar (NIHIN Studies). Kailuweit R, Staudinger E, Künkel L. (eds.) Freiburg, 2018. P. 355–78.

Perekhval'skaya 2017 — Perekhval'skaya E.V. Kauzativnye konstrucii v baškirskom jazyke [Causative constructions in Bashkir]. Acta Linguistica Petropolitana Works by Institute of Linguistic Studies, vol. 13(1), 2017. P. 231–254.

Petitjean et al. 2016 — Petitjean S., Duchier D., Parmentier Y. XMG-2: describing description languages. International Conference on Logical Aspects of Computational Linguistics. Springer, 2016. P. 255–72.

Say 2009 — Say S.S. Argumentnaja struktura kalmyckix kauzativnyx konstrukcij [Argument structure of Kalmyk causative constructions]. Acta Linguistica Petropolitana Works by Institute of Linguistic Studies, 2009. P. 387–464.

Song 2013a — Song J.J. Non-periphrastic causative constructions. The World Atlas of Language Structures Online. Dryer M.S., Haspelmath M. (eds.). Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

Song 2013b — Song J.J. Periphrastic causative constructions. The World Atlas of Language Structures Online. Dryer M.S., Haspelmath M. (eds.). Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

Van Valin: 2005 — Van Valin R.D. Exploring the syntax-semantics interface. Cambridge University Press, 2005.

Van Valin, LaPolla 1997 — Van Valin R.D., LaPolla R.J. Syntax: Structure, meaning, and function. Cambridge University Press, 1997.

Vydrina 2009 — Vydrina A. Upotreblenie pokazatel'a passiva v kalmyckom jazyke [Usage of passive marker in Kalmyk]. Acta Linguistica Petropolitana Works by Institute of Linguistic Studies, 2009. P. 347–386.

Валерия Антоновна Генералова

аспирант, Университет им. Генриха Гейне в Дюссельдорфе; участник проекта TreeGraSP, поддерживаемого грантом ERC Consolidator grant

Valeria Generalova

Ph.D. student, Heinrich Heine University of Düsseldorf; performer of the TreeGraSP project funded by the ERC Consolidator grant

generalo@hhu.de

Симон Петижан

Ph.D.; постдок, Университет им. Генриха Гейне в Дюссельдорфе; участник проекта TreeGraSP, поддерживаемого грантом ERC Consolidator grant

Simon Petitjean

Ph.D.; postdoctoral researcher, Heinrich Heine University of Düsseldorf; performer of the TreeGraSP project funded by the ERC Consolidator grant

petitjean@phil.uni-duesseldorf.de

Семантика и селективные ограничения лёгкого глагола *pǝrak* 'бросить' в пошкартском диалекте чувашского языка[*]

Ф. В. Голосов

Национальный исследовательский университет «Высшая школа экономики»

Статья посвящена семантики и дистрибуции лёгкого глагола *pǝrak* 'бросить' в пошкартском диалекте чувашского языка. В статье даётся детальное описание селективных ограничений лёгкого глагола и акциональных свойств образованных сложных глаголов. Кроме того, автор предлагает вариант формального структурно-событийного анализа конструкций с *pǝrak,* предсказывающий описываемые сочетаемостные особенности лёгкого глагола и его акциональную семантику.

Ключевые слова: сложные глаголы, лёгкие глаголы, сериализация, грамматикализация глагола со значением 'бросить', акциональность, структура события, чувашский язык.

SEMANTICS AND SELECTIONAL RESTRICTIONS OF THE LIGHT VERB *PƏRAK* 'THROW' IN POSHKART CHUVASH[*]

Fedor Golosov

National Research University Higher School of Economics

The article is dedicated to semantics and distribution of the light verb *pǝrak* 'throw' in Poshkart Chuvash. Author gives a detailed description of selectional restrictions of the light verb and actional properties of the formed complex predicates. Besides that, author suggests possible formal event-structure analysis of the constructions with *pǝrak* which would predict both established distributional properties of the light verb and its actional meaning.

Keywords: complex predicates, light verbs, serialization, grammaticalization of a verb with meaning 'throw', actionality, event structure, Chuvash.

1. Введение

В чувашском языке есть так называемые сложные глаголы — конструкции, состоящие из двух глагольных форм, но описывающие одно событие. Пример сложного глагола можно видеть в (1):

(1)   *vaçə*   **ç*ur-za***   ***ka-r*^{j}*-ə***
       Вася   спать-CV_SIM   уйти-PST-3SG
       'Вася **уснул**.'

Сложный глагол состоит из деепричастия, задающего лексические свойства ситуации (далее **лексического глагола**), и главного глагола, выступающего в роли структурно-семантического модификатора (далее **лёгкого глагола**). Так, в примере (1) сложный глагол *çurza kaj* 'уснуть' состоит из лексического глагола *çur* 'спать', описывающего состояние сна, и лёгкого глагола *kaj* 'уйти', выступающего в качестве акционального модификатора, образующего сложные глаголы, описывающие вхождение в некоторое состояние. В результате взаимодействия частей сложный глагол получает значение 'войти в состояние сна', то есть 'уснуть'.

В чувашском языке, как и во многих языках тюркской группы [Гращенков 2015], грамматикализовалось большое количество лёгких глаголов, функционирующих как акциональные, директивные и валентностные операторы, см, например, [Шлуинский 2006; Лебедев 2016; Golosov (in press)]. Настоящая статья посвящена семантике и дистрибуции лёгкого глагола *pərak* 'бросить' в пошкартском диалекте чувашского языка, образующего предельные сложные глаголы:

(2) a. *vaçə*   *avdan-a*   ***pёr   minut   xoʂ-in-ʥe***
         Вася   петух-OBJ   1   минута   период-P_3-LOC

         pula-za   *pərak-r*^{j}*-ə*
         душить-CV_SIM   бросить-PST-3SG
         'Вася задушил петуха **за минуту**.'

     b. **vaçə*   *avdan-a*   ***pёr   minut   xoʂ-i***
         Вася   петух-OBJ   1   минута   период-P_3

         pula-za   *pərak-r*^{j}*-ə*
         душить-CV_SIM   бросить-PST-3SG
         Ожид. зн.: 'Вася подушил петуха **минуту**.'

Материал для нашего исследования был собран в ходе экспедиции в село Малое Карачкино (Пошкарт) Ядринского района республики Чувашия в марте 2020 года.

Структура статьи имеет следующий вид. За вводным разделом 1 следует раздел 2, посвящённый краткому обзору литературы по лёгкому глаголу *pərak*. В разделе 3 мы приведём систематизированное описание данных, а в разделе 4 предложим формальный анализ сложных глаголов с *pərak*. Наконец, раздел 5 будет посвящён общим выводам.

2. Обзор литературы

Впервые чувашский глагол *пӑрах*[1] 'бросать' упоминается в работе Н. И. Ашмарина [1898], где перечисляется среди вспомогательных глаголов, передающих значение русского вида. Более подробный комментарий Ашмарин даёт в своей более поздней работе [Ашмарин 1923], в которой он пишет, что глагол *пӑрах* 'бросать' «сообщает смысл бросания, швыряния, резкого, грубого или поспешного выполнения действия, иногда также указывает на что-либо, случающееся некстати» [ibid: 49].

Лёгкий глагол *пӑрах* упоминается и в монографии Э. Е. Лебедева [2016], в которой автор подробно рассматривает различные акциональные значения лёгких глаголов, обращая внимания на то, как они ведут себя в сочетании с разными семантическими классами лексических глаголов. Глагол *пӑрах* 'бросить', согласно Лебедеву, имеет фазовые значения и сочетается с глаголами двух семантических групп: глаголами физического воздействия на объект (например, *пӑв* 'душить') и глаголами «чувственного» воздействия на объект (например, *юрат* 'любить').

Таким образом, благодаря уже имеющейся литературе, посвящённой лёгкому глаголу *пӑрах* 'бросить', мы можем предположить, что он грамматикализовался в некоторый аспектуальный показатель, который сочетается с двумя группами глаголов: глаголами физического и эмоционального воздействия на объект, а также передаёт значение быстроты действия и обладает негативными коннотациями. Однако изложенные выше обобщения сформулированы достаточно общо и нестрого, кроме того, никакие выводы о селективных ограничениях не подкрепляются отрицательным

---

[1] Ввиду фонетических различий между пошкартским и литературным чувашским и разницы в конвенциональной записи примеров литературный вариант глагола *pərak* выглядит как *пӑрах*.

материалом. Наконец, все упомянутые выше работы имеют скорее описательный характер и не ставят перед собой цели предложить теоретическое осмысление полученных данных.

В нашей работе мы постараемся продвинуться именно в этих двух направлениях, а именно предложим более детальное описание селективных ограничений лёгкого глагола *pərak* 'бросить' и его семантики и построим структурно-событийный формальный анализ сложных глаголов с *pərak*, который бы предсказывал установленные лексические ограничения лёгкого глагола и акциональные свойства образованных сложных глаголов.

3. Данные[2]

3.1. Общие сведения

Как уже было упомянуто во введении, лёгкий глагол *pərak* — телисизатор, он образует предельные сложные глаголы, повторим контраст, приводимый ранее в (2):

(3) a. *vaɕə avdan-a pёr minut xoʂ-in-ʥe*
Вася   петух-OBJ   1   минута   период-P_3-LOC

     pula-za pərak-rʲ-ə
     душить-CV_SIM   бросить-PST-3SG
     'Вася задушил петуха **за минуту**.'

   b. **vaɕə avdan-a pёr minut xoʂ-i pula-za pərak-rʲ-ə*
     Вася   петух-OBJ   1   минута   период-P_3   душить-CV_SIM   бросить-PST-3SG
     Ожид. зн.: 'Вася подушил петуха **минуту**.'

---

[2] Все примеры были собраны методом элицитации: мы либо просили консультанта перевести русское предложение на чувашский, либо конструировали чувашский стимул и просили консультанта оценить его приемлемость и, в случае положительной оценки стимула, перевести его на русский язык.

Если чувашский стимул был оценён как минимум 3 носителями как грамматичный, мы работали с ним как с грамматичным примером. В некоторых случаях предложение принималось большинством носителей, но не всеми, в таких случаях иногда используется обозначение вида $^{x/y}$, где x — количество носителей, признавших пример грамматичным, а y — общее количество спрошенных носителей.

Прочие обозначения: * — предложение грамматически неприемлемо, # — предложение прагматически неприемлемо, ? — предложение приемлемо, но не совсем естественно, % — предложение приемлемо для части носителей, OK — предложение грамматично (используется только при контрасте с *).

Лёгкий глагол *pərak* сочетается с двумя группами лексических глаголов — состояниями (4b) и переходными процессами (4a), но не сочетается с непереходными процессами, как в (4c):

(4) a. *vaçə   tarelkə   jori   ҫёmër-ʑe   pərak-rʲ-ə*
Вася   тарелка   специально   ломать-CV_SIM   бросить-PST-3SG
'Вася специально **разбил** тарелку.'

b. *vaçə   daʂ-a   jorat-sa   pərak-rʲ-ə*
Вася   Даша-OBJ   любить-CV_SIM   бросить-PST-3SG
'Вася **полюбил** Дашу.'

c. **pørt   ҫon-ʑa   pərak-rʲ-ə*
дом   гореть- CV_SIM   бросить-PST-3SG
Ожид. зн.: 'Дом **сгорел**.'

Дальнейшие обобщения касательно семантических особенностей лёгкого глагола *pərak* и его селективных ограничений будут касаться только контекстов типа (4a), то есть сложных глаголов от переходных процессов.

3.2. Затронутость объекта

В сочетании с непредельными процессами лёгкий глагол *pərak* проявляет чувствительность к степени затронутости объекта лексического глагола. Параметр затронутости используется в работах по лексической семантике, аргументной структуре и переходности глагола, см., например, [Fillmore 1968, Dowty 1991; Hopper, Thompson 1980]. Само понятие затронутости, однако, в большинстве работ определяется достаточно абстрактно и нестрого как степень полноты вовлечения объекта в событие. Решение этой проблемы было предложено в статье [Beavers 2011], в которой автор даёт затронутости следующее формальное определение:

(5)   Актант x затронут тогда и только тогда, когда существует событие e и шкала некоторого свойства s такие, что x достигает нового состояния на шкале s путём инкрементального, абстрактного движения по s.[3]

Сравнивая различные явления, в которых затронутость выступает в качестве селективного ограничения, Биверс приходит к выводу, что затрону-

---

[3] An argument x is affected iff there is an event e and a property scale s such that x reaches a new state on s through incremental, abstract motion along s [Beavers 2011: 3].

тость — градуальный параметр, у которого могут быть бо́льшие или меньшие степени:

(6)   Степень затронутости *x* определяется тем, насколько предикат определён относительно того, в какой точке на шкале *s* заканчивается изменение состояния *x*.[4]

Основываясь на принципе (6), Биверс [Beavers 2011: 21] вводит шкалу затронутости, состоящую из четырёх степеней:

(7)   a. x undergoes a quantized change iff $\varphi \rightarrow \exists e \exists s[\text{result}'\ (x, s, g\varphi, e)]$
        (e.g. accomplishments/achievements: break, shatter, destroy, devour x)

       b. x undergoes a non-quantized change iff $\varphi \rightarrow \exists e \exists s \exists g[\text{result}'(x, s, g, e)]$
        (e.g. degree achievements/cutting: widen, lengthen, cut, slice x)

       c. x has potential for change iff $\varphi \rightarrow \exists e \exists s \exists \theta[\theta(x, s, e)]$
        (e.g. surface contact/impact: wipe, scrub, rub, punch, hit, kick, slap x)

       d. x is unspecified for change iff $\varphi \rightarrow \exists e \exists \theta[\theta(x, e)]$
        (e.g. all others: see, smell, follow, play (as children), ponder x)

Остановимся подробнее на каждой из степеней затронутости. Так, наверху шкалы находятся предикаты класса (7a), описывающие квантованное изменение состояния объекта. Глаголы этой группы выделяются тем, что задают в своей семантике специфическую конечную точку изменения состояния объекта. Так, например, глагол *break* 'сломать, разбить' описывает процесс изменения состояния объекта, достигающий предела в конкретный момент — когда на объекте появляются следы повреждения. Похожее рассуждение, по-видимому, справедливо и для глаголов типа *shatter* 'разбить', *destroy* 'уничтожить' и *devour* 'пожирать' — все они описывают некоторый процесс изменения состояния объекта, стремящийся к конкретной конечной точке.

Следующий класс глаголов — глаголы класса (7b), описывающие неквантованное изменение состояния объекта — отличается от глаголов класса (7a) тем, что изменение состояния объекта не ограничено специфическим пределом. Рассмотрим в качестве примера предложение (8):

---

[4] How affected *x* is corresponds to how specific the predicate is about where *x* ends up on *s* [Beavers 2011: 3].

(8)   *I lengthened the scarf.*
      'Я удлинил шарф.'

Так, (8) может описывать массу разных событий удлинения шарфа: говорящий мог удлинить шарф как на 5 или 10 сантиметров, так и на 20 или 30 сантиметров — в любом случае эти ситуации будут удовлетворять условиям истинности предиката *lengthen the scarf*. Отсутствие конкретной конечной точки, специфицированной в семантике глагола, делает такие глаголы неквантованными в понимании Биверса: так, любая собственная часть события, входящего в экстенсионал предиката *lengthen the scarf* 'удлинить шарф' будет также удовлетворять условиям истинности этого предиката, поскольку для этого достаточно, чтобы изменение состояния объекта произошло **хотя бы до какой-то** степени. Основную массу глаголов класса (7b), таким образом, составляют градативы, или degree achievements, такие как *lengthen* 'удлинять', *dry* 'сушить', *widen* 'расширять' и др.

Несмотря на отличия с точки зрения специфичности относительно точки кульминации события, глаголы классов (7a) и (7b) объединяет между собой то, что они описывают события, в ходе которых объект обязательно претерпевает изменение состояния. Это отличает их от глаголов с меньшей степенью затронутости объекта, принадлежащим к классам (7c) и (7d). Так, глаголы класса (7c) описывают события, в ходе которых может произойти изменение состояния объекта, но это только один из возможных исходов. Рассмотрим предложение (9):

(9)   *John kicked the dog.*
      'Джон пнул собаку.'

Так, событие, описываемое в (9), может привести к тому, что собака изменит своё состояние, например, пинок окажет на неё деструктивный эффект или переместит её в пространстве. Однако возможен и альтернативный исход, в котором в ходе пинка с собакой ничего не происходит, кроме контакта с ногой Джона. Таким образом, в класс затронутости (7c) попадают глаголы типа *kick* 'пнуть', *wipe* 'протереть', *hit* 'ударить' и др., описывающие события с потенциальным, но не обязательным изменением состояния объекта.

Наконец, глаголы класса затронутости (7d) отличаются от глаголов с большей степенью затронутости тем, что их семантика вообще не предполагает изменение состояния объекта, даже в качестве потенциального исхода. Рассмотрим пример (10):

(10)   *John read the book.*
         'Джон прочитал книгу.'

Предложение (10) описывает событие, завершающееся в тот момент, когда книга прочитана. Несмотря на то, что книга необходимым образом вовлечена в процесс как предмет чтения, она никак физически не изменяется и не должна измениться в ходе процесса — во всяком случае, сам по себе процесс чтения этого не предполагает. Так устроены глаголы типа *read* 'читать', *smell* 'нюхать', *follow* 'преследовать' и др., описывающие события, в ходе которых объект не претерпевает изменения состояния.

Степень затронутости объекта лексического глагола оказывается ключевым параметром, определяющим селективные и семантические ограничения лёгкого глагола *pərak*, а именно можно сформулировать следующие обобщения:

(11) a. Лёгкий глагол *pərak* сочетается только с такими лексическими глаголами, в комбинации с которыми возможно образовать сложный глагол, описывающий квантованное изменение состояния объекта.

     b. Лёгкий глагол *pərak* образует сложные глаголы, описывающие квантованное изменение состояния объекта.

Так, лёгкий глагол *pərak* не сочетается или сочетается очень плохо с глаголами класса затронутости D, не предполагающими изменение состояния объекта, см. примеры в (12):

(12) a. **vaɕə*   *jor-a*   **etle-ʐe**   **pərak-r^j-ə**
        Вася   песня-OBJ   слушать-CV_SIM   бросить-PST-3SG
        Ожид. зн.: 'Вася **прослушал** песню.'

     b. *^{??}vaɕə*   *piɕmo*   **vula-ʐa**   **pərak-r^j-ə**
        Вася   письмо   читать-CV_SIM   бросить-PST-3SG
        'Вася **прочитал** письмо.'

С глаголами класса затронутости C лёгкий глагол сочетается лексически ограниченно, однако во всех случаях, когда *pərak* допустим, он образует сложный глагол со значением затронутости объекта, вхождения его в некоторое состояние:

(13) a. *vaɕə   petʲ-a   **ɕap-sa**   **pərak-rʲ-ə***
      Вася   Петя-OBJ   ударить-CV_SIM   бросить-PST-3SG
      'Вася **ударил** Петю (и Петя пострадал от этого).'

    b. *vaɕə-a   ёne   **tёk-se**   **pərak-rʲ-ə***
      Вася-OBJ   корова   толкать-CV_SIM   бросить-PST-3SG
      'Васю **забодала** корова.'

    c. **jɨdə   petʲ-a   **ɕɨrt-sa**   **pərak-rʲ-ə***
      пёс   Петя-OBJ   кусать-CV_SIM   бросить-PST-3SG
      Ожид. зн.: 'Пёс **искусал** Петю.'

    d. *%pɨl   xortɕ-ə   vaɕə   **sək-sa**   **pərak-rʲ-ə***
      мёд   пчела-P_3   Вася   ужалить-CV_SIM   бросить-PST-3SG
      'Пчела **ужалила** Васю.'

Так, например, предложение (13a) описывает событие, обязательно приводящее к некоторому деструктивному эффекту на объекте. Этим семантика сложного глагола *ɕapsa pərak* отличается от поведения лексического глагола *ɕap* 'ударить', который, как и ожидается от глаголов класса затронутости C, не предполагает обязательное изменение состояния пациенса, см. контраст в (14):

(14) a. *vaɕə   petʲ-a   **ɕap-sa***
      Вася   Петя-OBJ   ударить-CV_SIM
      'Вася **ударил** Петю.'

    b. *vaɕə   petʲ-a   **ɕap-sa**   **pərak-rʲ-ə***
      Вася   Петя-OBJ   ударить-CV_SIM   бросить-PST-3SG
      'Вася **ударил** Петю (и тот как-то пострадал от этого).' (= (13a))

Селективные ограничения, определяющие приемлемость примеров типа (13a–b) и проблематичность примеров типа (13c–d), пока не определены и требуют дополнительного изучения. На основании примеров выше можно предположить, что лёгкий глагол *pərak* сочетается только с такими глаголами класса затронутости C, которые предполагают не просто деструктивный эффект на объекте, а его падение, потерю равновесия. Это ограничение, если оно верно, может проистекать из лексического значения *pərak*, поскольку ситуация бросания предполагает падение объекта. Однако на настоящем этапе исследования мы не можем с уверенностью говорить о валидности этой рабочей гипотезы.

В сочетании с глаголами класса затронутости B, в первую очередь с градативами, лёгкий глагол *pərak* образует сложные глаголы с эксцессивным значением (действие развивается сверх меры):

(15) a. *vaçə*   *çyç-n-e*   *??(itla)*   **tibët-se**   **pərak-rʲ-ə**
       Вася   волосы-P_3-OBJ   слишком   сушить-CV_SIM   бросить-PST-3SG
       'Вася **пересушил** волосы.'

     b. *vaçə*   *şa*   *??(itla)*   **əʑət-sa**   **pərak-rʲ-ə**
       Вася   вода.OBJ   слишком   греть-CV_SIM   бросить-PST-3SG
       'Вася **перегрел** воду.'

Так, предложения (15) описывают ситуации, в которых изменение состояния объекта по некоторой шкале происходит больше ожидаемого. Интересно отметить, что для таких сложных глаголов принципиально употребление наречия *itla* 'слишком', без него или — по предварительным данным — другого эксплицитного указания на конкретную меру изменения состояния объекта употребление *pərak* затруднено, что также говорит в пользу того, что лёгкий глагол проявляет чувствительность к квантованности в понимании Биверса.

Наконец, в сочетании с глаголами класса A *pərak* образует сложные глаголы без каких-либо значимых эффектов с точки зрения затронутости:

(16) a. *vaçə*   *oraba-ja*   **joʑa-ʑa**   **pərak-rʲ-ə**
       Вася   телега-OBJ   чинить-CV_SIM   бросить-PST-3SG
       'Вася (быстро) **починил** телегу.'

     b. *vaçə*   *fajl-a*   **kəlar-ʑa**   **pərak-r-ʲə**
       Вася   файл-OBJ   выставить-CV_SIM   бросить-PST-3SG
       'Вася **удалил** файл.'

     c. ⁴/₅*vəl*   *tënʥe-je*   **oləʂtar-ʑa**   **pərak-rʲ-ə**
       он   мир-OBJ   изменить-CV_SIM   бросить-PST-3SG
       'Он **изменил** мир.'

Поведение телисизатора *pərak* в сочетании с лексическими глаголами разной степени затронутости можно резюмировать в виде следующей таблицы:

Таблица 1. Поведение телисизатора *pərak*
в сочетании с лексическими глаголами разной степени затронутости

Степень затронутости по [Beavers 2011]	Значение сложного глагола с *pərak*	
Квантованное изменение состояния объекта	Точка кульминации	
Неквантованное изменение состояния объекта	Эксцессив, изменение состояния сверх меры	
Событие, потенциально изменяющее состояние объекта	Изменение состояния объекта до некоторой серьёзной степени	
Отсутствие подразумеваемого изменения состояния		

Интересно, что способность лексического глагола образовать сложный глагол, относящийся к классу затронутости A, является необходимым, но, по всей видимости, не достаточным условием для сочетаемости с *pərak*. Так, согласно нашим данным, лёгкий глагол имеет некоторые проблемы в сочетаемости с такими глаголами максимальной затронутости, как, например, *ёҫ* 'пить', *ҫi* 'есть', *yger* 'рисовать', см. пример (17):

(17) **vaҫə   ҫa   ёҫ-se   pərak-rʲ-ə*
　　　Вася   вода.obj   пить-cv_sim   бросить-pst-3sg
　　　Ожид. зн.: 'Вася **выпил** воду.'

Эти глаголы не сочетаются с *pərak* в нулевом контексте, однако их приемлемость может быть улучшена с помощью аккомодации специфического контекста — контрастивности события, о котором пойдёт речь в следующем разделе.

3.3. Контрастивность события

Итак, как было замечено в предыдущем разделе, лёгкий глагол *pərak* в сочетании по крайней мере с некоторыми глаголами оказывается чувствителен к контексту. Рассмотрим пример (18):

(18) **Нулевой контекст (ср. (17)).
　　　ᴼᴷКонтекст. Васе предстоит операция, и ему нельзя пить воду. Но он её всё-таки выпивает.
　　　vaҫə   ҫa   ёҫ-se   pərak-rʲ-ə
　　　Вася   вода.obj   пить-cv_sim   бросить-pst-3sg
　　　'Вася **выпил** воду.'

Так, неграмматичное в нулевом контексте предложение (18) становится приемлемым, если навести специфический контекст, в котором выпивание воды становится ненормальным. Тот же эффект достигается в примере (19), который уже сам по себе предполагает некоторое выделяющееся событие:

(19) $^{3/4}$*vaɕə* *nargəməş* ***ëɕ-se***   ***pərak-rʲ-ə***
      Вася   яд   пить-CV_SIM   бросить-PST-3SG
      'Вася **выпил** яд.'

Какие именно требования к контексту предъявляет *pərak* в таких случаях? В докладе по [Голосов 2019а] было предположено, что лёгкий глагол требует случайности, ненамеренности действия, как в (20):

(20) a. *vaɕə*   *petʲ-a* ***podar-za***   ***pərak-rʲ-ə***
        Вася   Петя-OBJ утопить-CV_SIM бросить-PST-3SG
        'Вася **утопил** Петю (случайно столкнул его).'

   b. $^{?}$*vaɕə* *şørbe*   ***pëʑer-ze***   ***pərak-rʲ-ə***
        Вася   суп   готовить-CV_SIM бросить-PST-3SG
        'Вася ненароком **приготовил** суп (вместо другого блюда).'

На первый взгляд такой анализ можно распространить и на примеры в (20): действительно, в нулевом контексте говорить о том, что Вася случайно выпил воду, прагматически странно — другое дело, если он по невнимательности выпил воду вопреки рекомендациям врачей или по ошибке выпил яд. Однако, согласно нашим данным, *pərak* допустим и в контекстах, в которых эксплицитно указывается на намеренность действий агенса:

(21) a. *vaɕə*   *tarelkə* ***jori***   *ɕëmër-ze*   ***pərak-rʲ-ə***
        Вася   тарелка   специально разбить-CV_SIM бросить-PST-3SG
        'Вася **специально** разбил тарелку.' (= (4a))

   b. *vaɕə*   *avdan-a* ***jori***   *pula-za*   ***pərak-rʲ-ə***
        Вася   петух-OBJ специально душить-CV_SIM бросить-PST-3SG
        'Вася **преднамеренно** задушил петуха.'

Таким образом, если у лёгкого глагола *pərak* и есть какое-то общее требование к контексту, оно заключается не в случайности. Наше предположение состоит в том, что на самом деле телисизатор используется в контекстах, когда событие оказывается неожиданным, идущим вразрез с естественным, дефолтным ходом вещей (именно это значение мы имеем в виду, когда говорим о **контрастивности** события).

Такой анализ правильно предсказывает поведение примера (18): действительно, обычно в том, чтобы выпить воду, нет ничего удивительного, это совершенно ожидаемое повседневное событие, однако оно становится контрастивным, если аккомодировать контекст, в котором пить воду не положено. Аналогичным образом объясняется и пример (20b): в нулевом контексте приготовление супа — процесс, не противоречащий никаким ожиданиям, но он становится таковым, если планировалось приготовить другое блюдо. При этом примеры типа (21) не противоречат такому обобщению: Вася мог действовать специально, но неожиданно, более того, разбить тарелку и придушить петуха — действия, явно противоречащие естественному, нормальному ходу событий.

Другой аргумент в пользу такого анализа состоит в том, что по крайней мере некоторые глаголы уничтожения, такие как *vëler* 'убить', в нулевом контексте хорошо сочетающиеся с лёгким глаголом *pərak*, оказываются менее приемлемы для некоторых носителей в контекстах как в (22):

(22) Контекст: Убийство преступников на казни — обычная, «рутинная» работа для палача. Сегодня палач казнил очередного преступника.

$^{\#?}$*palaʨ*   *prestupnig-a*   **vëler-ze**   **pərak-r^{j}-ə**

палач   преступник-OBJ   убить-CV_SIM   бросить-PST-3SG

'Палач **казнил** преступника.'

Так, обычно неестественное, выделяющееся на фоне других событие убийства становится ожидаемым и, если так можно выразиться, «повседневным» в контексте работы палача. Некоторые из опрошенных носителей замечали, что такое предложение всё же можно сказать, если палач казнил преступника раньше, чем планировал, или более мастерски, чем обычно это делает. Отметим, что такие попытки «спасти» пример (22) только подтверждают наше предположение: они направлены на то, чтобы представить событие убийства как противоречащее их ожиданиям.

Ещё один косвенный аргумент в пользу нашего анализа основывается на поведении лексического глагола *sekter* 'лопнуть' в сочетании с *pərak*. Так, несмотря на то что этот глагол относится к классу затронутости А и может описывать незапланированные события, он плохо сочетается с лёгким глаголом:

(23) $^{2/5}$*vaʨə*   *ʂarig-a*   **sek-ter-ze**   **pərak-r^{j}-ə**

Вася   шарик-OBJ   лопнуть-CAUS-CV_SIM   бросить-PST-3SG

'Вася **лопнул** шарик.'

Согласно нашему предположению, неестественность примера (23) для большинства носителей заключается в том, что лопанье шарика является слишком незначительным событием для того, чтобы нарушить естественный ход вещей. Мы вернёмся к этой мысли в анализе, а пока отметим то, что двое носителей, которые всё же сочли предложение (23) приемлемым, указали на то, что событие было неожиданным для говорящего. Это можно рассматривать как косвенное указание на то, что, если уж говорящий счёл событие в (23) «достойным внимания», оно представляется для него неожиданным.

Перед тем, как перейти к следующему разделу, отметим, что, согласно предварительным данным, глаголы класса затронутости D не сочетаются с *pərak* на том же уровне приемлемости, что и глаголы других классов, даже в том случае, если описываемое событие удовлетворяет требованиям контрастивности, как в (24), ср. также (12b):

(24) Контекст: у Маши есть секретное письмо, и Васю попросили не читать его. Но он всё-таки прочитал.

<div style="margin-left:2em">

$^{??}$*vaçə*   *piçmo-nʲ-a*   ***vula-ʐa***   ***pərak-rʲ-ə***

Вася   письмо-P_3-OBJ   читать-CV_SIM   бросить-PST-3SG

'Вася **прочитал** письмо.'

</div>

3.4. Данные: обобщения

Итак, лёгкий глагол *pərak* грамматикализовался в акциональный модификатор, образующий предельные сложные глаголы в сочетании с состояниями и с некоторыми переходными процессами. Как минимум в последнем случае *pərak* накладывает требования на затронутость объекта и на контрастивность описываемого события. В следующем разделе мы предложим намётки формального анализа сложных глаголов с *pərak*, предсказывающего селективные ограничения телисизатора и его семантику.

4. Анализ

4.1. Введение во фреймворк: синтаксис первой фазы

Фреймворк, в рамках которого мы собираемся предложить анализ сложных глаголов с *pərak* — это синтаксис первой фазы [Ramchand 2008a]. Это формальная теория акциональной декомпозиции, согласно которой структура события глагола представлена в синтаксисе в виде некоторого набора

из 3 событийных вершин: *init,* ответственной за каузирующее событие, *proc,* вводящей подсобытие изменения состояния и/или процесса, и *res,* ответственной за результирующее состояние. У каждой из вершин есть спецификаторы, соответствующие основным участникам события: у *init* это Инициатор — участник-источник каузирующей деятельности, у *proc* это Претерпевающий — участник, вовлечённый в процесс и/или изменяющий состояние в ходе события, а у *res* это Носитель результирующего состояния. Комплементами вершин выступают каузально зависимые от них событийные группы (так, комплементом *init* обычно выступает *proc*P, а комплементом *proc* — *res*P) или несобытийные аргументы, некоторым образом влияющие на акциональные свойства предиката. Максимальную проекцию синтаксиса первой фазы можно видеть на следующем рисунке:
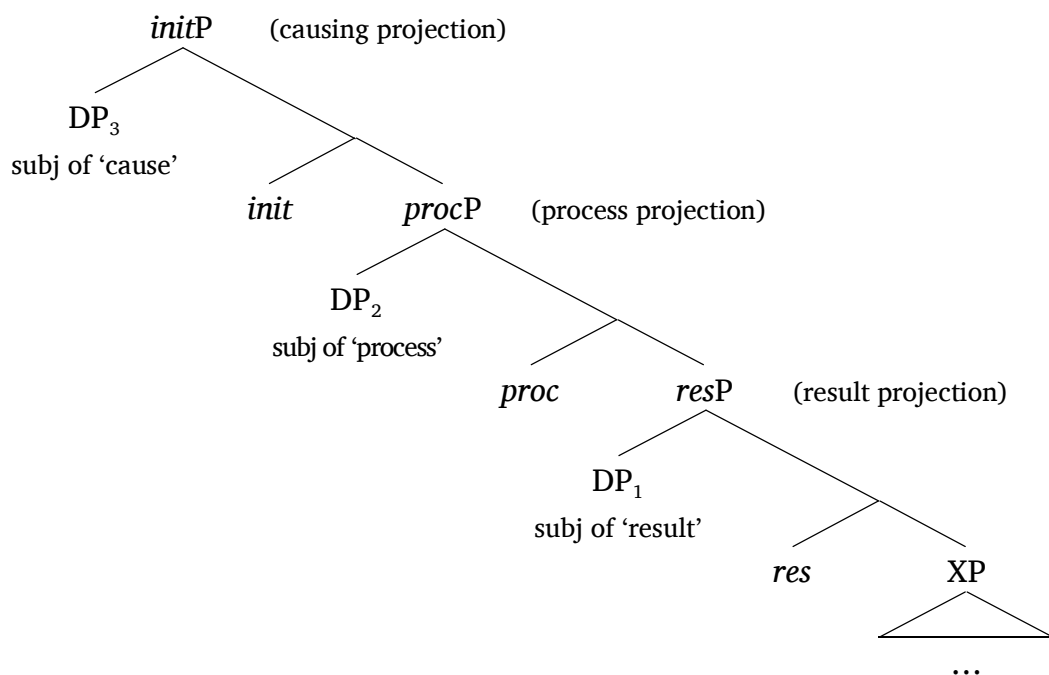

Рисунок 1. Синтаксис первой фазы [Ramchand 2008: 46]

В рамках настоящей статьи у нас нет возможности подробно разобрать общее устройство синтаксиса первой фазы, поэтому мы отсылаем читателя к первоисточнику. Отметим лишь, что, в отличие от классического генеративизма, в синтаксисе первой фазы семантические роли аргументов глагола не распределяются биективно по разным именным группам, а могут «наслаиваться» друг на друга. Так, например, единственный аргумент глагола *arrive* 'прибыть' в ходе деривации получает сложную семантическую роль Инициатора-Претерпевающего-Носителя результирующего со-

стояния, поскольку референт соответствующей ИГ является одновременно и (само)каузатором процесса, и участником, изменяющим своё положение в пространстве. Соответственно, ключевыми параметрами вариативности в структуре события у разных глаголов являются набор подсобытий и распределение семантических ролей между участниками события. Так, например, неаккузативы отличаются от неэргативов тем, что вершину *init* проецируют только последние, а неэргативы отличаются от транзитивов в общем случае тем, что у первых роли Претерпевающего и Инициатора получает одна и та же ИГ, а у вторых — нет.

Синтаксис первой фазы является удачным инструментом для анализа различных структурно-событийных языковых явлений, в частности, он широко используется в работах, посвящённых синтаксису и семантике сложных предикатов в индоарийских языках [Ramchand 2008a; Ramchand 2008b; Ozarkar, Ramchand 2018] и языках Поволжья [Голосов 2019b; Golosov (in print); Kashkin, Dyachkov 2018]. В рамках этой теории лёгкие глаголы — это такие глаголы, которые утратили лексический компонент своего значения, но сохранили некоторым образом структуру события, и именно она определяет их селективные ограничения и акциональные свойства образуемых сложных глаголов.

В рамках нашего анализа мы примем допущение, согласно которому группа лексического глагола выступает комплементом самой низкой событийной вершины лёгкого глагола (в нашем случае это будет *res*), как это было сделано в рамках анализа сложных глаголов языка маратхи в работе [Ozarkar, Ramchand 2018], а объединение экстенсионалов двух глагольных групп происходит с помощью такой семантической операции, как обобщённая идентификация событий [Татевосов, Киселёва 2019], к которой мы ещё вернёмся в анализе.

Таким образом, исходя из предположения, что селективные ограничения и семантика сложного глагола обусловлены структурой события лёгкого глагола, мы попробуем определить его набор подсобытий и распределение семантических ролей по аргументам с помощью данных, рассмотренных в предыдущем разделе.

4.2. Структура события лёгкого глагола *pərak*

Итак, то, что лёгкий глагол *pərak* сочетается только с переходными процессами и притом образует сложные глаголы с затронутым объектом, в рамках синтаксиса первой фазы можно объяснить только тем, что в его

структуре события есть событийные вершины *init* и *proc*, причём семантические роли Инициатора и Претерпевающего обязательно присваиваются разным именным группам.

Следующий вопрос заключается в том, есть ли у *pərak* в структуре события результирующее состояние. Несмотря на то, что в теории синтаксиса первой фазы *res* не является единственным источником предельной интерпретации [Ramchand 2008; Лютикова и др. 2006], только наличие этой вершины в структуре события может обуславливать её обязательность, см. рассуждение в [Лютикова и др. 2006]. Таким образом, структура сложного глагола с *pərak* может быть представлена в виде следующего дерева:
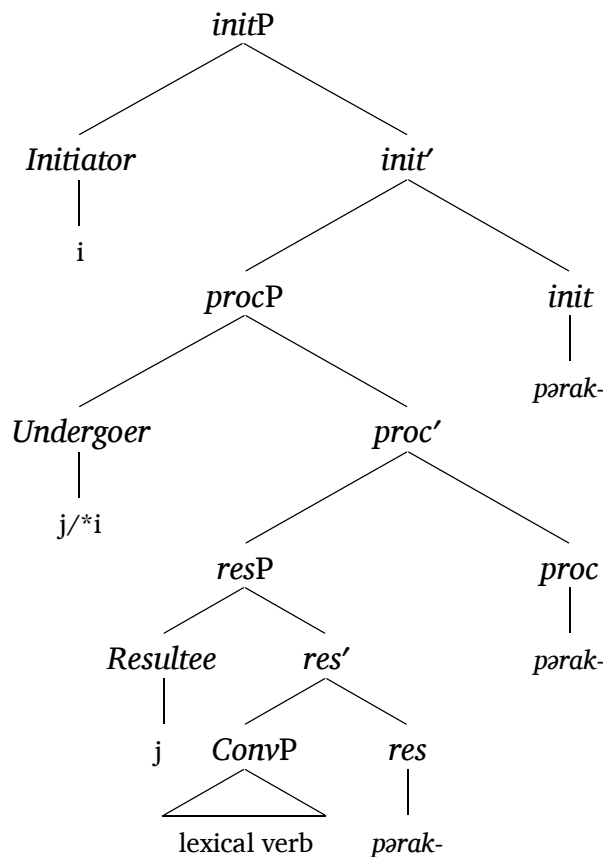

Рисунок 2. Структура события сложного глагола с *pərak*

К сожалению, в рамках этой статьи мы не имеем возможности подробно объяснить все технические детали деривации сложного глагола, поэтому обрисуем лишь основные фрагменты. Так, аргументы лексического глагола передвигаются в соответствующие аргументные позиции лёгкого глагола, а семантическое объединение происходит с помощью операции обобщённой идентификации событий, предложенной в [Татевосов, Киселева 2019].

Не имея желания нагружать читателя формальными деталями этой операции, перескажем общую идею: в ходе деривации экстенсионалы групп лексического и лёгкого глаголов пересекаются. Так, например, значение сложного глагола *jozaza pərak* ‘починить’ образуется с помощью пересечения множества ситуаций, удовлетворяющих условиям истинности глагола *joza* ‘чинить’, и множества ситуаций, удовлетворяющих событийному шаблону *pərak*: ‘Инициатор скаузировал Претерпевающего превратиться в Носителя некоторого результирующего состояния’. В результате *jozaza pərak* значит ‘заставить Претерпевающего путём починки войти в состояние починенности’, то есть, собственно, ‘починить’.

Синтаксическое дерево на рисунке 2 правильно предсказывает, что лёгкий глагол *pərak* может сочетаться только с переходными процессами и образует предельные сложные глаголы, описывающие изменение состояния объекта. Однако акциональная семантика телисизатора несколько сложнее: *pərak* описывает не просто изменение состояния объекта — оно должно быть квантованным в терминологии Биверса.

По всей видимости, решение этой проблемы состоит в том, чтобы дать более строгую семантику для событийной вершины res. На настоящем этапе исследования у нас есть две возможных теории того, как эта семантика должна быть устроена.

Первая гипотеза состоит в следующем. Следуя анализу горномарийского лёгкого глагола *šuaš* ‘бросить’ в [Голосов 2019b], можно предположить, что событийная вершина res телисизатора *pərak* вводит состояние локальной исчерпанности: объект входит в такое состояние, на протяжении которого с ним невозможно повторно совершить действие, описываемое сложным глаголом.

Такое предположение согласуется с тем, как *pərak* ведёт себя с глаголами разной степени затронутости. Так, лексические глаголы класса затронутости D, такие как *etle* ‘слушать’ и *vula* ‘читать’, не описывают изменения состояния в принципе, и, как следствие, объект не входит в состояние локальной исчерпанности: книга по прочтении по-прежнему может быть прочитана, а песня может быть заново прослушана.

Лексические глаголы класса затронутости C, такие как *ҫap* ‘ударить’ или *tĕk* ‘боднуть’, описывают событие с потенциалом к изменению состояния, однако требование *pərak* на локальную исчерпанность объекта сужает экстенсионал этих предикатов до таких событий, в которых пациенс получил серьёзный деструктивный эффект, например, умер или потерял сознание или равновесие.

Лексические глаголы класса затронутости B, такие как * əʐət* 'греть' и *tibёt* 'сушить', предсказуемо получают эксцессивное значение как способ достичь локальной исчерпанности. Так, если Вася **нагрел** воду, в каких-то случаях он может нагреть её ещё раз, однако если он **перегрел** воду, то он не сможет перегреть воду до тех пор, пока она не остынет, то есть не выйдет из состояния локальной исчерпанности.

Наконец, лексические глаголы класса затронутости A в большинстве своём устроены так, как требует *pərak*, поскольку описывают вхождение в специфическое состояние, войдя в которое, объект не может в том же смысле изменить своё состояние ещё раз. Так, например, если Вася задушил петуха, он не может повторить это действие над тем же объектом, так как петух умер и не может потерять жизнь заново. То же самое верно, например, для *joʐa* 'починить': если телега починена, то её нельзя починить ещё раз, пока она не сломается.

Проблемными для этой теории являются контексты типа (25) (= (16c)):

(25)   ⁴ᐟ⁵*vəl*   *tёndʑe-je*   **oləʂtar-ʐa**   *pərak-rʲ-ə*
      он   мир-OBJ   изменить-CV_SIM   бросить-PST-3SG
      'Он **изменил** мир.'

Так, в (25) описывается ситуация изменения мира, которая, строго говоря, не должна приводить к его локальной исчерпанности: мир не теряет способность изменяться дальше даже на время. Возможно, в (25) имеется в виду, что мир изменился настолько, что его можно считать другим объектом, и в таком случае мы просто теряем доступ к исходному референту, как в случае с разбитой чашкой.

В этом месте более удачной оказывается другая гипотеза о семантике событийной вершины res лёгкого глагола *pərak*, которая, однако, несколько теряет в строгости по сравнению с предыдущей: лёгкий глагол требует, чтобы объект изменил своё состояние до некоторой контекстно релевантной максимальной степени. Такая теория, по всей видимости, в большинстве случаев делает те же предсказания, что и теория локальной исчерпанности, поскольку достижение максимальной степени на шкале предполагает, что дальше изменять состояние некуда. Однако теория максимальной степени не вступает в такое явное противоречие с примерами типа (25), поскольку можно, например, изменить мир полностью, во всех аспектах.

На настоящий момент не представляется возможным отдать предпочтение одной из гипотез, поскольку в большинстве случаев они делают одинаковые предсказания. Для того, чтобы чётко развести две теории между собой, необходимо проверить, приемлем ли *pərak* в контекстах вроде 'Вася приоткрыл окно', в которых изменение состояния приводит к локальной исчерпанности (окно нельзя открыть ещё раз, пока оно открыто), но при этом степень изменения состояния не максимальная (окно открыто не до конца).

Итак, на настоящий момент наш анализ до некоторой степени справляется с тем, чтобы предсказать чувствительность *pərak* к таким параметрам, как предельная интерпретация сложного глагола, переходность лексического глагола и квантованность изменения состояния объекта в понимании Биверса. Нам осталось представить последний фрагмент анализа, предсказывающий ещё одно свойство телисизатора — его чувствительность к контрастивности события.

4.3. Анализ семантики контрастивности

Чувствительность *pərak* к тому, чтобы описываемое сложным глаголом событие оказывалось неожиданным с точки зрения естественного хода вещей, сближает его с русской конструкцией *взять и сделать,* см. пример (26):

(26) *Вася взял и машину купил.* [Стойнова 2007: 187]

В работе [Стойнова 2007] предлагается следующее толкование для этой конструкции:

(27) Ситуация резко изменяет естественный ход предшествующих событий, нарушая ожидания относительно дальнейшего развития событий. [ibid: 159]

Как кажется, такое толкование подходит и для чувашских данных, как мы постарались показать в предыдущем разделе. Однако оно слишком нестрогое: так, не совсем понятно, что значит «естественный ход предшествующих событий». Можно ли предложить такую же семантику в более строгом виде?

В статье [Alonso-Ovalle, Hsieh 2017] рассматривается ещё одна языковая единица, поведение которой напоминает требования лёгкого глагола *pərak* к контрастивности события. Так, в тагальском языке есть специальная па-

радигма глагольных форм со значением «успешности действия или его непроизвольности»[5] (далее AIA), см. минимальную пару из [Alonso-Ovalle, Hsieh 2017: 76], где нас прежде всего интересует (28b):

(28) a. *B<in>uks-an   ni   Lisa   ang   pinto.*
        <PFV.NTL>open-LV   GEN   Lisa   NOM   door
        'Lisa opened the door.'

    b. **Na**-*buks-an*   *ni*   *Lisa*   *ang*   *pinto.*
        PFV.AIA-open-LV   GEN   Lisa   NOM   door
        'Lisa managed to open the door. / Lisa accidentally opened the door.'

Авторы предлагают для таких на первый взгляд разнородных значений успешности и непроизвольности действия общую интенсиональную семантику, см. формулу (29) [Alonso-Ovalle, Hsieh 2017: 84]:

(29) a. For any e, w, $[[\text{aia } VP]]^c$ (e)(w) is defined iff
      i. c makes available a causal structure C and a stereotypical ordering source S and,
      ii. $C \rightarrow_{pVP, w} \cup S_w \cup \{p_{VP}\}$ is an inconsistent set of propositions.

    b. If defined, $[[\text{aia } VP]]^c$ (e)(w) = 1 iff $[[VP]]^c$ (e)(w) = 1

Попробуем изложить содержание формулы более простым языком. Так, в (29а) заключена пресуппозиция глагольных групп с маркером AIA, накладывающая на контекст два требования. Во-первых, в контексте должна быть доступна каузальная структура C (некоторый набор факторов-пропозиций, оказывающих каузальное влияние на пропозицию P, описываемую VP) и множество пропозиций, соответствующих стереотипным ожиданиям говорящего относительно возможного развития событий. Во-вторых, пропозиция P, описываемая предикатом, должна вступать в противоречие со стереотипными ожиданиями говорящего и/или с набором каузально релевантных факторов.

Как кажется, анализ в (29) можно распространить и на сложные глаголы с *pərak*. Рассмотрим, как такая семантика предсказывает разные контексты употребления лёгкого глагола. Так, понятно, почему пример (30) (=(18)) оказывается маргинален в нулевом контексте: по умолчанию пить

---

[5] 'Ability / involuntary action' [Alonso-Ovalle, Hsieh 2017: 75].

воду — событие ожидаемое и вытекающее из базовых потребностей человека. Однако в контексте, когда Васе предстоит операция, питьё воды входит в противоречие со стереотипными ожиданиями: нельзя пить воду, иначе будут проблемы с проведением операции.

(30) *Нулевой контекст.
ОККонтекст. Васе предстоит операция, и ему нельзя пить воду. Но он её всё-таки выпивает.

vaɕə	*ṣa*	*ёɕ-se*	*pərak-rʲ-ə*
Вася	вода.OBJ	пить-CV_SIM	бросить-PST-3SG

'Вася **выпил** воду.'

Напротив, глаголы уничтожения, такие как *vëler* 'убить' или *ɕёmёr* 'разбить', обычно описывают события, противоречащие нашим ожиданиям, и, кроме того, обычно не предполагают каузальных факторов, ведущих к их свершению, поэтому семантика контрастивности в нулевом контексте в соответствующих сложных глаголах может быть почти незаметна. Однако ситуация меняется, если аккомодировать контекст, в котором убийство становится ожидаемым, обыденным событием и следует из естественного хода вещей — поэтому проблематичны примеры типа (31) (=(22)):

(31) Контекст: Убийство преступников на казни — обычная, «рутинная» работа для палача. Сегодня палач казнил очередного преступника.

#?*palatɕ*	*prestupnig-a*	*vëler-ze*	*pərak-rʲ-ə*
палач	преступник-OBJ	убить-CV_SIM	бросить-PST-3SG

'Палач **казнил** преступника.'

Наконец, спорная приемлемость примера (32) (=(23)) может объясняться тем, что описываемое событие оказывается настолько незначительным в картине мира говорящего, что относительно него не формируется никаких стереотипных ожиданий и тем самым нарушаются прессупозиционные требования *pərak*:

(32) 2/5*vaɕə* | *ṣarig-a* | *sek-ter-ze* | *pərak-rʲ-ə*
| Вася | шарик-OBJ | лопнуть-CAUS-CV_SIM | бросить-PST-3SG |

'Вася **лопнул** шарик.'

При этом предложение (32) приемлемо для части носителей, что, согласно нашему предположению, свидетельствует о том, что незначительность события — субъективный параметр, зависящий от мнения говоря-

щего. В пользу этого также говорят, например, реакции разных носителей на предложения со сложным глаголом *vĕlerze pǝrak* 'убить' в случае, если объектом выступает комар: один из консультантов скептически заметил, что убийство комара является незначительным событием, которое не может затронуть чьих-либо чувств, в отличие, например, от убийства человека или домашнего животного, в то время как некоторые другие консультанты оценивали такие предложения как приемлемые и указывали на случайность или неожиданность события.

Перед тем, как перейти к выводам, выскажем свои соображения насчёт того, в каком месте в структуре семантика контрастивности вступает в деривацию. На этот вопрос у нас нет однозначного ответа, но исходя из чувствительности этого параметра к каузальным факторам события можно предположить, что контрастивность задана в семантике событийной вершины *init*. Косвенным аргументом в пользу такого анализа выступает устройство лексического источника телисизатора: глагол *pǝrak* 'бросить' обозначает событие с моментальной каузирующей стадией и потому не предполагает сколько-нибудь растянутой каузальной цепочки, см. русское *#Вася бросает камень уже несколько секунд*. Если это рассуждение верно, то можно предположить, что лёгкий глагол *pǝrak* сохранил своё требование к моментальности каузирующей стадии в виде семантики контрастивности. В противном случае можно предположить, что контрастивность включается в деривацию в некоторой функциональной вершине уже выше *v*P.

На этом раздел статьи, посвящённый анализу лёгкого глагола *pǝrak*, завершается, и мы переходим к общим выводам.

5. Выводы

Итак, мы установили, что чувашский глагол *pǝrak* 'бросить' грамматикализовался в акциональный модификатор, образующий предельные сложные глаголы от некоторых состояний и переходных процессов. В сочетании с последними лёгкий глагол проявляет чувствительность к параметру затронутости, образуя сложные глаголы, описывающие квантованное (в терминологии Биверса) изменение состояния объекта, и к параметру контрастивности, требуя от контекста, чтобы событие, описываемое сложным глаголом, противоречило ожиданиям говорящего или контекстно релевантным каузальным факторам.

Кроме того, мы предложили анализ сложных глаголов с *pərak* и переходными процессами в рамках синтаксиса первой фазы, согласно которому описанные выше селективные ограничения телисизатора и его семантика объясняются его структурой события. Так, лёгкий глагол проецирует событийные вершины *init*, *proc* и *res*, причём семантические роли Инициатора и Претерпевающего приписываются разным ИГ, *res* вводит некоторые дополнительные требования к исчерпанности процесса, а *init* (или, возможно, функциональная вершина над *v*P лёгкого глагола) ответственна за семантику контрастивности. Эта структура служит своего рода селективным и семантическим фильтром для лексического глагола при его соединении с телисизатором.

Отметим, что перечисленные выше свойства лёгкого глагола в значительной степени сопоставимы с устройством его лексического источника. Так, лексический глагол *pərak* сам переходный, сам описывает квантованное изменение состояния объекта и сам не предполагает длительного каузирующего подсобытия.

Список условных сокращений

1, 2, 3 — 1, 2, 3 лицо; AIA — форма со значением успешности/ непроизвольности действия; CAUS — каузатив; CV_SIM — нейтральное деепричастие; GEN — генитив; LV — локативный залог; NOM — номинатив; NTL — нейтральная форма; LOC — локатив; OBJ — объектный падеж; P_X — посессив лица x; PFV — перфектив; PST — прошедшее время; SG — единственное число.

Литература

Ашмарин 1898 — Ашмарин Н.И. Материалы для исследования чувашского языка. Казань: Типо-литография Императорского университета. 1898. [Ashmarin N.I. Materialy dlya issledovaniya chuvashskogo yazyka [Materials for research of the Chuvash language]. Kazan': Tipo-litografiya Imperatorskogo universiteta. 1898.]

Ашмарин 1923 — Ашмарин Н.И. Опыт исследования чувашского синтаксиса. Ч. 2. Симбирск. 1923. [Ashmarin N.I. Opyt issledovaniya chuvashskogo sintaksisa [Chuvash syntax research experience]. V. 2. Simbirsk. 1923.]

Голосов 2019a — Голосов Ф.В. Лёгкий глагол-деинкрементализатор: грамматикализация чувашского глагола *prak* 'бросить' // Шестнадцатая конференция по типологии и грамматике для молодых исследователей. Тезисы докладов. Санкт-Петербург, 21–23 ноября 2019 г. СПб.: ИЛИ РАН. 2019. С. 25–28. [Golosov F.V. Legkii glagol-deinkrementalizator: grammatikalizatsiya chuvashskogo glagola *prak* 'brosit'' [Deincrementalizing light verb: Grammaticalization of Chuvash verb *prak* 'throw']. Shestnadtsataya konferentsiya po tipologii i grammatike dlya molodykh issledovatelei. Tezisy dokladov. Sankt-Peterburg, 21–23 Nov. 2019. SPb.: ILI RAN. 2019. P. 25–28.]

Голосов 2019b — Голосов Ф.В. Грамматикализация глагола *šuaš* 'бросить': что осталось от бросков? // Типология морфосинтаксических параметров. 2019. Т. 2. Вып. 2. С. 54–78. [Golosov F.V. Grammatikalizatsiya glagola *šuaš* 'brosit'': chto ostalos' ot broskov? [Grammaticalization of the verb *šuaš* 'throw': What is left of throws?]. Typology of morphosyntactic parameters. 2019. V. 2. Iss. 2. P. 54–78.]

Гращенков 2015 — Гращенков П.В. Тюркские конвербы и сериализация: синтаксис, семантика, грамматикализация. М.: Языки славянской культуры. 2015. [Grashchenkov P.V. Tyurkskie konverby i serializatsiya: sintaksis, semantika, grammatikalizatsiya [Turkic converbs and serialization: Syntax, semantics and grammaticalization]. Moscow: Yazyki slavyanskoi kul'tury. 2015.]

Лебедев 2016 — Лебедев Э.Е. Акционсартовые значения сложновербальных аналитических форм в чувашском языке. Чебоксары: Чувашский государственный институт гуманитарных наук. 2016. [Lebedev È.E. Aktsionsartovye znacheniya slozhnoverbal'nykh analiticheskikh form v chuvashskom yazyke [Actional meanings of the complex predicate analytic forms in Chuvash]. Cheboksary: Chuvashskii gosudarstvennyi institut gumanitarnykh nauk. 2016.]

Лютикова и др. 2006 — Лютикова Е.А., Татевосов С.Г., Иванов М.Ю., Пазельская А.Г., Шлуинский А.Б. Структура события и семантика глагола в карачаево-балкарском языке. М.: ИМЛИ РАН, 2006. [Lyutikova E.A., Tatevosov S.G., Ivanov M.Yu., Pazel'skaya A.G., Shluinskii A.B. Struktura sobytiya i semantika glagola v karachaevo-balkarskom yazyke [Event structure and verb semantics in Karachay-Balkar]. Moscow: IMLI RAN, 2006.]

Стойнова 2007 — Стойнова Н.М. Конструкция *взять и сделать* в русском языке // Дудчук Ф.И., Ивлиева Н.В., Подобряев А.В. (ред.). Структуры и интерпретации: работы молодых исследователей по теоретической и прикладной лингвистике. М.: Издательство Московского Университета, 2007. С. 144–171. [Stoinova N.M. Konstruktsiya vzyat' i sdelat' v russkom yazyke [Construction *vzyat' I sde'at'* in Russian]. Dudchuk F.I., Ivlieva N.V., Podobryaev A.V. (eds.). Struktury i interpretatsii: raboty molodykh issledovatelei po teoreticheskoi i prikladnoi lingvistike. M.: Izdatel'stvo Moskovskogo Universiteta, P. 144–171.]

Татевосов, Киселева 2019 — Татевосов С.Г., Киселева К.Л. Идентификация событий в префиксально-постфиксальном словообразовании. Вестник Московского университета, Сер. 9, Филология, 2019 (1). С. 7–35. [Tatevosov S.G., Kiseleva K.L. Identifikatsiya sobytii v prefiksal'no-postfiksal'nom slovoobrazovanii [Event identification in the prefixal-postfixal derivations]. Vestnik Moskovskogo universiteta, Ser. 9, Philology, 2019 (1). P. 7–35.]

Шлуинский 2006 — Шлуинский А.Б. Акциональность и аспектуальный показатель: конструкция со вспомогательным глаголом *il-* в чувашском языке. Вестник Московского университета. Сер. 9, Филология, 2006 (1). С. 32–47. [Shluinskii A.B. Aktsional'nost' i aspektual'nyi pokazatel': konstruktsiya so vspomogatel'nym glagolom *il-* v chuvashskom yazyke [Actionality and aspectual marker: Constructions with *il-* auxiliary in Chuvash.]. Vestnik Moskovskogo universiteta. Ser. 9, Philology, 2006 (1). P. 32–47.]

Alonso-Ovalle, Hsieh 2017 — Alonso-Ovalle L., Hsieh H. Causes and expectations: On the interpretation of the Tagalog ability/involuntary action form. Semantics and Linguistic Theory. 2017. Vol. 27. P. 75–94.

Beavers 2011 — Beavers J. On affectedness. Natural Language and Linguistic Theory. 2011. 29(2). P. 335–370.

Dowty 1991 — Dowty D. Thematic proto-roles and argument selection. Language. 1991. V. 67, Iss. 3. P. 547–619.

Fillmore 1968 — Fillmore Ch.J. The case for case. Bach E., Harms R.T. (eds.). Universals in Linguistic Theory. New York: Holt. 1968. P. 1–90.

Golosov (in press) — Golosov F. Event structure of the complex predicates with light verb *lart* 'seat, put' in Poshkart Chuvash. Acta Linguistica Petropolitana. Труды Института лингвистических исследований РАН. In press.

Kashkin, Dyachkov 2018 — Kashkin E., Dyachkov, V. Complex verb constructions in Hill Mari: Semantics and event structure. 51st Annual Meeting of the Societas Linguistica Europaea, 29 August – 1 September 2018. Tallinn: Tallinn University, 2018. P. 129–130.

Ozarkar, Ramchand 2018 — Ozarkar R., Ramchand G. Structure matching and structure building in Marathi complex predicates. Journal of South Asian Linguistics, 2018. Vol. 8. Iss. 1.

Ramchand 2008a — Ramchand G. Verb meaning and the lexicon: A first-phase syntax. Cambridge: Cambridge University Press. 2008.

Ramchand 2008b — Ramchand G. Lexical items in complex predications: Selection as underassociation. Svenonius P., Tolskaya I. (eds.). Tromsø Working Papers on Language & Linguistics: Nordlyd 35, special issue on Complex Predication, 2008. P. 115–141.

Голосов Фёдор Валентинович

стажёр-исследователь, лаборатория по формальным моделям в лингвистике, Национальный исследовательский университет «Высшая школа экономики»

Fedor Golosov

research assistant, Laboratory of Formal Models in Linguistics, National Research University Higher School of Economics

golosovfv@gmail.com

Функциональные проекции и языковой контакт: маркирование множественного числа испанских имен в юкатекском майянском[*]

Родриго Гутьеррес-Браво[1]*, Мелани Ут*[2]
[1]*Мексиканский колледж,* [2]*Кёльнский университет*

В статье мы предлагаем анализ двух различных в типологическом отношении маркеров множественного числа, сосуществующих в юкатекском майянском в результате языкового контакта с испанским языком. Мы показываем, что в юкатекском майянском дистрибуция испанского маркера множественного числа идентична дистрибуции юкатекского маркера множественного числа (и не совпадает с дистрибуцией этого маркера в испанском) за исключением случаев, когда в именной группе присутствует испанское числительное — в последнем случае используется модель маркирования, характерная для испанского языка. В предлагаемом анализе эти факты следуют из стандартных представлений о структуре составляющих и лексической селекции при независимо аргументируемом в литературе допущении, что два анализируемых маркера множественного числа соответствуют разным формальным сущностям.

Ключевые слова: число, множественность, языковой контакт, юкатекский язык, язык майя, испанский язык.

FUNCTIONAL PROJECTIONS AND LANGUAGE CONTACT: THE CASE OF PLURAL MARKING OF SPANISH NOUNS IN YUCATEC MAYA[*]

Rodrigo Gutiérrez-Bravo[1], Melanie Uth[2]
[1]El Colegio de México, [2]Universität zu Köln

In this paper we present an analysis of two crosslinguistically different plural markers that coexist in Yucatec Maya (Mayan, Mexico) as a result of language contact with Spanish. We show that the distribution of the Spanish plural marker is exactly like the one of the Yucatec plural marker (unlike Spanish), with the exception that it is sensitive to the presence of a Spanish numeral, in which case it behaves like in Spanish. We develop an analysis where the relevant facts follow from standard assumptions about phrase structure and lexical selection, so long as an analysis is adopted in which the two plural markers are formally different, as has been suggested in the literature.

Keywords: number, plural, language contact, Yucatec Maya, Spanish.

1. Introduction

In this paper we present an analysis of two crosslinguistically different plural markers that coexist in Yucatec Maya (a Mayan language from Mexico) as the result of language contact with Spanish. We develop a formal analysis of one of the results reported in [Uth, Gutiérrez-Bravo 2018], which has to do with the interaction between Spanish numerals borrowed as loanwords and the mechanisms available in Yucatec Maya to express plurality. Plural marking is optional in Yucatec and this is also the case both for Spanish loanwords, and for the Spanish plural suffix which can appear with these loanwords. Yet plural marking in Spanish nouns, otherwise entirely optional, is obligatory in one specific context: in the presence of a numeral borrowed from Spanish. As a result of the analysis that we develop to account for these data, we submit the following hypothesis regarding the distribution of functional heads in contexts of language contact:

(1)   a. The adoption by a recipient language A of functional head X^0 from a language B does not entail that X^0 will behave in language A in the same way that it does in language B.

   b. In contrast, the adoption by a recipient language A of **both** a functional head X^0 **and** a lexical (or functional) head Y^0 that selects X^0 does bring with it a distribution of X^0 that is similar to the one observed in the contact language B.

We show that no special theoretical mechanisms or stipulations are needed to account for the state of affairs described in (1): as shown in the analysis below, this follows from the basic properties of Merge in conjunction with lexical selection. The paper is organized as follows; in Section 2 we provide some basic information about Yucatec Maya and of its nominal morphology. In Section 3 we provide a brief overview of the theoretical apparatus to be used later in the analysis, namely, the proposal developed in [Wiltschko 2008] and the specific analysis in [Butler 2012, 2013] of plural marking in Yucatec nouns using this proposal. In Section 4 we present the data corresponding to plural marking in Spanish loanwords, followed by the analysis we propose to account for these data. In Section 5 we discuss two alternative analyses of the data and conclude that they are problematic when compared to our analysis. Section 6 provides some brief discussion of further issues related to our proposal, and in Section 7 we present our conclusions.

2. Nominal morphology in Yucatec Maya

Yucatec Maya (henceforth Yucatec) is the Mayan language spoken in the Yucatan Peninsula, Mexico, by 824,670 people [2010 census: INEGI]. It is also spoken in some bordering districts of Belize and Guatemala. It is arguably the most studied and best described of all the 364 indigenous languages spoken in Mexico.

Figure 1. The Yucatan Peninsula

Nominal inflectional morphology is particularly simple in Yucatec Maya, like in other Mayan languages (see for instance [Coon 2016; England 2017]). Inflectional morphology basically reduces to the following suffixes: (a) an honorific suffix *-tsil* (currently very rarely used), (b) two relational suffixes, *-il* and *-el* (the latter suffix is used exclusively for inalienable relations) and (c) the plural suffix *-o'ob*, which has an allomorph *-ob* when it follows a glottalized vowel. The plural suffix *-o'ob*, which will be the focus of this paper, is transcategorial: it is observed not only in nouns, but also in verbs (as in (2)), adjectives, relational nouns and prepositions (in which case it marks agreement between the preposition and its complement when the complement is a null *pro* with a plural referent). There are also two prefixes for biological gender, *j-* [h] for male and *x-* [ʃ] for female entities; these are actually the only two prefixes in the language. With the exception of the relational suffixes in certain contexts, all of these inflectional affixes are essentially optional.

The optional nature of plural marking in Yucatec has been widely observed in the literature [Andrade 1955; Lehmann 1998; Lucy 1992; Briceño Chel 2002] and has more recently been the focus of considerable experimental research [Butler 2012, 2013, Butler et al. 2014]. Optional plural marking can be observed in subject-verb agreement, for instance.[1] Hence in (2) the plural subject triggers plural agreement on the verb, which is probably the most frequent option.

(2) *Tumen*   *le*   *paal-al-**o'ob**-o'* <…>   *t-u*   *k'áat-ik-∅-**o'ob***
    because   DET   child-RDP-PL-CL   PROG-ERG.3SG   ask.for-IND-ABS.3SG-PL
    'The children <…> ask for it.' [NM-189]

Plural agreement, however, is optional. Hence in (3) we have once again a plural subject (*things*), but the verb does not show the kind of plural agreement observed in (2). It is also possible, although unusual, to have the opposite situation, that is, to have a verb show plural agreement with a semantically plural subject which is not morphsyntactically marked as plural. This is observed in (4), presumably an example constructed for experimental purposes, and in the text example in (5), where the subject is expressed by a singular demonstrative pronoun, and not by the corresponding plural demonstrative pronoun *lelo'oba'* 'these'.

(3) *tuláakal*   *ba'al-**o'ob***   *k-u*   *taal*
    all   thing-PL   HAB-ERG.3SG   come
    'all (kinds of) things come.' [NM-191]

(4) *Táan*   *u*   *k'aay-**o'ob***   *le*   *x-ch'úupal-o'*.
    PROG   ERG.3SG   sing-PL   DET   FEM-girl-CL
    'The girls are singing.' [Butler 2013:107]

(5) *Lela'*   *teen*   *k-u*   *tomojchi'-t-ik-en-**o'ob***.
    this   1SG   HAB-ERG.3SG   evil.omen-TR-IND-ABS.1SG-PL
    'These (signs) are showing evil omen to me.' [NMC-71]

It is important to note that the optionality of the plural marker is not entirely random. As noted in a number of works that address this topic [Lucy 1992; Lehmann 1998; Uth, Gutiérrez-Bravo 2018, and others], there is a pref-

---

[1] All of the data in this paper are presented in standard Yucatec orthography. In this orthographical system, symbols have their expected values except for *ch* = [tʃ], *j* = [h], *x* = [ʃ], and ' = [ʔ]. In the Yucatec examples, NM corresponds to [Monforte et al. 2010] and NMC to [Can, Gutiérrez-Bravo 2016]. A list of abbreviations can be found at the end of this paper.

erence for animate and human nominal expressions to show the plural suffix, in contrast with inanimate ones. [Uth, Gutiérrez-Bravo 2018] further observe that there is also a preference to use the Yucatec plural suffix in Spanish loanwords when the loanword corresponds to the grammatical subject (see also (3)). Nevertheless, these are only tendencies and none of these preferences is categorical: all else being equal, the Yucatec plural suffix can be observed with any kind of noun, irrespective of its semantic properties or its grammatical relation. We refer the reader to [Uth, Gutiérrez-Bravo 2018] for further details on the interaction of these typological parameters and plural marking of Spanish loanwords in Yucatec.

Now, observe that the same optionality of plural marking is found in the presence of Yucatec numerals. In Yucatec, most of the original Yucatec numerals have been replaced by Spanish loanwords since at least the mid-20th century. The only remaining Yucatec numerals are the ones going from 1 to 4. As shown in (6)–(7) below, plural marking is also optional in this case. From 5 and onwards, Spanish numerals are always used, as in (8), where the Spanish numeral *cinco* 'five' is observed. Plural marking of the Yucatec nouns in this case is equally optional, as illustrated in (9) with the Spanish numeral *trece* 'thirteen'. Observe that plural marking in the presence of both Yucatec and Spanish numerals is attested even though Yucatec is a language with numeral classifiers.

(6)   *le   óox   túul   paal-al-**o'ob**-o'*
      DET   three   CLF   child-RDP-PL-CL
      'the three children' [NMC-85]

(7)   *ka'a   túul   nukuch   tso'*
      two   CLF   big   turkey
      'two big turkeys' [NM-38]

(8)   *cinco   chan   ba'as-**o'ob***
      five   little   suitcase-PL
   'five little suitcases' [NM-135]

(9)   *trece   mejen   luuch*
      thirteen   tiny   gourd
      'thirteen tiny gourds' [NM-138]

As shown by these data and in the references cited above, the optionality of plural marking is robust and widespread in Yucatec. It is therefore particularly
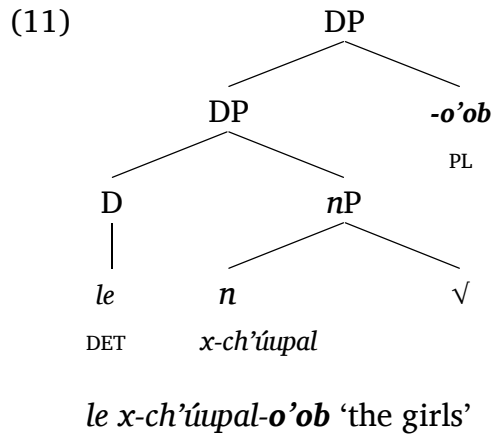
interesting that this optionality is not observed in the presence of the Spanish loanwords that have replaced the original Yucatec numerals in all instances starting with five and above. Before addressing this phenomenon, though, we present a brief sketch of the theoretical tools that we use in the analysis that follows.

3. Two different kinds of plural marking

[Wiltschko 2008] proposes that, crosslinguistically, there are two different kinds of plural marking. In English and Spanish where plural marking is obligatory, it is the result of the presence of a Number Phrase (NumP) above the *n*P headed by a [plural] feature [Bernstein 1991; Ritter 1991, and many others; see also Mathieu 2013]. This structure is illustrated in (10). Since the structure of Spanish nominal expressions is not a central topic of our analysis we assume this simplified version of Spanish DPs, but see [Roca 2015], [Höhn 2016] and references therein for more detailed descriptions of Spanish DP structure.

(10)

```
                DP
              /    \
            D       NumP
                   /    \
                 Num      nP
                [plur]
```
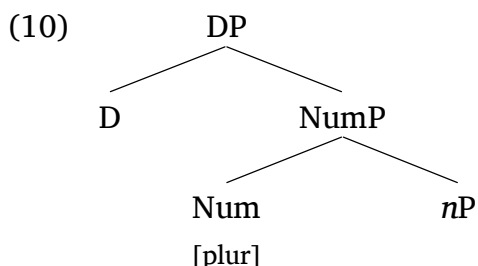
In contrast, in languages where plural marking is optional (for instance, Halkomelem), the plural marker is just an adjunct feature that adjoins somewhere in the N-projection, possibly without any categorial properties of its own. Like any other adjunct, its presence is entirely optional, and this accounts for the optionality of plural marking in those languages where it is observed. This is indeed the analysis proposed in [Butler 2012, 2013] and [Butler et al. 2014], following [Wiltschko 2008]. In this analysis, the Yucatec plural suffix **-o'ob** is an adjunct modifier of DP. We adopt Butler's analysis in its essentials, with one relevant difference: we assume that, all else being equal, there is no NumP in Yucatec nominal expressions. This is crucial for understanding why plural marking is not obligatory in Yucatec, but more importantly, it will be crucial in accounting for the data resulting from language contact with Spanish. As such, the basic structure of nominal expressions in Yucatec that we adopt in our analysis is the following:

(11)

```
                        DP
                     /      \
                  DP         -o'ob
                /    \        PL
              D       nP
              |      /   \
             le    n      √
            DET  x-ch'úupal
```

*le x-ch'úupal-**o'ob*** 'the girls'

With this theoretical background we now address the central topic of this paper. We begin by presenting a description of the relevant data.

4. Plural marking of Spanish nouns in Yucatec

4.1. Basic description

As previously mentioned, the inflectional morphology of Yucatec nouns is surprisingly simple when compared to its complex verbal morphology. Spanish nominal morphology is considerably more complex, but crucially, the only inflectional category in which the two languages intersect is precisely plural.

Table 1. Nominal morphology in Yucatec and Spanish

Yucatec	Spanish
plural	plural
honorific	gender
relational suffix	affective suffixes

According to some works on language contact [Thomason Kaufmann 1988; Klee, Lynch 2009], mutual translinguistic influence is more likely between typologically similar languages, since "when parallel structures exist in the two languages, it is much more likely that there is linguistic transfer or convergence between them" [Klee, Lynch 2009: 20, our translation]. Hence it is arguably not surprising that the complex interaction between the two languages described in what follows has to do with the inflectional category shared by both.

Recall that plural marking in Yucatec nouns is essentially optional. Nouns borrowed from Spanish, however, exhibit a puzzling behavior. Spanish loanwords can show plural morphology, both with the Spanish plural suffix -*s* (as in *clase* 'class' and *tejido* 'knitting' in (12)) **and** with the Yucatec plural suffix -*o'ob* (as in *abuelo* 'grandparent' in (13)).

(12)   *Clase-s-il   tejido-s   k-in   meen-t-ik-∅.*
       class-PL-RL   knitting-**PL**   HAB-ERG.1SG   make-TR-IND-ABS.3SG
       'I make different kinds of knittings.' [NM-31]

(13)   *Tumen   leti'ob,   u   abuelo-'ob   in   w-íicham-o'*
       because   them   his   grandparent-**PL**   my   EP-husband-CL
       'Because them, my husband's grandparents …' [NM-24]

More interestingly, in most cases plural marking on these loanwords is equally optional. This is shown in (14b), where *candado* 'padlock' shows no plural inflection in the presence of a Yucatec non-singular numeral (cf. 14a), something which is completely impossible in Spanish.

(14) a. *Yaan-∅   kex   ka'a p'éel   docena-s   u   y-alak' peek'.*
        EX-ABS.3SG about   two CLF   dozen-**PL**   his   EP-CLF dog
        'He had about two dozen dogs.' [NM-209]

   b. *Óox   p'éel   candado   yaan-∅-i'.*
      three   CLF   padlock   EX-ABS.3SG-LOC
      'It had three padlocks.' [NM-26]

Summing up, the general pattern illustrated so far is that plural Spanish nouns can show three different forms: with the Spanish plural suffix (12), with the Yucatec plural suffix (13), or with no plural morphology whatsoever ((14b); see also Fig. 2).[2] However, our results from a corpus study [Uth, Gutiérrez-Bravo 2018] show that Spanish nouns preceded by a Spanish numeral **always** show the Spanish plural suffix -*s* and **only** the suffix -*s*, as in (15)–(17), where Spanish loanwords are underlined.

(15)   *jach   diez metro-s   wal-e'.*
       very   ten meter-**PL**   perhaps-CL
       'at most perhaps ten meters' [NM-48]

---

[2] In §4.2 we will see that there is actually a fourth possibility.

(16) *diez*   *bolsa-s*   *lu'um*
      ten   bag-**PL**   soil
      'ten bags of soil' [NM-13]

(17) *Yaan-∅*   *kex*   *siete*   *rancho-s*   *ti'-∅.*
      EX-ABS.3SG   about   seven   ranch-**PL**   PREP-ABS.3SG
      'He had about seven ranches.' [NM-209]

In what follows, we only analyze the data from [Uth, Gutiérrez-Bravo 2018]. Since this is corpus data, it is subject to the same inherent limitations as any corpus study. Specifically, in what follows we are unable to provide minimal pairs and negative evidence to further support the descriptive generalizations outlined above. However, it is worth mentioning that in these data, plural marking of a Spanish noun in the presence of a Spanish numeral is categorical, as shown in Figure 2.[3] Categorical effects are not often found in corpus data, so we consider our data to be reliable. The complete quantitative results, originally reported in [Uth, Gutiérrez-Bravo 2018], are shown in the graph below (SN = Spanish numeral).
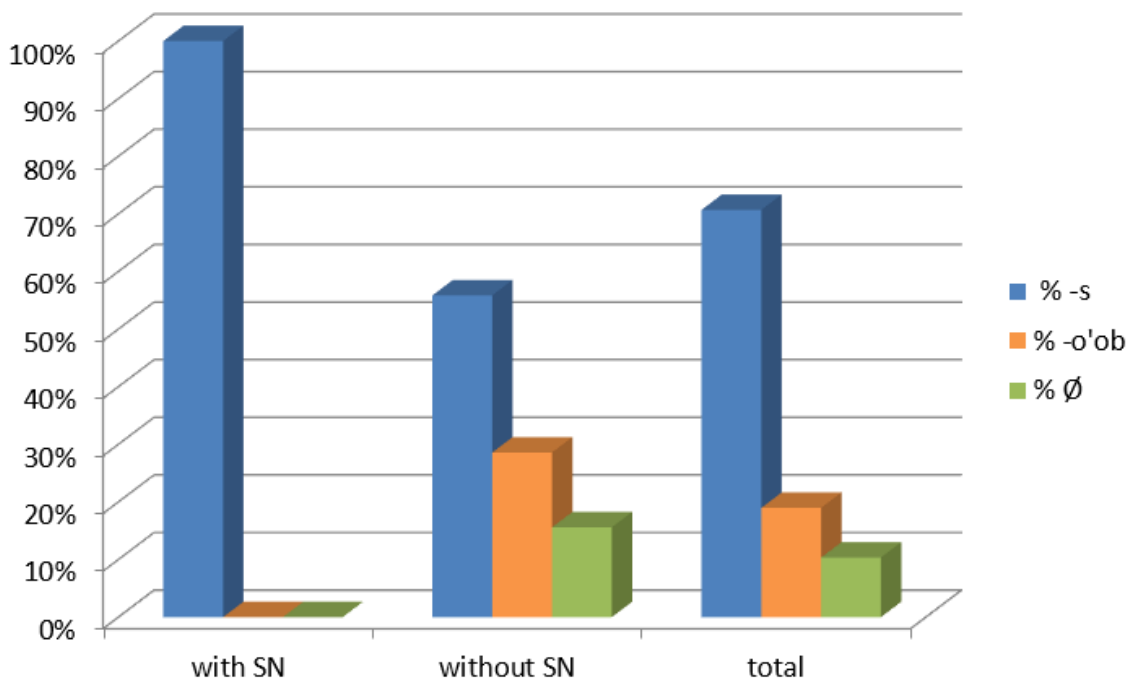
Figure 2. Plural marking of Spanish loanwords preceded by a Spanish numeral *vs*. Spanish loanwords without a Spanish numeral

---

[3] Christian Lehmann (personal communication) also found this effect to be categorical in his own corpus, which is one of the largest, and certainly the most detailed corpus of Yucatec narratives that we know of.

Although data in our corpus is too scarce to make any conclusions (N = 1), plural marking with the Spanish suffix is also observed in an alternative construction in Yucatec in which a possessed numeral classifier introduces the Spanish noun, as in (18). This construction is widely reported in descriptions of Yucatec from the 1950–80s, but is essentially out of use today.[4]

(18) *ocho*   u   *túul-ul*   *señora-s*   ts'-u   *k'uch-ul-o'ob-i'.*
      eight   ERG.3   CLF-RDP   lady-**PL**   TRM-ERG.3   arrive-IND-PL-LOC
      'eight ladies had already arrived.' [NM-273]

Crucially, the observed behavior is categorical in the presence of a Spanish numeral, but not in its absence: when there is no Spanish numeral, the presence of the Spanish plural suffix *-s* is not obligatory, as in (14b). This shows that the obligatory nature of plural marking in (15)–(17) cannot simply be the result of the fact that plural morphology is obligatory in Spanish. In what follows we propose that the relevant facts receive a straightforward account with Wiltschko's [2008] analysis of two formally different kinds of plural.

4.2. Analysis

To account for the Yucatec data, we propose an analysis where Spanish numerals have kept their original selectional properties, even after having been long borrowed by Yucatec. These numerals c-select a NumP, as they do in Spanish, and so plural marking becomes obligatory, as originally illustrated in (10). Because of the selectional properties of the Spanish numeral, the resulting structure in (19), based on (15), is no different from what it would be in Spanish.
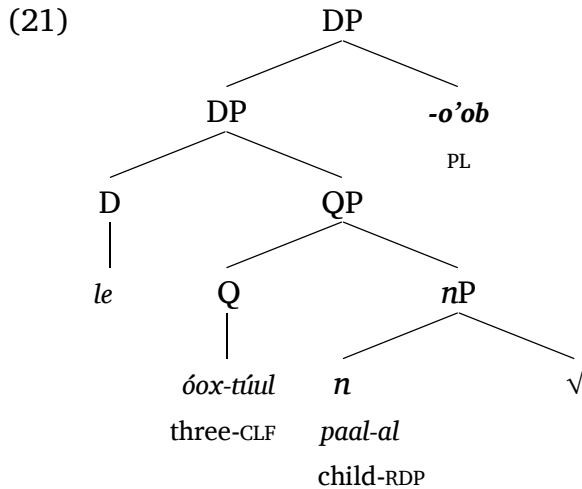
---

[4] We label this construction "alternative" because we assume that, when it was still productive, it coexisted (alternated) with the constructions that show the pattern that we analyze here, i.e. (8)–(9) and (15)–(17). We assume this based on the fact that the monolingual speaker that produced (18) also produced one other construction with a Spanish numeral, but with the current pattern of (8)–(9) and (15)–(17). This is only an assumption: it is of course impossible to conclude on the basis of only two data points that such an alternation existed when the construction in (18) was still productive. Since this alternative construction has all but disappeared в contemporary spoken Yucatec, it is unclear if further verification of this assumption is possible, but this is not in any way crucial for the analysis we propose of the contemporary construction in (8)–(9) and (15)–(17).

(19)

```
              QP
            /    \
         diez    NumP
         ten    /    \
              -s      nP
              PL     metro
                     meter
```

diez metro-s 'ten meters'
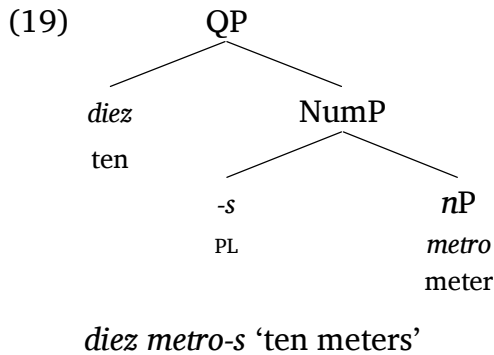
Yucatec numerals and other functional heads in the nominal domain, however, lack this lexical property: in Yucatec [plural] is an adjunct and not the head feature of a complement that can be selected for (§3). As such, NumP is not selected by any head in the nominal domain, and in this way the non-obligatory nature of plural marking, even in the presence of a Yucatec numeral, is accounted for. This is illustrated in (20)–(21).

Importantly, our analysis also provides specific and concrete evidence that morphological number marking in Spanish depends exclusively on a functional head distinct from *n*/N. This is because of two different considerations. The first one is that *-s* is subject to c-selection by another syntactic head, i.e. the numeral in (19), as is typical of syntactic (versus purely morphological) constituents. The second one has to do with semantic compatibility: in (20) there is no NumP, and consequently no source for morphological plural marking of the Spanish noun with the plural suffix *-s*. But clearly this absence does not bring with it a singular (atomic) interpretation of the noun (see [Wiltschko 2008]): if it did, we would expect a semantic anomaly to arise between the plural numeral *óox* 'three', and the noun *candado* 'padlock'. We conclude that whatever singular or plural specification the Spanish noun has must come from some other element that is *not* the noun itself, i.e., the head of NumP. This in turn supports the proposal in [Wiltschko 2008] that nouns are not lexically specified for [number], contra [Chierchia 1998].
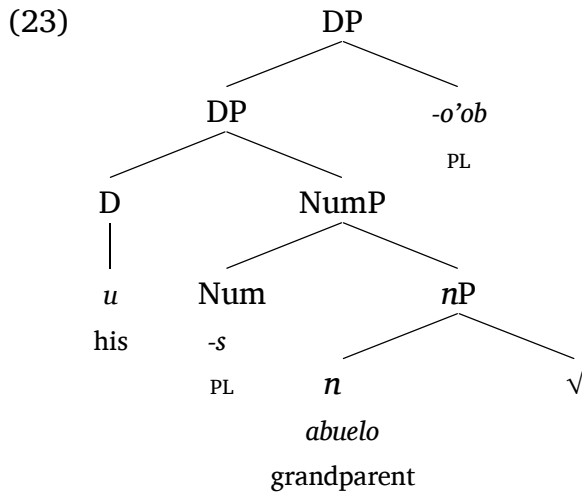
(20)

```
                QP
              /    \
        óox-p'éel    nP
        three-CLF   /    \
                   n      √
                candado
                padlock
```

óox-p'éel candado 'three padlocks'

(21)

```
                         DP
                ┌─────────┴──────────┐
               DP                  -o'ob
          ┌─────┴──────┐            PL
         D             QP
         │        ┌─────┴──────┐
         le       Q           nP
                  │       ┌─────┴─────┐
               óox-túul   n           √
               three-CLF  paal-al
                          child-RDP
```
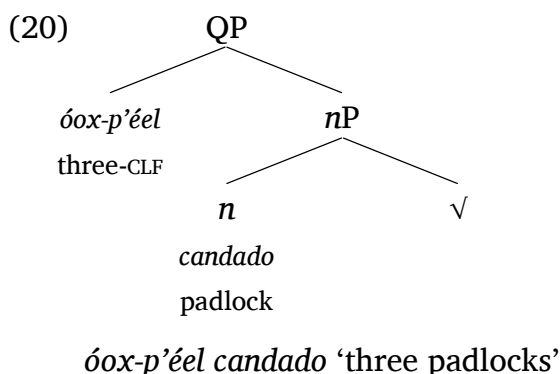
*le óox-túul-paal-al-**o'ob*** 'the three children'

Observe now that the two plural markers are independent from one another (one is the result of lexical selection, the other one of adjunction), so in principle it should be possible to find both of them simultaneously in the same language. This indeed has recently been argued to be the case in Blackfoot [Kim et al. 2017]. We suggest that Yucatec instantiates a similar kind of language, with the difference that one of the two plural markers originates from language contact. Furthermore, since the two different plural features are formally different (one is a head, the other one is just an adjunct), in principle it should be possible for a noun to simultaneously display them both. In contrast to Blackfoot, this is indeed attested in Yucatec, where Spanish loanwords can simultaneously show the plural suffix of Spanish and the plural suffix of Yucatec.

(22) *ayik'al   le   u   <u>abuelo-**s-o'ob**</u>*
     rich   DET   his   grandparent-PL-PL
     'his grandparents were rich' [NM-24]

The only attested order in this case is *-s* + *-o'ob* (see [Uth, Gutiérrez-Bravo 2018]). The inverse order (*-o'ob* + *-s*) is unattested in our data, and is not reported elsewhere in the literature. Our analysis accounts for this fact, since *-s* is the head of the NumP, but *-o'ob* right-adjoins to the DP which dominates NumP. This fact also provides an additional argument in favor of Butler's analysis of the Yucatec plural suffix as an adjunct and not a head.

(23)

```
                        DP
                   _____
                  DP              -o'ob
              _____            PL
             D        NumP
             |      _____
             u    Num        nP
            his    -s      _____
                   PL     n        √
                        abuelo
                     grandparent
```

u abuelo-s-o'ob 'his grandparents'

The Yucatec data are particularly relevant because they provide strong evidence in favor of Wiltschko's analysis that languages can show two structurally different kinds of plural. As mentioned, since the two kinds of plural are structurally different and completely independent from one another, in principle one would expect to find languages in which both plurals coexist. The Spanish loanwords of Yucatec Maya with double plural marking are exactly the case in point, and so they fulfill this expectation in a particularly convincing way. For instance, [Kim et al. 2017] propose that Blackfoot also has the two different kinds of plural of [Wiltschko 2008]. However, in Blackfoot the two plural suffixes have the same phonetic realization, i.e. the plural that is a head feature and the plural that is an adjunct feature are not phonemically distinct (they are homophonous). Additionally, although in principle possible, double plural marking is actually unattested in Blackfoot, a fact that [Kim et al. 2017] attribute to semantic anomaly. In contrast, the situation observed in Yucatec is much more straightforward, since the two different plural suffixes are clearly distinct and double plural marking by combining them is indeed possible.[5]

Our analysis of the coexistence of these two plural suffixes in Yucatec also relates in an interesting way to the Late Insertion Hypothesis. In principle, since Num^0 has no phonemic content before Spellout, we would expect that the [plur] feature could be realized as either the plural suffix of Spanish or the plu-

---

[5] Since we expect the semantics of plural marking in this case to be the same for Blackfoot and Yucatec, this also points to the conclusion that the absence of double plural marking in Blackfoot is not due to a semantic restriction, and must be accounted for in a different way.

ral suffix of Yucatec. However, as we have shown, this is not the case: Spanish loanwords preceded by a numeral always show the suffix *-s* and only this suffix. Crucially, this is precisely what is expected in an analysis that adopts Wiltschko's proposal, where the two different kinds of plurals are formally different. The plural suffix of Spanish has the possibility of functioning as a head, the plural suffix of Yucatec does not.[6] Consequently, if the [plur] feature is found in a head position at the point of Spellout, this feature can only be realized as the suffix *-s* of Spanish, and if it is found in an adjoined position, it can only be realized as the suffix *-o'ob* of Yucatec.

Our analysis of this language contact situation further makes an interesting prediction: if a language has its own plural marker, and it has also borrowed the plural marker from the contact language, it is predicted that (while optional, like in Yucatec) the simultaneous presence of both plural markers should never be obligatory, even if both plural markers are heads (and not adjuncts), as illustrated in (24).

(24) *Noun -PL   -PL
          obligatory   obligatory

This is because the obligatory presence of Num^0 is the result of selection by a higher head. However if there are two NumPs, the head of the higher one would have to select the lower NumP, but (to the best of our knowledge) we do not know of any cases reported in the literature where a Num^0 selects a second Num^0. In other words, we hypothesize that selection of another NumP is not part of the lexical information of either Num^0-*1* or Num^0-*2*. Because of this the following configuration cannot be obligatory in the way that it is when a numeral (Q) selects NumP.

(25)

```
              QP
            /    \
          Q      NumP-1
                /    \
            [plur]   NumP-2
                    /     \
                [plur]    ...
```

---

[6] Ultimately, it is reasonable to assume that this is specified as part of the lexical information of each of these suffixes.

In this structure, NumP-*2* can appear optionally, in the same way that the NumP headed by the Spanish suffix *-s* can appear optionally and is not obligatorily present unless a Spanish numeral selects it (we briefly address this specific point in §6). However, as mentioned, we submit that it should not be obligatory since it is not being selected by the higher Num head. Furthermore, if, as is most often assumed, the configuration in (25) can *only* be the result of Num⁰-*1* selecting NumP-*2*, then the prediction made by our analysis is even stronger: the specific configuration in (25) should quite simply never be attested at all.[7] Again, this is because the head of the higher NumP would have to select another NumP, again a situation theoretically unheard-of to the best of our knowledge.

5. Alternative analyses

In this section we discuss two alternative analyses of the data we have presented here, each of which was pointed out to us by an anonymous reviewer. The first alternative analysis has to do with our interpretation of the data. A reviewer points out that there are cases where it is not immediately obvious if the contact phenomena studied here corresponds to borrowing (as we claim in our analysis) or code-switching: see, for instance, [Treffers-Daller 1991]. More specifically to the point at hand, [Sebonde 2014] notes that this is particularly the case of single-word switches in numerical expressions, and furthermore, [Khomchenkova, Pleshak 2019] show that there are unequivocal cases of code switching of numerical expressions in Moksha and Hill Mari (Finno-Ugric, Russia), which additionally indicate that cardinal numerals seem to require from their complement to be a well-formed nominal expression of the same language, i.e., no switching is allowed within a numeral construction. The resulting switches to Russian, of which we present an example in (26), look very similar to the Yucatec data presented so far.

(26) Hill Mari

st'ip'end'ij-žǝ	*dv'enacat'*	*rubl'-ej*	*ǝl'-ǝ*	*v*	*m'es'ac*
scholarship-POSS.3SG	twelve	ruble-GEN.PL(rus)	be-AOR.3SG	in	month

'The scholarship was 12 rubles per month.' [Khomchenkova, Pleshak 2019]

There are numerous facts, however, that point to the conclusion that the Yucatec data is not the result of code-switching. First, the original numeral system

---

[7] We are thankful to an anonymous reviewer for bringing this latter point to our attention.

of Yucatec has not been conserved, except for the four cardinal numbers going from 1 to 4 (see §2). As such, it is unsurprising that our data set includes data from monolingual Yucatec speakers (see [Uth, Gutiérrez-Bravo 2018] for further details). However, these data (which do not show any relevant differences when compared to the data from bilingual speakers) could not possibly have been the result of code-switching. Secondly, and also in relation to this, it is common to distinguish borrowings from code switching on the basis of predictability: the occurrence of any given borrowing is mostly predictable, while code-switching never is. In this respect, the occurrence in Yucatec of every Spanish numeral above four is entirely predictable, since only the cardinal numerals from 1 to 4 have been conserved.

There is also morphosyntactic evidence which supports the conclusion that the constructions we have analyzed here are borrowings. We have extracted the relevant parts of the complete examples above and repeat them below for simplicity in (27)–(28), (30). Consider example (27) first. Both here and in [Uth, Gutiérrez-Bravo 2018] we assume the code-switching typology in [Poplack 1980]. According to this typology, words that have adapted morphology from the recipient language correspond to borrowings and not to code-switching. Based on this, (27), where the Yucatec plural suffix is used instead of the Spanish plural suffix, should be considered a borrowing (see also (22)).

(27) *u   abuelo-'ob   in   w-íicham-o'*
     his   grandparent-**PL**   my   **EP**-husband-**CL**
    'my husband's grandparents' (extraction from (13))

A word of caution is in order here, though. Since, as mentioned in §4.1, Spanish loanwords preceded by a Spanish numeral never show Yucatec plural morphology in our corpus, this evidence can only be taken to be indirect. Nevertheless, on the basis of the much larger data set of pluralization of Spanish loanwords analyzed in [Uth, Gutiérrez-Bravo 2018] (to which we refer the reader for further details), we do not find any reason to differentiate between the relevant data that shows only Spanish plural morphology and the data that shows Yucatec morphology or mixed Yucatec and Spanish morphology. Further observe that this limitation applies only to the Spanish nouns in the data, since, as mentioned above, the fact that the original Yucatec numeral system has not been conserved allows us to safely conclude that Spanish numerals are unquestionably loanwords.

More evidence can be found in data like (28), which corresponds to a complex nominal expression in which the embedded NP *tejidos* 'knittings' complements the head of the larger NP, *clases* 'kinds'.

(28) *Clase-**s**-il*   *tejido-**s***
     class-**PL**-RL   knitting-**PL**
     'different kinds of knittings' (extraction from (12))

Even though the two nouns of this nominal expression come from Spanish, this is certainly not a possible Spanish nominal expression. In this nominal expression we find the Yucatec relational suffix *-il*, which indicates that the first noun is related to the second one, playing a role somewhat similar to that of the Spanish preposition *de* 'of', in the Spanish nominal expression *clases de tejidos* '(different) kinds of knittings'. It is precisely the presence of the relational suffix and the absence of the Spanish preposition that indicates that this is a nominal expression composed of two loanwords, and not a switch. Additionally, the fact that the suffix *-il* intervenes between the two Spanish parts of the NP results in a pattern which is not normally considered as code-switching in the literature, since it violates (at least) the free morpheme constraint, according to which "[c]odes may be switched after any constituent in discourse provided that constituent is not a bound morpheme" [Poplack 1980: 585f]. As such, this constitutes independent evidence that the Spanish plural suffix is productively found in constituents that cannot be analyzed as switches. Crucially, the Yucatec relational suffix *-il* can be observed in the numeral constructions analyzed in this paper, as in (29).

(29) *seis* *mes-**es**-il*   *u*   *ts'o'ok-ok-∅*   *u*   *beel-e'*
     six   month-**PL**-RL   ERG.3   finish-IRR-ABS.3SG   his   path-CL
     'six months after he gets married' Lit.: 'six months (after) his path finishes.' [NM-222]

Finally, additional evidence can be found in the numeral construction in (30), originally presented in (18). Recall that at some point this construction was the canonical construction used with loanword numerals from Spanish.

(30) *ocho*   *u*   *túul-ul*   *señora-**s***
     eight   ERG.3   CLF-RDP   lady-**PL**
     'eight ladies' (extraction from (18))

In this example the Spanish numeral and the Spanish noun it quantifies are not syntactically adjacent. This represents a clear example against a code-switching analysis, since there is no evidence for chunks being interrupted in this way as the result of code-switching. While it is true that this construction is no longer productive in Yucatec, it still provides evidence in favor of our analysis, since it shows Spanish numerals and Spanish nouns functioning as loanwords (not code-switching) even at an earlier stage of the language.

The second alternative analysis pointed to us by a different anonymous reviewer corresponds to a different formalization of the analysis. The reviewer points out that there is an alternative way of analyzing multiple plural markers in loanwords, which does not need to appeal to the different syntactic status of the plural markers involved. Consider the following data from Russian, where the English loanwords have kept their original plural marker *-s* and also show the characteristic number/case declensions of Russian:

(31) Russian

    a. *čips-y*   b. *v MS Teams-ax*

       chips-NOM.PL   in MS Teams-PRP.PL

       'potato chips'   'in Microsoft Teams'

The reviewer points out that, since Russian and English are structurally much closer to one another than Yucatec and Spanish, it does not seem feasible to analyze the two plural markers as having a different syntactic status. We agree with the reviewer on this point. Now, since the two plural markers probably do not have a different syntactic status in this case, our proposal cannot be applied to the Russian data, and consequently a different analysis altogether is required.

The reviewer suggests an analysis in which the English plural suffix in these examples has been reanalyzed as belonging to the lexical root. As such, *čips* 'chips' enters the syntactic derivation as an inseparable lexical item which then merges with the Russian plural, which results in the apparent sequence of two adjacent plural suffixes observed in (31): we agree with the reviewer both with respect to the fact that this is a theoretically plausible analysis and that it is very likely the correct analysis for the relevant Russian data. Yet, if an analysis along these lines is independently needed for data like (31), it would be preferable to adopt this very same analysis for the Yucatec data, both in terms of having a unified analysis for the whole data set and in terms of having an analysis that does not depend on the different syntactic status of the two plural suffixes.

However, one specific characteristic of the Yucatec data makes it very problematic to adopt this analysis for the facts presented so far. Specifically, the Spanish plural suffix is optional (just like its Yucatec counterpart) in every context except in the presence of a Spanish numeral. This can be observed in (14), repeated here has (32), where the Spanish plural suffix -*s* is optional even in the presence of a Yucatec numeral.

(32) a. *Yaan-∅   kex   ka'a p'éel   docena-s u   y-alak'   peek'.*
       EX-ABS.3SG   about   two   CLF   dozen-**PL** his   EP-CLF   dog
      'He had about two dozen dogs.' [NM-209]

    b. *Óox   p'éel   candado   yaan-∅-i'.*
       three   CLF   padlock   EX-ABS.3SG-LOC
      'It had three padlocks.' [NM-26]

The optionality of the Spanish plural suffix indicates that in those loanwords where it does appear it has not been reanalyzed a part of the root, but instead remains a distinct morphological entity as in Spanish and also like the Yucatec plural suffix. Since the competing analysis laid out above crucially depends on the Spanish plural suffix having been reanalyzed as part of the root, we conclude that it would be problematic to adopt this competing analysis and consequently that an analysis along the lines of the one we propose is necessary for the Yucatec data.

Furthermore, it is worth observing that, irrespective of the language contact situation that we analyze here, the behavior of the Yucatec plural suffix is so different from that of the plural suffixes of Spanish and English that it seems necessary to conclude that it is an entirely different morphosyntactic entity, which is precisely the proposal of [Butler 2012, 2013]. Consequently, it is necessary to analyze the plural suffixes of Yucatec and Spanish as having a different syntactic status independently of the specific language contact data that we present here. As such, although it would certainly be appealing in analytical and theoretical terms to have a unified analysis of the Spanish and Yucatec plural suffixes, the data indicates that this possibility might simply not be available (see also [Wiltschko 2008] and [Kim et al. 2017]).

6. Discussion

Before concluding this paper, we briefly address two points which require a more thorough discussion than can be provided here. The first one relates to the specific architecture of the analysis that we have developed so far, the sec-

ond one relates to a much broader (and at this point, speculative) discussion of what exactly facilitates linguistic transfer like the one described in this paper.

With respect to the first point, recall that, when there is no Spanish numeral, Spanish plural morphology is optional in Spanish loanwords, as in (32). The relevant point is whether it is not problematic to propose that the presence of a functional head can be obligatory in some contexts (i.e. (15)–(17)), but optional in others, as in (32). With respect to this, it is worth pointing out that this very same situation is observed in other syntactic phenomena. One that is particularly well known is the presence or the absence of the complementizer *that* in English.

(33) a. *He thinks* [*that the visitors are not really comfortable*].

b. *He thinks* [*the visitors are not really comfortable*].

From the perspective of an extended projection analysis [Grimshaw 1997] the absence of a head with explicit phonetic content in most cases indicates that the head is not present in the syntactic representation/numeration at all, which is the analysis we assume for (33). Yet, as is well known, there are some contexts where these otherwise optional functional heads become obligatory. Hence, *that* deletion is not possible when the clause is a sentential subject or when the clause and the verb are separated by an adverbial expression:

(34) a. [*That Fred is feeding her rabbit*] *annoys Karen.*

b. *[*Fred is feeding her rabbit*] *annoys Karen.* [Baker 1989: 144]

(35) a. *We didn't know, until we read the story in the paper,* [*that he had resigned*].

b. *We didn't know, until we read the story in the paper,* [*he had resigned*]. [Radford 2016: 198]

The structures in (34)–(35) are obviously not parallel to what is observed in Yucatec, where there is a strictly local head-complement configuration between the numeral and NumP, but they clearly illustrate the more general situation in which the presence of an otherwise optional functional head becomes obligatory in some specific context. Observe now that *that* deletion is not possible or becomes degraded in a head-complement configuration where the matrix verb is a manner-of-speaking, factive, or response-stance verb [Stowell 1981, Baker 1989, Hiroe 1999].

(36) a. *Jack whispered* [*that the X-files were strictly confidential*].

b. **Jack whispered* [*the X-files were strictly confidential*]. [Hiroe 1999: 68]

In these cases, where the relevant syntactic configuration is just like the head-complement configuration in (19)–(20), the most plausible explanation of course is to attribute the absence of *that* deletion to the lexical properties of specific heads like the verb *whisper*. This is very much like the situation observed in Yucatec: NumP/*-s* can be optionally present in bare nominal expressions and when the preceding head is a Yucatec determiner or numeral, but its presence becomes obligatory when the preceding head is a Spanish numeral.

In this respect, an anonymous reviewer brings up an important point. The reviewer asks whether this part of our analysis is not problematic in view of the arguments presented in [Bruening 2009] and [Bruening et al. 2018] against Grimshaw's proposal that the clause is an extended projection of the verb. We do not think that this is problematic for our analysis for two reasons. First, even if correct, it is not immediately obvious that these authors' proposal can be extended across a broad typological spectrum. For instance, at first sight polysynthetic languages and languages where complement "clauses" are nominals and not clauses would appear to be a challenge to the proposal by these authors. More specifically for our purposes, Yucatec is clearly not typologically similar to the languages contemplated by these authors and so it doesn't seem possible to claim, without further analysis, if these authors' objections can be immediately carried through to Yucatec. The second reason why we think that this is not a problem for our analysis is that the basic proposal by these authors is that "clauses and nominals are not parallel at all" [Bruening et al. 2018: 6] and that this in turn indicates that the DP analysis is not correct. But crucially, in these authors' proposal one of the facts that shows that clauses and nominals are not parallel is that there is no sense in which the clause is an extended projection of the verb, while nominals clearly *are* a projection of the noun [Bruening 2009: 27; Bruening et al. 2018: 2, 12, 46]. This is the central claim of the extended projection analysis when applied to nominals. Our data and analysis correspond exclusively to the nominal domain, and so we do not consider that a critique of the notion of the clause as an extended projection of verb applies to our proposal.

The second point we would like to mention briefly in this final section has to do with what specifically facilitates or allows linguistic transfer as the result of

language contact. We previously mentioned that we follow those analyses which claim that the interaction between two languages in contact is most likely to occur in those areas that the grammars of these languages have in common. However, our primary goal in this paper has been to provide specific evidence that, morphosyntactically, the plural suffixes of Yucatec and Spanish are very different elements. Consequently, what the data presented here seems to indicate is that semantics, and not morphology and syntax, plays a more important role in linguistic transfer. More specifically, if the convergence illustrated in Table 1 were dependent on a concrete convergence of the formal morphological and syntactic properties of the elements in question, it is hard to see how this transfer from Spanish to Yucatec could have occurred in the first place. This leads us to speculate that it may be the case that originally the Spanish plural suffix was adopted as an adjunct, i.e., having morphological and syntactic properties closer to those of the Yucatec suffix, and that only at a later stage did this suffix begin to display a morpho-syntactic behavior more closely resembling the one observed in Spanish. This is not only an idea worth pursuing, but additionally one for which the relevant data probably exists, since there exist almost 500 years' worth of written Yucatec documents, going back to the beginning of the colonial era. Appealing as this idea is, it is without doubt beyond the scope of this study and should be left for further investigation in the future.

7. Conclusions

In this paper we have provided an analysis of plural marking in Spanish loanwords in Yucatec Maya. Our results strongly confirm the specific formalization in [Wiltschko 2008] of two different kind of plural affixes that can be observed in different languages. The two different kinds of suffixes coexist in Yucatec Maya as the result of language contact with Spanish. The hypothesis that we submit from the analysis of these data is that the adoption by a recipient language A of functional head X^0 from a language B does not entail that X^0 will behave in language A in the same way that it does in language B. Hence, by itself, the Spanish plural suffix behaves just like the Yucatec plural suffix in most respects. However, the adoption by a recipient language A of both a functional head X^0 and a lexical (or functional) head Y^0 that selects X^0 does bring with it a distribution of X^0 that is similar to the one observed in the contact language B. As such, in the presence of a Spanish numeral, the Spanish plural

suffix behaves in exactly the same way it would in Spanish. We have shown that in this analysis the relevant facts follow directly from the standard properties of Merge and lexical selection. Finally, we showed that further aspects of the interaction between the plural suffixes Spanish and Yucatec, including their simultaneous co-occurrence and their relative order, are also straightforwardly accounted for in the analysis we propose.

Abbreviations

ABS — absolutive; AOR — aorist; CL — clitic; CLF — classifier; DET — determiner; EP — epenthesis; ERG — ergative; EX — existential; FEM — female; GEN — genitive; HAB — habitual; IND — indicative; IRR — irrealis; LOC — locative; PL — plural; POSS — possession; PREP — preposition; PRP — prepositional case; PROG — progressive; RDP — reduplication; RL — relational; SG — singular; SUF — suffix; TR — transitive; TRM — terminative.

References

Andrade 1955 — Andrade M.J. A grammar of Modern Yucatec. Chicago: The University of Chicago microfilm collection of manuscripts, 1955 [1940].
http://www.christianlehmann.eu/ling/sprachen/maya/andrade/index.html.

Baker 1989 — Baker C.L. English syntax. Cambridge, MA: MIT Press, 1989.

Bernstein 1991 — Bernstein J. DPs in Walloon: Evidence for parametric variation in nominal head movement. Probus. 1991. Vol. 3. P. 101–126.

Briceño Chel 2002 — Briceño Chel F. Lengua e identidad entre los mayas de la península de Yucatán. Los investigadores de la cultura maya. 2002. Vol. 10. P. 370–379.

Bruening 2009 — Bruening B. Selectional asymmetries between CP and DP suggest that the DP hypothesis is wrong. Proceedings of the 32rd Annual Penn Linguistics Colloquium. MacKenzie L. (ed). 2009. P. 26–35.

Bruening et al. 2018 — Bruening B., Dinh X., Kim L. Selection, idioms, and the structure of nominal phrases with and without classifiers. Glossa: A Journal of General Linguistics. 2018. Vol. 3(1): 42. P. 1–46. DOI: http://doi.org/10.5334/gjgl.288.

Butler 2012 — Butler L.K. Crosslinguistic and experimental evidence for non-Number plurals. Linguistic Variation. 2010. Vol. 12. P. 27–56.

Butler 2013 — Butler L.K. Non-Number plurals. Proceedings of the 42nd Meeting of the North East Linguistic Society (GLSA Publications). 2013. Vol. 1. P. 105–118.

Butler et al. 2014 — Butler L.K., Bohnemeyer J., Jaeger T.F. Syntactic constraints and production preferences for optional plural marking in Yucatec Maya. Zwischen Kern und Peripherie: Untersuchungen zu Randbereichen in Sprache und Grammatik. Machicao y Priemer A., Nolda A., Sioupi A. (eds.). Berlin: De Gruyter, 2014. P. 181–207.

Can, Gutiérrez-Bravo 2016 — Can Canul C., Gutiérrez-Bravo R. Narraciones mayas de Campeche (Maayáaj tsikbalilo'ob Kaampech). Mexico City: Instituto Nacional de Lenguas Indígenas. 2016

Chierchia 1998 — Chierchia G. Reference to kinds across languages. Natural Language Semantics. 1998. Vol. 6. P. 339–405.

Coon 2016 — Coon J. Mayan morphosyntax. Language and Linguistics Compass. 2016. Vol. 10. P. 515–550.

England 2017 — England N.C. Mayan Languages. Oxford Research Encyclopedia of Linguistics (online). DOI: http://doi.org/10.1093/acrefore/9780199384655.013.60.

Grimshaw 1997 — Grimshaw J. Projection, heads and optimality. Linguistic Inquiry. 1997. Vol. 28. P. 373–422.

Hiroe 1999 — Hiroe A. Mood and complementizer deletion. English Linguistics. 1999. Vol. 16. P. 55–77.

Höhn 2016 — Höhn G. Unagreement is an illusion: Apparent person mismatches and nominal structure. Natural Language and Linguistic Theory. 2016. Vol. 34. P. 543–592.

Khomchenkova, Pleshak 2019 — Khomchenkova I., Pleshak P. Russian numerals in Moksha and Hill Mari speech. Poster presented at the Conference on the syntax of Uralic languages (SOUL 2019), Tartu, June 18–19, 2019.

Kim et al. 2017 — Kim K., Ritter E., Wiltschko M., Rullmann H. $2+2=3$: Number contrasts in Blackfoot. Glossa: A Journal of General Linguistics. 2017. Vol. 2, 96. P. 1–15.

Klee, Lynch 2009 — Klee C., Lynch A. El español en contacto con otras lenguas. Washington, D.C.: Georgetown University Press, 2009.

Lehmann 1998 — Lehmann C. Possession in Yucatec Maya. Munich: Lincom Europa, 1998.

Lucy 1992 — Lucy J.A. Grammatical categories and cognition: A case study of the linguistic relativity hypothesis. Cambridge: Cambridge University Press, 1992.

Mathieu 2013 — Mathieu E. On the mass/count distinction in Ojibwe. Count and mass across languages. Massam D. (ed.). Oxford: Oxford University Press, 2013. P. 172–198.

Monforte et al. 2010 — Monforte J., Dzul L., Gutiérrez-Bravo R. Narraciones Mayas. Mexico City: Instituto Nacional de Lenguas Indígenas. 2010.

Poplack 1980 — Poplack S. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching, Linguistics. 1980. Vol. 18. P. 581–618.

Radford 2016 — Radford A. Analysing English sentences, 2nd ed. Cambridge: Cambridge University Press, 2016.

Ritter 1991 — Ritter E. Two functional categories in modern Hebrew noun phrases. Perspectives on phrase structure: Heads and Licensing (Syntax and Semantics Vol. 25). Rothstein S. (ed.). New York: Academic Press, 1991. P. 37–60.

Roca 2015 — Roca F. Determinación y estructura del sintagma nominal. Perspectivas de sintaxis formal. Gallego A. (ed.). Madrid: Akal, 2015. P. 177–222.

Sebonde 2014 — Sebonde R.Y. Code-switching or lexical borrowing: Numerals in the Chasu language of rural Tanzania. Journal of Arts and Humanities. 2014. Vol. 3. P. 67–76.

Stowell 1981 — Stowell T. Origins of phrase structure. Ph.D. thesis. MIT. 1981.

Treffers-Daller 1991 — Treffers-Daller J. Towards a uniform approach to codeswitching and borrowing. ESF network on codeswitching and language contact. Papers for the workshop on constraints, conditions and models. European Science Foundation, Strasbourg. 1991. P. 259–279.

Thomason, Kaufman 1988 — Thomason S.G., Kaufman T. Language contact, creolization, and genetic linguistics. Berkley, CA: University of California Press, 1988.

Uth, Gutiérrez-Bravo 2018 — Uth M., Gutiérrez-Bravo R. La marcación de número en los préstamos españoles del maya yucateco: variación y restricciones. Cuadernos de Lingüística de El Colegio de México. 2018. Vol. 5. P. 183–224.

Wiltschko 2008 — Wiltschko M. The syntax of non-inflectional plural marking. Natural Language & Linguistic Theory. 2008. Vol. 26. P. 639–694.

Родриго Гутьеррес-Браво
профессор, Мексиканский колледж

Rodrigo Gutiérrez-Bravo
Professor, El Colegio de México

rogutierrez@colmex.mx

Мелани Ут
доцент, Кёльнский университет

Melanie Uth
Assistant Professor, Universität zu Köln

melanie.uth@uni-koeln.de

НЕОФОРМЛЕННЫЕ ПРЯМЫЕ ДОПОЛНЕНИЯ В УЗБЕКСКОМ ЯЗЫКЕ И ИХ АНАФОРИЧЕСКАЯ ДОСТУПНОСТЬ[*]

Э. Зайдель[1], З. Х. Леви-Форсайт[2]
[1]Кёльнский университет, [2]Университет им. Давида Бен-Гуриона в Негеве

В статье представлено первое экспериментальное исследование анафорического потенциала неоформленных прямых дополнений в узбекском языке. Результаты исследования суждений о приемлемости показывают, что беспадежные существительные в позиции прямого дополнения, анализируемые нами как псевдоинкорпорированные, доступны для дискурсивной анафоры. Статья преследует три цели: (i) рассмотреть растущие эмпирические разногласия в литературе относительно анафорической доступности неоформленного прямого дополнения; (ii) сформулировать анализ полученных данных и усовершенствовать анализ псевдоинкорпорации в узбекском языке; (iii) внести вклад в исследование феномена инкорпорации существительных посредством исследования тюркского языка, недостаточно изученного в рамках генеративной грамматики.

Ключевые слова: дискурсивная анафора, неоформленное прямое дополнение, псевдоинкорпорация, исследование приемлемости, узбекский язык, тюркские языки.

"Book" wants to be picked up:
Bare objects in Uzbek and their anaphoric accessibility[*]

Elyesa Seidel[1], Zarina Levy-Forsythe[2]
[1]*University of Cologne,* [2]*Ben-Gurion University of the Negev*

This paper presents the first experimental study of the anaphoric potential of bare objects in Uzbek. The results of the acceptability judgment study indicate that Uzbek bare nouns in the direct object position, analyzed as pseudo-incorporated, are accessible to discourse anaphora. The paper has three aims: (i) to address growing empirical disagreement in the literature with regards to the anaphoric accessibility of bare objects; (ii) to formulate an analysis of the data and to refine the analysis of pseudo-incorporation in Uzbek; (iii) to contribute to the research of noun incorporation by investigating a Turkic language that is understudied in the generative framework.

Keywords: discourse anaphora, bare object, pseudo-incorporation, acceptability judgment study, Uzbek, Turkic languages.

1. Introduction

Within the cross-linguistic study of (object) noun incorporation, there is growing empirical disagreement on one of the defining properties of the phenomenon: the anaphoric potential of bare objects, i.e., their (in)ability to antecede anaphoric pronouns ([Dayal 2011] for Hindi, [Farkas, de Swart 2003] and [Yanovich 2008] for Hungarian, [Espinal, McNally 2011] for Spanish and Catalan). This debate occurs in the literature on incorporation in Turkic languages. Thus, the common position that Turkish bare objects completely lack anaphoric potential [Öztürk 2005; Aydemir 2004; Ketrez 2005] has more recently been challenged [Bliss 2004; Kamali 2015; Seidel 2019, 2020a, 2020b]. Bare objects in Tatar are shown to have anaphoric potential [Lyutikova, Pereltsvaig 2015] while divergent views exist on whether Uzbek bare objects are anaphorically accessible [Levy-Forsythe 2018; Levy-Forsythe, Kagan 2020] or inaccessible [Türker 2019].

In this paper, we present empirical evidence from the first acceptability judgment study in Uzbek conducted on this topic, showing that bare objects, i.e., nouns lacking determiners, overt quantifiers, and case- and number-marking, are accessible for anaphoric uptake by covert anaphora (1).

(1) a. *Guli   kecha   uy-da   **kitob**   o'qi-di.*
         Guli   yesterday home-LOC book   read-PST.3SG
         'Guli did book-reading at home yesterday.'

     b. ***pro** yuz   betli   e-di.*
         hundred   pages   COP-PST.3SG
         'It was hundred pages long.'

We compare our results to those from an earlier study by E. Seidel [2019, 2020a, 2020b] on Turkish, placing our discussion into a broader context of the study of noun incorporation.

The paper is organized as follows. Section 2 provides relevant background on the cross-linguistic study of bare direct objects with a specific focus on Turkic languages and the research on noun incorporation. Section 3 discusses our acceptability judgment study, laying out our hypotheses and predictions, followed by the description of methods and results. Section 4 presents the discussion of our findings. Section 5 concludes the paper and suggests directions for future research.

2. Theoretical background

2.1. Differential Object Marking

Research interest in bare direct objects in Turkic languages originates from the discussion of 'differential object marking' (DOM), i.e., a cross-linguistic phenomenon of case-marking alternation, whereby some direct objects receive overt morphological marking while others do not [Bossong 1985].[1]

DOM in Turkic languages presents a rather consistent picture and primarily follows the Referentiality Scale (2) [Aissen 2003; Bossong 1985; Enç 1991; von Heusinger, Kornfilt 2005; Sinnemäki 2014; Kornfilt, von Heusinger 2017].[2] In particular, objects higher on the scale appear with the accusative suffix, while objects lower on the scale remain unmarked.

(2)   Referentiality Scale:
       Personal pronoun > Proper name > Definite NP > Indefinite specific NP > Non-specific NP [Silverstein 1976; Comrie 1989; Aissen 2002, a.o.]

In Uzbek, this renders the following contrast for singular objects (3). All definite direct objects, including personal pronouns, proper names, nouns preceded by demonstratives/ universal quantifiers, and definite nouns, obligatorily carry the accusative suffix *-ni* (3a). Objects can also contain the indefinite articles *bir* and *bitta*, glossed below as '*a*', which give rise to an indefinite reading (3b) [von Heusinger, Klein 2013]. Such objects may either carry case marking or occur unmarked. As a rule, the presence of the accusative marker indicates specificity while its absence signals the object's non-specificity. Finally, (3d) exemplifies a type of a direct object that obligatorily lacks case-marking and does not contain an indefinite article. Due to indefiniteness and generally assumed non-referentiality, bare nouns in the direct object position are considered lowest on the Referentiality Scale.

---

[1] DOM is reported in Turkic languages such as Azerbaijani [Key 2012; Kornfilt, von Heusinger 2017], Bashkir [Bossong 1998], Chuvash and Kazakh [Serdobolskaya, Toldova 2006], Kyrgyz [von Heusinger, Kornfilt 2017], Sakha [Baker, Vinokurova 2010; Podobrayev 2013], Tatar [Lyutikova, Pereltsvaig 2015; Lyutikova, Ibatullina 2015], Turkish [Erguvanlı-Taylan 1984; Knecht 1986; Enç 1991; Kelepir 2001; von Heusinger, Kornfilt 2005; Öztürk 2005], Tuvan [Muravyova 1992], Uyghur [Asarina 2011; Shklovsky, Sudo 2014], Uzbek [Serdobolskaya, Toldova 2006; Guntsetseg et al. 2008; von Heusinger, Kornfilt 2017; Levy-Forsythe 2018; Türker 2019; Kagan 2020].

[2] Research suggests that in Turkish and Uzbek, DOM is additionally triggered by the animacy, topicality, and affectedness of direct objects [Guntsetseg et al. 2008; Niyazmetova 2009; Dalrymple, Nikolaeva 2011; Laszakovits 2014; Krause, von Heusinger 2019; Kızılkaya et al. under review].

(3)   a. Definite

    Anvar   **kitob-ni**   *o'qi-di.*
    Anvar   book-ACC   read-PST.3SG
    'Anvar read the book.'

   b. Indefinite specific

    Anvar   **bir / bitta**   **kitob-ni**   *o'qi-di.*
    Anvar   a   book-ACC   read-PST.3SG
    'Anvar read a (specific) book.'

   c. Indefinite non-specific

    Anvar   **bir / bitta**   **kitob**   *o'qi-di.*
    Anvar   a   book   read-PST.3SG
    'Anvar read a (non-specific) book.'

   d. Bare noun

    Anvar   **kitob**   *o'qi-di.*
    Anvar   book   read-PST.3SG
    'Anvar read (a) book(s).' / 'Anvar did book-reading.'

In this study, we focus on bare objects illustrated in (3d).

2.2. Anaphoric accessibility of bare objects

2.1.1. Noun incorporation

For a number of world languages, including those in the Turkic group, a consistent trend exists to treat bare objects as undergoing '(object) noun incorporation', i.e., formation of a closely associated verbal unit via integration of a noun in the direct object position into its verbal predicate.

Starting with D. Massam's [2001] seminal work, the literature distinguishes between two types of noun incorporation.

True Incorporation (aka *canonical, classical, syntactic incorporation*)[3]**:** a morpho-phonological and/ or a morpho-syntactic fusion of a noun with a minimal nominal structure (N^0s) into a verb, yielding an incorporated verbal unit [Kroeber 1909; Sapir 1911; Sadock 1980; Mithun 1984; di Sciullo, Williams 1987; Baker 1988; Rosen 1989; van Geenhoven 1998]. Different versions of

---

[3] This analysis also appears in the literature under other terms as 'head-incorporation', 'syntactic head-movement', and 'head-movement' account of noun incorporation (see [Massam 2001; Scalise, Guevara 2005; Öztürk 2005; Borik, Gehrke 2015; Rosen 1989]).

true incorporation have been proposed to take place in Turkish [Mithun 1984; Knecht 1986; Kornfilt 2003; Aydemir 2004] and to a limited extend in Uzbek [Levy-Forsythe, Kagan 2020].

Pseudo-Incorporation (aka *pseudo noun incorporation, semantic incorporation*): a more distant morpho-syntactic interaction between a noun with a reduced but phrasal nominal structure (NPs or even NumPs) and its verbal predicate, characterized by persistent semantic properties of incorporation [Massam 2001; Dayal 2003, 2011, 2015; Farkas, de Swart 2003; Espinal, McNally 2011; Baker 2014; Modarresi 2014; Krifka, Modarresi 2016; Borik, Gehrke 2015]. In more recent years, research on Turkic languages suggests that this phenomenon is typical for Turkish [Öztürk 2005, 2009; Kamali 2015; Seidel 2019, 2020a, 2020b], Sakha [Baker 2014], Tatar [Lyutikova, Pereltsvaig 2015], and Uzbek [Levy-Forsythe 2018; Levy-Forsythe and Kagan 2020].

It is generally agreed that across languages, the two kinds of incorporated constructions exhibit a range of morpho-syntactic and semantic properties that do not pertain to 'regular direct object constructions', i.e., constructions that contain a fully transitive verb and a non-incorporated referential DP complement (see Table 1).

Table 1. Cross-linguistic properties of (pseudo-)incorporation identified in Uzbek
(based on [Dayal 2011, 2015; Levy-Forsythe 2018]).

Properties	Uzbek bare objects
Reduced morpho-syntax	Bare objects' functional architecture is reduced, i.e., NP and not DP.
Obligatory narrow scope	Bare objects are scopally inert and cannot receive a wide-scope interpretation.
Number-neutrality	Bare objects are morpho-syntactically singular but are compatible with both semantically singular and plural interpretations.
Name-worthiness — *Institutionalization* — *Gaps* — *Non-compositionality*	The entire *noun + verb* construction is interpreted as a conceptual whole. — interpretation of the activity as recognizable, culturally familiar, or habitual — impossibility of certain nominal-verb combinations — idiomatic interpretation of incorporated structures
Atypical discourse anaphora[4]	Bare objects' anaphoric potential contrasts with that of regular direct objects, and is characterized by both: — accessibility to covert and overt pronouns — accessibility to singular and plural pronouns

---

[4] This property has been discussed in the (pseudo-)incorporation literature under various terms, including 'discourse transparency / translucency' [Farkas, de Swart 2003; Modarresi 2014; van Geenhoven 1998], 'discourse anaphora' [Asudeh, Mikkelsen 2000; Dayal 2015], and '(non)referentiality' [Nilsson 1985; Schroeder 1999; Ketrez 2005; Öztürk 2005; Kamali 2015].

In this paper, we assume that bare objects in Uzbek are pseudo-incorporated, as argued by [Levy-Forsythe 2018 and Levy-Forsythe, Kagan 2020], and focus on a particular property of bare objects, namely their 'atypical discourse anaphora'.

2.2.2. Discourse anaphora

Given the cross-linguistically attested morpho-syntactic and/or semantic deficiency of incorporated objects, their anaphoric inaccessibility may seem an expected property, ascribed to the inability to introduce discourse referents. However, there has been an increased recognition that bare objects in incorporating languages present a more complex picture of anaphoric uptake by exhibiting various levels of accessibility to both covert and overt anaphora ([Sadock 1980; van Geenhoven 1998; Bittner 1994] for West Greenlandic, [Dayal 2011] for Hindi, [Asudeh, Mikkelsen 2000] for Danish, [Farkas, de Swart 2003; Yanovich 2008] for Hungarian, [Baker 2016] for Mapudungun, [Modarresi 2014; Krifka, Modarresi 2016] for Persian).

For instance, [Farkas, de Swart 2003] investigate the anaphoric potential of pseudo-incorporated objects in Hungarian. They argue that pseudo-incorporated nouns in Hungarian are number-neutral and 'discourse translucent' rather than 'discourse transparent', since they can only be picked up by covert but not overt pronouns (4). [Farkas, de Swart 2003] therefore conclude that the nature of the anaphoric expression affects the discourse transparency of pseudo-incorporated nouns.

(4) Hungarian [Farkas, de Swart 2003: 136]
  a. *János* **beteget** *vizsgált* *a* *rendelőben.*
      János   patient.ACC examine.PST   the   office.IN
      'János **patient**-examined in the office.'

  b. *pro* *Túl* *sulyosnak* *találta* **pro** / $^{??}$**őt** *és* *beutaltatta*
          too   severe.DAT   find.PST   he.ACC   and   intern.CAUS.PST

      pro *a* *korházba.*
            the   hospital.IN
      'He found **him** too sick and sent him to hospital.'

Based on [Farkas, de Swart 2003], F. Modarresi [2014] claims that pseudo-incorporated nouns in Persian show properties of discourse transparency since pseudo-incorporated nouns in Persian are accessible through covert and overt

pronouns. She argues that pseudo-incorporated nouns introduce number-neutral discourse referents, but depending on world-knowledge, the referent can be picked up by overt singular or plural pronouns (5).

(5) Persian [Modarresi 2014: 68, 81]

a. *mœn*   ***mobile***   *khœrid-œm.*
   I   cell.phone   bought-1SG

   Gozasht-œm-∅ / -esh / *-eshoon   *roo-ye-miz.*
   put-1SG   -it   -them   on-EZ-table
   'I bought a **cell phone**. I have put **it** / *them on the table.'

b. *Mœryœm*   ***havij***   *khœrid.*
   Maryam   carrot   bought.3SG

   Sepœs   *khoord-∅ / *-esh / -eshoon-kœrd.*
   then   cut   -it   -them-did.3SG
   'Maryam bought **carrots**. Then she cut *it / **them**.'

In contrast to [Farkas, de Swart 2003 and Modarresi 2014], V. Dayal [2011] claims that the number neutrality of pseudo-incorporated nouns in Hindi depends on the aspectual properties of the event. She argues that pseudo-incorporated nouns are not inherently number-neutral but rather that they are semantically singular. According to [Dayal 2011], the number-neutral interpretation of the pseudo-incorporated noun is a result of an interaction with aspectual operators, such as atelic predicates. Consider the examples in (6) and (7).

(6) Hindi [Dayal 2011: 159; the glosses were modified]

a. *anu-ne*   *apne*   *beTe*   *keliye*   ***laRkii***   *cun*   *lii.*
   Anu-ERG   self.GEN   son   for   girl   choose   take.PFV
   'Anu has **girl**-chosen for her son.'

b. *us-ne*   ***us-ko***   *ek*   *sone-kaa*   *cen*   *diyaa*   *hai.*
   she-ERG   her-DAT   one   gold-GEN   necklace   give.PFV   be.PRS
   'She has given **her** a gold necklace.'

(7) Hindi [Dayal 2011: 159; the glosses were modified]

a. *anu-ne*   *do*   *saal*   *tak*   *apne*   *beTe*   *ke-liye*   ***laRkii***   *dekhii.*
   Anu-ERG   two   year   till   self.GEN   son   for   girl   see.PFV
   'Anu **girl**-saw for her son for two years.'

b. *vo   hamesha   #us-se / laRkii-se ek   hii   savaal   puchtii thii.*
   she   always   her-INSTR   girl-INSTR   one   EMPH   question   ask.IPFV   be.PST
   'She always asked #**her / the girl** the same question.'

In case of a telic reading, the pseudo-incorporated noun *laRkii* 'girl' in (6a) can be referred back to by a singular pronoun, as in (6b). On the contrary, in case of an atelic reading, the pseudo-incorporated noun *laRkii* 'girl' in (7a) cannot be picked up by a singular pronoun, as shown in (7b). This is because the activity of looking at the same prospective bride repeatedly during a two-year interval conflicts with world knowledge. According to [Dayal 2011] a reading where individuals vary with sub-events of bride-looking is compatible with a definite noun phrase as a continuation for (7a).

In the literature on Turkic languages, there exists a considerable disagreement concerning the ability of bare objects to support discourse anaphora.

For instance, the example (8) reflects that the literature offers contradicting statements on whether bare objects in Turkish are accessible to overt pronouns. Thus, while some scholars have argued that bare objects' anaphoric uptake is completely ungrammatical (8a), others have claimed that it is grammatical with both singular and plural anaphora (8b) [Ketrez 2005; Bliss 2004].

(8)   Turkish
   a. *Ali   **kitap**   oku-du,   sonra   *****o-nu**   kütüphane-ye   geri   ver-di.*
      Ali   book   read-PST.3SG   then   it-ACC   library-DAT   back   give-PST.3SG
      'Ali did book-reading then (he) returned *it to the library.'
      [Ketrez 2005: 27]

   b. *Ali   **muz**   al-dı.   **On-u / On-lar-ı**   buzdolabın-a   koy-du.*
      Ali   banana   buy-PST.3SG   it-ACC   it-PL-ACC   fridge-DAT   put-PST.3SG
      'Ali bought a banana / bananas. He put it / them in the fridge.'
      [Bliss 2004: 24]

As shown in (9), bare objects in Tatar, a cognate language, are reported to antecede overt pronouns.

(9)   Tatar [Lyutikova, Pereltsvaig 2015: 308]
   a. *Sin   anarga   **kitap**   ala   ala-sıŋ.*
      you   that.DAT   book   take.IPFV   can.PRS-2SG
      'You can buy him a book.'

   b. *Häm **a-nı**   matur   it-ep   ter-ep   büläk   it-ergä   bula*
      and   that-ACC   beautifully   make-CONV   wrap-CONV   gift   make.INF   be.PRS
      'You can wrap it beautifully and give it to him as a gift.'

Next, there exists an empirical disagreement about bare objects' accessibility to covert anaphora. One line of work suggests that Turkish bare objects do not license covert anaphora (10a–b) [Ketrez 2005; Dede 1986]. Other scholars, however, argue the opposite (10c) [Erguvanlı-Taylan 1984; Schroeder 1999].

(10) Turkish

a. *Ali **kitap** oku-du, sonra ***pro** kütüphane-ye geri ver-di.*
Ali book read-PST.3SG then library-DAT back give-PST.3SG
'Ali did book-reading then (he) returned *it to the library.'
[Ketrez 2005: 27]

b. *Ali **kitap** oku-yor. ***pro** Çok ilginç.*
Ali book read-PRS very interesting
'Ali does book-reading. *It is very interesting.' [Dede 1986: 162]

c. *Ali kaçgündür **resim** yap-ıyor-du nihayet bugün **pro***
Ali how.many day picture make-PROG-PST.3SG finally today

bitir-di.
finish-PST.3SG
'Ali was picture-painting for days, finally he finished it today.'
[Erguvanlı-Taylan 1984: 23]

There are similarly divergent views on whether Uzbek bare objects are anaphorically accessible. Thus, [Türker 2019] claims that such objects lack referential interpretation and cannot serve as antecedents for pronouns. This point is illustrated in (11a) exemplifying the ungrammaticality of (covert) possessive pronominal anaphora. For contrast, direct objects marked with the accusative case are shown to be fully referential (11b).

(11) Uzbek [Türker 2019: 80]

a. *Hasan olma ye-di. #**Rang-i** yashil edi.*
Hasan apple eat-PST color-3POSS green was
'Hasan ate apple.' Int.: 'It was green.'

b. *Hasan olma-**ni** ye-di. **Rang-i** yashil edi.*
Hasan apple-ACC eat-PST color-3POSS green was
'Hasan ate the apple. It was green.'

[Levy-Forsythe 2018] and [Levy-Forsythe, Kagan 2020] argue the contrary. According to them, bare objects in Uzbek are discourse transparent and are

accessible for overt pronouns (12a), the possessive pronominal clitic (12b), and covert pronouns (12c). In addition, they show that, unlike non-incorporated objects, bare objects exhibit a unique ability to antecede both singular and plural anaphora. The latter point is evident from the fact that the first two types of anaphoric expressions (12a–b) are sensitive to number specification and carry morphological signs of number agreement.

(12) Uzbek [Levy-Forsythe, Kagan 2020: 55]
    Context:
    Anvar **rasm** *chiz-di.*
    Anvar   picture   draw-PST.3SG
    'Anvar drew (a) picture(s).'

    a. Overt pronominal anaphora
    Men **u-ni** / **ular-ni** *tezda sotdim.*
    I   it-ACC   them-ACC   quickly   sold.PST.1SG
    'I sold it / them quickly.'

    b. Possessive pronominal anaphora
    *Narx-**i*** / *narx-lar-**i*** *ancha baland.*
    price-POSS.3SG   price-PL-POSS.3SG   quite   high
    'Its price is quite high. / Their prices are quite high.'

    c. Covert anaphor
    Ra'no **pro** *sotib oldi.*
    Rano   buy.CVB   take.PST.3SG
    'Rano bought it / them.'

The cross-linguistic research suggests that apart from those discussed above, at least two more factors may influence anaphora resolution: lexical semantics of the incorporating verb and grammatical role parallelism.

It is well acknowledged that lexical semantics of verbs plays a crucial role in the phenomenon of object incorporation. For instance, [Chung, Ladusaw 2004] show for Chamorro that incorporation is limited to verbs of possession, which obligatorily incorporate their objects. [Espinal, McNally 2011] indicate that in Spanish and Catalan, pseudo-incorporation is restricted to the so-called *'have'-predicates*. [Van Geenhoven 1998] reports on two types of verbs in West Greenlandic, intensional verbs, which are inherently incorporating, and extensional verbs, which alternate between incorporating and non-incorporating instances.

[Dayal 2011] demonstrates that there are different lexical entries for incorporating and non-incorporating verbs in Hindi. [Lyutikova, Pereltsvaig 2015] for Tatar, and [Levy-Forsythe, Kagan 2020] for Uzbek, claim that the type of incorporation involved in the language depends on whether the bare object appears in a light or a regular ('heavy') verb construction. On the other hand, however, the interaction between lexical semantics of the incorporating verb and the anaphoric accessibility of the incorporated object have received little attention. [Dayal 2011] covers this point in her work on Hindi, arguing that the anaphoric accessibility of a pseudo-incorporated noun depends on the lexical semantics of the verb, namely, its aspectual properties.

The second factor — namely the influence of grammatical role parallelism on anaphora resolution, i.e., the idea that hearers tend to prefer to connect a subject anaphor to an antecedent in the subject position and an object anaphor to an antecedent in the object position [Smyth 1994; Stevenson et al. 1995; Chambers, Smyth 1998] — was not traditionally discussed in the literature on object incorporation. [Seidel 2019, 2020a, 2020b], the first to empirically examine the anaphoric accessibility of pseudo-incorporated objects in Turkish, considers the effect of grammatical role parallelism and lexical semantics of the incorporating verb in two separate studies.

In her first study, E. Seidel investigates overt pronouns in the object position, showing that pseudo-incorporated nouns prefer a singular uptake in the subsequent discourse, irrespective of the type of the anaphoric expression, cf.:

(13) Turkish [Seidel 2020b: 281]

       Context:

       *Ahmet bugün Taksim meydanın-da **hırsız** yakala-dı.*
       Ahmet today Taksim Square-LOC thief catch-PST.3SG
       'Ahmet did thief-catching at the Taksim Square yesterday.'

    a. Target sentences including overt pronouns in singular and plural
       On-u* / *On-lar-ı *rezil et-ti.*
       he-ACC he-PL-ACC embarrass do-PST.3SG
       'He embarrassed the him / them.'

    b. Target sentences including definite nouns in singular and plural
       Hırsız-ı* / *Hırsız-lar-ı *rezil et-ti.*
       thief-ACC thief-PL-ACC embarrass do-PST.3SG
       'He embarrassed the thief / the thieves.'

In her second study (see (14)), she investigates covert pronouns in the subject position and shows that (i) bare objects support anaphoric uptake with covert pronouns, and (ii) continuations with adnominal demonstratives are preferred over continuations with covert pronouns. She also examines different event types (usage events *vs.* creation events) and demonstrates that the accessibility of the pseudo-incorporated object varies depending on the event type. In particular, anaphoric reference to bare objects in creation events is more acceptable than it is to bare objects in usage events. Consequently, [Seidel 2019] argues that, in contrast to usage events, in creation events the discourse referent is "cumulatively created" out of the event, which is why the anaphoric expression refers to the result argument of the event.

(14) Turkish [Seidel 2020a: 263]

    a. Context sentence including a usage verb

      Nurten geçen gün ofis-de **mektup** **oku-du.**
      Nurten last day office-LOC letter read-PST.3SG
      'Nurten did letter-reading at the office yesterday.'

    b. Context sentence including a creation verb

      Nurten geçen gün çalışma odasın-da **mektup** **yaz-dı.**
      Nurten last day office.room-LOC letter write-PST.3SG
      'Nurten did letter-writing at the office yesterday.'

    c. Target sentences including a covert pronoun and a demonstrative NP

      pro / ***Bu mektup*** *üç sayfa-ydı.*
           this letter three pages-PST.3SG
      'It / This letter was three pages long.'

Overall, this empirical debate raises a natural question as to whether and how native speakers of Uzbek actually resolve discourse anaphora in reference to bare objects. In light of findings in Turkish, a closely related language, another question that requires examination is whether the anaphoric potential of bare objects may be influenced by the internal semantics of the event. In the next section, we present an acceptability judgment study conducted among native Uzbek speakers to investigate these issues.

3. Acceptability judgment study

3.1. Hypotheses and predictions

Based on the literature discussed above, we hypothesize that:

I.   Pseudo-incorporated or bare objects are less accessible than their non-incorporated counterparts [Farkas, de Swart 2003; Krifka, Modarresi 2016; Levy-Forsythe 2018; Seidel 2019].

II. Covert pronouns refer to highly accessible referents whereas adnominal demonstratives refer to less accessible referents [Ariel 1990]

III. Pseudo-incorporated objects in creation events are more accessible than pseudo-incorporated objects in usage events [Seidel 2019].

In accordance with the hypotheses above, we predict that covert pronouns referring to bare objects in Uzbek are less acceptable than adnominal demonstratives referring to bare objects. Additionally, according to Seidel (2019, 2020a, 2020b), we predict an effect of verb type, namely that covert pronouns referring to bare objects in creation events are more acceptable than covert pronouns referring to bare objects in usage events.

3.2. Method and procedure

To investigate the anaphoric potential of bare direct objects in Uzbek, we replicated an earlier acceptability judgment study from Turkish by [Seidel 2019]. The design of the experiment consisted of four conditions organized in a 2x2 factorial design. We manipulated the anaphoric expression (*pro vs. bu* + N 'this + N') and the verb type (usage *vs.* creation). A total of 48 critical items were created, 12 items for each context including usage verbs, cf. (15a), and creation verbs, cf. (15b). Each context sentence was followed by a target sentence, containing the anaphoric expression, cf. (16a) and (16b).

(15) a. *Gulruh o'tgan hafta libos-lar salon-i-da   uzuk   taq-di.*
       Gulruh last   week   dress.PL   salon-POSS.3SG-LOC ring   put.on-PST.3SG
       'Gulruh did **ring-putting on** in the dress salon last week.'

   b. *Kamola   o'tgan hafta   fabrikada uzuk   ishla-di.*
       Kamola   last   week   factory   ring   make-PST.3SG
       'Kamola did **ring-making** in the factory last week.'

(16) a. ***pro***   *kumush-dan*   *e-di.*
         silver-ABL   COP-PST.3SG
    '**It** was silver.'

    b. ***bu***   ***uzuk***   *kumush-dan*   *e-di.*
      this   ring   silver-ABL   COP-PST.3SG
    '**This ring** was silver.'

In addition to the critical items, we added 24 control filler sentences that were either grammatical, incongruent (i.e., sentences with agreement violation in number), or ungrammatical (i.e., sentences with ungrammatical case endings). These items were added in order to ensure that participants were attentive to the task and served as exclusion criteria. The grammatical items were constructed in such a way that the covert pronoun in the target sentence always referred to the subject of the context sentence.

All items were distributed onto four lists, each of which was balanced with respect to verb type and type of anaphoric expression. The items were presented in a pseudo-randomized order, showing only one item at a time. The results from 80 participants (mean age: 32.43 years, 38 women, 42 men, 0 of unknown gender) entered the statistical analysis.[5] Participants received a web-based questionnaire and were asked to rate sentence pairs on a scale from 1 (very bad) to 7 (very good) with regard to the naturalness of the link of the context sentence and the target sentence. Written informed consent was obtained from all participants prior to the experiment.

3.3. Data analysis and results

We fitted a linear mixed effects analysis in order to assess the relationship between type of anaphoric expression (*pro vs. bu* + N 'this' + N) and verb type (usage *vs.* creation), using the lme4 package [Bates et al. 2015] in R [R Core Team 2017] with the score as the dependent variable and the anaphoric expression and verb type as the predictors. Variability due to participant and item were taken into account by including them as random intercepts, as well as by-participant and by-item random slopes for the effects of anaphoric expression and verb type.

---

[5] Originally, we collected data from 124 participants. However, we had to exclude 44 participants from our analysis due to incorrect answers to control items. In particular, participants were excluded (i) when they rated ungrammatical sentences higher than 3, and / or (ii) when they rated grammatical sentences lower than 4.

Figure 1 presents the predicted means by the linear mixed effect model. Our data show that anaphoric uptake of bare objects in Uzbek is acceptable (mean > 4).[6] Statistical analyses reveal that contexts with overt anaphoric expressions are rated significantly better than contexts with covert anaphora ($b = 1.14$, $SE = 0.19$, $t = 6.00$, $p < 0.01$). We find no significant effect of verb type ($b = 0.12$, $SE = 0.19$, $t = 0.62$, $p > 0.05$) nor a significant interaction of verb type and anaphoric expression ($b = 0.03$, $SE = 0.24$, $t = 0.13$; $p > 0.05$).
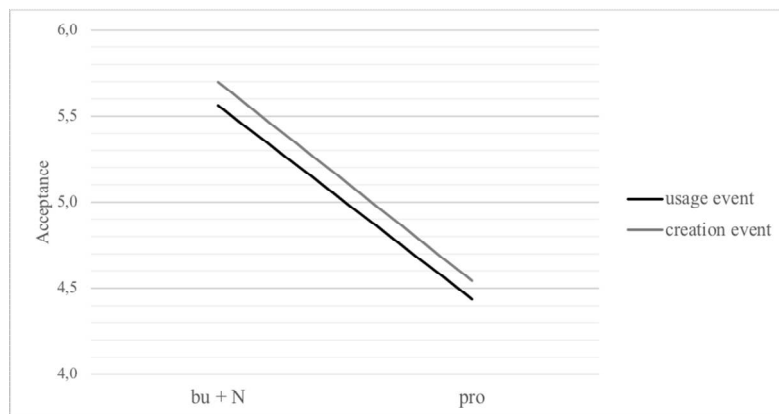
Figure 1. Predicted means showing the effect of anaphoric expression on verb type

Our findings indicate that Uzbek bare objects are visible to covert anaphora regardless of the verb type. Overall, the anaphoric accessibility of Uzbek bare objects patterns with that reported in Turkish [Bliss 2004; Kamali 2015; Seidel 2019, 2020a, 2020b]. However, we also observe a minor micro-variation (see Figure 2). In Turkish, the objects' anaphoric potential is dependent on the event type of the incorporated object, i.e., the accessibility for the incorporated object is higher in creation events than in usage events [Seidel 2019, 2020a, 2020b]. In Uzbek, the accessibility for the incorporated object is stable across both event types. As expected, contexts containing $bu + \text{N}$ ('this' + N) are more acceptable than those with covert subject pronouns.

These findings support the view that, unlike regular non-incorporated direct objects, (pseudo-)incorporated objects do not introduce referents into the discourse immediately. Rather, they achieve their anaphoric uptake indirectly, via complex event formation [Farkas, de Swart 2003; Dayal 2011; Krifka, Modarresi 2016].

---

[6] Compare the observed mean values in Figure 2 for the critical conditions and the control conditions. Sentences with continuations including the adnominal demonstrative were rated as good as the grammatical control conditions.
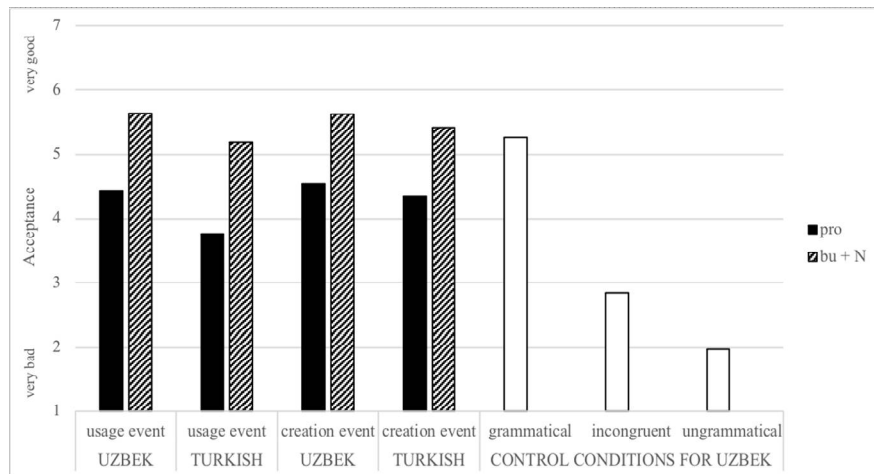
Figure 2. Observed means for the anaphoric uptake of bare objects in Uzbek in comparison to Turkish, including means for control items. Turkish data adopted from [Seidel 2019]

4. Discussion

We propose to treat complex event formation in Uzbek pseudo-incorporation in line with Krifka, Modarresi's [2016] account of pseudo-incorporation in Persian. In particular, following their analysis, formulated in terms of Discourse Representation Theory (DRT) [Kamp, Reyle 1994] in (17), we suggest that Uzbek bare objects are event-dependent definites. Crucially, however, bare objects' interpretation remains indefinite within the verbal phrase and is dependent on the Davidsonian event argument that undergoes existential closure (EC) (see (17a)). Thus, the bare object *kitob* 'book' in (17a) is interpreted as a function that identifies the unique 'book' of the reading event, introducing a discourse referent, i.e., x_2, for it. Due to its occurrence in a subordinated structure, i.e., $\exists[x_2 \mid \ldots]$, accessibility of this referent to anaphora is indirect. Anaphoric uptake of an embedded discourse referent requires an extra step, namely the Abstraction & Summation operation (17b) (see [Yanovich 2008; Krifka, Modarresi 2016]).

(17) *Guli   **kitob**   o'qidi.   **pro**   yuz   betli   e-di.*
Guli   book   read.PST.3SG   hundred   page   COP-PST.3SG
'Guli did book-reading. It was hundred pages long.'

    a. $K_0 + [Guli_1 \text{ EC } [_{vP} t_1 \textit{ kitob o'qidi}]]$
        $= [x_1 \mid x_1 = \text{GULI}, \exists[e_3 x_2 \mid x_2 = \text{BOOK-OF}(e_3), \text{READ}(x_1, x_2, e_3)]] = K_1$

    b. $K_1 + [\textit{pro yuz betli edi}]$
        $= [x_1 \mid x_1 = \text{GULI}, \exists[e_3 x_2 \mid x_2 = \text{BOOK-OF}(e_3), \text{READ }(x_1, x_2, e_3)]]$
        $\xi_4 = \Sigma x_2 [x_2 e_3 \mid x_2 = \text{BOOK-OF}(e_3), \text{READ }(x_1, x_2)], \text{100-PAGES-LONG }(\xi_4)]$

Anaphora to bare objects are analyzed as 'E-type' pronouns, namely "pronouns with quantifier antecedents that do not c-command them", which explains the restricted nature of anaphoric uptake (Krifka, Modarresi 2016: 878). The possibility of both singular and plural anaphoric uptake of bare objects is enabled via the availability of multiple reading events (17a).

The proposed analysis accounts for the atypicality of discourse anaphora in reference to bare objects in Uzbek. In particular, it explains why: (i) bare objects' anaphoric uptake is possible, yet more restricted than that of non-incorporated regular direct objects; and (ii) anaphoric reference to bare objects is characterized by number-neutrality.

5. Conclusion and future research

We conclude from our results that bare direct objects in Uzbek are anaphorically accessible. We could not confirm that their anaphoric potential is dependent on event types, suggesting that future research may take into account other event types, i.e. destruction verbs (i.e., *kitob yirtmoq* 'book-ripping'). The present investigation did not take into account the comparison between covert and overt pronouns. Therefore, we suggest conducting a forced choice task including both conditions within one experiment. Future studies should also investigate anaphoric reference to different types of full-fledged direct objects by considering different types of anaphora (covert *vs.* overt).

Abbreviations

1, 3 — 1st, 3rd person; ABL— ablative; ACC — accusative; CAUS — causative; COMPL — complementizer; CONV/CVB — converb; COP — copula; DAT — dative; EMPH — emphatic particle; ERG — ergative; EZ — ezafe; GEN — genitive; IPFV — imperfective; INF — infinitve; INSTR — instrumental; LOC — locative; PFV — perfective; PL — plural; POSS — possessive; PROG — progressive; PRS — present; PST — past; SG — singular.

References

Aissen 2003 — Aissen J. Differential object marking: Iconicity vs. economy. Natural Language and Linguistic Theory. 2003. 21(3). P. 435–483.

Asarina 2011 — Asarina A. Case in Uyghur and beyond. Ph.D. thesis. Massachusetts Institute of Technology. 2011.

Asudeh, Mikkelsen 2000 — Asudeh A., Mikkelsen L.H. Incorporation in Danish: Implications for interfaces. In Cann R., Grover C., Miller P. (eds.). A collection of papers on head-driven phrase structure grammar, Stanford: CSLI Publications, 2000. P. 1–15.

Ariel 1990 — Ariel M. Accessing noun-phrase antecedents. London, New York: Routledge. 1990.

Aydemir 2004 — Aydemir Y. Are Turkish preverbal bare nouns syntactic arguments? Linguistic Inquiry. 2004. 35(3). P. 465–474.

Baker 1988 — Baker M. Incorporation: A theory of grammatical function changing. Chicago: University of Chicago Press. 1988.

Baker 2006 — Baker M. Noun incorporation in Mapudungun. Paper presented at the Noun Incorporation and its Kind Conference. University of Ottawa, Canada. 2006.

Baker 2014 — Baker M. Pseudo noun incorporation as covert incorporation: Linearization and crosslinguistic variation. Language and Linguistics. 2014. No. 15. P. 5–46.

Baker, Vinokurova 2010 — Baker M., Vinokurova N. Two modalities of case assignment: Case in Sakha. Natural Language and Linguistic Theory. 2010. 28 (3). P. 593–642.

Bates et al. 2015 — Bates D., Maechler M., Bolker B., Walker S.. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 2015. 67(1). P. 1–48.

Beavers 2011 — Beavers J.T. On affectedness. Natural Language and Linguistic Theory. 2011. No. 29. P. 335–370.

Bittner 1994 — Bittner M. Case, scope, and binding. Dordrecht: Kluwer. 1994.

Bliss 2004 — Bliss H. The semantics of the bare noun in Turkish. Calgary Papers in Linguistics. 2004. 25(1). P. 1–65.

Borik, Gehrke 2015 — Borik O., Gehrke B. An introduction to the syntax and semantics of pseudo-incorporation. In Borik O., Gehrke B. (eds.). The syntax and semantics of pseudo-incorporation, Leiden: Brill, 2015. P. 1–43.

Bossong 1985 — Bossong G. Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen. Tübingen: Narr. 1985.

Chambers, Smyth 1998 — Chambers G.C., Smyth R.H. Structural parallelism and discourse coherence: A test of centering theory. Journal of Memory and Language. 1998. No. 39. P. 593–608.

Chung, Ladusaw 2004 — Chung S., Ladusaw W. Restriction and Saturation. Cambridge, MA: MIT Press. 2004.

Comrie 1989 — Comrie B. Language universals and linguistic typology. Syntax and morphology. Chicago: The University of Chicago Press. 1989.

Dalrymple, Nikolaeva 2011 — Dalrymple M., Nikolaeva I. Objects and information structure. Cambridge: Cambridge University Press. 2011.

Dayal 2003 — Dayal V. A semantics for pseudo incorporation. Ms., Rutgers University. 2003.

Dayal 2011 — Dayal V. Hindi pseudo-incorporation. Natural Language and Linguistic Theory. 2011. 29(1). P. 123–167.

Dayal 2015 — Dayal V. Incorporation: Morpho-syntaxtic vs. semantic considerations. In Borik O., Gehrke B. (eds.). The syntax and semantics of pseudo-incorporation, Leiden: Brill, 2015. P. 47–87.

Dede 1981 — Dede M. Grammatical relations and surface cases in Turkish. In Proceedings of the Berkeley Linguistic Society. 1981. Vol. 7. P. 40–49.

di Sciullo, Williams 1987 — di Sciullo A.M., Williams E. On the definition of word. Cambridge, MA: MIT Press. 1987.

Enç 1991 — Enç M. The semantics of specificity. Linguistic Inquiry. 1991. No. 22. P. 1–25.

Erguvanlı-Taylan 1984 — Erguvanlı-Taylan E. The function of word order in Turkish grammar. Berkeley, CA: University of California Press. 1984.

Espinal, McNally 2011 — Espinal M.T., McNally L. Bare nominals and incorporating verbs in Spanish and Catalan. Journal of Linguistics 2011. No. 47. P. 87–128.

Farkas, de Swart 2003 — Farkas D.F., de Swart H. The semantics of incorporation: From argument structure to discourse transparency. Stanford, CA: CSLI Publications. 2003.

Guntsetseg et al. 2008 — Guntsetseg D., von Heusinger K., Klein U., Niyazmetova D. Differential object marking in Mongolian and Uzbek. SFB 732 presentation. Stuttgart. 2008.

von Heusinger, Klein 2013 — von Heusinger K., Klein U. The distribution of two indefinite articles. In Ebert C., Hinterwimmer S. (eds.). Different kinds of specificity across languages 92. Dordrecht: Springer, 2013. P. 155–176.

von Heusinger, Kornfilt 2005 — von Heusinger K., Kornfilt J. The case of the direct object in Turkish: Semantics, syntax and morphology. Turkic Languages. 2005. No. 9. P. 3–44.

von Heusinger, Kornfilt 2017 — von Heusinger K., Kornfilt J. Partitivity and case marking in Turkish and related languages. Glossa: A Journal of General Linguistics. 2017. 2(1): 20. P. 1–40.

Kagan 2020 — Kagan O. The semantics of case (key topics in semantics and pragmatics). Cambridge University Press. 2020.

Kamali 2015 — Kamali B. Caseless direct objects in Turkish revisited. In Meinunger A. (ed.). Byproducts and side effects: Nebenprodukte und Nebeneffekte. ZAS Papers in Linguistics. 2015. No. 58, Berlin: ZAS. P. 107–123.

Kelepir 2001 — Kelepir M. Topics in Turkish syntax: Clausal structure and scope. Ph.D. thesis. Massachusetts Institute of Technology. 2001.

Key 2012 — Key G. Differential object marking in Turkic and Persian as a contact phenomenon. Proceedings of the Annual Meeting of the Berkeley Linguistics Society. 2012. No. 38, P. 239–254.

Ketrez 2005 — Ketrez F.N. Children's scope of indefinite objects. Ph.D. thesis. University of Southern California. 2005.

Knecht 1986 — Knecht E.L. Subject and object in Turkish. Ph.D. thesis. Massachusetts Institute of Technology. 1986.

Krause, von Heusinger 2019 — Krause E., von Heusinger K. Gradient effects of animacy on differential object marking in Turkish. Open Linguistics. 2019. 5(1). P. 171–190.

Krifka, Modarresi 2016 — Krifka M., Modarresi F. Number neutrality and anaphoric update of pseudo-incorporated nominals in Persian (and weak definites in English). In Moroney M., Little C.-R., Collard J., Burgdorf D. (eds.). Proceedings of the 26th Semantics and Linguistic Theory Conference (SALT 26), Austin, TX: University of Texas, 2016. P. 874–891.

Kroeber 1909 — Kroeber A. Noun incorporation in American languages. XVI Internationaler Amerikanisten-Kongress, 1909. P. 569–576.

Laszakovits 2014 — Laszakovits S. ACC as topic marker in Turkish. Wiener Linguistische Gazette 78A, 2014. P. 166–177.

Levy-Forsythe 2018 — Levy-Forsythe Z. Object incorporation in Uzbek. M.A. thesis. Ben-Gurion University of the Negev. 2018.

Levy-Forsythe, Kagan 2020 — Levy-Forsythe Z., Kagan O. Two types of object incorporation in Uzbek. In Danon G. (ed.). Proceedings of the 34th annual conference of the Israel Association for Theoretical Linguistics (IATL 34). Cambridge, MA: Working Papers in Linguistics. 2020. P. 45–66.

Lyutikova, Ibatullina 2015 — Lyutikova E., Ibatullina D. Case theory and case variation in Tatar. In Lyutikova E.A., Zimmerling A.V., Konoshenko M.B. (eds.). Typology of morphosyntactic parameters. Proceedings of the international conference "Typology of morphosyntactic parameters 2015". Iss. 2. Moscow: Moscow State University of Education, 2015. P. 228–244.

Lyutikova, Pereltsvaig 2015 — Lyutikova E., Pereltsvaig A. The Tatar DP. The Canadian Journal of Linguistics / La revue canadienne de linguistique. 2015. 60 (3). P. 289–325.

Massam 2001 — Massam D. Pseudo noun incorporation in Niuean. Natural Language and Linguistic Theory. 2001. No. 19. P. 153–197.

Mithun 1984 — Mithun M. The evolution of noun incorporation. Language. 1984. Vol. 60. P. 847–895.

Modarresi 2014 — Modarresi F. Bare nouns in Persian: Interpretation, grammar and prosody. Ph.D. thesis. Humboldt-Universität zu Berlin. 2014.

Muravyova 1992 — Muravyova I. Unmarked noun form in Turkic languages: A typological point of view. In Proceedings of the 33rd Meeting of the Permanent International Altaistic Conference, Budapest, 1992. P. 257–261.

Nilsson 1985 — Nilsson B. Case marking semantics in Turkish. Ph.D. thesis. University of Stockholm. 1985.

Niyazmetova 2009 — Niyazmetova D. Die differentielle Objektmarkierung im Usbekischen. M.A. thesis. Universität Stuttgart. 2009.

Öztürk 2005 — Öztürk B. Case, referentiality and phrase structure. Linguistik Aktuell/Linguistics Today 77. Amsterdam/Philadelphia: John Benjamins. 2005.

Podobryaev 2013 — Podobryaev A. Differential case marking in Turkic as intermediate dependent case. In Özge U. (ed.). Proceedings of the 8th Workshop on Altaic Formal Linguistics (WAFL8). MIT Working Papers in Linguistics 67, Cambridge, MA: MITWPL. 2013. P. 281–292.

R Core Team 2017 — R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org. 2017.

Rosen 1989 — Rosen S. Two types of noun incorporation: a lexical analysis. Language. 1989. Vol. 65. No. 2. P. 294–317.

Sadock 1980 — Sadock J. Noun incorporation in Greenlandic. Language. 1980. Vol. 56. P. 300–319.

Sapir 1911 — Sapir E. The problem of noun incorporation in American languages. American Anthropologist 13. 1911. P. 250–282.

Scalise, Guevara 2005 — Scalise S., Guevara E. The lexicalist approach to word-formation and the notion of the lexicon. In Štekauer P., Lieber R. (eds.). Handbook of Word-Formation. Studies in Natural Language & Linguistic Theory 64. Dordrecht: Springer, 2005. P. 147–187.

Schroeder 1999 — Schroeder C. The Turkish nominal phrase in spoken discourse. Wiesbaden: Harrassowitz. 1999.

Seidel 2019 — Seidel E. The Discourse Properties of Bare Nouns in Turkish. Ph.D. thesis. University of Cologne. 2019.

Seidel 2020a — Seidel E. Anaphoric potential of pseudo-incorporated nouns in Turkish. In Sofu H., Can C., Abik A.D., Yüceol M.Ö., Kilimci A. (eds.). Turkish Linguistics Across Boundaries: The Adana Meeting. Wiesbaden: Harrassowitz. 2020. P. 259–266.

Seidel 2020b — Seidel E. Bare direct objects in Turkish. Pseudo-Incorporated or Weak Arguments. In Bondarenko T., Davis C., Colley J., Privoznov D. (eds.). Proceedings of the 14th Workshop on Altaic Formal Linguistics (WAFL14). MIT Working Papers in Linguistics 90, Cambridge, MA: MITWPL. 2020. P. 277–287.

Serdobolskaya, Toldova 2006 — Serdobolskaya N., Toldova S. Direct object marking in Finno-Ugric and Turkic languages: Verb semantics and definiteness of direct object NP. Paper presented at the Workshop on Presupposition Accommodation, 13–15 October 2006, Ohio State University, Columbus, Ohio. 2006.

Shklovsky, Sudo 2014 — Shklovsky K., Sudo Y. The syntax of monsters. Linguistic Inquiry. 2014. 45(3). P. 381–402.

Silverstein 1976 — Silverstein M. Hierarchy of features and ergativity. In Dixon R.M.W. (ed.). Grammatical Categories in Australian Languages. Canberra: Australian National University. 1976. P. 112–171.

Sinnemäki 2014 — Sinnemäki K. A typological perspective on differential object marking. Linguistics. 2014. 52(2). P. 281–313.

Smyth 1994 — Smyth R.H. Grammatical determinants of ambiguous pronoun resolution. Journal of Psycholinguistic Research. 1994. No. 23. P. 197–229.

Stevenson et al. 1995 — Stevenson R.J., Nelson A.W.R., Stenning K. The role of parallelism in strategies of pronoun comprehension. Language and Speech. 1995. 38(4). P. 393–418.

Türker 2019 — Türker L. Noun phrases in article-less languages: Uzbek and beyond, Linguistik Aktuell/Linguistics Today 253. Amsterdam/Philadelphia: John Benjamins. 2019.

van Geenhoven 1998 — van Geenhoven V. Semantic incorporation and indefinite descriptions: Semantic and syntactic aspects of noun incorporation in West Greenlandic. Stanford, CA: CSLI Publications. 1998.

Yanovich 2008 — Yanovich I. Incorporated nominals as antecedents of anaphora, or how to save the thematic arguments theory. University of Pennsylvania Working Papers in Linguistics. 2008. No. 14. P. 367–379.

Эльеса Зайдель

Ph.D.; Кёльнский университет

Elyesa Seidel

Ph.D.; University of Cologne

elyesa.seidel@uni-koeln.de

Зарина Хамидовна Леви-Форсайт

магистр; Университет им. Давида Бен-Гуриона в Негеве

Zarina H. Levy-Forsythe

Master of Arts; Ben-Gurion University of the Negev

zarinale@post.bgu.ac.il

К ТИПОЛОГИИ МАРКЕРОВ КОСВЕННОГО ВОПРОСА[*]

В. А. Морозова[1]*, Н. В. Сердобольская*[2]
[1]*Национальный исследовательский университет «Высшая школа экономики»,*
[2]*Институт языкознания РАН,*
[2]*Государственный институт русского языка им. А. С. Пушкина*

　　Работа представляет типологическое исследование конструкций с косвенным вопросом в языках мира. Проанализировано 30 языков и выявлено 89 маркеров косвенного вопроса. Маркеры классифицируются по следующим группам: нулевая стратегия (без специального маркера); цитативы; подчинительные союзы и союзные слова, включая комплементайзеры, релятивизаторы и обстоятельственные средства подчинения (показатели условия и уступки); вопросительные частицы, разделительные союзы и частицы; двухпредикатные комплексы; показатели косвенного наклонения и средства маркирования фокуса. Полученные данные позволяют сформулировать предварительные типологические обобщения и представить список типологических параметров, релевантных для описания косвенного вопроса.

　　Ключевые слова: типология, сентенциальные актанты, вопрос, косвенный вопрос, косвенная речь, полипредикация, вопросительное предложение.

TOWARDS THE TYPOLOGY OF INDIRECT QUESTION MARKERS[*]

Valeriya Morozova[1], Natalia Serdobolskaya[2]
[1]National Research University Higher School of Economics,
[2]Institute of Linguistics RAS, [2]Pushkin State Russian Language Institute

The paper is aimed at providing a typological survey of indirect question (IQ) markers in languages of the world. The sample of 30 languages has been examined, and 89 IQ markers have been identified and classified into the following groups: null strategy (asyndetic construction); quotative markers; subordinators, including complementizers, relativizers and adverbial clause markers (conditionals and concessives); question particles; disjunctive particles and conjunctions; two-predicate complexes; oblique mood markers and focus marking devices. Based on this data, we formulate preliminary typological implications and provide a list of relevant typological parameters.

Keywords: typology, argument clauses, complementation, complement clauses, question, interrogative clause, indirect question, embedded question, subordination.

1. Introduction

Indirect questions (IQs) have been defined as complex sentences containing a dependent clause with interrogative semantics [Testelets 2001: 246–247]. [Kroeger 2005: 239] singles out complement clauses (with a complementizer) "typically referring to a proposition that would answer the embedded question":

(1)   *I don't know where he lives.*

Thus, in (1) the complement clause has interrogative semantics, and refers to a proposition 'He lives in New York' which is a felicitous answer to the embedded question.

These definitions are based on the semantic notions of proposition, interrogative semantics, and argument clause, thus they can be used as a comparative concept in terms of [Haspelmath 2010].

Despite the growing interest in the typology of complex sentences, IQs have received little attention in typological literature. Although they are traditionally described as a type of complement clauses, they are usually left out of research scope or very briefly described both in studies of complementation and studies of direct speech in languages of the world (see, for example, [Noonan 1985; Dixon, Aikhenvald 2006]). Therefore, theoretical works on IQs are constrained to the material of well-studied languages.

A typological study of IQs has been proposed in [Kahrel 1985] in the vein of cross-linguistic word order studies [Dik 1978, 1980]. Based on the sample of 30 languages, [Kahrel 1985] analyses the relative order of clause + linker (a subordinator or another device introducing indirect questions), on the one hand, and the order of the main and the dependent clause, on the other hand. He shows that the two parameters are correlated in the following way: the preferred order is the one where the linker is located between the main and the dependent clause. In case of violation of this rule the language develops a new "secondary" linker that occupies the preferred position. The exceptions are not numerous, namely, three constructions among 54 (25 polar IQs and 29 IQs).

Unlike [Kahrel 1985] the present study is aimed at identifying the inventory of morphosyntactic devices used to encode IQs in languages of the world. [Kahrel 1985] mentions the following strategies: quotatives, case affixes, question words, conditionals, dubitative particles and relativizers. We are going to complete this list and try to measure the frequency of each strategy in our lan-

guage sample. As well as [Kahrel 1985], we take the sample of 30 languages. However, our sample is different, with one intersection only (Basque), and the number of IQ markers is larger (89 database entries).

Another goals are to identify the strategies specific for polar and constituent IQs and to identify the "preferences" of IQ markers for polar/constituent question types.

The literature on root questions proposes a number of relevant typological parameters, which we are going to check for IQs. Thus, this pilot study will allow to figure out the set of parameters relevant for the analysis of IQs, to adjust the direction of the further research and identify the potential problems. Indeed, the sample size is not very large, and all the conclusions and implications we formulate are preliminary. However, they can be used as a start for an in-depth typological research.

2. The database

Our study is based on the language diversity value sampling method [Miestamo et al. 2016: 244] aimed at identifying the maximum set of different types of the phenomenon of interest. Our goal is to reveal the limits of variation of the discussed phenomenon in languages of the world, taking into account the genetic and areal diversity of languages. Therefore, we included one or two languages from 20 language families (in case of large families two languages from different sub-families were included) and two isolates. The choice of languages is partly based on the size and quality of relevant parts of grammatical descriptions (the so-called bibliographic bias, see [Bakker 2010]).

We use reference grammars as a source of the data (95 grammars have been consulted). For Chuvash, we rely upon the field trip notes taken by Anastasiia Egorova (partly presented in [Egorova 2020]). For English and Russian we have made a corpora search, and for some languages we have consulted small collections of texts in reference grammars.

Polar and constituent (wh) questions are described in detail in many grammars. As for alternative IQs, this information is most often scarce, and it is not possible to draw any solid generalizations. Thus, we only describe here some interesting features of alternative questions.

The analysis is presented in a database format in Google Sheets. If a language has more than one way to express indirect questions, they are all documented in a database. Therefore, we have 89 IQ markers for 30 languages.

Based on typological works on independent questions, we have elaborated a list of parameters relevant for the description of IQs in a given language:

— types of questions (polar, wh, alternative)
— synchronous polysemy of IQ markers (adverbial clauses, citations etc.)
— the make-up of the IQ construction with the marker under examination
— the use of IQ markers in root questions
— types of questions: focused *vs.* non-focused
— word order (comparing to word order in independent clauses and in embedded clauses)
— semantic type of the embedding predicate (speech *vs.* mental predicates etc.).

In the course of the study, the first four parameters turned out to be the most relevant, and we concentrated our attention upon them. The focused *vs.* non-focused distinction, word order and semantic type of the embedding predicate proved to be complicated features that require an in-depth study. Thus, in what follows they are not going to be considered.

3. Cross-linguistic types of IQ markers

3.1. IQ markers and constructions

It is important to distinguish between markers used in IQ constructions and the constructions themselves. For example, in Basque polar IQs can be formed with a special non-finite suffix *-en* plus an emphatic prefix *ba-* and a question particle *al*. All the three make an IQ construction.

(2)   Basque (Isolate) [Hualde, Ortiz de Urbina 2011: 483]
errege-ak [*zerbait   egiten   **ba   al**   zeki**en**]   *galdetu   zion*
king-ERG   something   do.IMPF   FOC   Q   knew.REL   ask   AUX
'The king asked him whether he knew [how] to do anything.'

However, an IQ can be formed without the particle *al* and without the prefix:

(3)   Basque (Isolate) [Hualde, Ortiz de Urbina 2011: 482]
*ez   dakit   [zure   laguna   etorri-ko   d-**en**].*
NEG   know   your   friend   come-FUT   AUX-REL
'I don't know whether your friend will come.'

IQs can also include the particle *ea* meaning 'let us see if, I wonder' [King 1994: 400]:

(4)   Basque (Isolate) [de Rijk 2008: 448]
  [***ea*** *diru-rik*   *falta*   *ote*   ***zen***]   *galdetu*   *zion*
  Q   money.PRTT   miss   maybe   AUX.REL   ask   AUX

  etxekoandre-ari   *detektibe-ak*
  house.owner-DAT   detective-ERG
  'The detective asked the lady of the house if any money happened to be missing.'

The particle *ea* and the prefix *ba* can cooccur in one and the same clause to form an IQ:

(5)   Basque (Isolate) [Hualde, Ortiz de Urbina 2011: 483]
  Bidrios-ek   [***ea*** ***ba***-*zu*-***en***   *Institutu-ko*   *berry-rik*]   *galdetu*   *zion*
  Bidrios-ERG   Q   FOC-AUX-REL   school-ATTR   news-PRTT   ask   AUX
  'Bidrios asked him whether he had any news from the school.'

Thus, there are four morphosyntactic means of making IQs (the prefix *ba-*, the particles *-al* and *ea*, the non-finite *-en*), and there are four constructions, 1) *ba al* V-*en*, 2) V-*en*, 3) *ea* -*en*, 4) *ea ba*-V-*en* (omitting two other dialectal and substandard constructions considered in [Hualde, Ortiz de Urbina 2011] and [de Rijk 2008]). However, the markers are distributed among constructions in a peculiar way. There are markers that co-occur to make an IQ construction, and there are markers that are used in more than one construction. There seems to be only one obligatory marker, i.e. the non-finite suffix *-en* (which may also be absent in some substandard variants).

Our paper aims to present a typology of IQ markers, which means that the number of available constructions for each language would differ.

The inventory of IQ markers includes the following:

A. Null strategy, i.e. no specific marking, or asyndetic subordinates.

B. Quotatives.

C. Subordinators, including complementizers, relativizers and adverbial clause markers (conditionals and concessives).

D. Question particles.

E. Disjunctive particles and conjunctions.

F. Two-predicate complexes (e.g. lit. "I don't know he comes (or) not" / "…he comes (or) doesn't come" / "…he comes (or) what-he-does" / "…he will come (or) won't").

G. Oblique mood markers.

H. Focus marking devices (particles/constructions).

3.2. Null strategy (19 cases)

The "null" strategy does not involve any specific markers of IQ. The interrogative semantics is defined contextually or intonationally, as in Russian alternative IQs (*YA ne znayu, svoboden on ili zanyat* 'I don't know if he is free or busy') and in the following example from Ossetic:

(6)　Ossetic (Iranian > Indo-European) [Ossetic National Corpus]

nəwwaj	*=šзm*	*зmз*	*Baraštər-ə*	*a-fзrš*	*bənat*
ride[IMP.2SG]	they.ALL	and	Barastyr-GEN	PV-ask[IMP.2SG]	place

=ma	*=jзm*	*iš*
more	he.ALL	EXST

'Ride to them and ask Barastyr[1] whether he has a place [for you].'

The root question corresponding to the IQ in (6) would be (7). Note that (6) involves the indexical shift of the pronoun: the speaker uses a 3rd person pronoun to refer to the main clause dative argument, while in direct speech he would use the 2nd person pronoun, as in the following:

(7)　Ossetic

bynat	*ma*	*=dзm*	*iš?*
place	more	you.ALL	EXST

'Do you have a place [for me]?'

Ossetic does not use any special marking to encode neither root polar questions, nor IQs, with the reservation that we do not take prosody into account. Thus, it is an example of a "true" null strategy that indeed does not have any specific markers indicating that it is an IQ. As far as it concerns *wh-* and alternative questions, they include *wh-*words and disjunctive conjunctions (as in the Russian example above) that mark the examined sentence as having interrogative semantics. We have only one example of the true null strategy in our sample (Ossetic), all other examples including alternative or *wh*-IQs (Hebrew, Tamashek, Russian, Udihe and others).

---

[1] The deity of the underworld in Ossetian mythology.

In the "true null" type the interrogative semantics of a complement clause is not explicitly expressed, and the embeddedness of the interrogative clause is only manifested by the indexical shift of pronouns and other deictics (like the 3[d] person pronoun in the Ossetic sentence). Whenever it is possible, we use the indexical shift as a criterion for distinguishing between direct and indirect speech, see 3.4 for details.

In some languages, the dependent status of the interrogative clause is signaled by a specific word order, e.g.:

(8)   *Where is he going? I don't know where he is going.*

In English, IQs differ from root questions by having *wh*-word fronting, and they differ from root questions by the absence of subject-verb inversion. This leads to a question whether such structures should be termed as null strategy given that the word order is different. At present, we are unable to answer this question, because for most of the languages of our sample we only have scarce information about the basic word order in root and IQs.

Obviously, the null strategy occurs much rarer in polar question than in *wh*-questions. The tentative typological implication would then be that if polar IQs are marked with the null strategy then it is also possible for *wh*-questions. A more solid implication may be formulated on the basis of the 17 languages that have null strategy in *wh*-questions:

(i)   if IQs are marked with the null strategy then it is also possible for root questions of the same type (polar, *wh*, alternative).

Thus, the null strategy "inherits" semantic restrictions on null root questions, even if involves changes in prosody pattern and morphosyntax (including indexical shift and word order).

3.3. Question particles (16)

The question particles are widely used both in root in IQs. Most often, root and IQs are introduced by the same question particles, as *yɛ* in Supyire:

(9)   Supyire (Atlantic-Congo > Niger-Congo) [Carlson 2011: 304, 456]
      a. *Dì faŋké   màha n-tuga   à   jwu yɛ?*
         how grave.DEF HAB INTR-dig SC   say Q
      'How is the grave dug?'

b. *U   a   cè   jò   u   sí   m̀-pà   yɛ.*
he   PERF   know   who   s/he   FUT   FP-come   Q
'He knows who will come.'

However, some languages possess specialized IQ particles, as Chuvash *şi* and Basque *ea* (used in both polar and *wh*-questions). Such particles are characterized by specific complicated semantics, which makes their use restricted even in IQs. An interesting case is observed in Russian, where the IQ particle *li* is possible both in root and IQs; however, in root questions its use is restricted to specific contexts of doubt.

Question particles are most often restricted to either polar or *wh*-questions. However, Chuvash, Jamsay and Basque have particles that occur unspecified in both of these question types.

The data permits us to formulate the following generalization:

(ii)   if a question particle is used both in root and IQs, it usually covers the same questions types.

3.4. Quotatives (2)

In some languages quotatives introducing direct speech are used to introduce IQs:

(10) Mian (Ok > Trans-New Guinea) [Fedden 2011: 449]

[*kwĕit=e   hei-b-ne   lowon-nab-e=a=ba?*]
sugar_cane=SG.N1   cut.PFV-give.PFV-1SG.IO.PFV   eat.PFV-NRPST-3SG.M-S=Q=QUOT

ge   baa-n-o=ta
say.PFV   say.PFV-SEQ-3SG.F.S=MED

'"Did he cut and eat my sugar cane a short while ago?" she asked and then…'

As well as [Kahrel 1985] we included quotatives in our sample. However, as we aim at limiting our research to indirect questions only, we have tried to use the indexical shift (*He said: "I'm broke" — He said **he was** broke*) as a criterion for distinguishing direct and indirect speech. Direct speech is not taken into account[2]. However, it seems that for some languages this approach is not justified, as they seem to lack indirect speech constructions at all. For example, in

---

[2] It has been largely claimed that direct speech is distinguished from indirect speech both structurally and semantically, see [Partee 1973; Munro 1982] and the overview in [Aikhenvald 2011: 309].

Hup the indirect speech is a simple sentence marked by a special reportative mood marker (reportive evidential). By contrast, direct quotes are complex sentences (see [Epps 2005: 685–687]). K. J. Olawsky states that "Urarina has no formal strategy to create indirect questions. In constructions that involve two clauses where one is a question, the two are simply juxtaposed. Thus, the question is not actually "indirect" or dependent from another clause… However, the position before the main clause is preferred, as with dependent clauses except subordinate clause with purposive function." [Olawsky 2006: 836].

[Noonan 1985: 72] reports that Agta, Punjabi and Shina lack or hardly ever make use of indirect quote constructions. Presumably, such cases must be included in our sample, with the reservation that direct quote structures are included only for languages lacking the indexical shift. However, even this decision is problematic for some languages, since the indexical shift can be observed in case of coreference of the arguments of the main clause and the IQ clause, and the sources often lack such examples.

A solution to this problem would be an in-depth work with the experts. At present, we suggest that the following parameters must be included into the database: 1) the presence of indexical shift in the given language; and 2) the presence of indexical shift in the context of the discussed marker. Possibly, some conclusions must only be based on languages that do have indexical shift (the generalizations about markers used both in root and IQs, for example, the null strategy and question particles), while others can be based on the whole sample.

3.5. Subordinators (28)

Among 28 subordinators used in IQ, 11 cases include adverbial clause markers (9 conditionals and 2 concessive clause markers), 12 are complementizers and 5 relativizers.

The conditionals are rather frequently used to encode polar IQs (e.g. *He asked if I was frightened*); in some languages they can encode both polar and *wh*-IQs (Lezgian). Adyghe and English (*I don't know **whether** I am right or I am wrong / You will pay **whether** you want it or not*) also use concessive clause markers to encode IQs:

(11) Adyghe (Northwest Caucasian)

    a. *se*   *s-ṣ̂e-r-ep*   *we*   *wə-qə-z-ʁe-š'ta-ʁe-**m-jə***
        I   1SG-know-DYN-NEG   you   2SG-DIR-1SG-CAUS-fear-PST-COND-COH

        'I don't know, whether I frightened you.' [Gerassimov, Lander 2008: 9]

b. *marat*   *tje.ḳ^wa-ʁe-**m**-jə,*   *se*   *sə-g^w*
   Marat   win-PST-COND-ADD   I   1SG.PR-heart

   r-jə-hə-r-ep   *a-r*   *zer-je-ṣ̂a-ʁe-r*
   LOC-3SG.A-carry-DYN-NEG   DEM-ABS   MNR-OBL-do-PST-ABS

   'Although Marat won, I didn't like the way he played.' [Arkadiev et al. 2009: 101]

So far, we can say that languages using concession clause markers to encode IQs, also use conditionals in the same function. However, this implication has to be verified on a larger language sample.

Complementizers may well be used to mark IQs, although they are not as frequent as it could have been expected (less than 15%), given that IQs are traditionally considered to be a subtype of complement clauses. Out of these 12 markers, 5 involve nominalizers. There are IQ markers that are better represented in our sample (e.g. question particles).

Complementizers are unlikely to introduce polar IQs only (only one example from Urarina, Peru). They either take both polar and *wh*-questions, or only *wh*-questions.

One language has a specific complementizer for polar IQs, apparently, not used anywhere else, *yélà* in Humburi Senni (Nilo-Saharan)[3].

Relativizers are used to encode both *wh*- and polar questions. In Humburi Senni (Songhay > Nilo-Saharan) the questioned constituent is replaced with indefinite noun phrase *bòr* 'person' with a relative morpheme, so according to [Heath 2014: section 8.2.5] literally the example is translated as 'I don't know the person who came' (see also 3.9 below):

(12) Humburi Senni (Songhay > Nilo-Saharan) [Heath 2014: 340]
     ì   *sù*   *béy* [**bòr**↓   **ká** *kà*]
     1SG.S   IMPF.NEG   know [person   REL   come]
     'I don't know who came.'

3.6. Disjunctives (2)
Two languages of the sample use disjunctive conjunctions to encode polar IQs:

---

[3] This marker can also encode factive complements of the verb 'know'; however, it is not listed among the basic complementizers in the grammar and seems to be marginal in such contexts.

(13) Supyire (Atlantic-Congo > Niger-Congo) [Carlson 2011: 454]

Kà	*uru*	*nàŋi*	<...>	*mìì*	*yígé*	*na*	*uru*	*ù*
and	this(EMPH)	man.DEF		me	ask	that	he(EMPH)	he

∅	*sá*	*mìì*	*yàha*	*mobílíge*	*e*	**làa.**
SBJV	go	me	leave	truck.DEF	in	or

'Then that man <...> asked me if he should go take (lit. 'leave') me in the truck.'

Note that we took into account the cases where there is only one alternative expressed (in case two alternatives are explicitly stated, the question should be interpreted as an alternative question taking the disjunctive).

These constructions are only used to mark polar questions.

3.7. Two-predicate complexes (14)

[König, Siemund 2007] note that Mandarin Chinese and some Papuan languages (Amele, Kobon) make use of a specific 'disjunctive-negative construction' to mark root polar questions, lit. "he came or didn't come". It can also be used in IQs, with or without the disjunctive. For example, in Chinese it is the main strategy of forming polar questions, root or IQs [Liing 2014]:

(14) Mandarin Chinese [Liing 2014: 11]

a.
nǐ	*xǐhuān*	*bù*	*xǐhuān*	*lánsè?*
you	like	not	like	blue.color

'Do you like blue?'

b.
yuēhàn	*wàngjì*	[*mǎlì*	*huì*	*bú*	*huì*	*lái*]
John	forgot	Mary	will	not	will	come

'John forgot whether Mary will come.'

There are languages that form similar constructions with a question particle and/or with a subordinator (Chuvash, Adyghe, Tundra Nenets). It can be used on a par with a simple one-predicate structure:

(15) Chuvash [Egorova 2020]

maʂə	**kil-ed=i**	**kil-mest=i**	*te-ʑe*
Masha	come-NPST[3SG]=Q	come-NEG.NPST[3SG]=Q	QUOT-CVB_SIM

vaɕə	*it-r^j-ë*
Vasya	asked-PST-3SG

'Vasya asked, whether Masha would come.' (*kilmesti* may be omitted)

In some languages such sentences are stylistically marked and have a clear 'colloquial' meaning, as in Russian *Ya ne znayu, pridet on, ne pridet...* 'I don't know whether he would come' (see English *I don't know whether he will like it or not* and also *I don't know whether or not he will like it*). The stylistic restrictions are not clear for languages that do not have a long tradition of standardization. At least, it can be claimed that the speakers are prompt to give such examples during their interview with the linguist when translating polar questions (without the second alternative "or not" explicitly stated in the stimulus), which means that they are not as marked as the Russian example.

The two predicates may take subordinators of various kind, including conditionals, concessives, complementizers and nominalizations. In some languages they show a lot of variation:

(16) Udihe [Nikolaeva, Tolskaya 2011: 50]

Nua-ni	*xauntasi-e-ni*	*Sonia*	*skola-du*	**bi:-we-n-de**
he-3SG	ask-PST-3SG	Sonia	school-DAT	be.PTCP.PRS-ACC-3SG-DIS

e-i-we-n-de
NEG-PTCP.PRS-ACC-3SG-DIS

'He asked whether Sonia was at school or not.'

For example, in Udihe the two predicates occur with and without negation, they can include the auxiliary or the proverb (placeholder verb). Thus, it is unclear whether we deal with one and the same 'loose' construction or with a number of constructions with different morphosyntactic properties and (possibly) semantic restrictions. Obviously, the answers to these questions require the extension of the language database.

Another problem is the differentiation of polar and alternative questions within this type. For example, L. Berghäll states about Mauwake that "polar questions, when indirect, have to be alternative questions." [Berghäll 2015: 365]. For Udihe, Nikolaeva and Tolskaya [2011] describe (16) as an example of an alternative question. However, the second alternative is actually the negation of the first, which makes the discussed example semantically close to polar questions ("…whether Sonia was at school"), unlike alternative questions of the type "whether Sonia was at school or at home". Such ambiguous exam-

ples are given in many sources, which makes it difficult to distinguish between polar and alternative questions in case of two-predicate complexes[4]. The adequate description of this type requires further research.

Our data permits us to formulate the following preliminary implication:

(iii) If a two-predicate complex is possible in root questions, then it is also possible in IQs (while the opposite is not true).

3.8. Oblique mood markers (3 markers in one language)

There are languages that use oblique moods in IQs, see also [Kahrel 1985]. For example, in Tundra Nenets one of the oblique moods must be used if the dependent verb is in the past tense (17), and with the verb 'ignore' in the present or in the past (18).

(17) Tundra Nenets [Nikolaeva 2014: 306]

n'anaq   ma-s'°,   xən'ana   Wera   yil'e-sa
DAT.1PL   say-PST   where   Wera   live-INTER
'He told us where Wera lived.'

(18) Tundra Nenets [Nikolaeva 2014: 307]

mən'°   yexaraə-d°m,   xən'ana   yil'e-naki°
I   ignore-1SG   where   live-PROB
'I don't know where he lives.'

See (17) with the interrogative mood marker and (18) with the probabilitive.

The use of oblique moods in IQs is a problematic issue, since the context of an embedded question ("I don't know if P"; "I doubt if P"; "I asked if P") often makes the statement about a proposition which is only true with a degree of probability. Thus, in languages with elaborated irrealis mood system such clauses are expected to take oblique mood markers, as dubitative in Tundra Nenets:

(19) Tundra Nenets [Nikolaeva 2014: 308 (shortened)]

mən'°   yexaraə-d°m,   tūt°-bə-ta=w°h
I   ignore-1SG   come.FUT-COND-3SG-DUB
'I don't know if he is coming or not'.

---

[4] The semantic distinction between polar and alternative questions in concrete examples is also obscure, since polar questions may either have the polar or the alternative semantics [Krifka 2013].

[Nikolaeva 2014] does not explicitly mention the dubitative mood among IQ markers; however, all the six two-predicate examples on page 308 include either the dubitative, or the inferential, or the conditional (the latter being listed among IQ markers). See also J. Barbour's comment on Neverver: "The category of utterance predicates is the only CTP category with independent mood marking in the complement. All others involve dependent mood marking of some sort." [Barbour 2012: 386].

To disambiguate between irrealis mood as an IQ marker *vs.* a side-effect of the semantics of the dependent clause we need to find specific context of the type "I don't doubt if P", and such examples are most often absent from reference grammars.

3.9. Focus marking devices (5 focus particles/constructions)

Udihe makes use of the focus particle *-dA* to make polar question with the finite verb, cf.:

(20) Udihe [Nikolaeva, Tolskaya 2011: 443]
    omo   kusige-we-de   xebu-je
    one   knife-ACC-FOC   take-IMP.2SG
    'Take (at least) one knife.'

(21) Udihe [Nikolaeva, Tolskaya 2011: 818]
    čai   xekui-we-ni-de   amtala-ja.
    tea   hot-ACC-3SG-FOC   try-IMP.2SG
    'Try whether the tea is hot.'

This particle is a clitic introducing emphatic focus, the meanings "even", "too, as well", and it also used to form indefinite pronouns. It can also introduce concessive clauses, which makes it close to the "subordination" type illustrated in 3.5. Similar to *-dA* in Udihe, all the focus markers attested in our sample show large polysemy. For example, the particle *ba-* in Basque, apart from IQs and predicate focus, can introduce conditional clauses [de Rijk 2008: 172–183; 413].

For Humburi Senni the author states that "the particle *gâ* (*the IQ marker — N.S.*) is a workhorse in Humburi Senni grammar. It is the relative morpheme (§8.3), a kind of adverbial conjunction (§8.3.11), and the focus morpheme (§8.1). It is also the 'that' conjunction in factive complements." [Heath 2014: 398]. There is evidence for different analysis of focus and relative constructions; however, the author himself is not sure whether they should be treated

as different or the same. Minor syntactic differences between relativization (which is used to make focus constructions and root questions, see [Sumbatova 2009: 569]) and IQ are also observed in Adyghe [Lander 2014: 253–254]; however, the author explicitly states that "IQs are a subtype of relative constructions" [Lander 2014: 253].

Thus, the focus markers/constructions attested in our sample are counted twice (except for Udihe, for the reason that the discussed particle seems to be a secondary means of introducing concession, the infinitive being the main means [Nikolaeva, Tolskaya 2011: 727–728]).

3.10. Distribution of IQ markers

The distribution of 89 IQ markers is illustrated in Figure 1.

IQ marking types

- Oblique moods 3,4%
- Focus marking devices 5,6%
- Two-predicate complexes 15,7%
- "Null" strategy 21,3%
- Disjunctive 2,2%
- Question particles 18,0%
- Subordinators: complementizers 13,5%
- Subordinators: conditionals 10,1%
- Subordinators: concessive clause mark… 2,2%
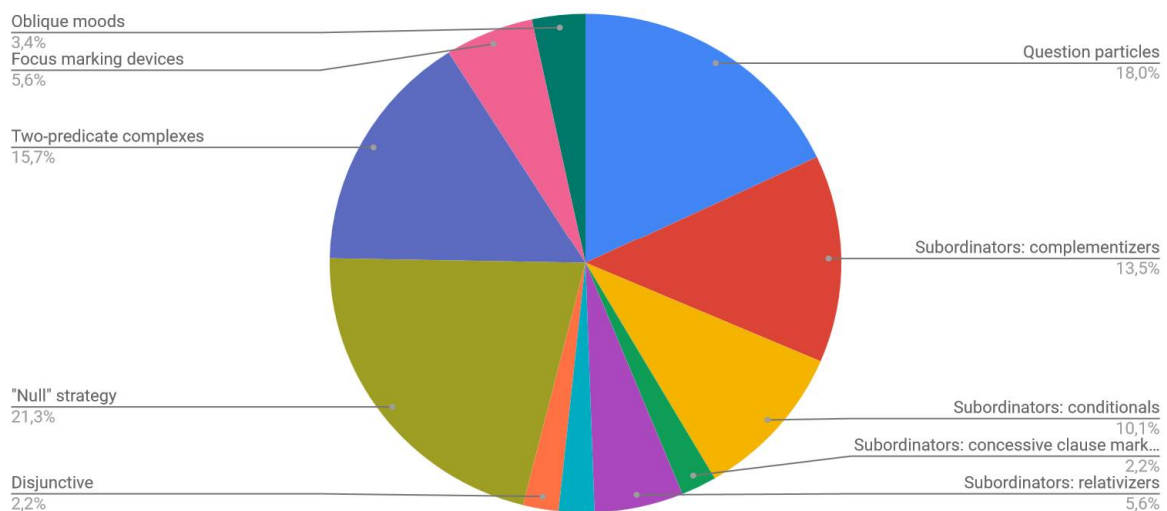- Subordinators: relativizers 5,6%

Figure 1. Distribution of IQ markers in languages of the sample

4. Preliminary typological implications

Comparing IQs to root questions, we can formulate the following preliminary implications:

1) If IQs are marked with the null strategy then it is also possible for root questions of the same type (polar, wh, alternative).

2) If a two-predicate complex is possible in root questions, then it is also possible in IQs (the opposite is not true).

3) If a question particle is used both in root and IQs, it usually covers the same questions types.

The present study has shown the relevance of the following typological parameters: types of questions (polar, wh, alternative); semantic restrictions of the examined markers in root questions; synchronous polysemy: the functions of IQ markers in root and subordinate clauses (adverbial clauses, citations etc.). Whenever possible, we have used deictic shift as a criterion for distinguishing between direct and indirect speech. However, our data shows that the presence of the deictic shift per se must be analyzed as a separate parameter (as it is not attested in a number of languages [Noonan 1985]), and another relevant parameter is the presence of the deictic shift in the context of the particular IQ marker.

We have hypothesized that the semantics of embedding predicate could be a significant parameter (e.g. speech *vs.* mental predicates). Indeed, there are quite a few languages, where concrete predicates have restrictions on IQ markers; however, the distribution has idiosyncratic, rather than systematic character.

Thus, this pilot study has enabled us to test the parameters relevant for the description of IQs in languages of the world; the next step is to expand our sample and provide an in-depth study of the complicated parameters.

Abbreviations

1, 2. 3 — 1st, 2nd, 3rd person; A — agent marker; ABS — absolutive; ACC — accusative; ADD — additive; ALL — allative; ATTR — attributive; AUX — auxiliary; CAUS — causative; COH — coherence marker; COND — conditional; CVB_SIM — converb of simultaneity; DAT — dative; DEF — definite; DEM — demonstrative; DIR — directive; DIS — disjunctive; DUB — dubitative; DYN — dynamicity; EMPH — emphasis; ERG — ergative; EXST — existential verb; F — female; FOC — focus; FP — future tense prefix; FUT — future; GEN — genitive; HAB — habitual auxiliary; IMP — imperative; IMPF — imperfective; INTER — interrogative mood; INTR — intransitive; IO — indirect object; IQ — indirect question; LOC — locative; M — male; MED — medial; MNR — manner; N1 — neuter 1; NEG — negative; NPST — non-past; NRPST — near past; OBL — oblique; PERF — perfect auxiliary; PFV — perfective; PL — plural; PQ — polar question; PR — person agreement with the possessor; PROB — probabilitive; PRS — present; PRTT — partitive; PST — past; PTCP — participle; Q — question marker; QUOT — quotative; REL — relativizer; S — subject; SBJV — subjunctive; SC — serial verb connective; SEQ — sequential; SG — singular; SS — same subject.

References

Aikhenvald 2011 — Aikhenvald A.Y. Speech reports: A cross-linguistics perspective. Aikhenvald A.Y., Dixon R.M.W. Language at large. Essays on syntax and sematics. Leiden – Boston: Brill, 2011.

Arkadiev et al. 2009 — Аркадьев П.М., Ландер Ю.А., Летучий А.Б., Сумбатова Н.Р., Тестелец Я.Г. Введение: основные сведения об адыгейском языке // Тестелец Я.Г., Аркадьев П.М., Летучий А.Б., Сумбатова Н.Р. (ред.). Аспекты полисинтетизма: очерки по грамматике

адыгейского языка. М.: РГГУ, 2009. С. 17–121. [Arkad'ev P.M., Lander Yu.A., Letuchii A.B., Sumbatova N.R., Testelets Ya.G. Vvedenie: osnovnye svedeniya ob adygeiskom yazyke. [Introduction: Basic information about Adyghe]. Testelets Ya.G., Arkad'ev P.M., Letuchii A.B., Sumbatova N.R. (eds.). Aspekty polisintetizma: ocherki po grammatike adygejskogo jazyka. Moscow: RSUH, 2009. P. 17–121].

Bakker 2010 — Bakker D. Language sampling. Song J.J. The Oxford handbook of linguistic typology. Oxford: Oxford University Press, 2010. P. 100–127.

Barbour 2012 — Barbour J. A grammar of Neverver. Berlin, Boston: De Gruyter Mouton, 2012.

Berghäll 2015 — Berghäll L. A grammar of Mauwake. Language Science Press, 2015.

Carlson 2011 — Carlson R. A grammar of Supyire. Berlin, Boston: Mouton De Gruyter, 2011.

de Rijk 2008 — de Rijk R.P.G. Standard Basque, a progressive grammar. Vol. 1. Cambridge, MA: MIT Press. 2008.

Dik 1978 — Dik S.C. Functional grammar. North-Holland, 1978.

Dik 1980 — Dik S.C. Studies in functional grammar. Academic Press, 1980.

Dixon, Aikhenvald 2006 — Dixon R.M.W., Aikhenvald A.Y. (eds.). Complementation. Oxford: Oxford University Press, 2006.

Egorova 2020 — Егорова А.Д. Косвенный вопрос в чувашском языке. Доклад на конференции «Малые языки в большой лингвистике» (МЯБЛ-2020), М.: 17–18 апреля 2020. [Egorova A.D. Kovsennyj vopros v chuvashskom yazyke. [Indirect question in Chuvash]. Talk at the conference «Malye yazyki v bolshoj lingvistike» (MYaBL-2020), Moscow, April 17–18 2020.]

Epps 2005 — Epps P. A grammar of Hup. Mouton de Gruyter, 2005.

Fedden 2011 — Fedden S.O. A grammar of Mian (Mouton Grammar Library 55). Berlin: Mouton de Gruyter, 2011.

Gerasimov, Lander 2008 — Gerassimov D.V. Lander Ju.A. Rel'ativizacija pod maskoj nominalizacii i faktivnyj argument v adygejskom jazyke [Relativization under the mask of nominalization]. Plungian V A., Tatevosov S.G. (eds.). Issledovanija po glagol'noj derivacii [Studies in verbal derivation]. Moscow: Jazyki slavjanskix kultur, 2008. P. 290–313.

Haspelmath 2010 — Haspelmath M. Comparative concepts and descriptive categories in crosslinguistic studies. Language. 2010. Vol. 86. № 3. P. 663–687.

Heath 2014 — Heath J. Grammar of Humburi Senni (Songhay of Hombori, Mali). Leipzig: Max-Planck Institute for Evolutionary Anthropology, 2014.

Hualde, Ortiz de Urbina 2011 — Hualde J., Ortiz de Urbina J. (eds.). A grammar of Basque. Mouton de Gruyter, 2011.

Kahrel 1985 — Kahrel P. Indirect questions. Bolkestein A.M., de Groot C., MacKenzie J.L. (eds.). Predicates and terms in functional grammar (Functional Grammar Series, 2). Mouton de Gruyter, 1985. P. 165–182.

King 1994 — King A.R. The Basque language. A practical introduction. Reno – Las Vegas – London: University of Nevada Press, 1994.

König, Siemund 2007 — König E., Siemund P. Speech act distinctions in grammar. Shopen T. (ed.). Language Typology and Syntactic Description. Vol. I. Cambridge: Cambridge University Press, 2007. P. 276–324.

Krifka 2013 — Krifka M. Questions. Maienborn C., Heusinger K., Portner P. (eds.). Semantics. Vol. 3. Berlin, Boston: De Gruyter Mouton, 2013. P. 1742–1785.

Kroeger 2005 — Kroeger P.R. Analyzing grammar. An introduction. Cambridge University Press, 2005.

Lander 2014 — Ландер Ю.А. Релятивизация в полисинтетическом языке: адыгейские относительные конструкции с типологической перспективе. Дисс… канд. филол. н.

РГГУ. М., 2014. [Lander Yu.A. Relyativizatsiya v polisinteticheskom yazyke: adygeiskie otnositel'nye konstruktsii s tipologicheskoi perspektive [Relativization in a polysynthetic language: Adyghe relative clauses in typological perspective]. Ph.D. thesis. Russian State University for the Humanities, Moscow, 2014.]

Liing 2014 — Liing W.-J. How to ask questions in Mandarin Chinese. CUNY Academic Works. 2014. https://academicworks.cuny.edu/gc etds/445.

Miestamo et al. 2016 — Miestamo M., Bakker D., Arppe A. Sampling for variety. Linguistic Typology. 2016. Vol. 20. № 2. P. 233–296.

Munro 1982 — Munro P. On the transitivity of "say" verbs. Hopper P.J., Thompson S.A. (eds.). Syntax and semantics. Vol. 15: Studies in transitivity. New York: Academic Press, 1982. P. 301–319.

Nikolaeva 2014 — Nikolaeva I. A grammar of Tundra Nenets. Berlin, Boston: De Gruyter, 2014.

Nikolaeva, Tolskaya 2011 — Nikolaeva I., Tolskaya M. A grammar of Udihe. Berlin, Boston: Mouton De Gruyter, 2011.

Noonan 1985 — Noonan M. Complementation. Shopen T. (ed.). Language Typology and Syntactic Description. Vol. 2: Complex Constructions. Cambridge: Cambridge University Press, 1985. P. 42–140.

Olawsky 2006 — Olawsky K.J. A grammar of Urarina. Berlin: Mouton de Gruyter, 2006.

Partee 1973 — Partee B.H. The syntax and semantics of quotation. Anderson S., Kiparsky P. (eds.). A Festschrift for Morris Halle. New York: Holt, Rinehart and Winston, 1973. P. 410–418.

Sumbatova 2009 — Сумбатова Н.Р. Коммуникативная структура адыгейского предложения: перспектива и фокус // Тестелец Я.Г., Аркадьев П.М., Летучий А.Б., Сумбатова Н.Р. (ред.). Аспекты полисинтетизма: очерки по грамматике адыгейского языка. М.: РГГУ, 2009. С. 559–612. [Sumbatova N.R. Kommunikativnaya struktura adygeiskogo predlozheniya: perspektiva i fokus [Information structure of Adyghe sentences: Perspective and focus]. Testelets Ya.G., Arkad'ev P.M., Letuchii A.B., Sumbatova N.R. (eds.). Aspekty polisintetizma: ocherki po grammatike adygejskogo jazyka. Moscow: RSUH, 2009. P. 559–612.]

Testelets 2001 — Тестелец Я.Г. Введение в общий синтаксис. М.: РГГУ, 2001. [Testelets Ya.G. Vvedenie v obshchii sintaksis [Introduction to general syntax]. Moscow: RGGU, 2001.]

Appendix: List of languages

Adyghe > Circassian > Northwest Caucasian
Alaskan Yupik > Eskimo > Eskimo-Aleut
Bardi > Nyulnyulan
Basque (Isolate)
Chuvash > Oghur > Turkic
Domari > Indo-Aryan > Indo-European
English > Germanic > Indo-European
Hdi > Chadic > Afro-Asiatic
Hebrew (Modern) > Semitic > Afro-Asiatic
Hindi > Indo-Iranian > Indo-European
Humburi Senni > Songhay > Nilo-Saharan
Jamsay > Dogon > Niger-Congo
Kayardild > Tangkic > Macro-Pama-Nyungan
Ket > Nothern Yeniseian > Dené-Yeniseian
Lango > Nilotic > Nilo-Saharan

Lezgian > Lezgic > Northeast Caucasian
Mauwake > Trans-New Guinea
Mian > Ok > Trans-New Guinea
Momu > Fas
Mongsen Ao > Ao > Sino-Tibetan
Mosetén > Moseten-Chonan
Neverver > Malayo-Polynesian > Austronesian
Ossetic > Iranian > Indo-European
Russian > Slavic > Indo-European
Saamáka (English based creole)
Supyire > Atlantic-Congo > Niger-Congo
Tamashek > Berber
Tundra Nenets > Samoyedic > Uralic
Udihe > Tungusic
Urarina > Macro-Jibaro

Валерия Алексеевна Морозова

Национальный исследовательский университет «Высшая школа экономики»

Valeriya Morozova

National Research University Higher School of Economics

tito_alba@mail.ru

Наталья Вадимовна Сердобольская

кандидат филологических наук; старший научный сотрудник, Институт языкознания РАН; старший научный сотрудник, Государственный институт русского языка им. А. С. Пушкина

Natalia Serdobolskaya

Ph. D.; senior researcher, Institute of Linguistics RAS; senior researcher, Pushkin State Russian language Institute

serdobolskaya@gmail.com

Грамматикализация и тип эвиденциальной системы[*]

А. Е. Постникова

Национальный исследовательский университет «Высшая школа экономики»

В статье представлены результаты типологического исследования эвиденциальных систем в девяноста двух языках мира. В работе выдвигаются некоторые фреквенталии относительно выражения эвиденциальности, а также рассматривается связь между степенью грамматикализации эвиденциальных систем и значениями, выражающимися с их помощью.

Ключевые слова: эвиденциальность, типология, грамматикализация.

Grammaticalization and the typology of evidential systems[**]

Anna Postnikova

National Research University Higher School of Economics

This article presents the results of a typological study of evidential systems in ninety two natural languages. The paper suggests some tendencies concerning the coding of evidentiality; it also considers the relationship between the grammaticalization and the semantics of evidential markers.

Keywords: evidentiality, typology, grammaticalization.

---

1. Введение

Эвиденциальность — грамматическая категория, выражающая указание на способ[1], которым говорящий получил информацию о сообщаемой им ситуации [Плунгян 2011: 449]. В настоящей статье рассматривается многообразие эвиденциальных значений и способов их выражения. Мы фокусируемся на типах эвиденциальных систем в языках мира, а также выдвигаем некоторые положения об устройстве этих систем и, в частности, показателей, представленных в них. Основная цель исследования — установить, есть ли связь между степенью грамматикализации эвиденциальной системы и значениями, которые выражаются с ее помощью.

Исследование проводилось на материале девяноста двух языков. Всего было проанализировано сто семь эвиденциальных систем[2]. В выборку включены языки из всех макроареалов, однако из-за неоднородности распространения эвиденциальности некоторые из ареалов представлены лучше, чем другие. Крайне малое количество языков с грамматическим выражением способа получения информации зафиксировано в Африке. По этой причине всего три африканских языка вошли в выборку. Лучше всего в выборке представлены языки Южной Америки, Северной Америки и Кавказа (полная языковая выборка представлена в приложении). При составлении выборки мы опирались на данные обзорной работы [de Haan 2005].

2. Типы эвиденциальных значений

А. Ю. Айхенвальд [Aikhenvald 2004: 63] выделяет шесть значений, кодирование которых может различаться в языке:

Визуальный — информация, полученная зрительно.

Сенсорный — информация, полученная посредством органов чувств помимо зрения.

Инферентивный — вывод, основанный на доказательствах, полученных посредством органов чувств.

Презумптивный — вывод, основанный на фоновых знаниях говорящего.

---

[1] В традиционной литературе об эвиденциальности (например, в [Aikhenvald 2004]) часто говорится о том, что эвиденциальность кодирует источник информации. Несмотря на то, что 'источник' и 'способ получения' информации — концептуально близкие понятия, второе понятие более точно отражает использование рассматриваемой категории. Поэтому, вслед за В. А. Плунгяном, в настоящей работе используется именно это определение.

[2] См. раздел 3.1 о том, почему число систем больше числа языков.

Репортативный — информация, полученная от третьего лица без упоминания источника.

Квотативный — информация, полученная от третьего лица с упоминанием источника.

Данная классификация не учитывает все эвиденциальные значения, существующие в языках мира. Так, в ней игнорируются 'партиципантное'[3] значение и значение 'общего знания'[4] [Плунгян 2011: 357], а также разделение на 'индивидуальное' и 'дуальное' знание[5].

В. А. Плунгян [2011: 349–350] классифицирует эвиденциальные значения в зависимости от типа доступа к информации: **прямой** доступ противопоставляется **непрямому**, **личный** доступ — **неличному**. Первая пара признаков характеризует наличие у говорящего доступа непосредственно к ситуации, а вторая — наличие доступа к информации о том, что ситуация произошла. Описанные типы доступа позволяют разделить эвиденциальные значения на три группы и расположить их на шкале от наибольшей степени засвидетельствованности до наименьшей.

визуальный > сенсорный > инферентивный > презумптивный > репортативный > квотативный

⟵——————————————————————————————————⟶

прямой доступ   *непрямой доступ*
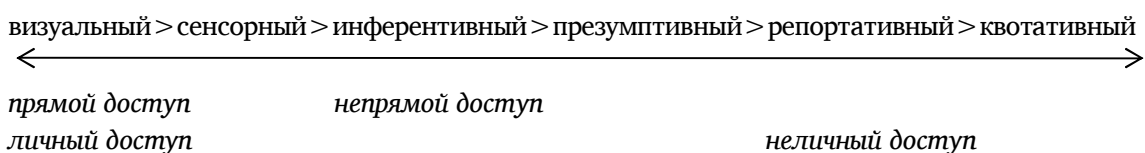*личный доступ*   *неличный доступ*

Рисунок 1. Иерархия эвиденциальных значений

Иерархия, представленная на рисунке 1, постулируется не только на основе классификации В. А. Плунгяна. На основании данных была сформулирована фреквенталия (1), которую нарушают всего два языка из выборки.

(1) Один показатель может покрывать только те два эвиденциальных значения, между которыми на шкале нет ни одного значения, которое в этой же системе выражается другим показателем.

---

[3] Показатели с партиципантным значением используются в том случае, когда говорящий сам принимал непосредственное участие в описываемом событии. Встречается, в том числе, в языках помо (Северная Америка, [McLendon 2003: 124]).

[4] Показатели с таким значением используются при описании общеизвестных фактов. Зафиксированы, например, в южном намбиквара (намбикварские, Южная Америка, [Eberhard 2018: 336]) и калмыцком (монгольские, Евразия, [Brosig, Skribnik 2018: 565]).

[5] Показатели индивидуального знания используются, если информацией владеет только говорящий, а дуального — если информация есть и у говорящего, и у собеседника. Встречается в языках намбиквара (Южная Америка, [Eberhard 2018: 336]) и кечуа (Северная Америка, [Hintz 2007: 66]).

Фреквенталия (1) не допускает, к примеру, существования такой системы, в котором маркер А кодировал бы визуальное и инферентивное значения, а маркер Б — сенсорное.

Один и тот же показатель может покрывать более двух значений — в таком случае правило (1) применяется к каждой паре значений. Так, суффикс *-па* в языке кева (энганские, Новая Гвинея, [Franklin 1964: 119])[6] покрывает все значения от сенсорного до репортативного (2).

(2) кева, энганские [Franklin 1964: 119]

> *pú-a-**ya***
> уйти-3SG.PST-**NVIS**
> 'Он ушел незамеченным.'

Квотативное значение нередко выражается тем же показателем, что и репортативное. Например, частица *négе* в языке гарифуна (аравакские, Южная Америка, Taylor 1956: 144) может появляться как с указанием источника (3a), так и без него (3b).

(3) гарифуна, аравакские [Taylor 1956: 144]

a. *leréga*   ***négе***   *tú*   *lúagu*   *láunaha*   ***négе***   *gárada*
3SG.M-сказать-INDF   **QUOT**   ей   о   3SG.M-послать   **QUOT**   письмо

   lú
   ему
   'Похоже, он сказал ей, что он отправит ему письмо.'

b. *ábạ* ***négе***   *ladógonụụ*
   и   **QUOT**   3SG.M-делать-3SG.INAN
   'Тогда, как они говорят, он написал его.'

Однако квотативное значение, как правило, не выражается показателями, кодирующими другие эвиденциальные значения кроме репортативного. В турецком (тюркские, Евразия) языке индирективный перфект покрывает все значения непрямой засвидетельствованности за исключением квотатива. Передать информацию, полученную от третьего лица, с указанием источника можно только с помощью конструкций с глаголами речи [Johanson 2003: 274].

---

[6] Здесь и далее в скобках указываются языковая семья и макроареал.

В выборке были зафиксированы две эвиденциальные системы, нарушающие правило (1). В языке богайа (изолят, Новая Гвинея, [Sarvasy 2018: 627]) эвиденциальные значения имеются у двух клитик: визуальной =*ki* и сенсорной =*ai*. Однако в некоторых контекстах =*ki* может также иметь инферентивное значение (4). Употребление этого показателя в инферентивном значении доступно только в том случае, если вывод сделан на основании видимого результата.

(4)   богайа, изолят [Sarvasy 2018: 627]

ho   *mabaro*   *wagan*   *mogona*=***ki***
3SG   свинья   охотиться   идти.PST=VIS
'Он ушел охотиться на свиней (я видел это/я сделал вывод).'

В языке паиван (австронезийские, Австронезия, [Pan 2018: 661]) противопоставлены две частицы: с инферентивно-презумптивным и с неинферентивным значениями. Частица *kaymaya* используется говорящим в том случае, если он сделал вывод на основе видимого результата или общих знаний. Все остальные эвиденциальные значения покрывает частица *aya* (5a–b)[7].

(5)   паиван, австронезийские [Pan 2018: 661]

a. *vaik*   *a*   *paucn*
   идти.AV   LINK   смотреть.AV

   na=*maca*=*anga*   ***aya***   *tua*   *zua*   *uʔalay*
   PFV=умереть.AV=COM   **NON.INFR**   OBL   этот   мужчина
   'Они ходили посмотреть. Этот мужчина умер.'

b. *'uzi*   *ki*   *limecav*=*anga*   *a*   *za*   *zalum'*   ***aya***   *timadju*
   хорошо   будет   чистый=COM   NOM   эта   вода   **NON.INFR**   3SG.NOM
   'Вода будет чистой? — сказал он.'

3. Основные параметры выражения эвиденциальности

3.1. Эвиденциальные системы

Язык может иметь как одну систему с эвиденциальными значениями, так и несколько. В западногренландском языке (эскимосско-алеутские, Север-

---

[7] На основании имеющихся данных нельзя однозначно утверждать, есть ли у частицы *aya* визуальное значение. Требуются дополнительные исследования, которые могут как подтвердить, так и опровергнуть нарушение языком паиван фреквенталии (1).

ная Америка) есть, по меньшей мере, четыре системы, в которых представлены показатели, кодирующие способ получения информации [Fortescue 2003: 291]. Так, в словоформе западногренландского глагола есть особый слот, который заполняется так называемыми сентенциальными аффиксами (категория, занимающая промежуточное положение между словообразованием и словоизменением). Два таких аффикса, *-gunar* и *-sima*, несут эвиденциальное значение, а остальные — эпистемическое. Ко второй системе относятся три словообразовательных аффикса с возможной репортативной интерпретацией: *-(r)palaar, -(r)pallaC* и *-(r)paluC*. Помимо этого, в западногренландском описаны еще два способа кодирования информации, полученной от третьего лица — клитика *=guuq* и частица *unnia*.

Разные системы одного языка различаются по своей сложности, составу и степени грамматикализации маркеров, представленных в ней. Нередко в одной клаузе сочетаются маркеры из двух разных систем. В примере (6) из аляскинского юпика в одной клаузе используются репортативная клитика *=qquq* и инферентивный аффикс *-llini*.

(6) аляскинский юпик, эскимосско-алеутские [Krawczyk 2012: 8]
     *tuai=llu=**qquq**   tauna   massiinaq   arulairr-**llini**-lria*
     и.тогда=и=**REP**   что   машина.ABS.3SG   остановиться-**INFR**-PART.3SG
     'И тогда, как они говорят, эта машина остановилась.'

Критерии разграничения самостоятельных эвиденциальных систем в одном языке требуют дальнейшего исследования. При описании систем мы руководствовались данными грамматик: если в грамматике два некоторых эвиденциальных показателя не объединяются в одну систему, в выборке они также относились к разным системам. В спорных случаях было принято техническое решение относить эвиденциальные показатели к разным системам, если они занимают разные позиции в словоформе и не находятся в парадигматических отношениях[8]. В некоторых языках одно и то же значение выражается несколькими показателями; тогда эти показатели также относились к разным системам.

---

[8] Некоторые языки позволяют совместное появление нескольких эвиденциальных маркеров в одной клаузе — например, языки помо [McLendon 2003: 112]. Тем не менее, во всех грамматиках эти показатели объединяются в одну парадигму. В нашем исследовании мы также отнесли их к одной системе.

3.2. Взаимодействие категорий эвиденциальности и времени

Эвиденциальные значения могут выражаться как в собственно эвиденциальных системах, так и в других системах языка. Из последних наиболее распространенными являются эвиденциальные расширения системы времен (см. подробнее раздел 4.3.2). В турецком языке [Johanson 2003: 274] представлены два прошедших времени. Одно из них покрывает все значения непрямой засвидетельствованности от инферентивного до репортативного; второе прошедшее, как и остальные времена в турецком, не кодирует способ получения информации. Наличие эвиденциальных расширений у системы времен также зафиксировано, например, в ингушском (нахско-дагестанские, Евразия, [Nichols 2011: 246]) и коми-зырянском (уральские, Евразия, [Skribnik, Kehayov 2018: 540]). В некоторых языках все показатели одного или нескольких времен кодируют эвиденциальные значения (см. подробнее раздел 4.3.1). Так, в тариана (аравакские, Южная Америка, [Aikhenvald 2003: 133]) в настоящем и прошедшем временах невозможно не указать на способ получения информации. В таких случаях можно говорить о том, что эвиденциальность и время выражаются кумулятивно.

В настоящем исследовании эвиденциальные расширения временных систем рассматриваются отдельно от собственно эвиденциальных систем.

3.3. Обязательность

Кодирование эвиденциальности в системе может быть как опциональным, так и обязательным. Эвиденциальность считается облигаторизованной, если способ получения информации невозможно не выразить. Облигаторизованные эвиденциальные системы встречаются относительно редко. Только в пятнадцати из ста семи рассмотренных систем способ получения информации кодируется обязательно.

В некоторых облигаторизованных системах на один или несколько типов информации может указывать немаркированная форма. В языке хуп (надахупские, Южная Америка, [Epps 2005: 623]) эвидециальные значения выражаются с помощью клитик. Отсутствие эвиденциальных показателей в предложении может означать только то, что информация была получена зрительно (7).

(7)   хуп, надахупские [Epps 2005: 623]
    manga   hɨd-an   tǝw-nɨh = kah
    Маргарита   3PL-OBJ   кричать.на-NEG = DISJ
    'На самом деле, Маргарита не кричала на них (я был там).'

В некоторых контекстах обязательность эвиденциальности может пропадать. Кодирование эвиденциальности в калашском (индоарийские, Евразия) является обязательным только в прошедшем времени [Bashir 1988]. В болгарском (славянские, Евразия) способ получения информации невозможно не указать в прошедшем и настоящем времени [Smirnova 2011: 279]. Очень малое количество языков позволяет кодирование источника информации в ирреалисе, и ни в одной системе оно не является обязательным. Часто эвиденциальные маркеры не сочетаются с первым лицом, особенно если в системе есть и другие ограничения на их употребление [Aikhenvald 2004: 219][9].

В настоящей работе система считается облигаторизованной, если кодирования источника информации невозможно избежать по крайней мере в наиболее характерном для этого контексте — в прошедшем времени и в третьем лице.

4. Способы кодирования эвиденциальности

Способы выражения эвиденциальных значений не менее разнообразны, чем сами значения. Эвиденциальность может кодироваться аффиксами, клитиками, частицами, а также выражаться с помощью глагольных времен.

4.1. Кодирование эвиденциальности аффиксами

Во многих системах источник информации кодируется глагольными аффиксами, чаще всего суффиксами. Так, эвиденциальность выражается с помощью суффиксов в абхазском (абхазо-адыгские, [Chirikba 2003: 262]), дияри (пама-ньюнгские, [Austin 1981: 180]), чероки (ирокезские, [Reyburn 1954: 64]) и других языках мира. Префиксы, кодирующие эвиденциальность, встречаются достаточно редко. В языке шиллук (нилотские, Африка, [Storch 2018: 610]), одном из немногих африканских языков с грамматикализованной эвиденциальностью, представлены префиксы *á-* со значением прямой засвидетельствованности (8a) и *ú-* со значением косвенной засвидетельствованности (8b).

---

[9] Подробнее о появлении эвиденциальных показателей в контекстах будущего времени и других ирреальных контекстах и о сочетании эвиденциальности и первого лица, а также о семантических эффектах, возникающих в этих случаях можно узнать из глав 7 и 8 работы [Aikhenvald 2004: 217, 241].

(8) шиллук, нилотские [Storch 2018: 610]
  a. *dhyáng`*    ***á-kwal`***   *yi*   *cvl*
     корова   DIR.PST-украсть   ERG   Кол
     'Кол украл корову (я видел это).'

  b. *dhyáng`*    ***ó-kwal`***   *yi*   *cvl*
     корова   INDIR.PST-украсть   ERG   Кол
     'Полагаю, Кол украл корову.'

В выборке зафиксирован всего один случай кодирования эвиденциальности циркумфиксом. В настоящем времени визуальное значение в языке фасу (изолят, Новая Гвинея, [Sarvasy 2018: 650]) выражается с помощью циркумфикса *a-...-re* (9). Все остальные значения выражаются суффиксами.

(9) фасу, изолят [Sarvasy 2018: 650]
     a*-pe-**re***
     VIS-приходить-CIRC
     'Я вижу, как оно приближается.'

Кодирование эвиденциальности инфиксами или трансфиксами не обнаружено ни в одной системе выборки.

4.2. Кодирование эвиденциальности клитиками и частицами

Клитики кодируют эвиденциальность во многих языках мира. Часто эвиденциальные клитики могут присоединяться только к глаголу — как в гитксане (цимшианские, Северная Америка, [Peterson 2018: 466]) и цахурском (нахско-дагестанские, Евразия, [Forker 2018: 505]). Синтаксическая дистрибуция клитик в других системах гораздо шире. В шипибо-конибо (пано-таканские, Южная Америка, [Valenzuela 2003: 35]) клитика *=ra* со значением прямой засвидетельствованности присоединяется к первому главному члену предложения (10).

(10) шипибо-конибо, пано-таканские [Valenzuela 2003: 35]
     westíora   *nete-n=**ra***   *ka-a*   *iki*   *nokon*   *yosi*
     один   день-TEMP=DIR   идти-PP   AUX   POSS.1SG   старший

     betan   *e-a,*   *piti*   *bena-i...*
     CONJ   1-ABS   рыба:ABS   искать-SSSI
     'Однажды мой дедушка и я пошли искать рыбу...'

В некоторых языках эвиденциальность может кодироваться частицами. В настоящем исследовании частицами называются просодически самостоятельные, но не изменяемые служебные слова. Эвиденциальность кодируется частицами во многих малых системах, состоящих из одного показателя. Такие системы можно встретить в абхазском (наряду с суффиксальным способом; абхазо-адыгские, Евразия, [Chirikba 2003: 258]), баскском (изолят, Евразия, [Korta, Zubeldia 2014]) и команчи (юто-ацтекские, Северная Америка, [Charney 1993: 188]).

Встречаются и сложные системы, способ получения информации в которых кодируется с помощью частиц. В языке кора (юто-ацтекские, Северная Америка) представлена система с четырьмя показателями, три из которых (инферентивный, репортативный и квотативный) являются частицами, а четвертый (показатель прямой засвидетельствованности) — клитикой [Thornes 2018: 426].

На материале грамматик не всегда представляется возможным различение клитик от частиц и клитик от аффиксов. Часто эти понятия смешиваются, а их характеристики определяются разными исследователями по-разному. При описании каждого языка мы делали выбор в пользу того понятия, которое используется автором грамматики.

4.3. Кодирование эвиденциальности во временных системах

Эвиденциальные значения часто выражаются во временных системах языка. Эвиденциальность в таких системах может кодироваться как обязательно, так и опционально[10]. В первом случае имеет смысл говорить о кумулятивном выражении времени и эвиденциальности; во втором — об эвиденциальных расширениях некоторых времен.

4.3.1. Системы с обязательным кодированием эвиденциальности

В языке тариана различаются пять типов источников информации в прошедшем времени и три в настоящем [Aikhenvald 2003: 133]. Все показатели представлены в таблице 1 и выражают кумулятивно способ получения информации и время. Согласование по лицу и числу кодируется другими показателями. Таким образом, эвиденциальность не может быть не выражена в тариана (по крайней мере, в настоящем и прошедшем временах).

---

[10] Даже если временная система, с помощью которой выражается эвиденциальность в некотором языке, облигаторизована, кодирование способа получения информации может не быть обязательным. Так, в турецком выражение времени обязательно в финитной клаузе, однако говорящий может избежать упоминания способа, которым он получил эту информацию, используя эвиденциально не маркированное время — аорист.

Таблица 1. Эвиденциальность в тариана [Aikhenvald 2003: 133]

типы значений	прошедшее время	настоящее время
визуальный	*-ka*	*-naka*
сенсорный	*-mahka*	*-mha*
инферентивный	*-nihka*	—
презумптивный	*-sika*	—
репортативный + квотативный	*-pidaka*	*-pida*

В некоторых системах эвиденциальность кодируется кумулятивно не только со временем, но и с лицом и числом, как, например, в языке туюка (туканские, Южная Америка) и некоторых других туканских языках [Barnes 1984: 258].

Облигаторизованные системы, в которых эвиденциальность кодируется кумулятивно со временем, можно также обнаружить в калашском [Bashir 1988], колымском юкагирском (изолят, Евразия, [Maslova 2003: 220]) и других языках мира.

4.3.2. Системы с необязательным кодированием эвиденциальности

Глагольные времена с эвиденциальными расширениями — достаточно распространенная стратегия кодирования способа получения информации. Наиболее типичный пример выражения эвиденциальности через время — турецкий индиректный перфект (раздел 3.2). Подобная система характерна для многих тюркских языков.

Более сложная система представлена в ингушском. Для ингушского описывается несколько серий времен (к примеру, простые времена, серия будущего, серия перфекта и другие). Некоторые времена обладают эвиденциальными расширениями, причем в одной серии может быть несколько времен, кодирующих один и тот же способ получения информации, но различающихся, к примеру, по аспектуальному значению. Не во всех сериях можно выразить все эвиденциальные значения, имеющиеся в ингушском. Так, в серии перфекта есть время, покрывающее все значения непрямой засвидетельствованности, и время с исключительно инферентивным значением. С помощью простых времен можно кодировать только инферентив. В серии будущего ожидаемо не зафиксировано ни одного эвиденциального расширения [Nichols 2011: 246].

5. Грамматикализация эвиденциальных показателей

5.1. В собственно эвиденциальных системах

В работе [Lehmann 2015 (1982): 15] предлагается шкала грамматикализованности. В соответствии с ней выводится иерархия, представленная на рисунке 2.

ноль	>	аффикс	>	клитика	>		частица
обязательное кодирование			>			опциональное кодирование	

←──→

высокая степень грамматикализации   *низкая степень грамматикализации*
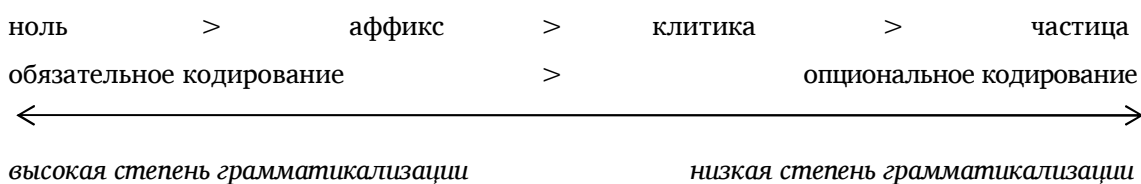
Рисунок 2. Шкала грамматикализованности

Помимо степени морфологической автономности показателей в определении степени грамматикализации большую роль играет облигаторизация [Lehmann 2015: 148]. Нулевой показатель может постулироваться только в облигаторизованных системах. По этой причине он помещен в начало иерархии как наиболее грамматикализованный.

Степень морфологической автономности показателя и облигаторизация — два отдельных, но коррелирующих параметра. В большинстве облигаторизованных систем эвиденциальность кодируется кумулятивно со временем. В двух языках — бора (бора-витотские, Южная Америка, [Thiesen 1996: 97]) и хуп [Epps 2005: 623] зафиксированы облигаторизованные системы, кодирующие источник информации клитиками. Ни в одном языке выборки не обнаружена облигаторизованная система, в которой эвиденциальность кодируется с помощью частиц.

Ожидается, что показатели, выражающие прямую засвидетельствованность, являются менее морфологически автономными и чаще встречаются в облигаторизованных системах, то есть более грамматикализованы[11].

В работе А. Ю. Айхенвальд [2004: 73] описывается близкая идея: если в системе ноль кодирует одно из значений, вероятнее всего, он кодирует

---

[11] Данное предположение основывается на идее, что прототипические для реального мира ситуации выражаются, как правило, наименее формально маркированными средствами [Comrie 1986]. Ситуацию, в которой говорящий получил информацию об описываемом событии через посредника или путем логических умозаключений, можно назвать более маркированной в терминах свойств реального мира относительно ситуации, в которой говорящий получил информацию прямым образом.

либо зрительную, либо одновременно зрительную и сенсорную засвидетельствованность. Мы предполагаем, что корреляция между степенью грамматикализации показателя и его семантикой не ограничивается этим утверждением.

На рисунке 3 представлена ожидаемая зависимость семантики показателя от степени его грамматикализации.

визуальный > сенсорный > инферентивный > презумптивный > репортативный > квотативный

$\longleftarrow \longrightarrow$

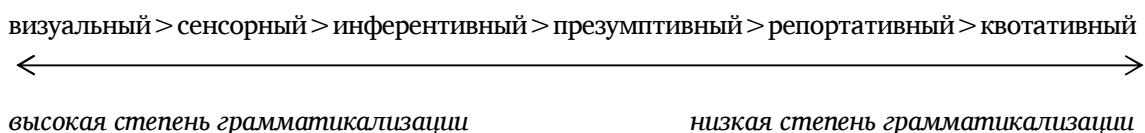*высокая степень грамматикализации*   *низкая степень грамматикализации*

Рисунок 3. Зависимость семантики показателя от степени грамматикализации

Данная гипотеза не предполагает, что визуальное значение не может кодироваться частицами, а репортативное и квотативное — аффиксами, а только предсказывает, что такой способ кодирования будет менее вероятным. Это подтверждается данными: почти половина визуальных и сенсорных маркеров — аффиксы; репортативное и квотативное значения, напротив, чаще всего выражаются частицами. Полное соотношение эвиденциальных значений и способов их выражений представлено в таблице 2.

Таблица 2. Эвиденциальные значения и показатели, которыми они выражаются[12]

показатели	визуальный и сенсорный	инферентивный и презумптивный	репортативный и квотативный
аффикс	49% (32)	42% (32)	27% (23)
клитика	27% (18)	22% (17)	21% (18)
частица	24% (16)	36% (28)	52% (45)
всего	**66**	**77**	**86**

Также на основе имеющихся данных постулируется фреквенталия (11), нарушение которой зафиксировано только в одном языке выборки.

(11) В одной системе значение более непосредственного доступа к информации не может кодироваться показателями с большей степенью автономности, чем в этой же системе кодируются значения с низкой степенью засвидетельствованности.

---

[12] Значение критерия χ^2 составляет 13,505; критическое значение χ^2 при уровне значимости p = 0,01 составляет 13,277. Связь между факторным и результативным признаками статистически значима при уровне значимости p < 0,01.

Согласно фреквенталии (11) если, к примеру, репортативное значение в некой системе кодируется аффиксом, то все значения, находящиеся на шкале левее него (визуальное, сенсорное, инферентивное и презумптивное), могут кодироваться только аффиксами или нулем, в то время как все значения правее (квотативное) могут кодироваться любым способом. Напротив, если визуальное значение в системе кодируется частицей, все остальные значения также будут кодироваться частицами.

Показатели с разной степенью грамматикализации нечасто встречаются в рамках одной системы: например, эвиденциальная система языка кора состоит из клитики =ku со значением прямой засвидетельствованности и частиц *séin, nū'u* и *yée,* выражающих инферентивное, репортативное и квотативное значение соответственно [Thornes 2018: 426], что соответствует нашей гипотезе.

Правило (11) справедливо не только для каждой отдельной системы, но и для всех систем, принадлежащих одному языку. Как упоминалось в разделе 3.1, аляскинский юпик использует две системы для кодирования способа получения информации. Инферентивное значение (более непосредственный доступ) выражается с помощью аффикса, а репортативное (менее непосредственный доступ) — клитикой (6). Язык дияри кодирует значение прямой засвидетельствованности глагольным аффиксом *-ku,* а презумптивное и репортативное — частицами *kara* и *pinthi* соответственно [Austin 1981: 180].

Зафиксирован по крайней мере один язык, нарушающий фреквенталию (11). В языке гитксан инферентивное значение выражается частицей *'nakw* (12a), а репортативное — клитикой =*kat* (12b). На основании имеющихся данных невозможно однозначно утверждать, находятся ли показатели в парадигматических отношениях. Но, так или иначе, в языке гитксан значение с менее высокой степенью засвидетельствованности выражается более грамматикализованным показателем.

(12) гитксан, цимшианские [Peterson 2018: 467]

    a. ***'nakw*** = *hl*   *ta'a-(t)* = *s*   *john*
       INFR = CND   дома-3 = PND   Джон
       'Похоже, Джон дома.'

    b. *siipxw* = ***ka****t* = *t*   *john*
       больной = REP = PND   Джон
       'Я слышал, что Джон болен.'

5.2. Во временных системах

Для систем, в которых кумулятивно кодируются эвиденциальность и время, облигаторизация является единственным параметром. В наименее грамматикализованных системах кодирование эвиденциальности опционально, как в гунзибском [van der Berg 1995: 103] и турецком [Johanson 2003: 274]. Языки с обязательным выражением эвиденциальности в рамках временной парадигмы, напротив, обладают высокой степенью грамматикализации. Среди них — тариана [Aikhenvald 2003: 133] и болгарский [Smirnova 2011: 279].

Анализ выборки показал, что значения прямой засвидетельствованности чаще всего возможно кодировать только в облигаторизованных системах. Эта зависимость представлена в таблице 3.

Таблица 3. Выражение прямой засвидетельствованности во временных системах[13]

системы	прямая и косвенная	только косвенная
облигаторизованные	82% (14)	0% (0)
необлигаторизованные	18% (3)	100% (6)
всего	**17**	**6**

Как уже упоминалось ранее, А. Ю. Айхенвальд [2004: 73] выдвигается следующая идея: если в системе ноль кодирует одно из значений, вероятнее всего, он кодирует прямую засвидетельствованность. Это полностью подтверждается данными и справедливо как для собственно эвиденциальных систем, так и для временных систем с эвиденциальными расширениями. Индиректив в арчинском (нахско-дагестанские) кодируется с помощью суффикса -*li*, являющимся частью временной парадигмы [Кибрик, 1977: 87]. Для выражения прямой засвидетельствованности используется немаркированная форма. Похожая система описана в болгарском [Smirnova 2011: 279].

6. Заключение

Сравнение более ста систем показало, что языки мира используют самые разнообразные грамматические средства для выражения способа получения информации. Системы, кодирующие эвиденциальность, различаются по своей сложности, степени грамматикализации и набору значений.

---

[13] Значение критерия χ^2 составляет 12,627; критическое значение χ^2 при уровне значимости p = 0,01 составляет 6,635. Связь между факторным и результативным признаками статистически значима при уровне значимости p < 0,01.

Анализ различных систем продемонстрировал, что эвиденциальные значения выстраиваются в иерархию в зависимости от доступа к информации, а степень грамматикализации эвиденциальных систем коррелирует со значениями, представленными в них. Так, показатели, выражающие прямую засвидетельствованность, встречаются в более грамматикализованных системах. Значения с более непосредственным доступом к информации склонны выражаться показателями с меньшей морфологической автономностью.

Список условных сокращений

1, 2, 3 — 1, 2, 3 лицо; ABS — абсолютивный падеж; AV — активный залог; AUX — вспомогательный глагол; CND — определитель нарицательных имен; COM — комитатив; CONJ — конъюнкт; DIR — прямая засвидетельствованность; DISJ — дизъюнкт; ERG — эргативный падеж; F — женский род; INAN — неодушевленный; INDF — неопределенный; INDIR — косвенная засвидетельствованность (индиректив); INFR — инферентив; M — мужской род; NEG — отрицание; NON.INF — не инферентив; NOM — именительный падеж; NVIS — информация, полученная не зрительно; OBJ — объект; OBL — косвенный падеж; PART — причастное наклонение; PFV — перфективный аспект; PL — множественное число; PND — определитель собственных имен; POSS — посессив; PP — комплетивное причастие; PST — прошедшее время; QUOT — квотатив; REP — репортатив; SG — единственное число; SSSI — спонтанное событие, тот же субъект, непереходная клауза; VIS — зрительно полученная информация.

Литература

Кибрик 1977 — Кибрик А.Е. Опыт структурного описания арчинского языка. Том III. Динамическая грамматика. М.: Московский государственный университет, 1977. [Kibrik A.E. Opyt strukturnogo opisaniya archinskogo yazyka. Tom III. Dinamicheskaya grammatika. [Structural approach to Archi language. Volume III. Dynamic grammar]. Moscow: Moscow State University, 1977.]

Плунгян 2011 — Плунгян В.А. Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира. М.: РГГУ, 2011. [Plungian V.A. Vvedenie v grammaticheskuyu semantiku: grammaticheskie znacheniya i grammaticheskie sistemy yazykov mira [The introduction into grammatical semantics: Grammatical meanings and grammatical systems of the world's languages]. Moscow: Russian State University for the Humanities, 2011]

Aikhenvald 2003 — Aikhenvald A.Yu. Evidentiality in Tariana. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 131–164.

Aikhenvald 2004 — Aikhenvald A.Yu. Evidentiality. Oxford: Oxford University Press, 2004.

Austin 1981 — Austin P.K. A grammar of Diyari, South Australia. Cambridge, Cambridge University Press, 1981.

Barnes 1984 — Barnes J. Evidentials in the Tuyuca verb. International Journal of American Linguistics. 1984. Vol. 50. No. 3. P. 255–271.

Bashir 1988 — Bashir E. Inferentiality in Kalasha and Khowar. Proceedings of the 24th regional meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society, 1988. P. 47–59.

Brosig, Skribnik 2018 — Brosig B., Skribnik E. Evidentiality in Mongolic. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 555–580.

Charney 1993 — Charney J.O. A grammar of Comanche. Lincoln: University of Nebraska Press, 1993.

Chirikba 2003 — Chirikba V. Evidential category and evidential strategy in Abkhaz. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 243–272.

Comrie 1986 — Comrie B. Markedness, grammar, people, and the world. Markedness. Eckman F., Moravcsik E., Wirth J. (eds.). New York: Springer Sciense + Business Media, 1986. P. 85–106.

de Haan — de Haan F. Coding of evidentiality. The World Atlas of Language Structures. Haspelmath M., Dryer M.S., Gil D., Comrie B. (eds.). Leipzig: Max Planc Institute for Evolutionary Anthropology, 2005.

Eberhard 2018 — Eberhard D. Evidentiality in Nambikwara languages. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 333–357.

Epps 2005 — Epps P. Areal diffusion and the development of evidentiality. Evidence from Hup. Studies in Language. 2005. Vol. 29. P. 617–50.

Forker 2018 — Forker D. Evidentiality in Nakh-Daghestanian languages. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 490–510.

Fortescue 2003 — Fortescue M. Evidentiality in West Greenlandic: A case of scattered coding. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 291–306.

Franklin 1964 — Franklin K.J. Kewa verb morphology. Verb studies in five New Guinea languages. Pence A. (ed.). OK: Summer Institute of Linguistics, University of Oklahoma, 1964. P. 100–130.

Hintz 2007 — Hintz D.M. Past tense forms and their functions in South Conchucos Quechua: time, evidentiality, discourse structure, and affect. Ph.D. thesis. University of California, 2007.

Johanson 2003 — Johanson L. Evidentiality in Turkic. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 273–290.

Korta, Zubeldia 2014 — Korta K., Zubeldia L. The contribution of evidentials to utterance content: Evidence from the Basque reportative particle omen. Language. 2014. Vol. 90. P. 389–423.

Krawczyk 2012 — Krawczyk E.A. Inferred Propositions and the expression of the evidence relation in natural language. Evidentiality in Central Alaskan Yup'ik Eskimo and English. Ph.D. thesis. Georgetown University, 2012.

Lehmann 2015 (1982) — Lehmann C. Thoughts on grammaticalization. 3rd edition 2015. Classics in Linguistics 1. Berlin: Language Science Press, 2015.

Maslova 2003 — Maslova E. Evidentiality in Yukaghir. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 219–236.

McLendon 2003 — McLendon S. Evidentials in Eastern Pomo with a comparative survey of the category in other Pomoan languages. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 101–130.

Nichols 2011 — Nichols J. Ingush grammar. University of California, Berkley, 2011.

Pan 2018 — Pan C.J. Evidentiality in Formosan Languages. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 657–674.

Peterson 2018 — Peterson T. Evidentiality and epistemic modality in Gitksan. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 463–489.

Reyburn 1954 — Reyburn W.D. Cherokee verb morphology: III. International Journal of American Linguistics. 1954. Vol. 20. P. 44–64.

Sarvasy 2018 — Sarvasy H. Evidentiality in the languages of New Guinea. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 630–657.

Skribnik, Kehayov 2018 — Skribnik E., Kehayov P. Evidentials in Uralic languages. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 525–553.

Smirnova 2011 — Smirnova A. The meaning of the Bulgarian evidential and why it cannot express inferences about the future. Proceedings of the 21st Semantics and Linguistic Theory Conference. Los Angeles, California, USA, 2011. P. 275–294.

Storch 2018 — Storch A. Evidentiality and the expression of knowledge: An African perspective. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 610–628.

Thiesen 1996 — Thiesen W. Gramática del idioma Bora. Peru: Instituto Lingüístico de Verano, 1996.

Thornes 2018 — Thornes T. Evidentiality in the Uto-Aztecan languages. The Oxford handbook of evidentiality. Aikhenvald A.Yu. (ed.). Oxford: Oxford University Press, 2018. P. 409–430.

Valenzuela 2003 — Valenzuela P.M. Evidentiality in Shipibo-Konibo, with a comparative overview of the category in Panoan. Studies in Evidentiality. Aikhenvald A.Yu., Dixon R.M.W. (eds.). Amsterdam, Philadelphia: John Benjamins Publishing Company, 2003. P. 33–62.

van der Berg 1995 — van der Berg H. A grammar of Hunzib. München, Newcastle: Lincom Europa, 1995.

Приложение. Языковая выборка

Австралия

Дияри, маранганджи, нгиямбаа, питьянтьятьяра.

Австронезия

Канаканаву, паиван, тагальский.

Африка

Африкаанс, лега, шиллук.

Европа

Абхазский, акха, арабский диалект персидского залива, арчинский, баскский, болгарский, гунзибский, думи, ингушский, калашский, калмыцкий, кантонский, коми-зырянский, кховар, лезгинский, македонский, нганасанский, пуми, северный помо, тамильский, тундровый ненецкий, турецкий, финский, халхасский, чеченский, шведский, эвенкийский, южноэстонский, юкагирский, японский.

Папуа

Богайа, дани из нижней долины, кева, фасу, экари, эмеле.

Северная Америка

Алабама, аляскинский юпик, билокси, ваиваи, винту, восточный помо, гитксан, запад-ногренландский, имбабура кечуа, истаютланский миштекский, каваиису, кайова, кашая, команчский, кора, марикопа, ньюмийский миштекский, оджибве, серрано, тлауитольте-пекский михе, тонкава, уанка кечуа, чемиуэви, чироки, шаста, юго-восточный тепеуа, южный кончукос кечуа, ючи.

Южная Америка

Апалаи, аруако, барасано, бора, варао, гарифуна, гуажажара, джаравара, лаконде, мапу-дунгун, ретуара, тариана, трумай, туюка, хуп, эпена педе, южный намбиквара.

Анна Евгеньевна Постникова

Национальный исследовательский университет «Высшая школа экономики»

Anna Postnikova

National Research University Higher School of Economics

apostnikova@hse.ru

Типология морфосинтаксических параметров
том 3, выпуск 2

Сайт журнала:

http://tmp.sc/

Оригинал-макет: Кс. П. Семёнова