

## Факультативное *se* в албанских придаточных изъяснительных: опыт корпусного исследования\*

М. М. Макартцев

Институт славяноведения РАН, Москва (Россия);

Институт славистики, Университет им. Карла фон Осецкого, Ольденбург  
(Германия); maxim.makartsev@uni-oldenburg.de

**Аннотация.** В албанском языке формально различаются комплементы-факты (комплементатор *që* или *se*), комплементы действия (комплементатор *si*, омонимичный союзному слову *si* ‘как’) и комплементы потенциала (комплементатор (*që*) *të*). В данной статье рассматривается факультативное употребление албанского комплементатора факта *se* совместно с союзными словами (матричные предикаты: глагол речи *them* ‘сказать’, глагол психической деятельности *di* ‘знать’ и глагол прямого вопроса *pyes* ‘спрашивать’; союзные слова: *çfarë* ‘что’, *ku* ‘где, куда’, *kur* ‘когда’, *kush* ‘кто’, *sa* ‘сколько’, *si* ‘как’). Использование комплементатора вместе с союзными словами позволяет говорить о дублированном маркировании синтаксической связи между главным предложением и комплементом, что трактуется как унифицирование оформления албанских комплементов-фактов. При этом если комплементы потенциала четко противопоставляются комплементам-фактам, то комплементы-действия дифференцируются от фактов недостаточно: в частности, с ними также может использоваться факультативный комплементатор факта *se*. Статистический анализ данных из Албанского национального корпуса показывает ряд тенденций в использовании комплементатора *se* в албанских придаточных изъяснительных, вводимых при помощи союзных слов. На использование комплементатора может влиять ближний контекст (сочетание матричного глагола и союзного слова) — например, положительно на него влияют *di* (*se*) *çfarë* и *di* (*se*) *kur*, а отрицательно — *di* (*se*) *si*. В диахроническом плане (при всех ограничениях, налагаемых непропорциональной представленностью в АНК текстов до 1989 г.)

---

\* Исследование проведено в рамках проекта «Корпусные и полевые исследования синтаксических дополнений в балканских языках и диалектах» (грант РФФИ № 18-012-00607). Выражаю глубокую благодарность Платору Гаши (Университет Приштины) за помощь в обработке материалов Албанского национального корпуса (см. далее), Олесе Палинской, Екатерине Даниловой и Иштвану Фекете (Институт славистики, Университет им. Карла фон Осецкого, Ольденбург) за консультацию по статистическому анализу данных.

значительных изменений не прослеживается. Показаны различия между регионами (наиболее похожи Албания и Косово; менее — Албания и Северная Македония; меньше всего — Косово и Северная Македония). Учет смешанных эффектов (дерево решений) позволяет также говорить о влиянии жанра текстов. Кроме того, при квазипрономинальном употреблении фразеологизмов с глаголом *di* ‘знать’ факультативное *se* используется значительно реже, чем после свободных сочетаний с этим же глаголом.

**Ключевые слова:** синтаксис албанского языка, албанский языковой корпус, комплементация, придаточные изъяснительные, языковая вариация и изменения языка.

## The facultative *se* in Albanian complement clauses: A corpus study

**M. M. Makartsev**

Institute for Slavic Studies, Russian Academy of Sciences, Moscow (Russia);  
Institute of Slavic Studies, Carl von Ossietzky University of Oldenburg (Germany);  
maxim.makartsev@uni-oldenburg.de

**Abstract.** Albanian formally differentiates between the following types of complements: Facts (complementizer *që* or *se* ‘that’), Activities (*si* ‘how’), and Potential (*(që) të* ‘(in order) to’). The article focuses on the facultative use of the Albanian complementizer of fact *se* in contexts with connective words (*çfarë* ‘that’, *ku* ‘where’, *kur* ‘when’, *kush* ‘who’, *sa* ‘how much; how many’, *si* ‘how’) after matrix predicates of non-visual access to information (a verb of speech *them* ‘say’, psychic activity *di* ‘know’, and a verb introducing direct questions *pyes* ‘ask’). Since the complementizer *se* in these contexts can be used together with connective words, it allows us to speak of double marking of the syntactic connection between the matrix clause and the complement or of unification of Albanian complements of Fact. While Fact-type complements in Albanian are clearly opposed to Potential-type complements, they are not fully differentiated from Activities. In particular, both Facts and Activities can be introduced by the facultative *se*. A statistical analysis of the Albanian National Corpus demonstrates several tendencies in the use of the facultative *se* with complements including connective words. One of the factors influencing the use of *se* is the closer context. Thus, *di (se) çfarë* and *di (se) kur* support while *di (se) si* suppresses the use of *se*. Diachronically (however limited a diachronic analysis on the basis of ANC can be, considering the disproportional representation of texts published before 1989), no significant developments were found. The paper demonstrates the differences between regions. The most similar in the use of the facultative *se* are Albania and Kosovo, while the most dissimilar are Kosovo and North

Macedonia. My mixed effects (condition tree) analysis shows text genres to be another factor in its distribution. Notably, in quasipronominal uses of phraseological units with the verb *di* 'know', the facultative *se* is much less frequent than after free word combinations with the same verb.

**Keywords:** Albanian syntax, Albanian language corpus, complementation, complement clauses, language variation and change.

## 1. Введение

В исследованиях по албанскому синтаксису, сделанных из албанской национальной лингвистической перспективы, виды предложений, обычно относимых в общетипологических исследованиях к компонентам (сентенциальным актантам), рассматриваются как частный случай придаточных предложений, внутренняя классификация которых проводится на основе не синтаксического, а семантического критерия. Например, так выглядит фрагмент оглавления грамматики албанского языка [Domí et al. 2002]: «Сложные предложения с зависимыми предложениями в функции подлежащего», «...сказуемого», «...дополнения», «...определения», «...обстоятельства места», «...обстоятельства времени»; «...с зависимым предложением с модальным значением», «...со значением цели», «...причины», «...следствия», «...условия», «...уступки», «...сравнения»<sup>1</sup>. Явление, которое в общетипологических исследованиях обычно описывается как комплементация («предикация в позиции аргумента»<sup>2</sup> [Horie, Comrie 2000: 1]), в албанской грамматике рассматривается как несколько подвидов «сложных

---

<sup>1</sup> «Periudha me fjali të varur kryefjalore», «...kallëzuesore», «...kundrinore», «...përcaktore», «...vendore», «...kohore», «...mënyrore», «...qëllimore», «...shkakore», «...trjedhimore», «...kushtore», «...lejore», «...krahasore». Рассматриваемые в настоящей статье работы на албанском языке терминологически и концептуально сильно отличаются как от общетипологических работ, так и от современных исследований по русистике. Поэтому при переводах с албанского языка я сохраняю внутреннюю форму терминов, привожу их также в оригинале и, если это не очевидно из предыдущего изложения, указываю, какое именно грамматическое явление имеется в виду.

<sup>2</sup> «Predication manifested in argument slots» (Здесь и далее перевод мой; курсив оригинала сохраняется. — М. М.).

предложений с придаточным дополнительным» («*periudha me fjali të varur kundrinore*»).

В то же время в общетипологических исследованиях последних десятилетий комплементации отводится важная роль на уровне таксономии. Она рассматривается как особый тип синтаксической связи в сложном предложении — наравне с координацией (*coordination*), придаточными, относящимися к одной из составляющих предложения и не имеющими обязательного характера (*relative clauses*), и другими типами (см. [Dixon 2006; Shopen 2007], см. также обсуждение и библиографию вопроса о месте комплементации среди прочих типов синтаксической связи [Schmidtke-Bode 2004: 21–26]).

Р. Диксон [Dixon 2006: 23–26] отмечает, что самыми распространенными типами придаточных-комплементов являются следующие: факт («*Fact*»), деятельность («*Activity*»)<sup>3</sup>, потенциал («*Potential*»). Эти типы выделяются на основе семантики комплемента, однако представляется возможным указать некоторые типологические тенденции в их структурном оформлении: референция по времени у **комплементов-фактов** обычно независима от референции главного предложения, как правило, такие придаточные вводятся при помощи специального союза-комплементатора (наподобие английского *that*); **комплементы деятельности** структурно часто похожи на именную фразу, возможны ограничения по времени, аспекту, модальности или возможности негирования, референция по времени может отличаться от референции главного предложения; **комплементы потенциала** меньше похожи на предложение, чем комплементы-факты, и меньше похожи на именную фразу, чем комплементы деятельности, обычно имеют ограничения по времени, аспекту и т. д., они имеют ту же референцию по времени, что и главное предложение, а глагол обычно стоит в особой форме (например, в инфинитиве).

На первый взгляд, в албанском структурно разграничиваются все три основных типа комплементов: факты (1), комплементы деятельности (2), комплементы потенциала (3), и для каждого из них можно было бы указать свой тип комплементатора.

---

<sup>3</sup> В [Hansen et al. 2016] для соответствующих русского и болгарского типов комплементов (вводимых при помощи рус. *как* / болг. *как* ‘то же’) используется более удачная формулировка *immediate perception* ‘непосредственное восприятие’, которую можно было бы ввести в обозначение таких комплементов: комплементы непосредственного восприятия. Но для простоты изложения в настоящей работе я буду придерживаться классической схемы Р. Диксона.

**Комплементы-факты (COMP.F), комплементатор *se* ‘что’:**

- (1) *Kryeparlamentari Krasniqi tha se*  
 спикер.парламента.NOM.SG.DEF Красники сказать.AOR.3SG COMP.F  
*me kryeministrin Thaçi komunikon*  
 с премьер.министр.ACC.SG.DEF Тачи общаться.PRS.3SG  
*vetëm zyrtarisht.*  
 только официально

‘Спикер парламента Красники рассказал, что с премьер-министром Тачи он общается только по официальным каналам’. [АНК: Zëri. 2013]<sup>4</sup>

**Комплементы деятельности (COMP.A), комплементатор *si* ‘как’:**

- (2) *Ai e= pa si shkëputi*  
 м.3SG.NOM 3SG.ACC= видеть.AOR.3SG COMP.A сорвать.AOR.3SG  
*nga gjerdani bombën e*  
 от гранатный.подсумок.NOM.SG.DEF бомба.ACC.SG.DEF и  
*si u= vërsul përjetë me*  
 COMP.A НАСТ= броситься.AOR.3SG вверх.по.склону с  
*grushtin lartë.*  
 кулак.ACC.SG.DEF навверх

‘Он увидел, как тот выхватил гранату из подсумка и как бросился вверх по склону, [сжимая ее] в кулаке над головой’. [Domi et al. 2002: 517]

**Комплементы потенциала (COMP.P), комплементатор (*që*) *të* ‘чтобы’:**

- (3) *Në këshillin e luftës Gjergj*  
 в совет.ACC.SG.DEF AGR война.GEN.SG.DEF Гергь.NOM.SG.INDF  
*Kastrioti propozoi që*  
 Кастриоти.NOM.SG.DEF предложить.AOR.3SG COMP.F  
*t’u binin turqve natën.*  
 COMP.P:3PL.DAT ударить.IPF.3PL турки.DAT.PL ночь.ACC.SG.DEF

‘На военном совете Гергь Кастриоти предложил напасть на турок ночью (досл. «чтобы они напали»)’. [Domi et al. 2002: 518–519]

<sup>4</sup> Если не указано иное, языковые примеры в настоящей статье взяты из Албанского национального корпуса, версия 2011–2016 гг. [АНК 2011–2016] (см. далее).

В таком виде описание албанской комплементации, пусть и с оговорками, соответствовало бы общетипологической схеме, предложенной Р. Диксоном (конечно, комплементы деятельности в албанском никак не похожи на именную фразу, но, с другой стороны, комплементы потенциала, как и следовало ожидать, показывают ограничения по аспекту и времени, глагол стоит в особой форме конъюнктива, и это вполне укладывается в схему Р. Диксона).

Однако в функционировании албанских комплементаторов довольно много неясного.

Так, *që* является избыточным средством маркирования, поскольку в предложениях типа (3) он легко может быть опущен и комплемент связывается с глаголом в главном предложении при помощи т. н. показателя балканского конъюнктива *të* (см. подробнее в статье [Морозова, настоящий сборник])<sup>5</sup>. Кроме того, *që* может использоваться и как комплементатор факта (в предложениях типа (1)).

В то же время комплементатор *se* неочевидным образом взаимодействует с другими средствами связи в албанских сложноподчиненных предложениях. *Se* не только может употребляться с союзными словами (см. далее), но и с *si*, который по крайней мере в отдельных контекстах семантически следовало бы считать комплементатором действия. Семантическая основа для этого варьирования не установлена (см. далее).

Корпусный анализ позволяет рассмотреть вопрос о возможной структурной обусловленности отсутствия / наличия комплементатора *se*. В настоящей статье я рассмотрю основные тенденции в структурной вариативности *se*, которые можно подтвердить статистически. Для этого я провел пилотное корпусное исследование на основе Албанского национального корпуса [АНК 2018], рассмотрев связь факультативного *se* с ближайшим контекстом в диахроническом и ареальном аспекте.

---

<sup>5</sup> Аналогичный показатель с похожим набором функций используется в других балканских языках; В. Фридман предложил назвать его «дентальным модальным субординатором», *dental modal subordinator*, или DMS, поскольку такие показатели «начинаются на дентальный или альвеолярный согласный и используются для введения глагольных форм в модальных конструкциях» [Friedman 1986: 381]. См. подробнее в [Joseph 2016; Manzini, Savoia 2018]. В последней работе описание балканской комплементации ограничивается комплементами потенциала — т. е. авторы имплицитно исходят из узкого толкования комплементации.

Таким образом, настоящая статья является опытом корпусного изучения албанского комплементатора *se* и преследует две основных цели. С одной стороны — изучение синтаксиса албанского сложного предложения. С другой — это попытка показать, насколько албанский комплементатор факта (не) вписывается в общетипологические ожидания.

Далее я рассмотрю проблематику факультативности *se* при совместном использовании с союзными словами (*Раздел 2*) и факторы, которые, возможно, влияют на использование *se* в описанных выше контекстах. Сначала я рассмотрю эти факторы как независимые переменные (*Раздел 3*), а потом опишу возможные смешанные эффекты (*Раздел 4*). *Разделы 5 и 6* посвящены двум частным вопросам, связанным с факультативным *se*: его взаимодействию с комплементатором действия *si* ‘как’ и проблеме фразеологизмов в функции матричного предиката. Таблицы и графики вынесены в *Приложения* к статье.

## 2. Факультативное *se* с союзными словами в грамматических описаниях

В современном албанском языке есть два основных способа оформления подчинительной связи<sup>6</sup> между главным предложением и придаточным изъяснительным (complement clauses, «Objektsätze im engeren Sinn» [Buchholz, Fiedler 1987: 513–515; «periudha me fjali të varur kundrinore» [Domi et al. 2002: 515–524; Rushiti 2015: 58–90])<sup>7</sup>:

- 1) при помощи союзов-комплементаторов *se*, *që*<sup>8</sup>, *si*<sup>9</sup>;

---

<sup>6</sup> Бессоюзные предложения в настоящей статье не рассматриваются.

<sup>7</sup> Р. Рушити приводит следующий список названий, под которыми описываемый тип упоминался в грамматических описаниях албанского языка: *fjali e caktimit* ‘определительное предложение’, *fjali e objektit* и *fjali objektore* ‘объектное предложение’, *fjali kundrinore ose objektore* ‘дополнительное или объектное предложение’, *fjali e nënrenditur kundrinore* ‘дополнительное придаточное предложение’ [Rushiti 2015: 58].

<sup>8</sup> Проблематика противопоставления *se* и *që* в настоящем исследовании рассматриваться не будет. Отмечу только, что в случае, который интересует меня, эта синонимия отсутствует. Когда комплементатор *që* употребляется перед *çfarë*, *kush*, *kur*, *ku*, *sa*, *si*, эти союзные слова обычно входят в относительное придаточное с опущенной вершиной в составе комплемента, соответственно, здесь нельзя

- 2) при помощи союзных слов (*çfarë* ‘что’, *ku* ‘где, куда’, *kur* ‘когда’, *kush* ‘кто’, *sa* ‘сколько’, *si* ‘как’ и др.)<sup>10</sup>.

Союз-комплементатор *se* может также факультативно ставиться в начале придаточного перед союзным словом во втором случае. Обычно это трактуется как плеонастическое использование («das pleonastische *se*» ‘плеонастическое *se*’ [Buchholz, Fiedler 1987: 514], или «*mjet plotësues pleonastik*» ‘дополнительное плеонастическое средство’ [Domí et al. 2002: 520]), поскольку союзные слова выполняют двойную функцию: они одновременно выступают и как союзы-комплементаторы, и как один из членов изъяснительного придаточного<sup>11</sup>. Таким образом, плеонастичность факультативного *se* можно было бы усматривать в том, что оно эксплицирует уже выраженную союзным словом синтаксическую связь придаточного с главным.

Однако это эксплицирование приводит к унификации оформления изъяснительных придаточных, что, строго говоря, упрощает понимание текста реципиентом. Это вполне можно трактовать как самостоятельную функцию, поэтому вряд ли уместно говорить о плеонастичности *se* в предложениях такого типа. Литература о факультативном *se*

---

говорить о совместном употреблении *që* и одного из союзных слов для введения комплемента:

*Të gjithë e dimë*

[*që*      [*ku*    *është*            *ai*]  
 COMP.F    где    быть.PRS.3SG    M.3SG.NOM

[*ka festë*], [*ka gëzim*], [*ka njerëz [që brohorasin e që këndojnë bashkë me atë]*].  
 ‘Все мы знаем, [что там [где он], там [праздник], там [радость], там [люди, которые аплодируют и поют вместе с ним]]’. [АНК: Koha.mk. 2012]

Такие случаи я не рассматриваю.

<sup>9</sup> В албанской грамматике это слово толкуется и как союз, и как союзное слово. Поэтому я включил его для простоты изложения в обе группы, принимая в качестве исходного пункта описание в [Domí et al. 2002: 515–524]. Ниже этот вопрос обсуждается более подробно.

<sup>10</sup> Грамматика [Domí et al. 2002: 515–524] предлагает опыт более подробной классификации, которая нерелевантна для конструкции, рассматриваемой в этой статье.

<sup>11</sup> Прочитую здесь соответствующий раздел албанской грамматики, к которому нам предстоит еще вернуться: «[союзные слова *çfarë*, *ku*, *kur*, *kush*, *sa*, *si* и другие] одновременно выступают и как средство связи (они функционируют как союзы), и как член зависимого предложения» [Domí et al. 2002: 520].

достаточно легко обозрима: помимо двух изданий — албанской грамматики [Domi et al. 2002] и немецкоязычной грамматики албанского языка [Buchholz, Fiedler 1987], основной не-албаноязычной работы, откуда можно получить сведения о грамматической структуре албанского языка — есть также монография Р. Рушити «Союзные предложения в албанском языке (связь сказуемых составляющих с союзами *që* и *se*)» [Rushiti 2015], которая содержит опыт описания албанских придаточных предложений с союзами *që* и *se*.

Примеры (4) и (5) иллюстрируют факультативность *se* перед союзным словом *kush* ‘кто’ в двух придаточных изъяснительных предложениях, максимально близких структурно:

- (4) *Ministri Hamza tregon kush*  
 министр.NOM.SG.DEF Хамза рассказывать.PRS.3SG кто.NOM  
*do të përfitojë nga rritja*  
 FUT COMP.P получить.выгоду.SBJV.PRS.3SG от рост.NOM.SG.DEF  
*e pagave.*  
 AGR зарплата.GEN.PL

‘Министр Хамза рассказывает, кому выгоден рост зарплат’. [Bota Sot. 2017]

- (5) *Ragip Behrami tregon se kush*  
 Рагип.NOM.SG.INDF Бехрами рассказывать.PRS.3SG COMP кто.NOM  
*do ta shoqërojë Valonin*  
 FUT COMP.P:3SG.ACC сопровождать.SBJV.PRS.3SG Валён.ACC.SG.DEF  
*në botëror.*  
 на чемпионат.мира.ACC.SG.INDF

‘Рагип Бехрами рассказывает, [что] кто поедет с Валёном на чемпионат мира’. [Telegrafi.com. 2018]

Отмечу, что Р. Рушити, которая посвятила значительную часть своего исследования разбору случаев с отсутствующим комплементатором (прежде всего *që*, вопрос об отсутствии *se* в изъяснительном придаточном поднимается, но ему уделяется несопоставимо меньше внимания), считает, что союзные слова не выражают связь между главным предложением и изъяснительным придаточным, и единственным носителем этой связи является комплементатор [Rushiti 2015: 109–114]. Для интерпретации случаев типа (4) Р. Рушити вводит понятие *lidhëza zero se / që* ‘нулевой союз *se / që*’. Иными словами, подчинительную связь между главным и изъяснительным придаточным, по ее мнению, всегда

оформляет комплементатор, который может быть выраженным или иметь нулевую форму. Такая интерпретация удачно соответствовала бы концепции Р. Диксона — это позволило бы считать, что албанские комплементы-факты вводятся единообразным способом (при помощи комплементатора, который может иметь нулевую форму), т. е. показывают структурное единство.

Наличие / отсутствию комплементатора *se* можно дать и более мягкую интерпретацию, связанную с общетипологической концептуальной рамкой, но не вводящую идею «нулевого союза», для которой Р. Рушити не приводит достаточных оснований<sup>12</sup>. По моему мнению, присутствие или возможность употребления *se* для введения комплементов-фактов (независимо от наличия других средств синтаксической связи, т. е. союзных слов) оправданы тем, что комплементы-факты можно трактовать на семантическом уровне как некий единый класс (четко противопоставленный по ряду параметров, скажем, комплементам потенциала). При этом *se* можно интерпретировать как факультативное средство подчеркивания этой синтаксической связи. В мягкой формулировке факультативность этого средства не исключает его корреляции, или статистической взаимосвязи без жесткого причинно-следственного характера, с некоторыми контекстуальными параметрами (структурными или семантическими). В этом случае факультативное *se* можно считать одним из второстепенных — но не обязательных — признаков албанских комплементов-фактов.

---

<sup>12</sup> Аргументом в пользу выделения нулевого комплементатора могло бы служить его «значащее отсутствие», т. е. существование примеров, которые бы меняли значение в зависимости от наличия / отсутствия комплементатора *se*. Так, по мнению Р. Рушити, вариация в использовании *se* зависит от актуального членения предложения. Если изъяснительное придаточное в постпозиции, то вариант с *se*, по ее мнению, выражает тему, а вариант без него — ремю. К сожалению, она не рассматривает параллельно структурно идентичные предложения наподобие (4) и (5), которые бы различались только наличием / отсутствием *se*, поэтому остается непонятным, на чем именно основано ее утверждение (к которому она возвращается несколько раз: [Rushiti 2015: 65, 75, 111]). В то же время наличие такого типа изоструктурных предложений, в которых статус придаточного не меняется (например, как в (4), так и в (5) оно является ремой), свидетельствует о том, что *se* нельзя считать средством выражения актуального членения.

### 3. Ближайший контекст и метаданные из АНК как независимые переменные

Поиск в Албанском национальном корпусе осуществлялся в конце 2018 — первой половине 2019 г., до его расширения и перехода на новую платформу (АНК 2019). Его параметры на тот момент описаны в [Морозова и др. 2016]. Эта версия АНК поддерживается на старой платформе до сих пор, но не пополняется. Она включает в себя литературные, публицистические, научные, официально-деловые и религиозные письменные тексты на стандартном албанском языке или в близких к нему формах общим объемом 19719338 словоупотреблений.

Область происхождения текстов — Албания, Косово или Северная Македония — к сожалению, нельзя учитывать при поиске (но ее, как правило, можно установить на основе места публикации текста для каждого отдельного примера, что и было сделано).

Я рассматриваю два периода — тексты, созданные до 1989 г. и с 1990 г. Этот рубеж установлен условно и связан с периодом затяжного общественно-политического кризиса в СФРЮ и НР Албании. Более дробное деление по периодам представляется нецелесообразным, поскольку это оставит слишком мало данных для каждого периода и статистический анализ не будет возможным (далее будет показано, что при комбинации некоторых параметров данных недостаточно даже при выбранном способе членения).

Жанр текстов, область происхождения (далее для простоты «регион») и период — независимые переменные, влияние которых на наличие / отсутствие факультативного *se* проверяется и интерпретируется в настоящей статье. В *Таблице 1* (см. *Приложение 1*) приводятся сведения об объеме подкорпусов по жанрам и по годам. Выбраны только те подкорпуса, поиск по которым дал придаточные изъяснительные с союзными словами (параметры поиска см. далее; первоначально поиск проводился по всему АНК, но в ряде подкорпусов не было найдено ни одного примера). В соответствующих ячейках таблицы указывается общее количество словоформ для подкорпусов и отношение к общему объему АНК. При анализе материала корпуса некоторые подкорпуса были объединены (см. столбец «Категории анализа»).

Данные подкорпусов очень неравномерны. В частности, для периода до 1989 г. пресса и религиозные тексты не представлены вовсе, а научные тексты представлены в минимальном объеме, из чего

следует, что диахронический анализ изменений в употреблении факультативного *se* по этим подкорпусам невозможен. Иными словами, в диахронии (периоды до 1989 г. и с 1990 г.) можно рассматривать только материал художественной литературы и научные тексты из Албании. Сопоставительное исследование по регионам возможно только на уровне синхронии (с 1990 г.).

Неравномерность данных по подкорпусам можно сгладить, если пересчитывать выдачу в количество словоформ на миллион (*ipm*). Отмечу, впрочем, что это не всегда необходимо и возможно. Во-первых, если я учитываю подкорпус как независимую переменную, меня интересует вероятность того или иного значения бинарной переменной «наличие / отсутствие факультативного *se*» внутри каждого подкорпуса исходя из общего количества примеров по этому подкорпусу. Существенно, что здесь я говорю о конечном количестве реальных примеров в корпусе и модель носит описательный характер. Во-вторых, пересчитать выдачу в *ipm* не всегда возможно технически, например, при анализе данных по регионам<sup>13</sup>. В-третьих, работа с *ipm* не всегда возможна при ряде статистических тестов, не только при построении дерева решений (Conditional Inference Tree, или *cree*, см. далее), но даже и при расчете  $\chi^2$ , поскольку в большинстве случаев при пересчете в *ipm* мои данные не достигают необходимых пороговых значений (см. далее). Кроме того, оба применяемых в настоящей статье теста ( $\chi^2$  и точный критерий Фишера) оперируют только целыми числами, соответственно, пересчет в *ipm* потребует округления, что будет снижать точность подсчетов. Поэтому далее, по возможности, я привожу данные как в *ipm*, так и в абсолютном количестве словоупотреблений, но все статистические расчеты делаю по реальным словоупотреблениям.

Ближайший контекст факультативного *se* включает в себя предикат в главном предложении, для которого изъяснительное придаточное выступает в роли дополнения (далее — глагол, управляющий компонентом) — левый контекст, а также союзное слово — правый контекст. Одним из условий моего поиска в корпусе было непосредственное соседство глагола, факультативного *se* (если есть) и союзного слова<sup>14</sup>.

<sup>13</sup> Хотя принципиальная возможность провести в АНК поиск по регионам есть, но он крайне трудозатратна и подразумевает ручной подбор авторов для каждого региона.

<sup>14</sup> Иными словами, поиск осуществлялся по следующим параметрам (приводятся леммы):  $(them \vee di \vee pyes) \wedge se \wedge (cfar\ddot{e} \vee ku \vee kur \vee kush \vee sa \vee si)$ ;  $(them \vee di \vee pyes) \wedge$

При отборе глаголов, управляющих компонентом, я опирался на грамматику О. Бухольц и В. Фидлера [Buchholz, Fiedler 1987: 513 и далее]<sup>15</sup>. С точки зрения семантики, я исходил из глаголов не-визуального доступа к информации, т. е. потенциально связанных с широкой зоной незасвидетельствованности. Таким образом было отобрано три семантических класса, для каждого из которых было выбрано по одному глаголу: глаголы **речи** (нем. Verben des Sagens: *them* ‘сказать’); глаголы **психической деятельности** (Verben der psychischen Tätigkeit: *di* ‘знать’) и глагол, вводящий косвенный **вопрос** (Verwendung in der indirekten Frage: *pyes* ‘спрашивать’). В поиске учитывались следующие союзные слова: *çfarë* ‘что’, *ku* ‘где, куда’, *kur* ‘когда’, *kush* ‘кто’, *sa* ‘сколько’, *si* ‘как’. Зависимой переменной является отсутствие / наличие комплементатора *se* (эта переменная бинарная), а независимыми переменными — помимо метаанных корпуса — глагол *them / di / pyes* и союзное слово *çfarë / kush / kur / ku / sa / si*<sup>16</sup>.

По этим критериям было найдено 1528 предложений без *se* и 1598 предложений с *se*. Такое практически равное (48,9 % против 51,1 %) соотношение представляет собой кумулятивный эффект от действия перечисленных переменных. Далее в тексте статьи я сначала описываю результаты статистического анализа этих переменных по отдельности. Затем я предлагаю модель описания, основанную на методе дерева решений (ctree). Большое преимущество этой модели — учет смешанных эффектов, т. е. совместного вклада нескольких независимых переменных в употребление факультативного *se*. Найденные в корпусе 3126 предложений были сведены в таблицу данных в SPSS<sup>17</sup>, и описа-

(*çfarë* ∨ *ku* ∨ *kur* ∨ *kush* ∨ *sa* ∨ *si*). Комплемент может также выноситься в позицию, предшествующую матричному глаголу, что обычно связано с эмфазой:

<i>Se</i>	<i>kush</i>	<i>i=</i>	<i>zotëron</i>	<i>këto</i>
COMP.F	кто.NOM	3PL.ACC=	владеть.PRS.3SG	эти.F.ACC.PL
<i>puse</i>	<i>ne</i>	<i>nuk</i>	<i>dimë</i>	<i>gjë</i> .
колодец.ACC.PL.INDF	1PL.NOM	NEG	знать.PRS.1PL	вещь.ACC.SG.INDF

‘Мы не знаем, кто владеет этими колодцами’. [Sot.com.al. 2017]

В настоящей статье такие случаи не рассматриваются.

<sup>15</sup> Аналогичный список можно эксцерпировать из примеров, приведенных в албанской грамматике [Domi et al. 2002: 515–524] и в монографии Р. Рушити [Rushiti 2015].

<sup>16</sup> Как было указано выше, *si* может выступать и как комплементатор и как союзное слово. Далее эта проблема будет разобрана на корпусном материале.

<sup>17</sup> Программа IBM SPSS Statistics, выпуск 26.0.0.0 (2020).

тельные статистики выполнялись на основе этого датасета. Для удобства чтения статьи таблицы и иллюстрации, отражающие ход анализа и содержащие статистические показатели для разных переменных, перенесены в *Приложения*.

Ближайший контекст для комплементатора — это матричный предикат (в настоящей статье один из глаголов *them* ‘сказать’, *di* ‘знать’ и *pyes* ‘спрашивать’) и одно из союзных слов (*çfarë* ‘что’, *kush* ‘кто’, *kur* ‘когда’, *ku* ‘где, куда’, *sa* ‘сколько’, *si* ‘как’). В этом разделе я проверяю, насколько похоже использование комплементатора в одинаковых контекстах в зависимости от таких параметров, как ареал и период. Как было указано выше, сравнение по регионам возможно только для текстов, опубликованных после 1990 г., а сравнение по периодам только для художественной литературы и научных текстов из Албании.

К сожалению, выборки неравнозначные: самым частотным является глагол *di*, в то время как *them* и *pyes* фиксируются в корпусе значительно реже (см. описательную статистику в *Таблице 2*); по союзным словам разброс не такой большой (*Таблице 3*).

Распределение комплементатора в зависимости от ближнего контекста представлено в *Таблице 4*. В столбцах АЛ, КС и СМ показаны абсолютные значения для Албании, Косова и Северной Македонии (здесь и далее цифра до косой черты — количество предложений для соответствующего матричного предиката и союзного слова, в которых отсутствует комплементатор *se*; цифра после косой черты — предложения, в которых используется комплементатор *se*). Это распределение сравнивается для каждой пары регионов в столбцах АЛ/КС, АЛ/СМ и КС/СМ с использованием точного текста Фишера (двустороннего) для выборок малого размера<sup>18</sup> и критерия  $\chi^2$  Пирсона для выборок большего размера. Основным параметром является р-критерий ( $p \leq 0,05$  — различие статистически значимо, такие случаи выделяются цветом;  $p > 0,05$  — различие статистически не значимо)<sup>19</sup>.

Данные, представленные в *Таблице 4*, показывают, что для глагола *them* с союзными словами нельзя говорить о региональной специфике (употребление факультативного *se* в этом контексте не показывает различий между регионами). Для глагола *pyes* региональные различия минимальны (комплементатор при матричном предикате *pyes* и союзном

<sup>18</sup> Если величина выборки  $n \leq 20$  или если значения в одной из ячеек  $n \leq 5$ .

<sup>19</sup> Степень свободы (degree of freedom) во всех случаях равна единице, поскольку наличие / отсутствие *se* является бинарной переменной.

слове *çfarë* значительно чаще используется в Албании, чем в Косово, но по остальным союзным словам никаких значимых различий не наблюдается). Возможно, это связано с относительно малым количеством предложений, найденных в корпусе. Глагол *di* показывает большее региональное разнообразие: стратегии использования комплементатора с этим глаголом наиболее похожи в Албании и Косово (они различаются всего по одному параметру), менее похожи в Албании и Северной Македонии (различия по трем союзным словам) и скорее различаются, чем похожи, стратегии в Косово и Северной Македонии (различия по четырем союзным словам).

При этом глагол *pyes* более или менее выраженно фаворизирует варианты с комплементатором во всех регионах, глагол *them* показывает некоторое разнообразие (*çfarë*, *kush*, *kur* при этом глаголе скорее фаворизируют вариант с комплементатором, для *ku* четких предпочтений не прослеживается, а *sa* и *si* ведут себя по-разному в разных регионах). *Di* показывает наибольшее разнообразие. При этом глаголе однозначно фаворизируют комплементатор *çfarë* и *kur*. *Si* фаворизирует вариант без комплементатора. Для *ku* варианты распределяются почти одинаковым образом (с тенденцией к предпочтению комплементатора в Северной Македонии). *Sa* и *kush* показывают разнонаправленные тенденции в разных регионах: *kush* поддерживает вариант с комплементатором в Албании и Северной Македонии и без комплементатора в Косово, а *sa* — вариант без комплементатора в Албании и Косово (в Северной Македонии количество вариантов достаточно близко к равному с тенденцией к фаворизации комплементатора).

В диахроническом плане ситуация почти не изменилась (Табл. 5)<sup>20</sup>. Только для двух контекстов можно говорить об изменениях: у *di (se) çfarë* и *pyes (se) si* значительно возросла частота вариантов с комплементатором.

#### 4. Учет смешанных эффектов при помощи дерева решений

Деревья решений (Conditional Inference Trees, или ctrees) позволяют учитывать сразу несколько независимых параметров и узнавать

---

<sup>20</sup> Напомню, что расчеты возможно провести только для научных и художественных текстов из Албании.

вклад каждого из них в итоговое распределение результатов. Этот метод был впервые опробован для лингвистического описания вариативности *was / were* в английском языке г. Йорка [Tagliamonte, Baayen 2012]. Я применяю его, следуя методологии, уточненной в статье [Hentschel et al. (в печати)]. Статистическая модель, лежащая в основе этого метода моделирования, обоснована в [Hothorn et al. 2006]. Это метод регрессивного анализа данных, основанный на классификации через бинарное рекурсивное деление [Levshina 2015: 291–300]: алгоритм выбирает ту независимую переменную, которая имеет наибольшую корреляцию с зависимой переменной, делит набор данных на два подмножества (обусловленное независимой переменной на заданном уровне р-критерия и не обусловленное на заданном уровне). Данные итерации повторяются для каждого подмножества, пока для всех конечных подмножеств не будет достигнут заданный уровень р-критерия.

При интерпретации результатов используется древовидный алгоритм: вклад каждого параметра соотносится с его иерархией в дереве (чем выше, тем больше вклад; параметр, находящийся на одной из ветвей, релевантен только при выполнении ближайшего параметра, находящегося выше по иерархии). Что важно, деревья решений позволяют работать с неполными данными (с отсутствующими значениями независимых переменных — например, когда примеру не сопоставлен период или ареал происхождения текста). Если какой-то параметр не появляется в дереве решений, это значит, что его релевантность для выбора зависимой переменной ниже, чем у отображенных параметров. Дерево решений также учитывает р-критерий (в моей модели я ввожу поправку Бонферрони для устранения эффекта множественных сравнений). Дерево решений для албанского факультативного *se* на материале АНК<sup>21</sup> (Рисунок 1) было построено средствами программной среды Rstudio 3.6.1 [Rstudio 2019] с использованием пакета *partykit* [Hothorn, Zeileis 2006]. Для построения модели было использовано 60 % данных. Затем прогностические свойства модели были протестированы на оставшихся 40 % при помощи матрицы ошибок (см. далее).

Модель интерпретируется следующим образом:

---

<sup>21</sup> Для того чтобы не перегружать модель, я использую только данные с 1990 г., поскольку в ином случае были бы внесены искажения из-за корреляции между ареалом (Албания) и периодом (до 1989 г.).

- 1) Стратегии выбора между наличием *se* (далее  $se^+$ ) и его отсутствием ( $se^-$ ) в албанском зависят прежде всего от жанра.
- 2) В литературных и религиозных текстах релевантен глагол *pyes* 'спрашивать', который фаворизует  $se^+$  (узел 3, *Рис. 1*), что согласуется с расчетом по индивидуальным переменным (см. *Раздел 3*).
- 3) Союзные слова *ku* 'где, куда', *sa* 'сколько', *si* 'как' при глаголах *them* 'сказать' или *di* 'знать' в литературных и религиозных текстах фаворизуют  $se^-$  (узел 6, *Рис. 1*), при этом регион оказывается не важен. Это дополняет наблюдения, полученные при анализе отдельных переменных из ближнего контекста, где жанровую специфику нельзя было учесть из-за неравномерности выборки.
- 4) В газетных и научных текстах значимым параметром является союзное слово. При этом *çfarë* и *kur* независимо от матричного глагола фаворизуют вариант с  $se^+$  (см. узел 8 на *Рис. 1*).

Алгоритм расчета показывает также, что для части групп (узел 5: союзные слова *çfarë*, *kush*, *kur* после глагола *them* или *di* в литературных и религиозных текстах; узел 9: союзные слова *kush*, *ku*, *sa*, *si* в газетах и научных текстах) выделить устойчивое предпочтение в (не)использовании факультативного *se* не представляется возможным.

Дерево решений показывает, что влияние ареала не является определяющим фактором при поиске самых устойчивых паттернов в использовании факультативного *se* (хотя в части контекстов это влияние просматривается очень хорошо, см. *Раздел 3*). Однако важен жанр текста, и для каждого жанра существует своя иерархия.

Эта модель была протестирована на остальных 40 % корпуса: с использованием пакета *caret* [Kuhn 2008, 2018] была рассчитана матрица ошибок (confusion matrix, см. *Рис. 2*). Матрица показывает верные и неверные предсказания (в 212 случаях верно предсказано  $se^-$ ; в 435 случаях верно предсказано  $se^+$ ; в 129 случаях неверно предсказано  $se^+$ ; в 253 случаях неверно предсказано  $se^-$ ). Матрица показывает, что дерево решений в два раза лучше предсказывает наличие компонента, чем его отсутствие. Точность предсказаний составляет 63 % (в интервале доверия [0,60; 0,66], р-критерий для сопоставления со случайным распределением по тесту МакНемара  $< 0,001$ ,  $\kappa \approx 0,23$ ). Отмечу, что на точности отрицательно сказывается большое количество примеров, попавших в узлы 5 и 9 дерева решений, для которых распределение примеров с факультативным *se* не отличается от случайного.



*dinim se si bëhet lufta.*  
 знать.IPF.3PL COMP как делать.NACT.PRS.3SG война.NOM.SG.DEF

‘К сожалению, мы не знали, как вести войну [досл. «как ведется война»]<sup>23</sup> [АНК: Rexhep Ferri. Njeriu, po kush tjetër. 1999–2001].

В качестве теста для различения *как*-комплементатора и *как* — союзного слова в статье [Hansen et al. 2016: 216] на болгарском материале предлагается модификатор «именно» (англ. *precisely*, болг. *точно*), поскольку «в конструкции с непрямым вопросом его [*как*, выступающего в качестве союзного слова — *М. М.*] семантическая функция такая же, как и в собственно вопросах: оно относится к „способу“, к „образу действия“, как именно происходила описываемая ситуация (...). В отличие от этого, в конструкции непосредственного восприятия *как* выступает в качестве комплементатора без компонента „способ осуществления“». Этот тест работает и в албанском языке: *si* в примере (2) нельзя модифицировать наречием *saktësisht* ‘точно, именно’ без значительного изменения значения (ср. (2a)), в то время как в примере (6) это вполне возможно (ср. (6a)). Таким образом, если в (2) *si* является союзным словом, то в (6) — комплементатором.

(2a) *Ai e= pa si saktësisht*  
 м.3SG.NOM 3SG.ACC= видеть.AOR.3SG COMP.A именно  
*shkëputi nga gjerdani bombën.*  
 сорвать.AOR.3SG от гранатный.подсумок.NOM.SG.DEF бомба.DEF.ACC.SG  
 ‘Он увидел, как именно [ТОТ] выхватил гранату из подсумка’.

(6a) *Ne nuk dinim se si saktësisht*  
 1PL.NOM NEG знать.IPF.3PL COMP как именно  
*bëhet lufta.*  
 делать.NACT.PRS.3SG война.NOM.SG.DEF  
 ‘Мы не знали, как именно вести войну’.

Второй тест на различение — подстановка комплементатора факта. С комплементатором действия такая подстановка работает без изменения значения (2б), в то время как с союзным словом такая подстановка без изменения значения невозможна (6б).

<sup>23</sup> Ср. болгарский пример из статьи [Hansen et al. 2016: 215]: *Знам как работи тази система* (знать.PRS.1SG как работать.PRS.3SG этот.F.SG система.SG) ‘Я знаю, как работает эта система’.

- (26) *Ai e= pa se shkëputi*  
 M.3SG.NOM 3SG.ACC= видеть.AOR.3SG COMP.A сорвать.AOR.3SG  
*nga gjerdani bombën.*  
 от гранатный.подсумок.NOM.SG.DEF бомба.DEF.ACC.SG

‘Он увидел, что [тот] выхватил гранату из подсумка’.

- (66) *Ne nuk dinim se bëhet*  
 1PL.NOM NEG знать.IPF.3PL COMP делать.NACT.PRS.3SG  
*lufta.*  
 война.NOM.SG.DEF

‘Мы не знали, что ведется война’.

Кроме того, есть и структурные ограничения: *si*-комплементатор не может вводить придаточные с конъюнктивом в качестве сказуемого, поскольку любой пример такого рода однозначно трактуется как придаточное образа действия:

- (7) *Më= thuaј si t'ia bëjmë*  
 1SG.DAT= сказать.IMP.2SG как COMP.P:3SG.DAT:3SG.ACC делать.PRS.1PL  
*me punën e këtyre festave...*  
 с работа.ACC.SG.DEF AGR этот.GEN.PL праздник.GEN.PL

‘Скажи мне, как нам поступить с этими праздниками...’ [АНК: Lebit Murtishi. Vargu magjik. 2009]

Факультативное *se* возможно как с *si*-комплементатором, так и с *si*-союзным словом, ср. примеры (2) (комплементатор без факультативного *se*) vs (8) (комплементатор с *se*), (6) (союзное слово без *se*) vs (9) (союзное слово с *se*):

- (8) *Do të shohësh, xhaxha, se si nuk*  
 FUT COMP.P видеть.SBJV.PRS.2SG дядя COMP.F COMP.A NEG  
*do të lëmë këmbë fashisti.*  
 FUT COMP.P разрешить.PRS.1PL нога.ACC.SG.INDF фашист.ABL.SG.INDF

‘Ты увидишь, дядя, что мы не дадим фашисту ступить и шагу’.  
 [Domi et al. 2002: 517]

- (9) *Në muajin maj kanë caktua[r]*  
 в месяц.ACC.SG.DEF май иметь.PRS.3PL определить.PTCP  
*zgjedhje lokale, por kush*  
 выборы.ACC.PL.INDF местный.F.PL но кто.NOM  
*me siguri nuk mund të thotë*  
 с уверенность.ACC.SG.INDF NEG PSB COMP.P сказать.PRS.3SG

*se si do të mbarojnë ato.*  
 COMP.F как FUT COMP.P закончиться.PRS.3PL F.3PL.NOM

‘В мае определили [дату] местных выборов, но кто же может быть уверен, что не знает их исхода’ (= все знают исход выборов). [АНК: Koha.mk. 2011]

*Si*-комплементатор оказывается довольно редким явлением в современном албанском языке (во всяком случае, в контекстах после трех указанных глаголов), всего в корпусе зафиксировано 14 (0,73 ipm) таких случаев. В основном ( $N = 813$ ) *si* используется как союзное слово.

Сравнение этих двух групп при помощи таблицы сопряженности (Таблица 6) показывает два разнонаправленных тренда. Так, перед *si*-комплементатором в большинстве случаев ставится *se* (11 случаев из 14), в то время как перед *si*-союзным словом наоборот, в большинстве случаев *se* не ставится ( $p$  (точный тест Фишера<sup>24</sup>)  $< 0,05$ ). Можно предположить, что это противопоставление основано на стремлении избежать омонимии двух типов придаточных изъяснительных: комплементативная функция *si* дублируется при помощи *se*; при *si*-союзном слове дублирование избегается.

Возможность дублирования *si*-комплементатора при помощи *se* (более того, его предпочтение) позволяет предположить, что в албанском языке можно говорить о слабой дифференциации комплементов действия и комплементов факта. Употребление факультативного *se* с обоими типами комплементов (с комплементами-фактами и с комплементами действия) показывает их структурную близость в албанском языке, противопоставляя их комплементам потенциала. Впрочем, поскольку *si*-комплементатор для глаголов не визуального доступа к информации является периферийным, подробное описание того, как с ним используется факультативное *se*, не входит в задачи настоящей статьи.

## 6. Факультативное *se* после фразеологизмов с компонентом-глаголом *di*

В ряде случаев главное предложение, структурно требующее комплемента, по сути представляет собой фразеологизированное сочетание.

<sup>24</sup> Поскольку в одной из ячеек таблицы значение меньше 5,  $\chi^2$  рассчитать нельзя.

В моей выборке к таким случаям относятся *zoti (një zot) e di* ‘(один) бог знает’, *dreqi (një dreq) e di* ‘(один) черт знает’, *kush e di* ‘кто (это, его) знает’ (возможен вариант со слитным написанием *kushedi*), *ku ta dish (ku ta di, ku me ditë)* ‘где (тебе, мне) это знать’ (глагол может стоять во втором или первом лице конъюнктива, а также в форме так называемого североалбанского инфинитива)<sup>25</sup>.

- (9) *Historinë tonë edhe ashtu po e=*  
 история.ACC.SG.DEF наш.ACC.SG и так PROG 3SG.ACC=  
*shkruajnë të tjerët, zoti*  
 писать.PRS.3PL AGR другой.M.NOM.PL.DEF ГОСПОДЬ.NOM.SG.DEF  
*e= di ku mbesim.*  
 3SG.ACC= знать.PRS.3SG где оставаться.PRS.1PL

‘Нашу историю и так пишут другие, мы бог весть где уже находимся’. [АНК: Zëri. 2013]

В современном албанском языке приведенные сочетания слов трактуются по-разному. Например, *kush e di* считается наречием или частицей (и пишется слитно: *kushedi*), модифицирующей союзное слово, к которому оно относится [Rushiti 2015: 76–77], а *zoti / një zot e di, dreqi / një dreq e di* — фразеологизмами. Для унификации я называю все сочетания такого рода фразеологизированными. Эти фразеологизмы допускают некоторую свободу (могут употребляться формы с артиклем и без, глагол в составе фразеологизма может спрягаться и так далее), которая теоретически может распространяться и на выбор лексики (ср. русск. *Аллах / Перун / Один его знает* и под.). Отмечу, что в данном случае следует отличать фразеологизмы от свободных сочетаний слов, ср. следующие примеры:

- (10) *Sigurisht, Allahu e= di se çfarë*  
 несомненно Аллах.NOM.SG.DEF 3SG.ACC= знать.PRS.3SG COMP.F что  
*adhurojnë ata në vend të Tij.*  
 почитать.PRS.3PL они.M.NOM в место.ACC.SG.INDF AGR его

‘Несомненно, Аллах ведает, что почитают они вместо него’.  
 [АНК: Kur’ani. Surja 29. Merimanga. 2006]<sup>26</sup>

<sup>25</sup> Ср. русские *бог весть, (один) бог знает, (один) черт знает, черт-те-* (напр. *черт-те что, черт-те где* и другие подобные).

<sup>26</sup> «Поистине, Аллах знает всякую вещь, которую они призывают помимо него» (Коран в переводе И. Ю. Крачковского).

- (11) **Zoti**                    *e*=                    **di**                    *si*                    *t'i*  
 господь.NOM.SG.DEF    3SG.ACC=    знать.PRS.3SG    как    COMP.P:3PL.ACC  
*shpëtojë*                    *të*                    *përshpirtshmit*                    *nga*  
 спасти.SBJV.PRS.3SG    AGR    благочестивый.ACC.PL.DEF    от  
*tundimi*.  
 искушение.NOM.SG.DEF

‘Господь знает, как избавить благочестивых от искушения’.  
 [АНК: Dhjata e Re. 2 Pjetri. 1994<sup>27</sup>]

Таблица сопряженности (Табл. 7) показывает, что с фразеологизмами *se* употребляется почти в два раза реже, чем с омонимичными им свободными сочетаниями ( $\chi^2(1, N = 2574) \approx 5,13$ ;  $p < 0,05$ ). Это вполне ожидаемо, поскольку, строго говоря, при фразеологическом употреблении таких сочетаний «придаточное изъяснительное» фактически становится главным предложением, а статус «главного» предложения понижается (например, в (9) *zoti e di ku* можно трактовать как обстоятельство места).

## 7. Заключение

Факультативный комплементатор *se* в албанском языке унифицирует оформление изъяснительных придаточных и эксплицирует их синтаксическую связь с главным предложением. Как показывает анализ материала Албанского национального корпуса (Разделы 3–4), вариативность в его использовании может быть связана как с ближайшим контекстом, так и с жанром текста; кроме того, наблюдаются некоторые незначительные изменения в диахронии. Из трех рассмотренных глаголов (глагол речи *them* ‘сказать’, глагол психической деятельности *di* ‘знать’ и глагол прямого вопроса *pyes* ‘спрашивать’) *them* показывает некоторое разнообразие (*çfarë*, *kush*, *kur* при этом глаголе скорее фаворизируют вариант с комплементатором, для *ku* четких предпочтений не прослеживается, а *sa* и *si* ведут себя по-разному в разных регионах); *pyes* фаворизирует комплементатор; а *di* сильно варьирует по ареалам. Контексты *di (se) çfarë* и *di (se) kur* фаворизируют комплементатор

<sup>27</sup> «Знает Господь, как избавлять благочестивых от искушения» (Синодальный перевод, 2 Пет., 2:9).

во всех регионах; *di (se) ku* не показывает четких предпочтений (хотя в Северной Македонии намечается тенденция к использованию комплементатора в этом контексте); *di (se) sa* и *di (se) kush* показывают разнонаправленные тенденции по регионам. На основе наблюдаемых тенденций регионы можно кластеризовать: больше всего похожи стратегии употребления факультативного комплементатора в Албании и Косово, менее похожи — в Албании и Северной Македонии (различия по трем союзным словам) и скорее различаются, чем похожи стратегии в Косово и Северной Македонии (различия по четырем союзным словам).

Диахронические изменения минимальны (*y di (se) çfarë* и *pyes (se) si* значительно возросла частота вариантов с комплементатором). В то же время, как показывает моделирование дерева решений (Раздел 5), использование комплементатора имеет жанровую специфику: литературные тексты показывают тенденции к отсутствию комплементатора, а газетные — к его использованию (что, возможно, связано с большей консервативностью языка в художественной литературе по сравнению с языком СМИ).

Разные модели взаимодействия факультативного *se* с ближним контекстом дополняют друг друга. Так, многофакторная модель дерева решений, согласуясь с расчетами по индивидуальным переменным, позволяет найти такие случаи, которые могли бы пройти мимо оптики, нацеленной только на индивидуальные переменные. В то же время, внимание к индивидуальным переменным позволяет увидеть частные тенденции, которые не попадают в комплексную модель.

Ожидаемым образом у явлений, на синтаксическом уровне связанных с комплементацией, но имеющих глубокие семантические различия, наблюдаются различия в функционировании факультативного *se*. Так, после фразеологизмов, которые используются в квазипрономинальной функции после глагола *di, se* употребляется почти в два раза реже, чем с омонимичными им свободными сочетаниями (Раздел 7).

На материале Албанского национального корпуса просматриваются различия между комплементатором действия *si* ‘как’ и союзным словом *si*. *Si*-комплементатор (с тремя рассмотренными глаголами не визуального доступа к информации) обычно используется с факультативным *se*, а союзное слово — без. Дублирование комплементатора действия при помощи комплементатора факта, более того, функционирующее как предпочитаемая стратегия, является свидетельством слабой дифференциации комплементов действия и комплементов факта в албанском языке.

## Список условных сокращений

алб. — албанский; англ. — английский; болг. — болгарский, рус. — русский; 1, 2, 3 — лицо глаголов и местоимений; AVL — аблатив; ACC — аккузатив; AGR — маркер согласования; AOR — аорист; COMP.A — комплементатор деятельности; COMP.F — комплементатор факта; COMP.P — комплементатор потенциала; SBJV — конъюнктив; DAT — датив; DEF — определенная форма; FUT — будущее; INDF — неопределенная форма; IPF — имперфект; NACT — неактивный залог; NEG — отрицание; NOM — номинатив; PL — множественное число; PRS — настоящее время; PROG — прогрессив; PSV — POSSIBILITY; SG — единственное число.

## Литература

- Морозова, настоящий сборник — М. С. Морозова. Конструкции с комплементаризмом *që* и конъюнктивом в албанском языке: корпусное исследование // Настоящий сборник.
- Морозова и др. 2016 — М. С. Морозова, Т. А. Архангельский, М. А. Даниэль, А. Ю. Русаков. Албанский национальный корпус: основные направления работы // *Acta Linguistica Petropolitana. Труды Института лингвистических исследований*. Т. XII. Ч. 3. С. 169–189.
- Buchholz, Fiedler 1987 — O. Buchholz, W. Fiedler. *Albanische Grammatik*. Leipzig: Enzyklopädie, 1987.
- Dixon 2006 — R. M. W. Dixon. Complement clauses and complementation strategies // R. M. W. Dixon, A. Yu. Aikhenvald (eds.). *Complementation (A cross-linguistic typology)*. Oxford: Oxford University Press, 2006. P. 1–48.
- Domi et al. 2002 — M. Domi (kryered.), M. Çeliku, S. Floqi, S. Mansaku, R. Përnaska, S. Prifti, M. Totoni. *Gramatika e gjuhës shqipe*. V. II. Sintaksa. Tiranë: Akademia e Shqipërisë, 2002.
- Friedman 1986 — V. Friedman. Balkan Romani modality and other Balkan languages // H. I. Aronson (ed.). *Da, na, sa, te, te*: Constructions with subordinating complementizers in the Balkans. Columbus: Slavica Publishers, 1986. P. 381–389.
- Hansen et al. 2016 — B. Hansen, A. Letuchiy, I. Błaszczuk. Complementizers in Slavic (Russian, Polish, and Bulgarian) // K. Boye, P. Kehayov (eds.). *Complementizer Semantics in European Languages*. (Empirical Approaches to Language Typology 57). Berlin, Boston: De Gruyter, 2016. P. 175–224.
- Hentschel et al. (в печати) — G. Hentschel, O. Palinska, I. Fekete. Dialect-leveling or colonial hybridization: On determining the impact of various donor codes on mixed subvarieties — The infinitive in Ukrainian-Russian “Surżyk” // *Language Variation and Change*.
- Horie, Comrie 2000 — K. Horie, B. Comrie. Introduction // K. Horie, B. Comrie (eds.). *Complementation (Cognitive and functional perspectives)*. Amsterdam; Philadelphia: John Benjamins, 2000. P. 1–10.

- Hothorn et al. 2006 — T. Hothorn, K. Hornik, A. Zeileis. Unbiased recursive partitioning: A conditional inference framework // *Journal of Computational and Graphical Statistics*. 2006. Vol. 3. № 15. P. 651–674.
- Hothorn, Zeileis 2006 — T. Hothorn, A. Zeileis. partykit: A modular toolkit for recursive partitioning in R // *Journal of Machine Learning Research*. 2006. Vol. 16. P. 3905–3909.
- Joseph 2016 — B. D. Joseph. The semantics and syntax of complementation markers as an areal phenomenon in the Balkans, with special attention to Albanian // K. Boye, P. Kehayov (eds.). *Complementizer Semantics in European Languages*. (Empirical Approaches to Language Typology 57). Berlin; Boston: De Gruyter, 2016. P. 265–292.
- Kuhn 2008 — M. Kuhn. Building predictive models in R using the caret package // *Journal of Statistical Software*. 2008. Vol. 28. № 5. P. 1–26. DOI: 10.11649/ch.2018.010.
- Kuhn 2018 — M. Kuhn. The caret package. Available at <http://CRAN.R-project.org/package=caret> (accessed on 18.12.2019).
- Levshina 2015 — N. Levshina. How to do linguistics with R. Data exploration and statistical analysis. Amsterdam; Philadelphia: John Benjamins, 2015.
- Manzini, Savoia 2018 — M. R. Manzini, L. M. Savoia. *The Morphosyntax of Albanian and Aromanian Varieties: Case, Agreement, Complementation*. Berlin; Boston: Walter de Gruyter, 2018.
- RStudio 2019 — RStudio: Integrated development environment for R (Version 3.6.1). Boston, MA. Available at: <http://www.rstudio.com/> (accessed on 18.12.2019).
- Rushiti 2015 — R. Rushiti. *Frazat lidhëzore në gjuhën shqipe (lidhja e përbërëseve kallëzuesore me lidhëzat që dhe se)*. Tiranë: Classic Print, 2015.
- Schmidtke-Bode 2014 — K. Schmidtke-Bode. *Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization*. PhD dissertation, University of Jena, 2014.
- Shopen 2007 — T. Shopen. *Language typology and syntactic description*. Vol. II: Complex constructions. Cambridge: Cambridge University Press, 2007.
- Tagliamonte, Baayen 2012 — S. A. Tagliamonte, R. H. Baayen. Models, forests and trees of York English: Was / were variation as a case study for statistical practice // *Language Variation and Change*. 2012. Vol. 24. P. 135–178.

## Источники

- АНК 2016 — М. С. Морозова, А. Ю. Русаков, Т. А. Архангельский. Албанский национальный корпус (версия 2011–2016 гг.) (электронный ресурс). URL: <http://web-corpora.net/AlbanianCorpus> (дата обращения 29.11.2019).
- АНК 2019 — М. С. Морозова, А. Ю. Русаков, Т. А. Архангельский. Албанский национальный корпус (электронный ресурс). URL: [albanian.web-corpora.net](http://albanian.web-corpora.net) (дата обращения 29.11.2019).
- Bota Sot — Bota Sot (электронный ресурс). URL: [botasot.info](http://botasot.info) (дата обращения 29.11.2019).

Sot.com.al — Sot (электронный ресурс). URL: sot.com.al (дата обращения 29.11.2019).  
Telegrafi.com — Telegrafi.com (электронный ресурс). URL: telegrafi.com (дата обращения 29.11.2019).

## References

- Buchholz, Fiedler 1987 — O. Buchholz, W. Fiedler. Albanische Grammatik. Leipzig: Enzyklopädie, 1987.
- Dixon 2006 — R. M. W. Dixon. Complement clauses and complementation strategies. R. M. W. Dixon, A. Yu. Aikhenvald (eds.). *Complementation (A cross-linguistic typology)*. Oxford: Oxford University Press, 2006. P. 1–48.
- Domi et al. 2002 — M. Domi (kryered.), M. Çeliku, S. Floqi, S. Mansaku, R. Përnaska, S. Prifti, M. Totoni. Gramatika e gjuhës shqipe. Vëll. II. Sintaksa [Grammar of Albanian. Vol. II. Syntax]. Tiranë: Akademia e Shkencave e Shqipërisë, 2002.
- Friedman 1986 — V. Friedman. Balkan Romani modality and other Balkan languages. H. I. Aronson (ed.). *Da, ná, sã, tẽ, te: Constructions with subordinating complementizers in the Balkans*. Columbus: Slavica Publishers, 1986. P. 381–389.
- Hansen et al. 2016 — B. Hansen, A. Letuchiy, I. Błaszczuk. Complementizers in Slavonic (Russian, Polish, and Bulgarian). K. Boye, P. Kehayov (eds.). *Complementizer Semantics in European Languages*. (Empirical Approaches to Language Typology 57). Berlin, Boston: De Gruyter, 2016. P. 175–224.
- Hentschel et al. (forthcoming) — G. Hentschel, O. Palinska, I. Fekete. Dialect-leveling or colonial hybridization: On determining the impact of various donor codes on mixed subvarieties — The infinitive in Ukrainian-Russian “Suržyk”. *Language Variation and Change*.
- Horie, Comrie 2000 — K. Horie, B. Comrie. Introduction. K. Horie, B. Comrie (eds.). *Complementation (Cognitive and functional perspectives)*. Amsterdam; Philadelphia: John Benjamins, 2000. P. 1–10.
- Hothorn et al. 2006 — T. Hothorn, K. Hornik, A. Zeileis. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*. 2006. Vol. 3. No. 15. P. 651–674.
- Hothorn, Zeileis 2006 — T. Hothorn, A. Zeileis. partykit: A modular toolkit for recursive partitioning in R. *Journal of Machine Learning Research*. 2006. Vol. 16. P. 3905–3909.
- Joseph 2016 — B. D. Joseph. The semantics and syntax of complementation markers as an areal phenomenon in the Balkans, with special attention to Albanian. K. Boye, P. Kehayov (eds.). *Complementizer Semantics in European Languages*. (Empirical Approaches to Language Typology 57). Berlin; Boston: De Gruyter, 2016. P. 265–292.
- Kuhn 2008 — M. Kuhn. Building predictive models in R using the caret package. *Journal of Statistical Software*. 2008. Vol. 28. No. 5. P. 1–26. DOI: 10.11649/ch.2018.010.
- Kuhn 2018 — M. Kuhn. The caret package. Available at: <http://CRAN.R-project.org/package=caret> (accessed on 18.12.2019).

- Levshina 2015 — N. Levshina. How to do linguistics with R. Data exploration and statistical analysis. Amsterdam; Philadelphia: John Benjamins, 2015.
- Manzini, Savoia 2018 — M. R. Manzini, L. M. Savoia. The Morphosyntax of Albanian and Aromanian Varieties: Case, Agreement, Complementation. Berlin; Boston: Walter de Gruyter, 2018.
- Morozova, this volume — M. S. Morozova. Konstruktsii s komplementayzerom *që* i konyunktivom v albanskom yazyke: korpusnoe issledovanie [Constructions with complementizer *që* and subjunctive in Albanian: A corpus study]. This volume.
- Morozova et al. 2016 — M. S. Morozova, T. A. Arkhangel'skiy, M. A. Daniel, A. Yu. Rusakov. Albanskiy nacionalnyy korpus: osnovnye napravleniya raboty [Albanian National Corpus: main areas of work]. *Acta Linguistica Petropolitana. Trudy Instituta lingvisticheskikh issledovaniy*. 2016. Vol. XII. Pt. 3. P. 169–189.
- RStudio 2019 — RStudio: Integrated development environment for R (Version 3.6.1). Boston, MA. Available at: <http://www.rstudio.com/> (accessed on 18.12.2019).
- Rushiti 2015 — R. Rushiti. Frazat lidhëzore në gjuhën shqipe (lidhja e përbërësve kallëzuesore me lidhëzat *që* dhe *se*) [Conjunction clauses in Albanian (sentential complements with conjunctions *që* and *se*)]. Tiranë: Shkronjë pas Shkronje, 2015.
- Schmidtke-Bode 2014 — K. Schmidtke-Bode. Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization. PhD dissertation, University of Jena, 2014.
- Shopen 2007 — T. Shopen. Language typology and syntactic description. Vol. II: Complex constructions. Cambridge: Cambridge University Press, 2007.
- Tagliamonte, Baayen 2012 — S. A. Tagliamonte, R. H. Baayen. Models, forests and trees of York English: Was / were variation as a case study for statistical practice. *Language Variation and Change*. 2012. Vol. 24. P. 135–178.

## Sources

- ANC 2011–2016 — M. Morozova, A. Rusakov, T. Arkhangel'skiy. Albanian National Corpus (version 2011–2016). Available at: <http://web-corpora.net/AlbanianCorpus> (accessed on 29.11.2019).
- ANC 2019 — M. Morozova, A. Rusakov, T. Arkhangel'skiy. Albanian National Corpus. Available at: [albanian.web-corpora.net](http://albanian.web-corpora.net) (accessed on 29.11.2019).
- Bota Sot — Bota Sot [The World Today]. Available at: [botasot.info](http://botasot.info) (accessed on 29.11.2019).
- Sot.com.al — Sot [Today]. Available at: [sot.com.al](http://sot.com.al) (accessed on 29.11.2019).
- Telegrafi.com — Telegrafi.com. Available at: [telegrafi.com](http://telegrafi.com) (accessed on 29.11.2019).

## Приложение 1. Таблицы

Таблица 1. Объем АНК по подкорпусам

Table 1. Albanian National Corpus size by subcorpora

Подкорпусы	Категории анализа	до 1989	с 1990
Художественные тексты: рассказы, романы, пьесы, песни, народные, для детей	литература	913 512; 4,63 %	1 753 333; 8,89 %
Пресса	пресса	0	12 230 215; 62,02 %
Нехудожественные тексты: научные тексты	научные тексты	17 082; 0,09 %	1 917 650; 9,72 %
Нехудожественные тексты: религиозные тексты	религиозные тексты	0	588 595; 2,98 %
Нет информации		0	432 109; 2,19 %
Общий объем корпуса по периодам <sup>28</sup> :		1 462 246; 7,42 %	17 352 771; 88,00 %

Таблица 2. Комплементатор vs глагол

Table 2. Complementizer vs. Verb

	<i>them</i> ‘сказать’	<i>di</i> ‘знать’	<i>pyes</i> ‘спрашивать’	Итого:
<i>se</i> <sup>-</sup>	111 (5,63 ipm)	1323 (67,17 ipm)	94 (4,77 ipm)	1528 (77,56 ipm)
<i>se</i> <sup>+</sup>	129 (6,55 ipm)	1251 (63,5 ipm)	218 (11,07 ipm)	1598 (81,12 ipm)
Итого:	240 (12,8 ipm)	2574 (130,66 ipm)	312 (15,84 ipm)	3126 (158,68 ipm)

<sup>28</sup> Поскольку у части текстов, вошедших в корпус, нет метаданных о годе публикации, сумма слов по периодам не дает ста процентов корпуса. Объем таких текстов: 904321 словоупотреблений или 4,58 % АНК.

Таблица 3. Комплементатор vs союзное слово  
Table 3. Complementizer vs. connecting pronoun or adverb

	<i>se</i> <sup>-</sup>	<i>se</i> <sup>+</sup>	<b>Итого:</b>
<i>çfarë</i> ‘что’	227 (11,95 ipm)	426 (22,42 ipm)	653 (34,37 ipm)
<i>kush</i> ‘кто’	218 (11,47 ipm)	267 (14,05 ipm)	485 (25,53 ipm)
<i>kur</i> ‘когда’	34 (1,79 ipm)	118 (6,21 ipm)	152 (8,00 ipm)
<i>ku</i> ‘где, куда’	288 (15,16 ipm)	278 (14,63 ipm)	566 (29,79 ipm)
<i>sa</i> ‘сколько’	244 (12,84 ipm)	183 (9,63 ipm)	427 (22,47 ipm)
<i>si</i> ‘как’	517 (27,21 ipm)	326 (17,16 ipm)	843 (44,37 ipm)
Итого:	1528 (77,56 ipm)	1598 (81,12 ipm)	3126 (158,68 ipm)

Таблица 4. Одновременное употребление глагола и союзного слова: сравнение данных по регионам (после 1990)<sup>29</sup>

Table 4. Verbs and connective words: a comparison of the regions (after 1990)

		АЛ	КС	СМ	АЛ/КС	АЛ/СМ	КС/СМ
<i>them</i>	<i>çfarë</i>	8/10	11/17	3/7	0,766; p>0,05	0,688; p>0,05	0,715; p>0,05
	<i>kush</i>	6/7	1/6	0/3	0,328; p>0,05	1,000; p>0,05	0,250; p>0,05
	<i>kur</i>	1/3	0/4	0/3	0,500; p>0,05	1,000; p>0,05	1,000; p>0,05
	<i>ku</i>	6/5	6/5	2/2	1,000; p>0,05	1,000; p>0,05	1,000; p>0,05
	<i>sa</i>	7/9	7/5	1/5	0,477; p>0,05	0,351; p>0,05	0,151; p>0,05
	<i>si</i>	6/5	6/11	6/4	0,440; p>0,05	1,000; p>0,05	0,256; p>0,05
<i>di</i>	<i>çfarë</i>	73/107	65/79	30/109	$\chi^2 = 0,687$ ; p>0,05	$\chi^2 = 12,913$ ; p<0,05	$\chi^2 = 17,599$ ; p<0,05
	<i>kush</i>	52/88	52/31	34/73	$\chi^2 = 13,623$ ; p<0,05	$\chi^2 = 0,769$ ; p>0,05	$\chi^2 = 17,984$ ; p<0,05
	<i>kur</i>	3/22	7/25	9/25	0,486; p>0,05	0,207; p>0,05	0,776; p>0,05
	<i>ku</i>	78/74	58/46	40/55	$\chi^2 = 0,491$ ; p>0,05	$\chi^2 = 1,987$ ; p>0,05	$\chi^2 = 3,708$ ; p<0,05

<sup>29</sup> Поскольку здесь учитываются только те примеры, для которых есть привязка по регионам и по ареалам, общее количество меньше 3126. Это же относится и к ряду последующих таблиц: при введении любого параметра, относящегося к метаинформации, не учитываются данные, для которых этот параметр не отражен в АНК (например, отсутствует год издания).

		АЛ	КС	СМ	АЛ/КС	АЛ/СМ	КС/СМ
<i>di</i>	<i>sa</i>	75/53	54/22	41/52	$\chi^2 = 3,183$ ; $p > 0,05$	$\chi^2 = 4,546$ ; $p < 0,05$	$\chi^2 = 12,356$ ; $p < 0,05$
	<i>si</i>	163/95	66/37	98/87	$\chi^2 = 0,025$ ; $p > 0,05$	$\chi^2 = 4,635$ ; $p < 0,05$	$\chi^2 = 3,327$ ; $p > 0,05$
<i>pyes</i>	<i>çfarë</i>	1/31	3/8	3/23	0,045; $p < 0,05$	0,316; $p > 0,05$	0,335; $p > 0,05$
	<i>kush</i>	5/13	2/4	2/10	1,000; $p > 0,05$	0,669; $p > 0,05$	0,568; $p > 0,05$
	<i>kur</i>	1/9	2/5	1/4	0,536; $p > 0,05$	1,000; $p > 0,05$	1,000; $p > 0,05$
	<i>ku</i>	10/13	0/7	3/14	0,063; $p > 0,05$	0,102; $p > 0,05$	0,529; $p > 0,05$
	<i>sa</i>	1/4	1/2	2/7	1,000; $p > 0,05$	1,000; $p > 0,05$	1,000; $p > 0,05$
	<i>si</i>	15/18	5/5	8/12	1,000; $p > 0,05$	0,779; $p > 0,05$	0,705; $p > 0,05$

Таблица 5. Одновременное употребление глагола и союзного слова: сравнение данных по периодам (Албания, литературные и научные тексты)

Table 5. Verbs and connective words: a comparison of the periods (Albania, belles-lettres and academic texts)

		до 1989	после 1990	
<i>them</i>	<i>çfarë</i>	0/1	5/8	1,000; $p > 0,05$
	<i>kush</i>	6/1	4/3	0,314; $p > 0,05$
	<i>kur</i>	0/1	0/1	1,000; $p > 0,05$
	<i>ku</i>	1/2	5/4	0,590; $p > 0,05$
	<i>sa</i>	2/1	4/5	0,590; $p > 0,05$
	<i>si</i>	11/4	6/2	1,000; $p > 0,05$
<i>di</i>	<i>çfarë</i>	16/6	28/31	$\chi^2 = 4,123$ ; $p < 0,05$
	<i>kush</i>	26/11	23/15	$\chi^2 = 0,785$ ; $p > 0,05$
	<i>kur</i>	4/5	1/8	0,161; $p > 0,05$
	<i>ku</i>	30/27	43/35	$\chi^2 = 0,082$ ; $p > 0,05$
	<i>sa</i>	21/13	29/18	$\chi^2 = 3,270$ ; $p > 0,05$
	<i>si</i>	45/24	83/34	$\chi^2 = 0,662$ ; $p > 0,05$
<i>pyes</i>	<i>çfarë</i>	0/5	1/15	1,000; $p > 0,05$
	<i>kush</i>	3/1	1/6	0,087; $p > 0,05$
	<i>kur</i>	0/3	1/2	0,500; $p > 0,05$
	<i>ku</i>	2/3	5/11	1,000; $p > 0,05$
	<i>sa</i>	1/1	0/1	1,000; $p > 0,05$
	<i>si</i>	12/2	9/11	0,030; $p < 0,05$

Таблица 6. Таблица сопряженности для *si* в функции  
комплементатора и союзного слова

Table 6. Cross table for *si* as a complementizer and a connective word

	$se^-$	$se^+$	Всего
Комплементатор	3 (0,16 ipm)	11 (0,58 ipm)	14 (0,74 ipm)
Союзное слово	503 (26,47 ipm)	310 (16,32 ipm)	813 (42,79 ipm)

Таблица 7. Таблица сопряженности для *se* после фразеологизмов  
с компонентом-глаголом *di*

Table 7. Cross table for *se* after phraseologisms with *di* as a component

	$se^-$	$se^+$	Всего
Свободное сочетание	1280 (67,37 ipm)	1228 (64,63 ipm)	2508 (132,00 ipm)
Фразеологизм	43 (2,26 ipm)	23 (1,21 ipm)	66 (3,47 ipm)

## Приложение 2. Рисунки

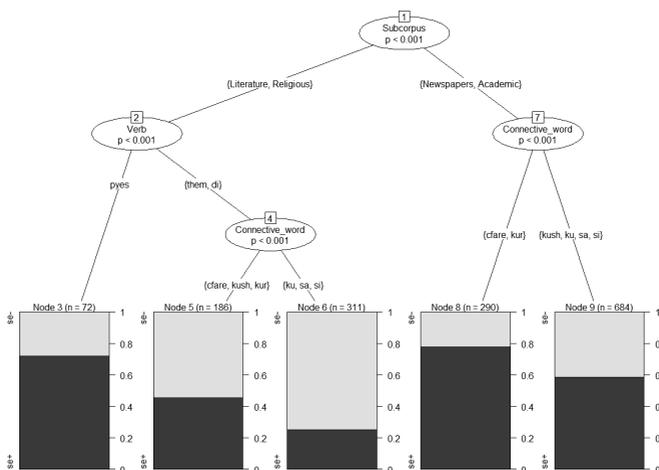


Рисунок 1. Дерево решений для выбора комплементатора (1990–)

Figure 1. Decision tree for the choice of complementizer (1990–)

Примечание к рисунку. При импорте данных в Rstudio я использую английский язык: *Verb* — глагол; *Connective\_word* — союзное слово; *Subcorpus: Academic, Literature, Newspapers, Religious* — Подкорпус (научные тексты, литература, газета, религиозные тексты).

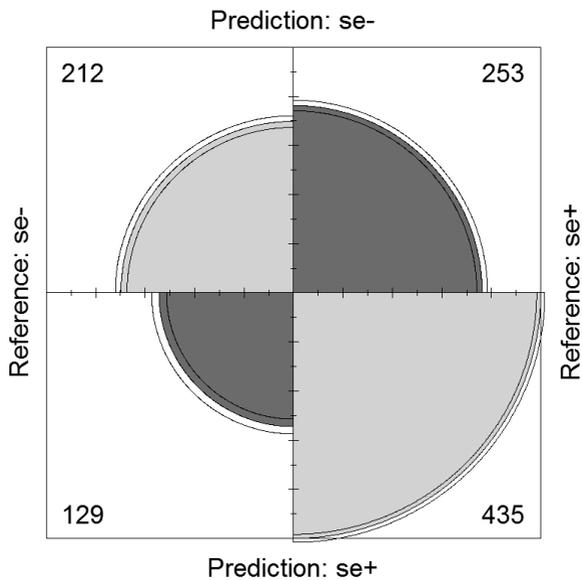


Рисунок 2. Матрица ошибок для дерева решений

Figure 2. Confusion matrix for the ctree

Примечание: *Prediction* — предсказано, *Reference* — зафиксировано.