

ВЛИЯНИЕ ФАКТОРА СИНТАКСИЧЕСКОЙ РОЛИ АНТЕЦЕДЕНТА НА РАЗРЕШЕНИЕ РЕФЕРЕНЦИАЛЬНОЙ НЕОДНОЗНАЧНОСТИ В РУССКОМ ЯЗЫКЕ¹

Деликишкина Е. А. (delik.kate@gmail.com),
Федорова О. В. (olga.fedorova@msu.ru)

Московский государственный университет имени
М. В. Ломоносова, Москва, Россия

Согласно принятой точке зрения, референциальный выбор осуществляется на основании механизма подсчета активации референта в когнитивной системе. По мнению ряда исследователей, выбор между полной и редуцированной именной группой, осуществляемый говорящим, может быть достоверно предсказан на основании фактора активации. Очевидно, аналогично механизм установления референции устроен и у слушателя: местоимение отождествляется с наиболее активированным в данный момент референтом. Однако ситуация осложняется тем, что в дискурсе может находиться одновременно несколько высоко активированных референтов, и случается так, что ни контекст, ни лексико-грамматическая информация не помогают в разрешении неоднозначности. Пытаясь спрогнозировать референциальный выбор адресата в подобных ситуациях, лингвисты выдвигали различные факторы. В данной работе показано, что носители русского языка более склонны отождествлять неоднозначные местоимения с грамматическими подлежащими, чем с дополнениями. Результаты эксперимента на русском материале согласуются с данными, полученными для других языков (в частности, английского и финского), и свидетельствуют в пользу универсальности фактора синтаксической роли при разрешении референциальной неоднозначности.

Ключевые слова: референция, неоднозначность, анафора, местоимение, подлежащее

¹ Работа выполнена при финансовой поддержке гранта РФФИ №12-06-00268.

THE EFFECT OF THE SYNTACTIC ROLE OF THE ANTECEDENT ON AMBIGUOUS PRONOUN RESOLUTION IN RUSSIAN

Delikishkina E. A. (delik.kate@gmail.com),

Fedorova O. V. (olga.fedorova@msu.ru)

Lomonosov Moscow State University, Moscow,
Russian Federation

According to a common view, referential choice is determined by the degree of activation of the referent in the cognitive system. It is even argued that our choice of referential device can be predicted solely on the basis of the activation factor. However, sometimes anaphora resolution is complicated by the presence of two or more competing referents, each of which enjoys activation high enough to be coreferent to the pronoun. If neither the context, nor the grammar can help us identify the antecedent, we are confronted with referential ambiguity. It has long been the subject of debate what determines the choice of referent in such cases. Various factors have been shown to affect pronoun resolution. According to one account, the preferred antecedent of an ambiguous pronoun is the first-mentioned NP. The first-mention advantage is attributed to language-independent, general cognitive processes. Alternatively, it is argued by some researchers that pronoun resolution is affected by specific linguistic factors such as the antecedent's grammatical role. According to this account, the most favoured antecedent is the grammatical subject. However, distinguishing between the two hypotheses using the data from languages such as English is impossible as the first-mentioned referent is also usually a subject of the clause.

In the present study we describe an experiment that has shown that speakers of Russian are more likely to corefer ambiguous pronouns to grammatical subjects regardless of whether they are mentioned first or not. We presented our subjects with ambiguous texts consisting of two sentences. The first one contained two referents, and the second one started with a pronoun that could refer to both of these referents. The first sentence had two versions: each of the referents could be either a subject or an object, which was attained by changing the diathesis of the transitive verb. Each of the 32 subjects that participated in the experiment was assigned to one of the two experimental lists that contained only one version of each stimulus text. The results showed that speakers of Russian considered the subject of the sentence to be an antecedent of the pronoun in 58,4% of cases, which is statistically significant (binomial test, $p < 0.01$), thus proving the importance and independence of the syntactic role factor.

In conclusion, we reason that the apparent universality of the subject-preference strategy (the similar results were obtained for other languages, e. g. Finnish) supports the idea that the choice of grammatical structure is conditioned by the constraints on one's cognitive system, or rather working memory.

Key words: reference, ambiguity, anaphora, pronoun, subject

1. Введение: референциальная неоднозначность и факторы, влияющие на ее разрешение

Порождая высказывания, говорящий, как правило, заботится о том, чтобы быть понятым адресатом. Поэтому наивному носителю языка может показаться удивительным, что исследователи усматривают потенциальную неоднозначность на каждом шагу. Большое количество лингвистических работ последних лет посвящено изучению неоднозначности и способов ее разрешения в лексике, синтаксисе, семантике и прагматике, а также коммуникативной функции неоднозначности в целом (свежий обзор см. в [Piantadosi et al. 2012]). Исследователей дискурса в первую очередь интересует такое явление, как референциальная неоднозначность², то есть наличие у анафора нескольких конкурирующих antecedентов.

Согласно свидетельству ряда исследований (см., в частности, [Kibrik 2011]), выбор между полным и редуцированным референциальным выражением осуществляется на основании активированности референта в рабочей памяти говорящего³: местоимения и анафорические нули используются только для отсылки к референтам, имеющим в данный момент высокий уровень активации, в то время как слабо активированные референты называются при помощи полных ИГ. Аналогичным образом механизм референции, очевидно, осуществляется и в процессе обработки речевых сообщений при понимании: адресат отождествляет местоимение с наиболее активированным в его рабочей памяти референтом. Очень часто, однако, сразу несколько высоко активированных референтов претендуют на роль antecedента местоимения. В большинстве случаев эта неоднозначность снимается за счет контекста или различных лексико-грамматических средств (подробнее о классификации средств снятия потенциальной референциальной неоднозначности см. в [Федорова и Успенская 2011]).

Тем не менее, иногда адресат сталкивается с глобальной референциальной неоднозначностью, когда несколько присутствующих в дискурсе референтов с равным успехом могут быть выбраны в качестве antecedента анафорического местоимения. Примером такого контекста является пара предложений:

- (1) *Tony Blair shook hands with George Bush in the White House. He wanted to discuss the situation in Iraq.*⁴

² В работах А. А. Кибрика используется синонимичный термин «референциальный конфликт».

³ Отметим, что ряд исследователей описывает процесс референциального выбора, используя термин «фокус внимания». Мы, вслед за [Kibrik 2011], принимаем модель, согласно которой выбор референциального средства когнитивно связан с уровнем активации референта в памяти.

⁴ Пример цитируется по [Järvikivi et al. 2005].

Здесь оба упоминающихся в первом предложении референта — ‘Tony Blair’ и ‘George Bush’ (в этом и последующих примерах они выделены подчеркиванием) — могут выступать в качестве antecedента местоимения *he* (выделено полужирным шрифтом), стоящего в начале второго предложения.

Вопрос о том, на каком основании адресат осуществляет выбор в подобных случаях, является предметом активного исследования с 1980-х годов. Лингвистами выдвигались различные гипотезы относительно того, какие факторы оказывают влияние на разрешение анафоры в контекстах с референциальной неоднозначностью. В частности, в работе [Gernsbacher and Hargreaves 1988] постулируется Принцип приоритета первого упоминания (англ. first-mention account). Согласно этому принципу, при выборе antecedента предпочтение отдается тому референту, который был упомянут в предыдущей клаузе первым, вне зависимости от других факторов. Действие этого принципа объясняется общим устройством когнитивной системы человека. В своей аргументации авторы опираются на концепцию построения ментальных структур (англ. structure building framework): для построения связной ментальной структуры необходимо, чтобы сначала была заложена основа, некий стержень, на который будет нанизываться поступающая в дальнейшем информация. Поэтому всякая информация, фигурирующая в начале, является более доступной.

Альтернативная гипотеза изложена в работах [Frederiksen 1981] и [Crawley et al. 1990]. Авторы данных исследований приводят аргументы в пользу Принципа приоритета подлежащего (subject-preference account), согласно которому при осуществлении кореференции предпочтение отдается референту, который был подлежащим предыдущей клаузы. Таким образом, по мнению авторов, разрешение референциальной неоднозначности оказывается обусловлено чисто лингвистическим фактором, а не общими особенностями устройства когнитивной системы человека⁵.

Однако разграничить две конкурирующие гипотезы, опираясь на материал такого языка, как английский, оказывается невозможным. В английском, имеющем фиксированный порядок слов, подлежащее почти всегда оказывается упомянутым первым. Таким образом, и Принцип приоритета первого упоминания, и Принцип приоритета подлежащего предсказывают, что в качестве antecedента местоимения в примере (1) с большой вероятностью будет выбран Тони Блэр. Чтобы выяснить, какой из двух факторов в действительности имеет большее значение, авторы работы [Järvikivi et al. 2005] использовали материал финского языка, имеющего свободный порядок слов и кодирующего синтаксические роли с помощью падежного маркирования. Исследователи сконструировали ряд экспериментальных контекстов, состоящих из двух предложений. Первое предложение вводило двух персонажей и могло иметь порядок слов SVO (пример 2) либо OVS (пример 3). Второе предложение начиналось с местоимения, которое могло отсылать к любому из этих персонажей.

⁵ Ср., однако, мнение о связи понятия подлежащего с понятием фокуса внимания ([Myachykov and Tomlin 2008]).

- (2) *Tony Blair kätteli George Bushia valkoisessa talossa. Hän halusi keskustella Irakin tilanteesta.*
[Тони Блэр пожал руку Джорджу Бушу в Белом доме. Он хотел обсудить ситуацию в Ираке.]
- (3) *George Bushia kätteli Tony Blair valkoisessa talossa. Hän halusi keskustella Irakin tilanteesta.*
[Джорджу Бушу пожал руку Тони Блэр в Белом доме. Он хотел обсудить ситуацию в Ираке.]

Эксперимент проводился с помощью методики регистрации движения глаз. Испытуемых просили слушать истории и одновременно с этим смотреть на экран, на котором предьявлялись изображения двух персонажей, в то время как движения их глаз регистрировались с помощью специальной аппаратуры. Статистический анализ полученных данных подтвердил наличие обоих эффектов. С одной стороны, испытуемые, услышав местоимение, чаще фиксировали взгляд на изображении персонажа, выступавшего в первом предложении в роли подлежащего, что свидетельствует о том, что местоимение интерпретировалось ими скорее как кореферентное подлежащему, нежели дополнению. С другой стороны, фиксаций было значительно больше на изображениях персонажа, упомянутого первым, чем вторым, а значит, эффект приоритета первого упоминания тоже имеет место. При этом корреляции между двумя эффектами обнаружено не было, что говорит об их независимости друг от друга. Таким образом, авторы приходят к выводу, что гипотезу, высказанную в [Gernsbacher and Hargreaves 1988], о том, что разрешение референциальной неоднозначности происходит за счет общих когнитивных процессов и непосредственно не связано с лингвистическими факторами, по всей видимости, следует отвергнуть⁶. В заключении авторы подчеркивают, что процесс разрешения анафоры осуществляется под влиянием одновременно нескольких факторов, поэтому от однофакторных моделей референциального выбора следует отказаться⁷.

2. Эксперимент М. Г. Степиной на русском материале

Первый подобный эксперимент на русском материале был проведен в 2005 году ([Степина 2005]). Стимульный материал конструировался по тому же принципу, что и описанный в работе [Järvikivi et al. 2005]: экспериментальный контекст состоял из двух предложений, в первом из которых вводились два одушевленных референта одного пола, а второе начиналось с анафорического местоимения *он/она*, которое по смыслу могло относиться к любому из этих двух референтов. Первое предложение имело два варианта — SVO и OVS.

⁶ См., однако ссылку 5.

⁷ См. аналогичный многофакторный подход в [Kibrik 2011].

Принципиальное отличие от эксперимента на финском материале было в том, что порядок упоминания референтов сохранялся, изменялось лишь их падежное маркирование, для чего дополнительно изменялась диатеза глагола. Это было сделано с целью сохранить тема-рематическую структуру предложения. Так, в примере (4) упоминаемая первой 'пожилая нянечка' выступает в функции подлежащего, а 'новая медсестра' является дополнением, выражаемым творительным падежом; в примере (5) 'нянечка' понижается до статуса прямого дополнения и маркируется винительным падежом, в то время как 'медсестра' получает статус подлежащего.

- (4) *Пожилая нянечка нашей поликлиники уже давно возмущалась новой медсестрой. Она всячески выказывала свое недовольство.*
- (5) *Пожилую нянечку нашей поликлиники уже давно возмущала новая медсестра. Она всячески выказывала свое недовольство.*

В данном исследовании, проведенном с 36 испытуемыми, использовалась методика чтения с саморегуляцией скорости. Тексты предъявлялись по одному на экране компьютера. После каждого текста появлялся слайд с вопросом, на который предлагалось на выбор два варианта ответа. Испытуемый должен был читать про себя предложение, нажимать на «пробел» и отвечать на вопрос, после чего заносить ответ в специальный бланк. 16 экспериментальных текстов в квазислучайном порядке чередовались с 36 текстами-филлерами, призванными отвлечь испытуемого от истинной цели исследования. Все вопросы к экспериментальным текстам запрашивали референта (*Кто выказывал недовольство?*), вопросы к филлерам относились также к какой-либо фактической детали в тексте.

В результате эксперимента получилось, что на долю первого упоминания приходится 47% всех ответов, на долю второго — соответственно, 53%. Как отмечает автор, «мы наблюдаем небольшую тенденцию к предпочтению выбора последнего из введенных референтов в качестве антецедента неоднозначного местоимения, но расхождение в 6% не может рассматриваться как статистически значимое» [Степина 2005: 48]. Процентное же отношение референтов-подлежащих к референтам-дополнениям оказалось равно 53,6% против 46,4%, что, по словам автора, тоже нельзя считать статистически значимой разницей. Таким образом, проанализировав полученные результаты на предмет соответствия гипотезам о приоритете первого упоминания и о приоритете подлежащего, автор исследования не обнаружила эффектов, выявленных в эксперименте [Järvikivi et al. 2005] для финского языка.

Объясняя полученные результаты, Степина отмечает, что «в эксперименте встретилось несколько примеров, которые оказались нечувствительны к видоизменению структуры» [Степина 2005: 52]: в 4 из 16 экспериментальных контекстов местоимение как при порядке SVO, так и при порядке OVS трактовалось в пользу определенного референта. Так, в приведенных выше примерах (4) и (5) большинство испытуемых посчитало, что антецедентом местоимения

являлась 'нянечка'. Очевидно, семантический «перевес» в пользу одного из референтов в данном случае настолько силен, что фактор варьирования синтаксической структуры оказался нерелевантным.

3. Новая серия экспериментов на русском материале

Учитывая семантическую несбалансированность как минимум четверти из используемых в исследовании [Степина 2005] экспериментальных контекстов, можно предположить, что отсутствие ожидаемых эффектов объясняется не специфичностью русского языкового материала, а низкой экологической валидностью эксперимента. Кроме того, несмотря на примерно равное распределение случаев выбора в пользу первого/второго референта, а также подлежащего/дополнения, отсутствие формальных статистических подсчетов делает результаты не вполне надежными. Наконец, в работе [Степина 2005] была допущена грубая методологическая ошибка: в отличие от примеров на финском языке, русские примеры (4) и (5) не различаются с точки зрения порядка упоминания референтов, вследствие чего собранный материал не может быть использован для тестирования Принципа приоритета первого упоминания.

3.1. Создание стимульного материала

Чтобы получить более валидные свидетельства за или против Принципа приоритета подлежащего, мы воспроизвели эксперимент Степиной, предварительно проверив стимульный материал на предмет семантической сбалансированности. Для этого мы искусственно создали новые стимульные тексты, местоимения в которых, по нашей интуиции, можно было с равной вероятностью отнести к каждому из двух потенциальных antecedентов, см. (6) и (7).

(6) *Моя младшая сестра очень полюбила свою новую учительницу. Она всегда приходила в класс за десять минут до звонка.*

(7) *Моей младшей сестре очень понравилась ее новая учительница. Она всегда приходила в класс за десять минут до звонка.*

3.2. Тест на приемлемость

Затем для подтверждения данных интроспекции мы провели тест на приемлемость. Мы предъявляли испытуемым 16 экспериментальных предложений и просили оценить, насколько вероятной является трактовка местоимения в пользу того или иного референта в каждом из них, используя четырехбалльную шкалу Ликерта (от 0 — 'такое значение не существует для данного местоимения', до 3 — 'именно это значение приходит в голову первым, сразу после

прочтения предложения'). 32 испытуемых были распределены так, чтобы каждый читал только один вариант текста. Приемлемыми были признаны тексты, средние оценки которых варьировались от 1,1 до 2,9. Один текст с референтом-дополнением, получившим оценку 0,8, был отбракован и заменен на другой.

Результаты проведенного теста на приемлемость косвенно свидетельствуют в пользу верности гипотезы о приоритете подлежащего: так, из 32 экспериментальных контекстов только в 5 случаях средняя оценка приемлемости референта-дополнения оказывалась выше оценки референта-подлежащего⁸.

3.3. Основной эксперимент

После оценки приемлемости созданных нами текстов мы провели основной эксперимент с 32 испытуемыми, распределенным поровну по двум сформированным экспериментальным листам. Помимо 16 экспериментальных текстов в каждый лист было включено 24 филлера. Процедура эксперимента в целом была аналогичной использованной в [Степина 2005]. Принципиальным отличием являлось то, что испытуемому не предлагалось вариантов ответов на поставленные вопросы, чтобы не акцентировать его внимание на потенциальной неоднозначности текстов. Кроме того, испытуемый не записывал ответы, а произносил их вслух.

Количество выбранных в качестве antecedентов подлежащих составило 58,4%, дополнений — 41,6%. Полученные различия являются статистически значимыми (биномиальный тест, $p=0.00017$), а значит, синтаксический статус antecedента следует признать важным фактором при разрешении референциальной неоднозначности в русском языке. Кроме того, общая тенденция к предпочтению подлежащих видна невооруженным глазом: в подавляющем большинстве текстов (в 24 из 32) испытуемые чаще выбирали в качестве antecedента местоимения подлежащие.

Характерно, что 4 из 8 текстов, ответы на вопросы к которым идут вразрез с гипотезой, показали аналогичные результаты в тесте на приемлемость — дополнения в них получили в среднем более высокие оценки, чем подлежащие. Это, по всей видимости, свидетельствует о том, что в данных текстах одна из трактовок имела преимущество перед другой в силу семантического фактора, то есть при дальнейшем усовершенствовании стимульного материала мы можем ожидать еще более показательные результаты.

⁸ В то же время, общий разброс в оценках приемлемости (от 1,3 до 2,9 для референтов, упоминающихся первыми; от 1,1 до 2,9 для референтов, упоминающихся вторыми) не позволяет сделать аналогичные выводы относительно гипотезы о приоритете первого упоминания.

4. Заключение

Итак, результаты эксперимента подтвердили верность выявленной на материале английского и финского языков гипотезы о влиянии синтаксической структуры предложения на разрешение референциальной неоднозначности и для русского языка. Универсальная тенденция к выбору подлежащего в качестве antecedента максимально активированного референта, выраженного местоимением, свидетельствует, по всей видимости, в пользу того, что когнитивные ограничения накладывают отпечаток на грамматическую структуру языка. Ср., однако, альтернативное мнение из работы [Муачыков and Tomlin 2008], в которой отрицается универсальность Принципа приоритета подлежащего и подчеркивается важность Принципа приоритета первого упоминания⁹.

Предполагается, что дальнейшая работа по исследованию референциальной неоднозначности будет вестись по следующим двум направлениям. С одной стороны, для проверки полученных результатов и установления природы когнитивных механизмов, стоящих за обнаруженным эффектом, мы планируем проведение исследований с применением методов регистрации мозговой активности (вызванных потенциалов и функциональной магнитно-резонансной томографии) и движений глаз.

С другой стороны, важной представляется также задача экспериментального тестирования других факторов, способных оказывать влияние на разрешение референциальной неоднозначности¹⁰, а также построения модели, предсказывающей выбор референциального средства на основании этих факторов.

⁹ Валидность результатов, полученных авторами исследования, с нашей точки зрения, представляется спорной. В качестве стимульного материала эксперимента выступал фильм, в котором красная рыба съедает синюю; при этом внимание испытуемого привлекалось к одной из рыб с помощью стрелки. Задача испытуемого состояла в том, чтобы описать происходящее на экране. Большинство участников англоязычного эксперимента делали рыбу, выделенную стрелкой, подлежащим, меняя активный залог на пассивный в зависимости от того, к какой из рыб было привлечено внимание (Ср. *The red fish ate the blue fish; The blue fish was eaten by the red fish*). Однако в аналогичном эксперименте на русском материале испытуемые, описывая стимулы, в которых привлекалось внимание к синей рыбе (пациенсу), предпочитали конструкции типа «Синюю рыбу съела красная рыба» конструкциям с пассивом («Синяя рыба была съедена красной рыбой»). Авторы делают вывод: “One initial hypothesis — that there is a universal tendency to assign subject to the perceptually prominent referent regardless of language — is not supported. Indeed, attention to the patient of the event significantly improved its chances of becoming the starting point of a Russian transitive sentence but not its chances of becoming the sentential subject.” ([Муачыков and Tomlin 2008: 44]). Мы считаем, однако, что несоблюдение Принципа приоритета подлежащего в данном случае вероятней всего объясняется нетипичностью для устной русской речи пассивных конструкций, а не нарушением универсальной тенденции к приписыванию наиболее активированному («перцептивно выделенному») референту синтаксической роли субъекта.

¹⁰ В частности, важным представляется исследование на русском материале т. н. Принципа структурного параллелизма, описываемого в ряде зарубежных работ (см., например, [Chambers and Smith 1998]). Согласно этому принципу, предпочтительным

References

1. *Crawley R. A., Stevenson R. J., Kleinman D.* (1990), The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research*, Vol. 19 (4), pp. 245–64.
2. *Chambers C., Smyth R.* (1998). Structural Parallelism and Discourse Coherence. *Journal of Memory and Language*, 39, pp. 593–608.
3. *Fedorova O. V., Uspenskaia A. M.* Experimental Analysis of Discourse: The Impact of a Potential Referential Conflict on the Choice of the Referring Expression (On the Material of Russian) [Eksperimental'nyi Analiz Diskursa: Referentsial'nyi Vychor v Situatsii Potentsial'nogo Referentsial'nogo Konflikta (Eksperimental'noie Issledovaniie na Materiale Russkogo Iazyka)]. *Komp'juternaia Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii "Dialog 2011"* [Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2011". Bekasovo, 2011, pp. 207–18.
4. *Frederiksen J.* (1981), Understanding Anaphora: Rules Used by Readers in Assigning Pronominal Referents. *Discourse Processes*, Vol. 4, pp. 323–47.
5. *Gernsbacher M. A., Hargreaves D. J.* (1988), Accessing Sentence Participants: The Advantage of First Mention. *Journal of Memory and Language*, Vol. 27, pp. 699–717.
6. *Järvikivi J., van Gompel R. P., Hyönä J., Bertram R.* (2005), Ambiguous Pronoun Resolution: Contrasting the First-Mention and Subject-Preference Accounts. *Psychological Science*, Vol. 16 (4), pp. 260–4.
7. *Kibrik A. A.* (2011), *Reference in Discourse*. Oxford University Press, Oxford.
8. *Myachikov A., Tomlin R. S.* (2008), Perceptual Priming and Structural Choice in Russian Sentence Production. *Journal of Cognitive Science*, Vol. 9 (1), pp. 31–48.
9. *Piantadosi S. T., Tily H., Gibson E.* (2012), The Communicative Function of Ambiguity in Language. *Cognition*, Vol. 122 (3), pp. 280–91.
10. *Stepina M. G.* (2005), The Impact of Working Memory Span on Language Behaviour (on the Example of the Process of Resolving Anaphora) [Vliianie Ob'ema Operativnoi Pamiati Cheloveka na ego Iazykovoie Povedenie (na Primere Vyboru Antetsedenta Lichnogo Mestoimeniia)]. Unpublished degree thesis, Moscow State University.