

Проблема номинации при повторном упоминании референта: опыт решения на материале японского языка¹

Ефимова Зоя Викторовна

Настоящая работа посвящена проблеме выбора референциальных средств при повторном упоминании референта в дискурсе. Исследование проводится на материале устных и письменных нарративов на японском языке. Нами были получены следующие результаты. 1) Были выделены три наиболее частотных класса референциальных выражений в японском языке: однозначные именные группы без показателя темы ВА, однозначные именные группы с показателем темы ВА и референциальные нули. 2) Среди потенциально значимых для референциального выбора факторов были определены релевантные и отфильтрованы нерелевантные факторы. 3) Далее среди релевантных были выделены две группы факторов, влияющих на выбор референта, и показано, что реакция именных групп с ВА на первую группу сближает ИГ с ВА с референциальными нулями, а реакция на вторую группу сближает ИГ с ВА с полными именными группами. Первую группу факторов можно условно назвать “центральность в нарративе и локальная центральность”, т.е. с помощью нулевой анафоры или топикальных именных групп обозначаются референты, являющиеся главными героями, при этом в локальной структуре (предикации) референт имеет семантическую роль “Актор” и занимает синтаксическую позицию подлежащего. Факторы второй группы облегчают или усложняют интерпретацию референциального выражения. Например, специфическим для японского языка является фактор наличия бенефактивной конструкции, при которой предпочтительно использовать референциальные нули. Также к этой группе относятся факторы расстояния до антецедента (предыдущего упоминания), которое может измеряться в предикациях, предложениях, абзацах и т.п.

1. Введение

Настоящее исследование посвящено проблеме выбора референциальных средств при повторном упоминании референта в дискурсе. Эта проблема довольно хорошо изучена для языков европейского стандарта, однако для языков, типологически отличных от европейских, данная проблема исследована недостаточно. Поэтому в качестве материала исследования нами был выбран японский язык - топикальный язык левого ветвления.

В японистической литературе обсуждались некоторые факторы выбора референциальных средств в японском языке: частота упоминания референта [6], синтаксическая роль именной группы, отсылающей к референту [12], нахождение предыдущего упоминания референта в другом дискурсивном единстве (абзаце) [10, 11]. Тем не менее, до сих пор не было проведено комплексного исследования всех факторов, которые влияют] или могли бы повлиять на выбор референциальных средств в японском дискурсе, а также не был оценен их относительный вес. На восполнение этого пробела и была направлена настоящая работа.

В качестве материала исследования были использованы устные и письменные тексты на японском языке. Письменные тексты были представлены современными переложениями японских сказок. Образцы устных текстов были специально собраны для целей исследования и представляют собой пересказы текстов из

¹ Работа выполнена при поддержке гранта Минобразования России «Правила выбора референциальных средств в связном тексте» № А03-1.5-188 и гранта РГНФ «Средства организации дискурса в типологическом освещении: корпусные исследования структуры нарратива» №04-04-00220а.

письменной части корпуса. Было проанализировано десять текстов – шесть письменных и четыре устных текста (всего около четырехсот повторных упоминаний различных референтов).

2. Основные понятия

В своем докладе мы будем опираться на такие понятия, как референт, референциальная цепочка, antecedent, референциальный класс, элементарные дискурсивные единицы.

Референтом называется объект действительности, который обозначается данным языковым выражением. В процессе порождения связного текста говорящий обычно некоторые референты упоминает не один раз. Например, в (1) референт ‘ворона’ упоминается во всех трех предикациях.

(1)²

1. **Sono koro karasu_k wa taihenna oshare de,**
те времена ворона TOP ужасный франт COP.INF
В те времена ворона была ужасным франтом
2. **0_k itsumo masshiroī kimono o ki-te**
. всегда белоснежный одежда DO носить-INF
и всегда щеголяя в белоснежных одеждах,
3. **0_k tobi-arui-te i-mashi-ta.**
. летать-прогуливаться-INF быть-ADDR-PAST
летала и прогуливалась туда-сюда.

Все упоминания одного референта в тексте составляют референциальную цепочку. Фрагмент референциальной цепочки для референта *karasu* ‘ворона’ в примере (1) можно представить следующим образом.

1. karasu_k wa
2. 0_k
3. 0_k

Для каждого повторного упоминания референта в тексте предыдущие упоминания называются антецедентами.³ Так, для 0_k в (1).3 антецедентами являются, например, karasu_k wa в (1).1 и 0_k в (1).2.

В тексте референт может обозначаться словами из разных референциальных классов. Традиционно выделяются три основных референциальных класса: полные именные группы, местоимения, референциальные нули. Решение о том, какие референциальные классы существенно выделять для японского нарратива мы предложим в разделе 4.

Важным понятием для нашего исследования является понятие элементарной дискурсивной единицы (ЭДЕ) – минимального кванта порождения текста. ЭДЕ как правило совпадает с предикацией и была принята в качестве элементарной единицы текста в нашем корпусе. (1) состоит из одного предложения и разбивается на три ЭДЕ. Подробнее о проблемах разбиения текста на ЭДЕ см. в [1].

3. Теоретическая база. Многофакторная модель референциального выбора

В данной работе используется когнитивный подход, т.е. мы будем исходить из предположения о том, что языковые явления можно понять только как встроенные в другие когнитивные процессы. Явление референции связывается во многих исследованиях с активацией референта ([3], [7]).

² В примерах нижними индексами помечаются кореферентные именные группы. С помощью «0» с соответствующими индексами обозначаются референциальные нули.

³ Анафора, т.е. отсылка к линейно предшествующим упоминаниям референта – явление наиболее частотное. Однако бывают и случаи, когда референциальное выражение отсылает и к линейно последующему упоминанию референта. Такой вид отсылки называется катафорическим или катафорой.

Проблема понятия активации состоит в том, что оно отражает состояние мозга говорящего, и невозможно непосредственно измерить степень активации референта. Однако, в ([7], [8]) был предложен “когнитивный количественный подход”, позволяющий измерить степень активации референта опосредованно. В соответствии с этой моделью, коэффициент активации референта в некоторой точке дискурса можно выразить через значение целого ряда измеряемых факторов. Каждому из значений релевантных факторов приписан свой вес. По сумме этих весов для каждого упоминания референта можно вычислить коэффициент активации. Коэффициент активации, в свою очередь, определяет допустимость употреблений референциальных выражений для обозначения данного референта в данной точке дискурса.

Цель нашего исследования применить предложенную модель к материалу японского языка – найти факторы, определяющие референциальный выбор в японском нарративе, а также корреляцию значений факторов с типом используемого для обозначения референта выражения.

4. Референциальные классы

Прежде чем приступить к решению проблемы референциального выбора необходимо классифицировать референциальные выражения. В традиционной теории референции выделяются три класса референциальных выражений – полнзначные именные группы, местоимения и референциальные нули. Для языков европейского стандарта наиболее часто употребляемые референциальные классы – это полные именные группы и местоимения.⁴ В отличных от них языках, например, в языках дальневосточного ареала, довольно широко распространена нулевая анафора. В ряде работ по японскому дискурсу было показано, что местоимения в японском языке по синтаксическим и семантическим свойствам сближаются с полными именными группами (см., например, [6]).

В [6: 52] высказывается мнение, что употребление местоимений третьего лица не типично для традиционного нарратива: «traditional narratives never employ the forms *kare* and *kanojo* (‘он’ и ‘она’)». Данные нашего исследования подтверждают это предположение и для других местоимений, только в более мягкой форме: ненулевые местоимения используются в нарративах крайне редко. Так, из 426 упоминаний встретившихся более одного раза референтов на местоимения приходится всего лишь 14, что составляет около 3%. Для сравнения, на класс референциальных нулей приходится около 50% упоминаний референтов. Таким образом, наиболее существенным оказывается противопоставление полная именная группа – референциальный ноль. Ввиду того, что количество встреченных местоимений не дает статистически достоверных результатов, они пока что не будут фигурировать в нашем исследовании.

В связи с тем, что японский – это язык с формально выраженной темой, для данной работы целесообразно сделать более дробную классификацию референциальных выражений:

- полные именные группы без показателя темы;
- полные именные группы с показателем темы;
- референциальные нули.

Каждое вхождение в референциальную цепочку относится к одному из этих классов. Цель нашей работы – понять, какие факторы влияют на выбор между референциальными выражениями разных классов.

5. Факторы, влияющие на референциальный выбор

В этом разделе приводится общий список факторов, которые мы рассматривали как потенциально релевантные для выбора класса референциального выражения в японском языке. Этот список в значительной степени опирается на исходный список факторов, предложенный в [7], однако в связи с особенностями японского языка к списку был добавлен фактор «Информация в глаголе», а в интерпретацию некоторых факторов был внесен ряд изменений. Факторы разбиты на три группы.

5.1 Факторы, связанные со свойствами референта

Протагонизм. Данный параметр имеет два значения: *протагонист/не-протагонист*. Протагонистами признаются все центральные герои повествования (см., напр., [5], [7]).

⁴ Референциальные нули также встречаются в этих языках, однако чаще всего правила употребления этого класса референциальных выражений довольно хорошо описывается с помощью синтаксических правил.

Одушевленность. Этот фактор напрямую не отражает центральности референта в нарративе, однако, как правило, коррелирует с фактором «протагонизма»: протагонистами чаще всего бывают одушевленные референты. Значения фактора: *одушевленный/неодушевленный референт*.

Синтаксическая роль референциального выражения в предикации. Мы выделяем два значения синтаксической роли: *подлежащее, не-подлежащее*.

Семантическая гиперроль референта в предикации. У параметра «Семантическая роль» бывает два значения: *актор* и *не-актор*. *Актор* – это гиперроль, в которую могут попасть несколько семантических ролей в зависимости от свойств предиката. *Актором* считается агенс (если он есть), а если его нет, то – экспериенцер или пациенс.

5.2 Факторы, связанные со свойствами antecedента

Расстояния от данного упоминания референта до предыдущего его упоминания, antecedента, могут измеряться разными способами. Мы рассматриваем несколько факторов.

Линейное расстояние – количество ЭДЕ до ближайшего antecedента (значения 1 и больше 1).

Синтаксическое расстояние – количество предложений до ближайшего antecedента (0 и больше 0).

Риторическое расстояние – расстояние, учитывающее иерархическое устройство текста и тот факт, что одни единицы текста могут быть связаны более тесно, чем другие. Это расстояние измеряется в относительном расположении референта и его antecedента в так называемой риторической структуре (подробнее см. [9], [4]).

Синтаксическая роль antecedента и Семантическая гиперроль antecedента. Определения и множество значений этих факторов такие же, как и для этих же факторов для референтов.

Референциальный статус antecedента. Для каждого употребления референта мы проверяем, к какому референциальному классу относится его antecedент: *полная ИГ без wa, полная ИГ с wa, местоимение без wa, местоимение с WA, референциальный ноль*.

5.3 Факультативные факторы

Факторы этой группы отличаются от всех остальных параметров тем, что они встречаются спорадически. Для этих факторов мы считаем релевантным распределение референциальных средств при положительном значении, то есть при переходе к новому абзацу, наличии бенефактивной конструкции или неполной кореферентности.

Antecedent находится в другом абзаце – данный параметр в общем-то относится к расстояниям однако в силу его необязательности мы отнесли его к факультативным факторам. Данный фактор пока что определен только для письменных текстов – задача выделения аналога абзаца в устной речи является предметом отдельного исследования.

Информация в глаголе. Информация о референте не всегда находится в ИГ, к нему отсылающей, но может выражаться, в частности, в синтаксических конструкциях и в формах глагола, связанных с соответствующей ИГ в предикатно-аргументной структуре предложения. При анализе нашего корпуса текстов чаще всего встречались бенефактивные конструкции, поэтому мы включили в число рассматриваемых факторов фактор «информация в глаголе». В [2, с.90-91] было показано, что обсуждаемая конструкция различает главного героя повествования (который является как бы “своим” для автора) и второстепенных героев.

Неполная кореферентность. “Неполная кореферентность” – это случаи неполного совпадения референта данной ИГ с референтом ее antecedента. Между референтом antecedента ИГ и ее собственным antecedентом могут быть отношения: объект – родовое понятие, действующее лицо – группа, в которую входит действующее лицо, и т.п.

6. Результаты

6.1 Представление данных

Перед исследованием стояла задача с помощью корпусного метода среди факторов, описанных в 5, выделить релевантные и отфильтровать нерелевантные факторы, а также выяснить корреляцию выбора референциального класса со значением параметра.

Для этого для каждого упоминания референта, встречающего в тексте более одного раза, записывались значения факторов, и то, к какому классу относится данное упоминание референта. Далее подсчитывалось, сколько раз встретились значения некоторого фактора при определенном классе референциальных выражений. Так, например, среди полных ИГ с *ва* одушевленные референты встретились 47 раз, а неодушевленные всего 4 раза. В таблице 1 представлена процентная корреляция всех параметров, описанных в п.5, с каждым из классов референциальных выражений.⁵

Таблица 1. Корреляции дискурсивных параметров и классов референциальных выражений.

Параметры / классы референциальных выражений		Полная ИГ без <i>ва</i>	Полная ИГ с <i>ва</i>	Референциальный ноль
Количество в корпусе		151	51	210
<i>Одушевленность</i>	+	55 %	93 %	90 %
	-	45 %	7 %	10 %
Протагонизм	+	47 %	75 %	81 %
	-	53 %	25 %	19 %
<i>Синтаксическая роль ИГ</i>	подл.	42 %	96 %	82 %
	не-подл.	58 %	4%	18 %
<i>Семантическая гиперроль референта</i>	актор	44 %	93 %	81 %
	не-актор	56 %	7 %	19 %
Количество референциальных групп, имеющих antecedent		123	44	209
<i>Синтаксическое расстояние</i>	0	12 %	7 %	66 %
	>0	88 %	93 %	34 %
<i>Линейное расстояние</i>	1	26 %	25 %	66 %
	>1	74 %	75 %	34 %
<i>Риторическое расстояние</i>	1	54 %	54 %	79 %
	>1	46 %	46 %	21 %
Референциальный статус antecedenta	ИГ	56 %	30 %	28 %
	ИГ <i>ва</i>	2 %	8 %	11 %
	Ноль	38 %	58 %	58 %
<i>Синтаксическая роль antecedenta</i>	подл.	53 %	50 %	82 %
	не-подл.	47 %	50 %	18 %
<i>Семантическая гиперроль antecedenta</i>	актор	53 %	58 %	79 %
	не-актор	47 %	42 %	21 %
<i>Antecedent находится в другом абзаце</i>		+	72 %	11 %
<i>Информация в глаголе</i>		+	3 (17%)	1 (6%)
<i>Неполная кореферентность</i>		+	38(79%)	5 (10.5%)

Статистически существенным мы считаем уровень корреляции, превышающий 66%. Все процентные показатели преодолевшие этот уровень выделены в таблице жирным шрифтом.

По результатам полученных статистических данных можно сделать следующие выводы.

6.2. Интерпретация результатов

Релевантные/нерелевантные факторы

Как видно из таблицы, нерелевантным для референциального выбора оказался фактор «Референциальный статус antecedenta».

По остальным параметрам получены показательные распределения значений хотя бы для одного из референциальных классов. Тем не менее, можно выделить более сильные и менее сильные факторы.

⁵ Распределение значений факторов по референциальным классам для устных и письменных нарративов в целом совпадают, поэтому в приводимой статистике мы объединили результаты.

Так, среди расстояний наиболее важным оказалось синтаксическое расстояние. Этот результат интересен тем, что согласно [8] для русского нарратива наиболее существенными из расстояний являются риторическое и линейное расстояния. Таким образом, для референциальной структуры японского языка более существенной является граница предложения, нежели граница предикации.

Факторы, связанные со свойствами референта, дают показательное распределение только для референциальных классов «ИГ с *wa*» и «референциальные нули». При этом самыми сильными являются факторы «Одушевленность» и «Синтаксическая роль ИГ».

В дальнейшем полученные результаты можно будет использовать для построения более четкой иерархии факторов. Кроме того, использованный метод дает возможность типологического сравнения референциальных структур разных языков.

В целом, все релевантные факторы можно разбить на две группы.

Параметры, связанные со свойствами референта и его синтактико-семантической функцией. Объединяют референциальные типы ИГ с *wa* и нулей и отражают центральность референта для повествования и его центральность на уровне предикации. Значение фактора «Протагонизм» прямо указывает на центральность референта в рассказе, а семантическая и синтаксическая роли референта и отсылающей к нему ИГ говорят о центральности референта на уровне конкретного высказывания. Параметр «Одушевленность» коррелирует и с «Протагонизмом», т.к. протагонистами чаще всего бывают одушевленные референты, и с семантико-синтаксической функцией, так как для одушевленных референтов более типична активная роль в продвижении событий.

Таким образом, появление ИГ с *wa* и нуля связано с глобальной и локальной центральностью их референтов.

Параметры, связанные с расстояниями до антецедента и с его семантико-синтаксическими функциями. Объединяют референциальные типы ИГ без *wa* и ИГ с *wa*. Это множество параметров скорее связано с ограничениями на интерпретацию референциальных классов. Короткие расстояния и центральность антецедента для нулей необходимы для правильного выбора референта. Для полных ИГ обычно нет сложности с выбором референта, однако частое повторение этого класса референциальных выражений оказывается прагматически не оправдано из-за появляющейся речевой неэкономичности. Более того, появление полных ИГ при повторном употреблении в пределах одного предложения часто имеет другую функцию – обозначение неполной коррелентности с антецедентом.

Таким образом, параметры, отсылающие к предшествующему дискурсу, накладывают довольно существенные ограничения на употребление нулей на не минимальных расстояниях, а также некоторые ограничения на употребление полных ИГ на минимальных расстояниях.

Факультативные факторы

Они отличаются от остальных параметров тем, что они не определены для большинства случаев. Это скорее параметры-фильтры, на наличие которых нужно проверять каждое употребление. Информация в глаголе облегчает употребление нулей, а появление неполной коррелентности или требует употребления полных именных групп, или обозначается с помощью лексических средств в других фрагментах фразы.

Интерпретация результатов исследования схематично изображена в блок-схеме 1.

Литература

1. Кибрик А.А., Подлеская В.И. К созданию корпусов устной русской речи: принцип транскрипции. // НТИ, Серия 2, №10. М., 2003.
2. Холодович А.А. Проблемы грамматической теории. Ленинград, 1979.
3. Chafe, Wallace. Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing. Chicago: University of Chicago Press, 1994.
4. Fox, Barbara. Discourse structure and anaphora. Cambridge: Cambridge University Press, 1987.
5. Givón, Talmy. Syntax: A functional-typological introduction. Vol.2. Amsterdam: Benjamins, 1990.
6. Hinds, J. Topic continuity in Japanese. // Topic continuity in discourse: a quantitative cross-linguistic study (Typological studies in language, vol. 3). – Amsterdam: Benjamins, 1983.

7. *Kibrik, Andrei A.* Anaphora in Russian narrative discourse: A cognitive calculative account. // *Studies in Anaphora* / ed. Barbara Fox. Amsterdam: Benjamins, 1996.
8. *Kibrik, Andrei A.* Cognitive inferences from discourse observations: Reference and working memory. // *Discourse studies in cognitive linguistics. Proceedings of the 5th International cognitive linguistic conference.* / Ed. Karen van Hoek, Andrei A. Kibrik, and Leo Nordman, P.29-52. Amsterdam: Benjamins, 1999.
9. *Mann, William, and Sandra A. Thompson.* Rhetorical structure theory: toward a functional theory of text organization. // *Text*, №8, 1988. P. 243-281.
10. *Maynard, Senko.* Principles of Japanese discourse: a handbook. Cambridge: Cambridge University Press, 1998.
11. *Nomoto, Tadashi, and Yoshihiko Nitta.* Resolving zero anaphora in Japanese. // *Sixth conference of the European chapter of the Association for computational linguistics: Proceedings of the conference.* Utrecht, 1993. P. 315-321.
12. *Walker, Marilyn A., Masayo Iida and Sharon Cote.* Japanese discourse and the process of centering. // *Computational Linguistics*, №20 Vol. 2, 1994. P.193-232.

Блок-схема 1

