

Автоматическое разрешение анафоры в русском тексте: случай нулевого субъекта

Анна Маракасова
НИУ ВШЭ, ИППИ РАН

Аннотация

Модуль семантического анализа системы ЭТАП-3 предполагает разрешение анафоры в два этапа: на начальном этапе обрабатываются случаи так называемой грамматической анафоры, а далее – сложные случаи, требующие умения производить логический вывод. Данная статья посвящена частной проблеме первого этапа, а именно проблеме поиска антецедента в случае нулевого субъекта.

1. Введение¹

Автоматическое распознавание кореферентных цепочек, или в других терминах, разрешение анафоры, является неотъемлемой составляющей систем семантического анализа текста (вопросно-ответных систем, систем извлечения информации, автоматического реферирования, семантического поиска). Так, при решении задачи извлечения фактов из текста необходимо установить все различные упоминания одного субъекта (т.е. соотнести все именованные сущности с объектами реального мира), чтобы корректно определить состав участников ситуации и их роли.

Методы разрешения анафоры в английском и некоторых других языках активно разрабатываются на протяжении последних двадцати пяти лет (отчасти это объясняется наличием для этих языков корпусов с анафорической разметкой). Современные системы в основном используют алгоритмы машинного обучения и включают следующие этапы: 1) поиск всех единиц, которые могут являться членом кореферентной цепочки (часть

работ посвящена отдельно выявлению объектов, которые упомянуты в тексте только один раз (singletons), с тем чтобы затем можно было исключить их из списка возможных антецедентов); 2) поиск анафора (в первую очередь местоимения) и его наиболее вероятного антецедента (\approx независимо употребленного слова, указывающего на объект действительности) [Poesio, Ponzetto & Versley 2010]. В качестве признаков для обучения используются различные морфологические, синтаксические и семантические характеристики, а также признаки, приписываемые системами распознавания именованных сущностей.

В силу особенностей грамматики русского языка методов разрешения анафоры в английских текстах (в частности, разработанных наборов признаков для классификатора) недостаточно для разрешения анафоры в русских. Во-первых, в русском языке нет категории определенности/неопределенности, которая позволяла бы с приемлемой степенью надежности определить, упоминается ли объект в первый раз или нет. Во-вторых, в русском языке относительно свободный порядок слов, так что, например, местоимение *свой* часто оказывается в препозиции к своему антецеденту (случай так называемой катафоры): *Свои фотографии Петров никому не показывает* [Toldova et al. 2015]. В-третьих, субъект в русском языке может иметь нулевую реализацию. В связи с этим для русского и других языков pro-drop со свободным порядком слов представляется оправданным использование синтаксиса зависимостей (а не составляющих, как до последнего времени было принято для английского языка) и правилых или гибридных алгоритмов [Dakwale, Mujadia & Sharma 2013, Bogdanov et al 2014, Kutuzov & Ionov 2014]. Машинное обучение также применяется, но, как правило, дает низкую полноту при сравнительно высокой точности –

¹ Настоящая работа выполнена при поддержке гранта РФФИ (№16-18-104220).

порядка 60-70% [Топегин, Ветров & Кропотов 2006, Мальковский, Старостин & Шилов 2013, Kamenskaya, Khramoin & Smirnov 2014].

2. Разрешение анафоры и семантический анализ в системе ЭТАП-3

Одним из способов проверить, что система способна осуществлять качественный семантический анализ текста (т.е. способна “понимать текст”), являются так называемые схемы Винограда [Levesque, Davis & Morgenstern 2012]. Они представляют собой специальным образом подобранные предложения с вопросами. Вопросы составлены так, что для ответа на них требуется понимание значения предложения (для человека) или умение производить логический вывод (для машины).

Фактически задание состоит в идентификации antecedenta анафорического местоимения: *Ann asked Mary what time the library closes, [but/because] she had forgotten. Who had forgotten?* (ответ: [Mary (в первом варианте) / Ann (во втором варианте)]). Однако вопросы исключают возможность разрешения анафоры с помощью статистического анализа размеченного корпуса или с использованием онтологического знания. Так, предложение *The women stopped taking pills because they were [pregnant/carcinogenic]*, в отличие, например, от предложения *The city councilmen refused the demonstrators a permit because they [feared/advocated] violence*, не стоит включать в список предложений схемы Винограда, поскольку установить референта местоимения they можно проще – на основании представления о том, что женщины никогда не вызывают рак, а таблетки не бывают беременными. В русском языке примером такого случая может послужить известное предложение *Мы дали обезьянам бананы, потому что они были [неспелые/голодные]*.

Заметим, что в некоторых контекстах неоднозначность they при переводе на другой язык либо сохраняется (как в русском, немецком, хинди и др.), либо нивелируется (как во французском, исландском, арабском и др. языках, различающих две и более формы личного местоимения множественного числа); см., например, предложение Т. Винограда: “The city councilmen refused to give the women a permit for a demonstration because they [feared/advocated] violence” [Winograd 1972].

Данная работа посвящена разрешению анафоры на начальном этапе семантического анализа в системе ЭТАП-3 (Iomdin et al. 2012, Богуславский и др., 2015). Распознавание анафорических связей производится с помощью специальных правил и основывается на готовой

синтаксической структуре предложения. Такие правила пока не разрешают неоднозначность antecedenta в том виде, как того требуют схемы Винограда – это возможно лишь на завершающей стадии семантического анализа. Сейчас в неоднозначных ситуациях выбор делается на основании данных статистического анализа. Например, близость предполагаемого antecedenta к анафору и одушевленность (90% местоимений ссылаются на одушевленный объект) повышают его вероятность действительно быть antecedentом [Мальковский, Старостин & Шилов 2013]. Проблема разграничения личной глагольной формы во множественном числе и имперсонального употребления глагола в неопределенно-личных предложениях также остается нерешенной: *Недавно минчане восстановили ратушу, которую снесли по личному приказу «распорядителя» Николая I, поскольку она, мол, «напоминала о былых вольностях»*, (минчане снесли или другие снесли) но и сомнительно, что её возможно решить компьютерными методами.

3. Правила разрешения анафоры и восстановления нулевого субъекта

Анализ корпусов с размеченными анафорическими и кореферентными связями показал, что существенный процент предложений в языках, допускающих нулевое подлежащее (pro-drop languages), не имеет выраженного субъекта: в испанском языке — 41,17%, в китайском — 36%, в итальянском — 30,42%, в польском — около 30% [Zhao & Ng 2007, Russo, Loáiciga & Gulati 2012, Корец 2014]. Поэтому для таких языков обнаружение и восстановление нулевых анафорических местоимений является важным предварительным этапом в решении задач разрешения анафоры и кореферентности. Под нулевым анафорическим местоимением (pro) мы понимаем отсутствующее местоимение, свободно заменяемое на ненулевое. Как было показано на материале польского языка в Корец 2014, качество системы распознавания кореферентности без использования модуля распознавания нулевых местоимений падает с 82,9% до 38,13%. Учитывая данный факт, ряд правил системы ЭТАП-3 предполагает не просто нахождение antecedenta того или иного личного местоимения, а выявление сначала контекста не наличествующего, но подразумеваемого местоимения и уже затем поиск его antecedenta.

При создании модуля семантического анализа текста, в отличие, например, от русско-английского машинного перевода, помимо pro необходимо восстанавливать и субъект нефинитных клауз (причастных, деепричастных и инфинитивных оборотов) – PRO, поскольку в

семантике важно получить представление, содержащее информацию о субъекте каждого действия. Для машинного перевода в указанной языковой паре восстановление PRO в целом не является обязательным. Так, предложение *Родители попросили сына, Φ_i вернуться домой вовремя* может быть адекватно переведено на английский с помощью соответствующего правила перевода без восстановления нулевого подлежащего: *The parents asked their son to come back home in time* (а, например, предложение *Родители попросили сына, чтобы Φ^{3sg}_i вернулся домой вовремя* требует такого восстановления: *The parents asked their son that HE should come back home in time*).

Далее мы представим два правила восстановления pro и PRO. С помощью этих же правил разрешается анафора в сочинительных, сравнительных и подчинительных конструкциях с наличествующим субъектом.

3.1. Правило восстановления субъекта в сочинительных и сравнительных конструкциях

Данное правило восстанавливает субъект:

(1) В союзных сложносочиненных предложениях с различными по значению союзами: противительными (*Его отец_i владел несколькими успешными фабриками, но [он_i] был не в состоянии оплатить скопившиеся долги*), градационными (*На каникулах (Маша с Катей)_i не только весело проводят время, но и [(Маша с Катей)_i] узнают много нового*), союзом «компенсации» (*Греция_i выполнила условия для вступления в зону евро, зато [она_i] подорвала доходную базу бюджета на годы вперед*) и др.

(2) В бессоюзных сочинительных предложениях, где с помощью наречий, частиц и предложно-падежных форм могут вноситься дополнительные оттенки значения: следствия (*Я_i никогда не отстаивал подобные интересы, поэтому [я_i] могу себе позволить не брать их в расчет*), ограничительности (*Карина_i хорошо подготовилась к юбилею, вот только [она_i] про меня позабыла*), присоединения (*Читатель_i ошибётся, при этом [он_i] сочтёт весь этот эпизод за случайный ляпсус*) и др.

(3) В некоторых сравнительных конструкциях, в том числе:

– в предложениях, в которых сравнительный союз (*как, словно, будто, как будто, точно, как если бы*) напрямую подчиняется предикату главной клаузы: *Она_i ответила грубо, словно [она_i] обиделась, Без введения стопроцентной оплаты коммунальных услуг он_i никакой холдинг создавать бы не стал, как [он_i] не стал бы вкладывать в этот планово-убыточный бизнес.*

– в предложениях с союзами *чем* и *нежели*, в которых зависимая клауза подчиняется сравнительной степени наречия или прилагательного главной клаузы: *Ловкий теннисист_i иногда зарабатывает за одну съемку больше, чем [теннисист_i] может получить, выиграв престижный турнир, Доход_i по-прежнему примерно раз в пять меньше, нежели [доход_i] был до 1992 года.*

3.2. Правило восстановления субъекта в нефинитных конструкциях

Правило обрабатывает, с одной стороны, нулевые подлежащие в причастных, деепричастных и инфинитивных оборотах, с другой стороны, нулевые анафорические местоимения в относительных, причинных, временных, условных, уступительных конструкциях, конструкциях меры и степени и некоторых других.

В результате работы первых двух подправил связь между причастием и управляемым им словом стирается, но устанавливается противоположно направленная связь, которая определяется залогом: предикативная или 1-я комплетивная. Рассмотрим предложения *На следующий день президент подписывает Указ, позволяющий полиции производить аресты без судебного контроля* и *Конструкция имеет "интеллектуальную" начинку и корпус, легко трансформируемый пользователем*. В первом случае в семантической структуре появляется предикат *позволять*, субъектом которого является *Указ*, а во втором случае – появляется предикат *трансформировать*, прямым объектом которого является *корпус*.

На рисунке 1 представлена предварительная семантическая структура первого из рассмотренных предложений. В первой колонке указаны вершины-хозяева, во второй колонке узлы, которые подчиняются этим вершинам; в последней колонке указано имя отношения, соединяющего главное слово (вершину) с зависимым. Номера соответствуют порядковому номеру словоформы в предложении, что также позволяет различать. Как видно из рисунка, существительное *указ*, с одной стороны, является прямым дополнением при глаголе *подписывать*, а с другой – подлежащим при глаголе *позволять*.

Подправила на деепричастия добавляют между субъектом некоторой клаузы и зависимым от её вершины деепричастием связь, которая, как и в случае с причастиями, зависит от залога: предикативная (*До сих пор депутаты лишь дискредитировали себя, обращаясь к проблемам IT-индустрии*) или 1-я комплетивная (*Леды столкнулись с проблемой, решить которую*

1.1 НА1	3.3 ДЕНЬ	▲	ПРЕДЛ
3.3 ДЕНЬ	2.10 СЛЕДУЮЩИЙ	▲▲	ОПРЕД
5.1 ПОДПИСЫВАТЬ	1.1 НА1		ОБСТ
	4.1 ПРЕЗИДЕНТ	▲	ПРЕДИК
	6.2 УКАЗ1	▲	1-КОМПЛ
7.2 ПОЗВОЛЯТЬ	6.2 УКАЗ1		ПРЕДИК
	8.2 ПОЛИЦИЯ	▲	2-КОМПЛ
	9.1 АРЕСТОВЫВАТЬ	▲	1-КОМПЛ
9.1 АРЕСТОВЫВАТЬ	8.2 ПОЛИЦИЯ		ПРЕДИК
	10.1 БЕЗ		АТРИБ
10.1 БЕЗ	12.1 КОНТРОЛЬ	▲▲	ПРЕДЛ
12.1 КОНТРОЛЬ	11.1 СУДЕБНЫЙ	▲▲	ОПРЕД

Рисунок 1.

можно было, либо перейдя на другой язык, либо подвергнувшись полному уничтожению).

Следующие четыре подправила нацелены на обработку инфинитивных оборотов. Первые два из них в качестве агента действия, названного инфинитивом, выбирают субъект глагола, им управляющего: *Начальник управления собирался позвонить Савельеву, но не успел, Он не был незаменимым, биржа не лопнула, когда ему пришлось уйти* (предикативным отношением связываются *начальник* и *позвонить*, *он* и *уйти* соответственно); а два других – выбирают объект глагола: *Вы взяли мою анкету и предложили мне прийти сегодня, Из-за этого в прошлую среду Семен Еремеевич и попросил её зайти через неделю* (предикативным отношением связываются *я* и *прийти*, *она* и *зайти* соответственно).

3.3. Правило восстановления субъекта в подчинительных конструкциях

Данное правило описывает различные по семантике подчинительные конструкции, предложения с сентенциальными актантами и связывает предикативным отношением подлежащее главной клаузы с вершиной зависимой. Правило не восстанавливает субъект там, где предикат зависимой клаузы имеет субъект, в том числе неканонический, как, например, в *Юля забыла в гостях телефон, без которого ей и дня не прожить*. Конструкции представлены следующими вариантами:

1) Предложения с подчинительными союзами *когда, если, потому что* и др.: *Человек_i может*

совершать многое, если [человек_i] посвятит себя труду на избранном поприще, Александр_i не желал пятиться, хотя [Александр_i] не причислял себя к друзьям прогресса, Косметологи_i обожают растительные масла, потому что [косметологи_i] знают о них все, Я_i позвонил начальнику, как только [я_i] добрался до назначенного места;

2) Предложения, в которых придаточное - сентенциальный актант напрямую подчинено вершине главной части: *Юля_i не могла и представить, вместе с кем [Юля_i] будет работать, Он_i почувствовал, как [он_i] от нее устал, Миша_i; пока не знает, пойдет [Миша_i] с нами в театр или нет, Дедушка_i; уже не помнил, записывался ли [дедушка_i] на прием;*

3) Предложения с относительными придаточными: *Зять_i чувствовал цену благополучия, которым [зять_i] наслаждался, Новый руководитель_i приводит только тех, в кого [руководитель_i] по-настоящему верит, Врач_i хочет встретиться с пациентом, которому [врач_i] назначил операцию, Мы_i прибыли в Санкт-Петербург, где [мы_i] хотели бы и остаться жить навсегда.* Этим подправилом также связывается анафорической связью союзное слово и его антецедент. На рисунке 2 изображена предварительная семантическая структура предложения *Зять чувствовал цену благополучия, которым наслаждался*. В представленной на рисунке схеме уже отсутствует союзное слово *который*, оно заменено на свой антецедент – *благополучие*; а при предикате зависимой клаузы восстановлено подлежащие *зять*.

4) Предложения с придаточными меры и

2.1 ЧУВСТВОВАТЬ	1.1 ЗЯТЬ	▲▲	ПРЕДИК
	3.1 ЦЕНА	▲	1-КОМПЛ
3.1 ЦЕНА	4.1 БЛАГОПОЛУЧИЕ	▲	1-КОМПЛ
4.1 БЛАГОПОЛУЧИЕ	6.1 НАСЛАЖДАТЬСЯ		РЕЛЯТ
6.1 НАСЛАЖДАТЬСЯ	1.1 ЗЯТЬ		ПРЕДИК
	5.2 БЛАГОПОЛУЧИЕ	▲	1-КОМПЛ

Рисунок 2.

степени; в роли опорного слова выступают наречия меры и степени (*настолько, так*): *Многие банки_i настолько осмелели, что [банки_i] сами выходят с кредитными предложениями на клиентов, Он_i переживал так отчаянно, словно [он_i] признался ему в сотрудничестве с французской разведкой, Этот удар_i был так удачно рассчитан, что [удар_i] прошел без малейшей ошибки*; указательное местоимение-прилагательное *такой*: *Петя_i был такой глупый, что [Петя_i] не понимал очевидных вещей, Акция_i оказалась такой успешной, что [акция_i] была продлена еще на неделю.*

Наконец, последним подправилом восстанавливается субъект, который соотносится с объектом действия предиката главной клаузы (а не субъектом, как в предыдущих случаях). Зависимая клауза является сентенциальным актантом предиката, который может быть как в личной форме: *Мама наказала сыну_i, чтобы [сын_i] разогрел обед сам*, так и в форме повелительного наклонения: *Напомните Катю_i, чтобы [Катя_i] не забыла купить цветы.*

3.4. Восстановление неоднократно опущенного субъекта

Правила и подправила применяются в такой последовательности, что система успешно восстанавливает субъект, опущенный сразу в нескольких клаузах (но в рамках одного предложения): *Иван_i долго не соглашался, но потом [Иван_i] все-таки нарисовал портрет, хотя [Иван_i] и опоздал с доставкой адресату, Миша_i думал, что когда [Миша_i] уходил, [Миша_i] сказал ей_j, чтобы [она_j] приходила завтра.*

Поскольку задача разрешения анафоры является подзадачей разрешения кореферентных связей, мы считаем, что все личные местоимения, antecedent которых был найден с помощью правил, должны заменяться на него в семантической структуре предложения. Таким образом мы сокращаем количество неоднозначно интерпретируемых элементов в кореферентной группе и несколько упрощаем выбор её вершины. Выше рассмотренные правила при обнаружении нулевого подлежащего сразу используют в семантической структуре узел, соответствующий

его antecedенту. Для случаев, когда в рамках одного предложения установлена анафорическая связь, а анафором является личное местоимение, специальное правило удаляет узел, соответствующий личному местоимению, и заменяет его на узел, соответствующий antecedенту.

Это правило также применяется для одного типа конструкций с сентенциальным актантом, где опущенное анафорическое местоимение восстанавливается на этапе синтаксического, а не семантического анализа: *Наш разведчик так и не смог понять, что [он] ошибся в решающий момент.* На данном этапе работы в кореферентной группе остаются только личные местоимения, antecedent которых определяется по широкому контексту (т.е. за рамками исходного предложения) или по ситуации.

На рисунке 3 приведена предварительная семантическая структура предложения *Когда уходила, Юля наказала сыну, чтобы разогрел обед сам.* В этом примере восстанавливается нулевой субъект в двух различных конструкциях, причем в одной из них мы имеем случай не анафоры, а катафоры. Разрешение катафоры осуществляется теми же правилами, что и разрешение анафоры. Это возможно благодаря тому, что правила в основном основаны на синтаксической структуре предложения и не используют информацию о линейном порядке следования словоформ в предложении.

На рисунке 4 – структура предложения *Лиза забыла, что она попросила Мишу купить цветы.* Здесь восстанавливается субъект инфинитивной конструкции, а также разрешается местоименная анафора. На схеме можно видеть, что теперь в семантической структуре предложения нет первоначального местоимения *она* – оно заменено на свой antecedent *Лиза*.

Отметим, что для случаев эллиптического сокращения подлежащего главной клаузы (характерных в русском языке преимущественно разговорной речи) имеется синтаксическое правило, восстанавливающее субъект при глаголах 1-го и 2-го лица (субъект при других глагольных формах невозможно однозначно восстановить на синтаксическом уровне): [Я] *думаю, что ему стоит обратиться к врачу.*

1.2 КОГДА2	2.1 УХОДИТЬ1	ПОДЧ-СОЮЗН
2.1 УХОДИТЬ1	3.2 ЮЛЯ	ПРЕДИК
4.2 НАКАЗЫВАТЬ2	1.2 КОГДА2	ОБСТ
	3.2 ЮЛЯ	▲▲ ПРЕДИК
	5.1 СЫН	▲ 2-КОМПЛ
	6.1 ЧТОВЫ	▲ 1-КОМПЛ
6.1 ЧТОВЫ	7.1 РАЗОГРЕВАТЬ	ПОДЧ-СОЮЗН
7.1 РАЗОГРЕВАТЬ	5.1 СЫН	ПРЕДИК

Рисунок 3.

2.1 ЗАБЫВАТЬ	1.1 ЛИЗА	▲	ПРЕДИК
	3.1 ЧТО1	▲	1-КОМПЛ
3.1 ЧТО1	4.1 ПРОСИТЬ1		ПОДЧ-СОЮЗН
4.1 ПРОСИТЬ1	1.1 ЛИЗА		ПРЕДИК
	5.1 МИША	▲	1-КОМПЛ
	6.1 ПОКУПАТЬ1	▲	2-КОМПЛ
6.1 ПОКУПАТЬ1	5.1 МИША		ПРЕДИК
	7.2 ЦВЕТЫ	▲	1-КОМПЛ

Рисунок 4.

Поиск антецедента таких восстановленных местоимений требует специальных методов обнаружения анафорических связей, учитывающих контекст нескольких предыдущих предложений. Решение этой проблемы, а также обработка с помощью онтологии и результатов семантического анализа случаев неоднозначного выбора антецедента составляют предмет дальнейшей работы.

4. Заключение

В данной работе был представлен алгоритм автоматического разрешения анафоры для случаев, когда анафором является нулевой субъект (а именно, про или PRO). Кроме того, разработанные правила разрешают анафору субъекта в сочинительных, сравнительных и различных по семантике подчинительных конструкциях. В относительных придаточных предложениях союзное слово также соединяется со своим антецедентом.

Литература

Богуславский И.М., Диконов В.Г., Иомдин Л.Л., Лазурский А.В., Сизов В.Г., Тимошенко С.П. 2015. Семантический анализ и ответы на вопросы: система в стадии разработки. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». Вып. 14 (21): В 2 т. Т. 1: Основная программа конференции. — М.: Изд-во РГГУ, 2015, с. 62-79.

Мальковский М.Г., Старостин А.С., Шилов И.А. 2013. Метод разрешения местоименной анафоры в процессе синтаксического анализа. В журнале *Сборник научных трудов SWorld по материалам международной научно-практической конференции*, том 11, № 4, с. 41-49.

Толпегин П.В., Ветров Д.П., Кропотов Д.А. Алгоритм автоматизированного разрешения анафоры местоимений третьего лица на основе методов машинного обучения. Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» / Под ред. Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея. — М.: Изд-во РГГУ, 2006, с. 504-507.

Bogdanov, A. V., Dzhumaev, S. S., Skorinkin, D. A., Starostin, A. S. 2014. Anaphora analysis based on ABBYY Compreno linguistic technologies. In proceedings of the International Conference on Computational Linguistics and

Intellectual Technologies Dialogue-2014. Issue 13 (20). М. 2014. P. 218-231.

Dakwale, P., Mujadia, V., Sharma D. 2013. A Hybrid Approach for Anaphora Resolution in Hindi. In Proceedings of International Joint Conference on Natural Language Processing (IJCNLP) 2013.

Iomdin L.L., Petrochenkov V.V., Sizov V.G., Tsinman L.L. 2012. ETAParser: state of the art, Computational Linguistics and Intellectual Technologies. International Conference (Dialog'2012). Moscow: RGGU Publishers, Issue11(18). pp. 830–843.

Kamenskaya M.A, Khrainin I.V., Smirnov I.V. Data-driven methods for anaphora resolution of russian texts //Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2014). Issue 13 (20). – 2014. – pp. 134-136.

Kutuzov A. B., Ionov M. The impact of morphology processing quality on automated anaphora resolution for Russian. In proceedings of the International Conference on Computational Linguistics and Intellectual Technologies Dialogue-2014. Issue 13 (20). М. :., 2014. P. 232-241.

Levesque, H., Davis, E., Morgenstern, L. The Winograd Schema challenge. In proceedings of 13th the International Conference on the Principles of Knowledge Representation and Reasoning, KR 2012. – pp. 552-561.

Poesio, M., Ponzetto, S. & Versley, Y. 2010. Computational models of anaphora resolution: A survey.

Russo, L., Loáiciga, Sh. & Gulati, A. Improving machine translation of null subjects in Italian and Spanish // Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2012. – pp. 81-89.

Toldova S., Lyashevskaya O., Bonch-Osmolovskaya A. A., Ionov M. Evaluation for morphologically rich language: Russian NLP, in: Proceedings on the International Conference on Artificial Intelligence (ICAI) Vol. 1. Las Vegas : CSREA Press, 2015. P. 300-306.

Winograd, T. *Understanding Natural Language*, (191 pp.) New York: Academic Press, 1972. Also published in *Cognitive Psychology*, 3:1 (1972), pp. 1-191.

Zhao, Sh & Ng, H. T. Identification and Resolution of Chinese Zero Pronouns: A Machine Learning Approach // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007). Prague, Czech Republic, 2007. – pp. 541 – 550.