

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ПЕДАГОГІЧНИЙ УНІВЕРСИТЕТ
імені М.П. ДРАГОМАНОВА

СИСТЕМА І СТРУКТУРА
СХІДНОСЛОВ'ЯНСЬКИХ МОВ:
ПАМ'ЯТІ АКАДЕМІКА Л.А. БУЛАХОВСЬКОГО

Збірник наукових праць

НАЦІОНАЛЬНА
ПАРЛАМЕНТСЬКА
БІБЛІОТЕКА
УКРАЇНИ

КІЇВ
ЗНАННЯ УКРАЇНИ
2006

29086888
"ІІІ" ЧУ.

Редакційна колегія:

В.І. Гончаров, канд. філол. наук, професор (відповідальний редактор);
А.М. Григораш, канд. філол. наук, доцент (відповідальний секретар);
М.Я. Брицин, докт. філол. наук, професор;
Л.П. Іванова, докт. філол. наук, професор;
Ф.О. Нікітіна, докт. філол. наук, професор;
Н.І. Озерова, докт. філол. наук, професор;
М.Я. Плющ, докт. філол. наук, професор;
О.О. Тараненко, докт. філол. наук, професор;
Л.С. Віннічук, канд. педагог. наук, доцент;
О.І. Дащенко, канд. філол. наук, доцент;
М.П. Дворжецька, канд. філол. наук, професор;
Т.І. Плужнікова, канд. філол. наук, доцент;
Т.В. Слободянюк, канд. філол. наук, доцент;
М.М. Тягунова, канд. філол. наук, доцент;
Н.В. Хруцька, канд. філол. наук, доцент.

Рецензенти:

Н.Л. Іваницька, докт. філол. наук, професор;
Є.А. Карпіловська, докт. філол. наук;
Т.Ю. Ковалевська, докт. філол. наук;
Л.М. Полюга, докт. філол. наук, професор;
О.А. Сербенська, докт. філол. наук, професор.

С40 Система і структура східнослов'янських мов: Пам'яті академіка Л.А. Булаховського: Зб. наук. праць (Редкол.: В.І. Гончаров (відп. ред.) та ін. — К.: Знання України, 2006. — 316 с. Бібліогр. в кінці ст. — Укр., рос.
ISBN 966-7201-26-0

У збірнику вміщено наукові праці докторантів, аспірантів, професорсько-викладацького складу мовознавчих кафедр Національного педагогічного університету ім. М.П. Драгоманова та інших вузів України з актуальних проблем східнослов'янського мовознавства. Тематика статей охоплює питання граматичної будови і лексичного складу східнослов'янських мов, етимології і словотвору, історії і взаємодії мов, культурології і стилістики. Об'єктом лінгвістичного аналізу виступають структурні одиниці різних рівнів мовної системи. У полі зору мовознавців питання фонетики, морфології, синтаксису, лексикології, термінології та лексикографії.

Для викладачів, аспірантів, студентів філологічних факультетів, учителів української та російської мов середніх шкіл та гімназій.

ББК 81.2я43

ISBN 966-7201-26-0

© Національний педагогічний університет
імені М.П. Драгоманова, 2006
© Автори статей, 2006

валентным для слова *вьюга* являются: *веялица* — болгарское¹, *вејавица*, *виулица* — македонское², *вејавица* — сербохорватское³, *zawieja* — польское⁴. Слово *завіруха* употребляется в украинском и белорусском языках, кроме того, известно функционирование *завія* в украинском языке, *завея* — в белорусском.

На основе наших наблюдений можно сделать краткие выводы:

1. Развитие безличных предложений однокомпонентной структуры со значением «стихийные явления природы» отличается неравномерностью в славянских языках. Принимая при анализе за исходный русский язык, мы отметили, что из 7 семантических групп 5 имеют соответствующие лаголы в тех или других славянских языках: *ветреть*, *грешить*, *дуть*, *мести*, *штормить*. Глагол *вьюжить* характерен только для русского языка. С другой стороны, в русском языке отсутствует глагол в безличной форме со значением 'сверкает молния', за исключением употребления в неоднокомпонентных конструкциях.

2. Однокомпонентным предложениям с безличным употреблением глагола соответствуют эквивалентные или синонимичные конструкции: а) однокомпонентные предложения с безлично-предикативными словами: *ветреет* — *ветрено*; *вьюжит* — *вьюжно*; б) номинативные предложения: *ветреет* — *ветер*; *грешит* — *гром*; *метёт* — *метель*; *вьюжит* — *вьюга*; *штормит* — *шторм*; в) двусоставные предложения с подлежащим, выраженным существительным того же корня, который имеет глагол в безличной форме: *ветрело* — *был ветер*; *грешело* — *был гром*; *мело* — *была метель*; *вьюжило* — *была вьюга*; *штормило* — *был шторм*.

В некоторых случаях глагол и существительное образуют тавтологические сочетания: *ветер веет*; *метель метёт*; *гром гремит*.

3. Сопоставление русского языка с другими славянскими языками, а также переводы словарных единиц с одного языка на другой позволяют выявить эквивалентные и синонимичные конструкции в пределах того или другого языка.

А.-Р. Валипур

СИНТАКСИЧЕСКИЕ И СЕМАНТИЧЕСКИЕ СВЯЗИ ЧЛЕНОВ ПРЕДЛОЖЕНИЯ ПРИ МАШИННОМ ПЕРЕВОДЕ

Проблема глубинного и поверхностного синтаксиса, т. е. соотношение семантической и синтаксической структуры предложения, соотношение семантических синтаксических актантов изучается лингвистами в разных областях языкознания, в том числе в прикладной лингвистике. Системы машинного перевода обеспечивают связный перевод текстов, учитывающий морфологические, синтаксические и семантические связи членов предложения. Сегодня программы-переводчики умеют строить осмысленные фразы, и за последние несколько лет качество перевода улучшилось. Однако компьютер еще плохо разбирается в грамматических нюансах и жаргоне, поэтому его главное назначение — переводы деловых бумаг, руководств, писем из электронной почты, страничек из Интернета. Другое применение систем — облегчение рутинной работы переводчиков, выполнение за них подстрочного перевода, который

¹ Бернштейн С.Б. Болгарско-русский словарь. — М., 1953.

² Толковский Д., Иллич-Свитыч В.М. Македонско-русский словарь / Под ред. Н.И. Толстого. — М., 1963.

³ Иванович С., Петранович И. Русско-сербохорватский словарь. — М., 1976.

⁴ Большой русско-польский словарь. — М.; Варшава, 1970.

можно легко подправить. Для улучшения качества машинного перевода необходимо продолжать исследование в синтаксисе и семантике предложения при машинном переводе.

Синтаксический анализ при машинном переводе предназначен для:

- разрешения морфологической омонимии слов;
- установления связей между словами-членами предложения;
- определения синтаксических категорий членов предложения.

Синтаксические зависимости между членами предложения отражают связи между отдельными понятиями, о которых идет речь в тексте, следовательно, определение связей в тексте необходимо для адекватного описания действия. На этапе собственно синтаксического анализа последовательность текстовых единиц, который совершается после предсинтаксической¹ обработки текста, подвергается комплексу процедур грамматического разбора в соответствии с правилами согласования и управления в русском языке, в ходе которого используется словарь моделей управления предикатов (глаголов и отглагольных существительных). Результатом синтаксического разбора предложения является список его семантически значимых текстовых единиц и различных типов отношений между ними. В число значимых единиц включаются все классы имен существительных, глаголов и прилагательных, и не включаются знаки препинания, а также ряд служебных частей речи. Отношения между текстовыми единицами (словами) выдаются в одну сторону — от главной единицы к зависимым, ввиду чего результат анализа представляет собой дерево синтактико-семантических зависимостей, с входами от любой из текстовых единиц. Например, результат анализа предложения «Отдел новостроек желает арендовать у нашего комбината малую строительную и погрузочную технику» описывается такой структурой: Отдел (существительное); генитивное отношение с «новостроек»; Новостроек (существительное); Желает (глагол); предикативное отношение с «отдел» в роли «Субъект действия»; предикативное отношение второго порядка с «арендовать»; Арендовать (глагол); предикативное отношение с «комбината» в роли «Источник»; предикативное отношение с «технику» в роли «Объект действия»; Нашего (местоименное прилагательное); Комбината (существительное); атрибутивное отношение с «нашего»; Строительную (прилагательное); Погрузочную (прилагательное); Технику (существительное); атрибутивное отношение с однородным членом «погрузочную»; атрибутивное отношение с однородным членом «строительную»; атрибутивное отношение с «малую». Все типы выделяемых отношений между текстовыми единицами разделяются на две группы. Первую представляют стандартные отношения общего вида между парой слов, которые предопределены общими правилами грамматики языка (атрибутивные, сравнительные, количественные и т. д.). Ко второй группе относятся предикативные отношения между предикатом и его аргументами в определенных семантических ролях. Именно предикатно-аргументная структура характеризует основной семантический (пропозициональный) компонент плана содержания высказывания, описывая

¹ На этапе предсинтаксической обработки каждое предложение преобразуется в последовательность базовых текстовых единиц, которые обычно соответствуют отдельным словам и знакам препинания. Каждая текстовая единица характеризуется грамматическим описанием, которое в общем случае получается на основании морфологического анализа соответствующей словоформы в предложении. Это описание позволяет установить все грамматические характеристики текстовой единицы, необходимые для правильного связывания ее с другими единицами — частями речи, родом, числом, падежами и др.

общую схему ситуации во фразе. Отношения этого вида описываются в словаре моделей управления, где позволяет задавать способ синтаксической реализации отношения (падеж и предлог аргумента предиката), а также семантическую роль каждого аргумента. Например, для существительных в роли прямого дополнения, которое выражается винительным падежом без предлога, при глаголе «арендовать» можно ввести название «предмет аренды», а для косвенного дополнения, выражаемого родительным падежом с предлогом «у», можно ввести роль «арендодатель».

Синтаксические и семантические связи членов предложения

Приемы и методы анализа языков разного грамматического строя разрабатывались лингвистами на протяжении многих десятилетий. В последнее время эти результаты стали использоваться для компьютерной обработки текстовой информации. Для языков, основанных на алфавитных знаковых системах, к настоящему времени созданы достаточно эффективные поисковые технологии для широкого класса практических применений. В области компьютерной обработки иероглифических текстов существенных продвижений пока не достигнуто. Это обусловлено, очевидно, существенным своеобразием иероглифической письменности, а также сложностью и особенностями грамматического строя иероглифических языков, громоздкостью используемых знаковых систем. В то же время, несмотря на существенные отличия грамматического строя языков, использующих иероглифическую письменность, от грамматического строя европейских языков, в ходе лингвистических исследований были выявлены грамматические категории, обладающие универсальной общностью, применимой к языкам обеих групп и создающей объективные предпосылки для универсализации методов машинного анализа текстов независимо от их принадлежности к тому или иному языковому типу. Их краткое описание сводится к следующему. В любом тексте можно выделить ряд часто повторяющихся устойчивых единиц (в случае языков с иероглифической письменностью это одно-, двух- и, реже, трех- и четырехсложные иероглифы).

Эти устойчивые единицы могут быть выделены на основе графической записи вне зависимости от фонетического представления (доказательством чему служит тот факт, что носители разных диалектов часто вынуждены прибегать к графической записи для того, чтобы понять друг друга). Несмотря на определенные трудности сегментации иероглифического текста, когда трудно бывает провести грань между основной единицей морфологического анализа — словом и основной синтаксической единицей — синтагмой, само наличие морфологических категорий — существительное, глагол и т. д. — не отрицается. Однако ввиду почти полного отсутствия аффиксов, которые могли бы служить для определения части речи, к которой относится та или иная единица, эти морфологические категории определяются почти исключительно на основании той роли, которую они играют в предложении (их синтаксической роли). Основные понятия, которыми оперирует синтаксис — подлежащее, сказуемое, определение применимы, таким образом, и к языкам, использующим иероглифическую письменность.

Полностью автоматизированный высококачественный машинный перевод следует, видимо, считать некоторым идеалом, подобно многим абстракциям науки. Нет необходимости доказывать, что недостижимость такого идеала не должна служить препятствием к попыткам добиться реально пригодного МП. Ближайшей аналогией здесь является положение с «человеческим» переводом: неизбежность потери информации при переводе, препятствующая получению абсолютно адекватного полностью эквивалентного перевода, не могла

и не может служить основанием для отказа от перевода, и профессия переводчика является одной из древнейших (Марчук, 1985).

Общие задачи исследования могут быть определены следующим образом:

- исследовать теоретические подходы к моделированию перевода и определить принципиальные позиции, с которых целесообразно подходить к такому моделированию в современных технологических условиях переработки информации;
- создать лингвистическую модель машинного перевода с использованием уровня переводных соответствий;
- построить действующую систему для конкретной пары языков, для чего определить конкретное «наполнение» теоретических компонентов модели;
- изучить основные особенности функционирования системы на текстах и закономерности, полученные в ходе работы;
- разработать принципы реализации технических решений в рамках общей концепции приближенных лингвистических вычислений;
- оценить технико-экономическую эффективность полученной модели в ее промышленной реализации в заданных конкретных условиях эксплуатации;
- определить таким образом, основные этапы построения системы машинного перевода и круг связанных с этим теоретических и практических проблем.

Морфологическая часть речи определяется традиционным образом за исключением того, что некоторые специфические морфологические формы также получают статус части речи. Индексы в названии части речи обозначают тип морфологического словаря. Граммема — это элементарный морфологический описатель, относящий словоформу к какому-то морфологическому классу, например, словоформе *стол* с леммой СТОЛ будут приписаны следующие наборы граммем: «мр, ед, им, но», «мр, ед, вн, но». Таким образом, морфологический анализ выдает два варианта анализа словоформы *стол* с леммой СТОЛ внутри одной морфологической интерпретации: с винительным (вн) и именительным падежами (им). Ниже перечислены все используемые граммемы: мр, жр, ср — мужской, женский, средний род; од, но — одушевленность, неодушевленность; ед, мн — единственное, множественное число; им, рд, дт, вн, тв, пр — падежи: именительный, родительный, дательный, винительный, творительный, предложный; #1, #2 — соответствуют пометам «параметр» в словаре А.А. Зализняка; св, нс — совершенный, несовершенный вид; пе, нп — переходный, непереходный глагол; дст, стр — действительный, страдательный залог; нст, прш, буд — настоящее, прошедшее, будущее время; пвл — повелительная форма глагола; 1л, 2л, 3л — первое, второе, третье лицо; 0 — неизменяемое. Как уже было сказано, одной словоформе может соответствовать несколько морфологических интерпретаций. Например, у словоформы *стали* две интерпретации: (СТАЛЬ, NOUN, {«жр,ед, рд,но», «жр,ед, дт,но»}, «жр, мн, им, но», «жр, мн, вн, но»); (СТАТЬ, VERB, {«мн, св, нп, дст, прш»}).

В реальном тексте на последующих шагах анализа выбор одной морфологической интерпретации одного слова может повлиять на выбор морфологической интерпретации соседнего слова, поэтому программы работают с вариантами морфологических интерпретаций. Общий механизм работы системы на лямбда-исчислении для одного предложения выглядит так: всем словам приписывается по формуле, составленной по законам лямбда-исчисления, которые в результате составляют одну большую формулу; последовательно применяя *b*-конверсию и другие законы логики первого порядка, упрощаем формулу; в результате должна получиться формула без *I*-операторов и

операторов конкатенации. В качестве примера, припишем формулы словам этой фразы:¹

every: IP.IQ."x(P@x -> Q@x)
boxer: ly.BOXER(y)
walks: lx.WALK(x)

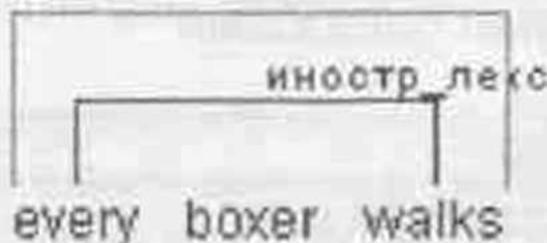
Применяя b-конверсию для формул every boxer, получаем:

IP.IQ."x(P@x -> Q@x)@ ly.BOXER(y) = IQ."x(Iy.BOXER(y)@x -> Q@x)
= IQ."x(BOXER(x)-> Q@x).

Применяя еще раз b-конверсию, получаем:

IQ."x(BOXER(x)-> Q@x) @ lx.WALK(x) = "x(BOXER(x)-> lx.WALK(x)@x) =
"x(BOXER(x)-> WALK(x))

Последняя формула точно отражает смысл предложения *every boxer walks*, точнее смысл кванторного слова *every*.



Задача приписывания номеров значений словам сформулирована такова: дано синтаксическое представление одного предложения, нужно приписать им номера значений, если для каждого номера значения указано, с какими другими номерами он может употребляться. То, что номер значения какого-то слова обуславливает выбор номера значения другого слова, мы будем называть ограничением. Граф, в узлах которого стоят слова, а на стрелках — находятся ограничения, будем называть графом ограничений.

Стрелки графа ограничений получаются из синтаксического анализа, а сами ограничения — из лексикона и онтологии, которые в этой системе называются селективными ограничениями. Например, по предложению: *Петя ест уху*, строится синтаксическое представление: подл (есть, Петя), к_доп (есть, уха), по которому строится граф ограничений: подл (есть, Петя): «Петя — одушевленный субъект» к_доп (есть, уха): «уха — съедобный объект». Кроме стрелок, возникших из синтаксиса, ограничения могут быть записаны в словаре. Например, может быть сказано, что прямой объект данного предиката находится в такой-то семантической связи с подлежащим. Понятно, что между этими объектами прямой синтаксической связи нет. Каких-либо полных и авторитетных перечней семантических и грамматических характеристик в книге Dr.Uchida² не дается, но зато есть проработанный список семантических отношений с определениями и комментариями. Этот список представлен ниже в сокращенном варианте.

В перечне вышеуказанной таблицы в книге Dr.Uchida нет семантических и грамматических характеристик некоторых самых общих отношений исходной и конечной точки. Вместо этого есть только конкретные реализации: конечная точка (gol, plt, pur, tmt, to); исходная точка (frm, plf, tmt); отрезки (dur, fmt). Отношения via (маршрут) и reg (периодичность) — очень хорошее новое предложение. Кажется нужно, что эти отношения обязательно войдут в будущие системы автоматической обработки текста.

¹ Здесь уместно поблагодарить Н.Н.Леонтьеву, из работ которой по автоматическому обработку тексту я заимствовал.

² Основной документ по UNL (Uchida, 1996), в котором содержатся спецификация и идеология языка, доступен в Интернете. На данный момент над проектом UNL работают пятнадцать университетов и научных институтов, в том числе часть сотрудников ИППИ РАН, разрабатывающих систему ЭТАП (Boguslavsky et al. 2000).

Отношение	Определение	Примеры
agt	Определяет участника, который контролирует ситуацию (АГЕНС)	<i>John breaks; Computer translates</i>
aoj	Связывает объект с его признаком или состоянием	<i>Leaf is red; John is a teacher</i>
ben	Связывает событие с участником, который либо что-то приобрел в результате, либо пострадал от этого события	<i>John give... for Mary</i>
cag	Нефокусный агент ситуации	<i>Walk with John; lives with aunt</i>
cnt	Равенство	
cob	Нефокусный объект ситуации	<i>Died with Mary; have a pen with a pencil</i>
fmt	Промежуток между двумя полюсами	<i>From a to z; from Osaka to New York</i>
Frm	Происхождение объекта	<i>Man from Japan</i>
Gol	Конечное состояние процесса или заключительное состояние	<i>Change ... to red; is transformed ... to strong</i>
Ins	Инструмент	<i>Look... with telescope</i>
man	Способ действия или аспект состояния	<i>Look quickly; very beautiful</i>
met	Средство, используемое для достижения целей	<i>... solve ... with dynamics; ... solve ... using algorithm</i>
mod	Любое ограничение объекта	<i>Whole story; master plan</i>
obj	Объект, на который напрямую направлено действие	<i>Cure patient; snow melt</i>
opl	Место, на которое направлено действие	<i>Put ... on shoulder; cut ... in middle</i>
plc	Место	<i>Cook ... in kitchen</i>
plf	Исходная точка	<i>Go from home</i>
plt	Конечная точка	<i>Talk until Boston</i>
pof	Отношение ЧАСТЬ	<i>Bird's wing</i>
pos	Отношение принадлежности	<i>John's dog</i>
ptn	Партнер, с которым вы вместе действуете	<i>Collaborate with machine</i>
tim	Время	<i>Look on Tuesday</i>
tmf	Время начала	<i>Look since morning</i>
via	Промежуточное состояние, маршрут	<i>Go ... via New York</i>

Заключение

Как уже было сказано, средства машинного перевода никогда не смогут улавливать все смысловые нюансы оригинального текста. Различия в синтаксисе и семантике, особенно между западными и восточными языками, — даже сторонники машинного перевода признают, что он способен в лучшем случае передать основную суть документа. Но уместно сказать, что ошибки часто случаются и у обычных переводчиков, и наивно ожидать от системы машинного перевода, что она способна исправить ошибки оригинала и выдаст грамотный перевод. Безусловно, тезис о том, что мозг человека совершенней компьютера, не требует никаких доказательств. Хотя честно сказать, после общения на выставках с некоторыми пользователями в этом появляются сомнения. Из вышеуказанного таблицы видно, что в перечне довольно много традиционных семантических отношений (agt, obj, tim, aoj, plc и т. д.), но кажется есть, некоторые излишние новшества: например отношения cag отличается от agt тем, что cag связывает нефокусного агента с процессом, который возник в результате основного процесса. Правда, примеры свидетельствуют

о том, что фокусный и нефокусный агенты скорее задействованы в одном процессе (см. в таблице). Такие же соображения относятся к нефокусному объекту (соб) и состоянию (сао), поэтому, наверно, проще было бы ввести специальный атрибут @фокусность, а не плодить отношения.

ЛИТЕРАТУРА¹

1. Аношкина Ж.Г. Морфологический процессор русского языка // Альманах «Говор», Сыктывкар, 1995. — С. 17-23.
2. Апресян Ю.Д. Избранные труды, Том 1. Лексическая семантика: 2-е изд., испр. и доп. — М.: Школа «Языки русской культуры», 1995.
3. Арутюнова Н.Д. Время: модели и метафоры. Логический анализ языка: язык и время. — М., 1979.
4. Баскакова И.В. Классификация текстов по заданному набору тематик. Дни науки НГТУ-2003. — Новосибирск: Изд-во НГТУ, 2003.
5. Борисова Е.Г. Коллокации. Что это такое и как их изучать. — М.: Филология, 1995.
6. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения. — М., 1985.
7. Марчук Ю.Н. Вычислительная лексикография. — М.: ВЦП, 1976. — 183 с.
8. Марчук Ю.Н. Синтактико-семантический анализ в системе машинного перевода АМПАР. — Т. Марчук Ю.Н. О моделировании перевода.— Вопросы информационной теории и практики. — Т. 5 — М.: ВИНИТИ, 1980.
9. № 43 — М.: ВИНИТИ, 1980.
10. Ножов И.М. Прикладной морфологический анализ без словаря. КИИ-2000. Труды конференции. — Т. 1. — М.: Физматлит, 2000.
11. Зализняк А.А. Грамматический словарь русского языка. — М.: 1980.
12. Семенова С.Ю. Алгоритмический и лексикографический аспекты Сокирко А.В. Программная реализация русского общесемантического словаря. Дипломная работа, ФТИПЛ, РГГУ, 1997.
13. Uchida Hiroshi, Zhu Meiyang, Tarcisio Della Senta. A Gift for a Millennium. <http://www.unl.ias.unu.edu/>. 1996.

О.С. Волынчик

ПРОБЛЕМЫ ИЗУЧЕНИЯ ГЕНДЕРНОЙ ИДЕНТИФИКАЦИИ РЕЧИ

Исследования мужской и женской речи входят в круг проблем лингвистической гендерологии с момента ее зарождения в недрах феминистской критики языка в 70-х годах 20 столетия на западе. Данные исследования, эволюционировав в от программ выявления и преодоления андроцентризма языка до идентификации авторства текста в криминалистике и исследования реализации гендерных стереотипов в литературоведении, продолжают оставаться актуальными во всех сферах лингвогендерологии, о чем свидетельствует целый пласт работ последних лет, посвященных данной проблеме. [6; 8; 11; 17; 22 и мн. др.]

В разное время для обозначения гендерных различий в речи использовались термины женский язык, гендерплект, мужская и женская речь/язык/вербальное поведение.

Женский язык — условное название явления дифференциации единиц языка в зависимости от пола говорящего, избирательное употребление тех или иных единиц языка женщинами. [13, с.153]

Гендерплект — постоянный набор признаков мужской и женской речи. Возникновение этого понятия обусловлено формированием гипотезы гендерных субкультур. [19]

¹ В статье изложены результаты исследований, которые проводились в Тегеранском университете.