

УДК: 378:81'42

Е.В. ШКУРКО,
*кандидат филологических наук,
доцент кафедры перевода и лингвистической подготовки иностранцев
Днепропетровского национального университета имени Олеся Гончара*

ЛИНГВИСТИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ ОСОБЕННОСТИ СИНТАКСИЧЕСКОГО АНАЛИЗА НЕОДНОЗНАЧНЫХ ПРЕДЛОЖЕНИЙ

В статье рассматривается явление синтаксической омонимии, затрудняющей понимание текста человеком и негативно влияющей на эффективность работы компьютерных систем. Исследуются лингвистические и психологические особенности восприятия неоднозначных предложений. Анализируются ошибки в определении синтаксических связей между компонентами конструкций и их влияние на понимание семантики высказываний.

Ключевые слова: двусторонняя синтаксическая связь, синтаксическая омонимия, омонимизирующий компонент, автоматический синтаксический анализ, закономерности первоначально-го восприятия текста.

Постановка проблемы. XX век стал эпохой бурного развития науки, внедрения новейших информационных и компьютерных технологий. Однако для того чтобы воспринимать и обрабатывать информацию машина, должна знать язык, на котором эта информация записана, следовательно, человек должен ввести необходимые знания в понятном для компьютера виде. Когда перед учеными встала задача обучения машины языку, оказалось, что традиционная грамматика не может справиться с ней, поскольку общепринятые лингвистические описания не являются строго структурированными. Возникла необходимость создания более точных описаний языка, понятных вычислительным машинам. Так в 60-е годы прошлого столетия появилось новое научное направление, одной из задач которого и было создание технологий понимания языка для общения человека с машиной, – компьютерная лингвистика.

Однако при построении систем автоматической обработки текста ученые столкнулись с проблемой многозначности, присущей всем языкам мира. Полисемия обусловлена самой сущностью устройства языка, а также биологическими и психологическими предпосылками мышления людей: сознание человека содержит бесконечное количество образов, понятий и тому подобных единиц, однако число языковых знаков ограничено, в силу чего один и тот же звуковой или графический комплекс может стать многофункциональным [4, с. 13–14]. Но для автоматической обработки текста любая многозначность противопоказана, поскольку машина может эффективно работать только с конструкциями, имеющими строго однозначную структуру. Вследствие этого возникла настоятельная необходимость изучения такого лингвистического явления, как синтаксическая омонимия.

Анализ последних исследований. В наши дни одним из основных направлений компьютерной лингвистики является создание программ автоматической обработки текстов на естественном языке, который, по мнению специалистов, служит оптимальной формой диалога человека и машины [1, 2, 6]. «Научить компьютер «понимать» текст и означает наделить его способностью добывать из него необходимую для выполнения того или ино-

го задания информацию. Такое «понимание» текста заключается в умении анализировать его на разных уровнях представления информации: морфологическом, синтаксическом, логико-семантическом – и обобщать полученные в ходе подобного анализа результаты в определенной форме» [3, с.126].

Формулировка цели. Цель статьи заключается в рассмотрении явления синтаксической омонимии, затрудняющей понимание текста человеком и негативно влияющей на эффективность работы компьютерных систем, анализе неоднозначных конструкций с использованием традиционных методов исследования и методики автоматического синтаксического анализа высказываний, определении лингвистических и психологических особенностей восприятия предложений, содержащих синтаксические омонимы.

Изложение основного материала. В современном русском языке достаточно широко представлены конструкции, в основе построения которых лежит двусторонняя синтаксическая связь, при которой некоторое слово или словосочетание может быть подчинено нескольким доминантам. Иногда такая множественная связь обуславливает появление синтаксически неоднозначных предложений, поскольку возможность соотношения определенных компонентов с любой из доминант приводит к изменению семантики высказываний.

Причинами возникновения омонимичных конструкций в русском языке являются порядок следования компонентов высказывания, возможность различного членения предложения на синтагмы или, чаще всего, совокупность указанных факторов, поскольку зачастую синтаксическая неоднозначность бывает обусловлена наличием в предложении компонента, занимающего такую позицию, которая позволяет соотносить его либо с предыдущей, либо с последующей синтагмой. Такой компонент мы называем *омонимизирующим*.

Так, в предложении *После рассмотрения дела в суде стали происходить странные события* двойная связь возникает потому, что компонент *в суде* занимает в высказывании такую позицию, которая позволяет соотносить его с любой из двух возможных доминант. При этом существует возможность двоякого членения предложения на синтагмы: 1) *После рассмотрения дела / в суде стали происходить странные события*, при котором фраза будет иметь следующее значение: после того, как дело было рассмотрено, в суде стали происходить странные события; 2) *После рассмотрения делав суде / стали происходить странные события*, при котором предложение приобретает иной смысл: странные события стали происходить после того, как дело было рассмотрено в суде.

Как видим, наличие в тексте синтаксических омонимов затрудняет процесс коммуникации, поскольку ставит адресата речи перед необходимостью выбирать одно из нескольких значений высказывания. Неоднозначность представляет собой серьезную проблему и при проведении автоматической обработки текстов, в частности автоматического синтаксического анализа.

Автоматический синтаксический анализ (АСА) – это анализ, направленный на выявление в тексте синтаксических структур и их формальное представление. Поскольку при анализе текста компьютерная система имеет дело с синтаксическими единицами разной сложности, используются различные стратегии выделения в предложениях минимальных синтагм – слов, связанных определенными видами синтаксических отношений. Локальные системы, предназначенные для анализа отдельных частей синтаксической структуры высказываний, применяют метод непосредственных составляющих или анализ контактных слов; интегральные системы, рассматривающие синтаксическую структуру всей конструкции, направлены на выявление главного и зависимого компонентов без учета их расположения в предложении.

Результаты автоматического синтаксического анализа, полученные в этих системах, имеют разные способы графического представления: в локальных системах используется запись пар непосредственных составляющих с помощью скобок или стрелок, в интегральных системах результаты анализа чаще всего представляются в виде дерева зависимостей, то есть в виде ориентированного графа.

Специалисты, занимающиеся разработкой интеллектуальных программ, утверждают, что современные экспертные системы могут не только обрабатывать информацию, но и оценивать ее, а также делать определенные выводы. Кроме того, компьютеры последне-

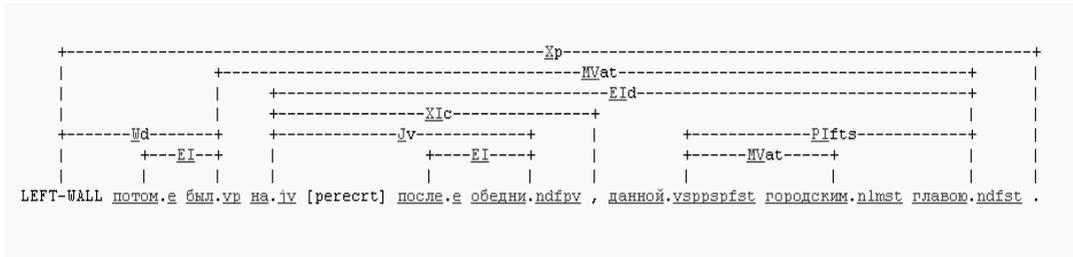
го поколения способны, как люди, выдавать разного рода предположения, опираясь на неполные или неточные сведения.

Мы решили провести эксперимент: с помощью системы автоматического синтаксического анализа проанализировали несколько предложений, содержащих синтаксические омонимы. Эти же конструкции были предложены для анализа студентам специальности «Прикладная лингвистика». И компьютерная система, и студенты получили идентичные задания – установить связи между компонентами анализируемых предложений и изобразить эти отношения графически (студенты – с помощью стрелок, компьютер – в виде дерева зависимостей).

1. *Потом был на закуске после обедни, данной городским главою...* (Н.В. Гоголь «Мертвые души»)

Это предложение может быть понято двояко, в зависимости от того, с какой из двух возможных доминант – 1) *закуска* или 2) *обедня* – будет соотнесен омонимизирующий компонент *данной*: в первом случае фраза приобретает следующее значение – Чичиков после обеденной службы в церкви закусывал у городского главы; читатель, не знающий точного значения слова «обедня», соотнесет слово *данной* со второй доминантой и поймет предложение по-другому – после обедни, которую организовал городской глава, герой был еще и на закуске.

Проанализировав указанное предложение, машина выдала следующий результат:



Для более удобного визуального представления полученных данных воспользуемся другим способом иллюстрации информации:

((потом) (был) (на закуске) (данной городским главою)) (после обедни)).

Компьютерная система правильно установила синтаксические отношения, существующие между компонентами данного высказывания, поскольку имела доступ к корпусам, в которых содержится этот текст, следовательно, владела контекстуальной информацией и абсолютно точно понимала значение слова «обедня», в отличие от студентов-филологов, которые *единодушно* соотнесли компонент *данной* со словом «обедня», тем самым допустив ошибку в определении связей между указанными компонентами предложения.

2. *Торговка вяленой воблой торчала между ящиками* (В. Катаев «Белеет парус одинокий»)

В данном случае омонимизирующим компонентом является словосочетание *вяленой воблой*, которое занимает в предложении позицию, позволяющую отнести его к словам, стоящим как в препозиции, так и в постпозиции. Если соотнести указанный омонимизирующий компонент с существительным «торговка», предложение получит следующее значение: женщина, торговавшая вяленой воблой, торчала между ящиками; если же отнести словосочетание *вяленой воблой* к глаголу «торчала», высказывание приобретет другое значение: торговка, имевшая вид вяленой воблы, торчала между ящиками.

Зная контекст и отношение автора к описываемым событиям, можно допустить, что верным с точки зрения адресанта высказывания является второй вариант прочтения данного предложения. Мы предположили, что компьютер, лишённый человеческой интуиции и чувства юмора, соотнесет словосочетание *вяленой воблой* с ближайшим словом – *торговка*. И действительно, машина проиллюстрировала отношения между элементами данного высказывания так, как мы и предполагали:

((торговка) (вяленой воблой)) (торчала (между ящиками))).

Студенты определили синтаксические связи по-разному: 80% соотнесли словосочетание *вяленой воблой с доминантой «торговка»*, 20% – с доминантой «торчала».

3. *Взгляд горожан радовали который год подряд с любовью и заботой высаживаемые цветочно-кустарниковые ансамбли («Днепр вечерний», 19.07.2011).*

Это предложение содержит компонент *который год подряд*, стоящий в позиции, позволяющей по-разному разделить высказывание на синтагмы: *Взгляд горожан радовали который год подряд / с любовью и заботой высаживаемые цветочно-кустарниковые ансамбли* и *Взгляд горожан радовали / который год подряд с любовью и заботой высаживаемые цветочно-кустарниковые ансамбли*.

При анализе высказывания все студенты выбрали первый вариант актуального членения предложения, соотнеся омонимизирующий компонент с глаголом «радовали». Машина не смогла определить связи словосочетания *который год подряд* с возможными доминантами.

4. *Центр социально-экономического развития обеспечит подготовку пакета документов для открытия предприятия в кратчайшие сроки («Днепр вечерний», 5.02.2014).*

В этом предложении омонимизирующий компонент *в кратчайшие сроки* дистанцирован от доминанты «обеспечит», что делает возможным его соотнесение со словоформой «для открытия», вследствие чего высказывание может быть понято двояко: 1) центр в кратчайшие сроки обеспечит подготовку пакета документов; 2) центр обеспечит подготовку документов, которые позволят открыть предприятие в кратчайшие сроки.

90% студентов соотнесли омонимизирующий компонент с ближайшей из возможных доминант – словоформой «для открытия», 10 % – с глаголом «обеспечит». Компьютер же, как и в предыдущем случае, не смог определить связи словосочетания *в кратчайшие сроки* ни с одной из предполагаемых доминант.

Выводы. Полученные результаты синтаксического анализа неоднозначных предложений позволяют нам утверждать следующее:

1. Несмотря на грандиозные открытия в области высоких технологий, наука до конца не познала механизмы человеческого мышления и речи. Как показывают результаты нашего и ряда других исследований, на современном этапе развития компьютерных технологий эффективным является автоматический синтаксический анализ предложений с прямым порядком слов. Кроме того, для правильного понимания синтаксической структуры высказывания и его семантики программа должна иметь доступ к корпусам текстов, что позволяет компьютерной системе опираться на контекст, хотя и это не всегда позволяет правильно определить истинный смысл, заложенный в высказывание его автором (как, например, в рассмотренном выше третьем предложении). И еще – какой бы высокоинтеллектуальной ни была экспертная система, на современном уровне развития искусственного интеллекта машине пока не доступны человеческая интуиция и чувство юмора, которые в определенных ситуациях могли бы помочь ей в процессах обработки текстов на естественном языке.

2. Ошибки, допущенные студентами при определении синтаксических связей между компонентами рассмотренных неоднозначных предложений, объясняются закономерностями первоначального восприятия текста и, в частности, закономерностями смыслового объединения слов, выявленными и детально исследованными Б.С. Мучником: 1) каждое слово, которое может быть связано с несколькими словами, стоящими по одну сторону от данного, объединяется при первоначальном восприятии с *ближайшим* из этих двух слов, даже если пишущий связывал его с другим, более отдаленным; 2) каждое слово, которое может быть объединено в предложении или с предыдущим словом, или с последующим, объединяется при первоначальном восприятии с *предыдущим*, даже если пишущий относил его к последующему [2, с. 228–229].

Как показал эксперимент, обнаружить синтаксическую омонимию бывает непросто, следовательно, целесообразно не только ознакомить учащихся школ и вузов с этим негативным языковым явлением, но и обучить их приемам выявления синтаксических омонимов. На наш взгляд, эти знания и умения особенно необходимы студентам-филологам и журналистам, творчество которых в перспективе может стать достоянием миллионов читателей.

Список использованных источников

1. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения / А.В. Гладкий. – М.: Наука, 2007. – 152 с.
2. Ермаков А.Е. Компьютерная лингвистика и интеллектуальные технологии / А.Е. Ермаков // Труды Международного семинара Диалог'2002: в 2 т. – М.: Наука, 2002. – Т. 2. – С. 23–35.
3. Карпіловська Є.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник / Є.А. Карпіловська. – Донецьк: Юго-Восток, Лтд, 2006. – 188 с.
4. Колшанский Г.В. Контекстная семантика / Г.В. Колшанский. – М.: Наука, 1980. – 149 с.
5. Мучник Б.С. Человек и текст: Основы культуры речи / Б.С. Мучник – М.: Наука, 1985. – 256 с.
6. Ножов И.М. Морфологическая и синтаксическая обработка текста (модели программы) / И.М. Ножов. – М.: Наука, 2003. – 140 с.

References

1. Gladkij, A.V. (2007). *Sintaksicheskie struktury estestvennogo jazyka v avtomatizirovannyh sistemah obshchenija* [Syntactic structure of natural language in the automated communication systems]. Moscow, Nauka Publ., 152 p. (In Russian).
2. Ermakov, A.E. (2002). *Komp'yuternaja lingvistika i intellektual'nye tehnologii. Trudy Mezhdunar. seminar Dialog'2002: v 2 t., t. 2* [Computational Linguistics and Intellectual Technologies. Proceedings of the International Workshop, Dialog'2002: in 2 volumes, volum 2], Moscow, Nauka Publ., pp. 23–35 (in Russian).
3. Karpilovska E.A. (2006). *Vstup do prykladnoi lingvistyky: kompjuterna lingvistyka: Pidruchnyk* [Introduction to Applied Linguistics: Computer Linguistics: Textbook]. Donetsk, Yugo-zapad Ltd Publ., 188 p. (In Ukrainian).
4. Kolshanskij, G.V. (1980). *Kontekstnaja semantika* [Contextual semantics]. Moscow, Nauka Publ., 149 p. (In Russian).
5. Muchnik, B.S. (1985). *Chelovek i tekst: Osnovy kultury rechi* [Man and Text: The Basics of a Culture of Speech]. Moscow, Nauka Publ., 256 p. (In Russian).
6. Nozhov, I.M. (2003). *Morfologicheskaja i sintaksicheskaja obrabotka teksta (modeli programy)* [Morphological and syntactical text processing (software model)]. Moscow, Nauka Publ., 140 p. (In Russian).

У статті розглянуто явище синтаксичної омонімії, яка ускладнює розуміння тексту людиною і негативно впливає на ефективність роботи комп'ютерних систем. Досліджено лінгвістичні та психологічні особливості сприйняття неоднозначних речень. Проаналізовано помилки у визначенні синтаксичних зв'язків між компонентами конструкцій і їх вплив на розуміння семантики висловлювань.

Ключові слова: двосторонній синтаксичний зв'язок, синтаксична омонімія, омонімізуючий компонент, автоматичний синтаксичний аналіз, закономірності первинного сприйняття тексту.

The article deals with the phenomenon of syntactic homonymy which complicates the understanding of text by a person and negatively affects the efficiency of computer systems. The linguistic and psychological peculiarities of comprehension of ambiguous sentences are investigated. The errors in the determination of syntactic relationships between components of structures and their influence on understanding of the semantics of utterances are analyzed.

Key words: double-sided syntactic tie, syntactical homonymy, homonymizing component, automatic syntactic analysis, law of primitive text perception.

Одержано 3.01.2018.