

СИСТЕМНОСТЬ СИНТАКСИСА РУССКОГО ЯЗЫКА. АНАЛИЗ УПОТРЕБЛЕНИЯ РАЗЛИЧНЫХ КОМБИНАЦИЙ СУБЪЕКТА, ПРЕДИКАТА, ОБЪЕКТА В ПРОСТОМ ПРЕДЛОЖЕНИИ

Объект и предмет исследования

Объектом исследования являются простые предложения и их упрощённая синтаксическая модель, отражающая порядок субъекта, предиката и объекта. Это – ядерные композиции. Согласно В.А. Карпову, ядерная композиция – это структура из двух первичных словоформ. Данное исследование признаёт конструкции из трёх элементов – субъекта, предиката и объекта – ядерными на том основании, что грамматически и логически объект зависим от предиката и может считаться частью его состава.

В данном исследовании рассматриваются одноядерные композиции (структурные модели простых предложений, имеющих одну грамматическую основу). Выясняются их возможные изомерийные варианты и встречаемость в тексте. На основе собранной статистики делаются практические выводы, применимые для автоматической обработки текстов на ЕЯ.

Метод и аппарат исследования

В основе исследования лежит системный подход. Активно используются также методы количественной лингвистики (сбор и анализ статистики для подтверждения выдвинутого предположения).

Была осуществлена сплошная выборка из 4 корпусов текстов: по 2 текста прозаических и поэтических произведений. В каждом блоке (проза и поэзия) – около 1050 предложений. Примерно одинаковое количество позволяет адекватно сравнивать удельный вес различных конструкций в текстах. Конечно, небольшие объёмы и узкая выборка делают полученные в результате исследования цифры неточными, но общая тенденция и направление дальнейших исследований могут быть прослежены уже на этом материале.

Внутри блоков тексты различаются между собой стилистически, и это делает статистику в целом более универсальной, а в частности показывает, как различаются между собой различные стили художественной речи, как прозаической, так и поэтической.

Для анализа были выбраны прозаические произведения В.В. Бианки и М.М. Пришвина. Оба писали о природе, для широкого круга читателей, в т.ч. для детей. Но проведённый анализ показывает, что их стили весьма сильно различаются. Возможно, это связано с разницей жанров (у Бианки это в основном повесть для детей, у Пришвина – этюды, зарисовки).

В качестве поэтического материала исследовались японская поэзия хайку серебряного века и ранние стихотворения А.А. Блока. Разница в культурной и литературной традиции приводит к разнице в манере письма, что обнаруживается и при анализе собранной статистики.

Теоретическое предположение

Ядерная композиция состоит из двух элементов, могущих сочетаться в любой последовательности. Для предложения это *субъект* и *предикат*. Все остальные члены предложения можно отнести к группе субъекта либо предиката. Во многих лингвистических традициях из группы предиката на основе противопоставления субъекту отдельно выделяется *объект*. В данном исследовании объект также выделялся как особый элемент композиции,

однако учитывалась возможность наличия у глагола не только прямого (Вин.п.), но и косвенного объекта, выраженного словом в другом падеже.

Учитывая то, что предложения могут содержать как все три элемента (субъект, объект, предикат), так и два, и даже один, всего мы получаем 15 перестановочных вариантов: 3 для 1-элементных конструкций и по 6 для 2- и 3-компонентных.

Задачей исследования является определить, все ли из имеющихся изомеров встречаются в текстах на ЕЯ, с какой частотой они встречаются и какие особенности текста влияют на их встречаемость.

Практические результаты

Исследуются одноядерные композиции, т.е. простые предложения. Нас интересует, помимо их структуры, их удельный вес в текстах.

Одноядерные и остальные композиции

На нашем корпусе были получены следующие результаты: в прозе доля одноядерных композиций составляет 44%, в поэзии – 61%. Для поэзии процент высок за счёт очень высокого аналогичного показателя в блоке хайку – там он равен 68%. Это обусловлено спецификой поэзии хайку, традиционной формой стиха в три строки, т.е. маленьким объёмом законченного произведения. Средний процент для прозы и поэзии составляет 53%.

Частая встречаемость одноядерных композиций говорит прежде всего об их важном месте в языке. Примерно половина всех предложений в текстах является одноядерными композициями. Этот факт объясняется структурой высказывания, которая соотносится с одним мыслительным актом, этапами которого являются называние предмета и присваивание/признание за ним некоторых свойств/отношений, статических либо динамических.

1-, 2- и 3-элементные композиции

Проза: 1 – 15%, 2 – 60%, 3 – 25%.

Если принять объект как часть группы предиката, то получим 1 – 15%, 2 – 85%.

Поэзия: 1 – 33%, 2 – 55%, 3 – 12%.

Если принять объект как часть группы предиката, то получим 1 – 33%, 2 – 67%.

В целом: 1 – 26%, 2 – 56%, 3 – 18%.

Если принять объект как часть группы предиката, то получим 1 – 26%, 2 – 74%.

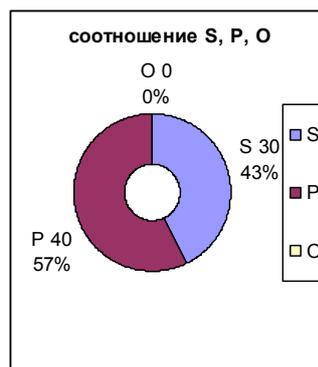
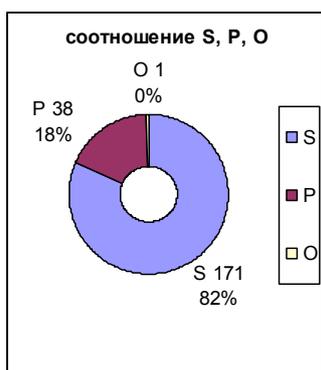
Как видим, наиболее распространёнными в любом случае являются именно 2-элементные ядерные конструкции, как с выделением объекта, так и с отнесением объекта в группу предиката. Это опять же объясняется соответствием структуры высказывания структуре речемыслительного акта.

Подробнее о каждом виде композиций.

1-элементные вызывают большой интерес именно за счёт своего "усечённого" состава. Статистика показывает, что в прозе преобладают Р-композиции, а в поэзии – S-композиции (57% и 82% соответственно). Это можно объяснить отчасти различиями в поэтической и прозаической речи, отчасти особенностями поэзии хайку, где доля S-композиций составляет 87%.

Поэзия:

Проза:



S-композиции. Конструкция, в которой представлен только субъект, формально является одноэлементной. Обычно в русском языке таким композициям соответствуют односоставные назывные предложения. Однако при более внимательном рассмотрении становится ясным, что термин "назывные" не очень уместен: помимо названия субъекта, высказывание как минимум констатирует его наличие, т.е. несёт в себе внутреннюю предикативность. Наиболее ярко это видно на предложениях типа "Дождь. Поздняя осень. Холодный ветер." – в них номинатив сам по себе обозначает действие (дождь = идёт дождь, дождит; ветер = дует ветер) или состояние (поздняя осень = здесь и сейчас поздняя осень). В подобных конструкциях мы можем увидеть свёрнутую двухэлементную композицию, которая в очень сжатой форме выражает богатое содержание. Возможно, именно этой особенностью можно объяснить частое использование таких конструкций в хайку и в поэзии в целом: когда форма материально ограничена, сохранить многогранность смысла позволяет подобная компрессия.

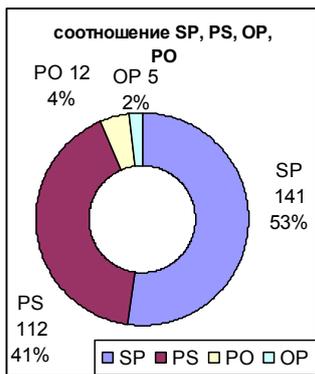
P-композиции. Конструкция, состоящая только из предиката, типична в тех случаях, когда в односоставных безличных предложениях идёт речь о явлениях природы, т.е. субъект как таковой не присутствует. Здесь мы говорим о субъекте в грамматическом, формальном понимании. С точки зрения логики действие не может проходить бессубъектно. Поэтому здесь мы также можем говорить о фразеологичности подобных конструкций и компрессии смысла.

Также часты конструкции, в которых субъект опущен из-за его очевидности (определённоличные предложения). Субъект легко восстанавливается из грамматической формы предиката или контекста.

O-композиции. В корпусе текстов из более чем 2000 предложений была встречена всего 1 композиция такого вида: [Помнишь ли город тревожный?] Синюю дымку вдали? (Блок). Подобные парцеллированные композиции довольно легко представить в разговорной речи, однако в литературном языке они, как видим, практически не употребляются. Приведённый пример был обнаружен в поэтическом тексте, допускающем использование необычных, малоупотребительных оборотов.

2-элементные композиции, как уже было отмечено, составляют большинство встреченных в тексте одноядерных композиций.

Всего для 2-элементных композиций возможно 6 изомерийных вариантов. Опираясь на утверждение о свободном порядке слов, можно предположить, что все они будут встречаться в тексте с одинаковой частотой. Однако статистика опровергает это предположение.



Для прозы:

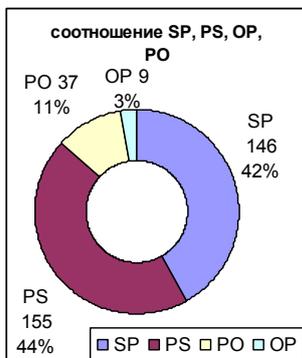
Как видим, и в том, и в другом случае безусловное большинство составляют композиции, содержащие субъект и предикат (94% для прозы и 86% для поэзии). По 6% и 14% соответственно приходится на композиции, состоящие из предиката и объекта. При этом нет ни одной композиции, содержащей субъект и объект.

Очевидно, что основой композиции является предикат, т.к. без него композиция теряет смысл, различить субъект и объект становится практически

невозможным.

К слову, большинство конструкций, состоящих из предиката и объекта, содержат предикат в такой форме, которая не требует отдельного слова для выражения субъекта (личная форма глагола).

Для поэзии:



Если принять объект как часть группы предиката, то конструкции типа OP и PO сводятся к типу P. В то же время конструкции типа SO и OS являются спорными: могут ли они сводиться к SP и PS? С одной стороны, в них не выражается предикативность, отношения между двумя объектами реальности остаются невыраженными/неописанными. Примеров таких композиций в нашем корпусе найдено не было. С другой стороны, такие конструкции легко себе представить в разговорной речи, например: – "Сын очень любит варенье. – А дочь? – Дочь – мороженое". Подобные конструкции эллиптические и в скрытой форме содержат предикат.

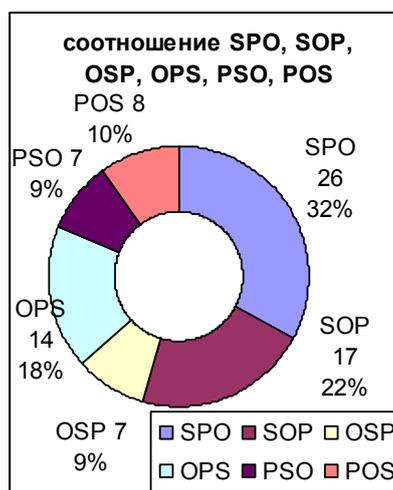
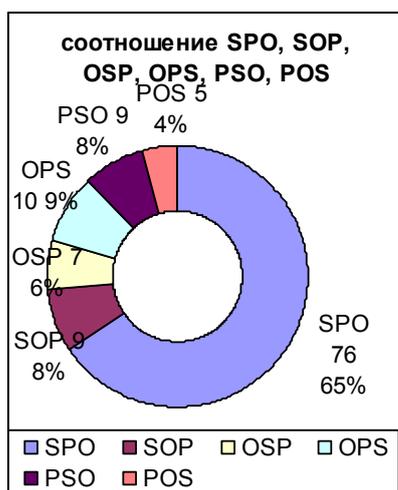
Рассмотрим более подробно соотношение композиций типа PS и SP. В поэзии их доли примерно равны – 44% и 42% соответственно. В прозе конструкции, где субъект предшествует предикату, встречаются чаще: 53% против 41%. Обычно порядок SP называют прямым, а его изомер PS – обратным порядком слов. При этом обращается внимание на то, что второй случай, инверсия, обладает большей экспрессивностью. Этим можно объяснить то, что процент инверсированных конструкций в поэзии выше, чем в прозе.

В целом же полученный процент конструкций типа SP ненамного превышает процент встречаемости конструкций типа PS: 47% против 43%. Исходя из этих данных, мы можем сделать вывод о том, что в смысловом значении последовательность предиката и субъекта ничего не меняет, однако порядок элементов способен изменять эмоциональные акценты в высказывании.

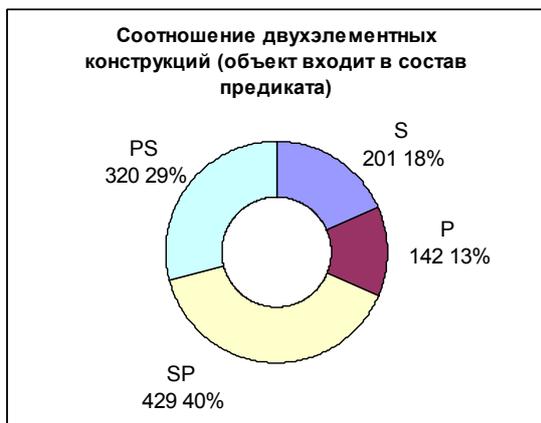
3-элементные композиции представляют особый интерес для исследования, т.к. именно здесь можно наиболее подробно рассмотреть вопрос о прямом/инверсированном и фиксированном/нефиксированном порядке слов.

Проза:

Поэзия:



Прямым традиционно считается порядок элементов в композиции типа SPO. Как видим, конструкции такого типа составляют 65% 3-элементных композиций в прозе и 32% в поэзии. В целом они составляют 53% 3-элементных композиций, или примерно 9% всех одноядерных композиций, или около 5% всех композиций исследованных текстов. Как видим, абсолютная их доля мала, однако в своём классе они являются несомненными лидерами. Малый удельный вес в общей статистике объясняется тем, что большинство конструкций составляют 2-элементные композиции. Если же пересчитать соотношение, приняв объект за часть состава предиката, получим следующее:



На долю SP теперь выпадает 40% одноядерных конструкций, PS – 29%.

Предпочтительным снова оказывается прямой порядок элементов.

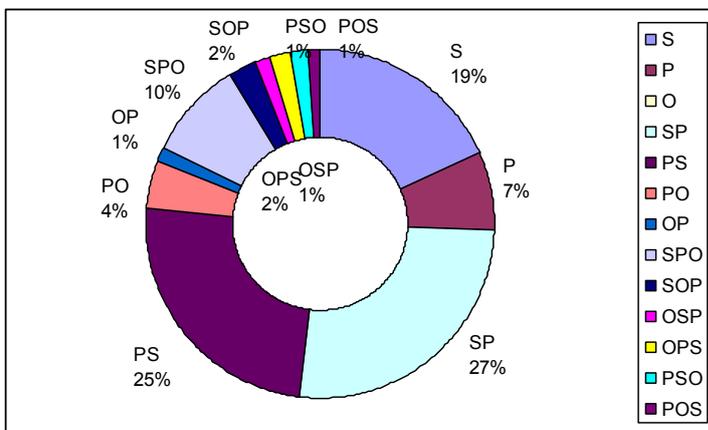
Однако вернёмся к рассмотрению 3-элементных композиций. Как уже отмечалось, самым распространёнными являются конструкции типа SPO. Далее "рейтинг" можно выстроить следующим образом:

Проза	Поэзия
1. SPO	1. SPO
2. OPS 9%	2. SOP 22%
3. PSO SOP 8%	3. OPS 18%
4. OSP 6%	4. POS 10%
5. POS 4%	5. PSO OSP 9%

Эти рейтинги не совпадают между собой. Исходя из небольшой разбежки процента встречаемости конструкций кроме SPO в прозе (4–9%), а также сравнив эти показатели с аналогичным для поэзии (9–22%), можно сделать вывод, что в поэзии представлено больше разнообразных конструкций. Этот факт можно объяснить потребностью поэтической речи в выразительных средствах. Многообразие конструкций, таким образом, предполагает многообразие оттенков смысла, что очень важно в поэзии.

Основные выводы

1. Из 15 возможных вариантов одноядерных композиций реализуются в языке не все. В исследованном корпусе текстов ни разу не встретились конструкции типа SO и OS, конструкция типа O встретилась 1 раз в поэтическом тексте. Это говорит о том, что подобные объектные конструкции невозможны без предиката либо содержат предикат в скрытом виде.



2. Основным элементом ядерной композиции является предикат, который скрыто присутствует и в конструкциях типа S (как констатация наличия/присутствия какого-то предмета/лица/явления).

3. Объект может быть отнесён к составу предиката, при этом все ядерные композиции могут рассматриваться как

двухэлементные.

4. Порядок элементов в ядерных композициях может быть различным, однако некоторые конструкции встречаются чаще остальных. Такими конструкциями являются конструкции типа SP (27%), PS (25%), S (19%), SPO (10%), P (7%). В сумме они составляют 88% одноядерных композиций.

Полученные данные показывают, что исследования в этой области могут привести к получению результатов, важных для компьютерного анализа текста на естественном языке. Уже это небольшое исследование показывает, что, получив статистику на большом корпусе текстов, можно улучшить показатели автоматического синтаксического анализа, который является одним из самых сложных и трудоёмких этапов NLP.