

О микросинтактической конструкции, образованной по модели «всем X-ам X»

Гофман Сергей

НИУ МИЭТ / Москва, Россия

gofmansergeyy@mail.ru

Abstract

This paper focuses on a microsyntactic construction “Vsem X-am X” formed by the adjective pronoun and lexical repetition of a noun. A corpus-based research showed that linguistic processors fail to separate such phrasemes from occasional sequences that only resemble the sought construction. Its idiomatic nature is another reason for mistakes in the work of linguistic tools, e. g. machine translators. The solution might be in applying a rule-based approach to provide a clear description of the phraseme.

Key words: microsyntax, microsyntactic constructions, syntactic phrasemes, corpus research, superlativity, rule-based approach.

1 Исследование работы корпуса и систем машинного перевода с синтаксической фраземой “всем X-ам X”

О наличии в синтаксисе идиоматических выражений лингвисты говорят уже не одно десятилетие. Этому посвящены, в частности, теория грамматики конструкций, описанная в книге Е. В. Рахилиной, в которой исследуются конструкции - совокупности единиц, выступающие как единое целое [4], труды Д. Н. Шмелева [7], и Н. Ю. Шведовой [6], в которых фокус смещен на семантическую единость некоторых словосочетаний. Ряд ученых, среди которых Иншакова Е. С., Иомдин Л. Л., Митюшин Л. Г., Сизов В. Г., Фролова Т. И., Цинман Л. Л., предложили термин микросинтаксис, который формально разграничивает подходы к изучению свободных синтаксических конструкций и несвободных, которые получили название микросинтаксических [2]. Такая идея получила развитие в других статьях Л. Л. Иомдина (напр., [3]). Это положительно сказалось в том числе на компьютерной лингвистике, поскольку, если отдельно задать правила определения микросинтаксических конструкций в тексте, качество его автоматической обработки улучшится, что повлияет, например, на системы машинного перевода и на разметку текстов. Улучшения произошли в токенизации текстов. Как отмечено в [2], опираясь на словарь микросинтаксических конструкций, система может автоматически относить к одному токену больше одного слова. Кроме того, так как в микросинтаксическую разметку входит описание лексических значений конструкций, можно предположить, что точнее стал производиться и семантический анализ.

В настоящей статье рассматривается синтаксическая фразема, образованная определительным местоимением и лексическим повтором: «Всем X-ам X», так что формы X-а должны иметь одну лемму, напр.:

(1) *Всем ягодам ягода.*

Данная конструкция, несомненно, обладает идиоматичностью, по функции приближаясь к суперлятиву прилагательных: сравните предложения *Всем врачам - врач* в значении “каждому врачу полагается свой врач” и *Арск - всем городам город*, под которым подразумевается, что Арск - лучший город. В статье А. П. Ушаковой эта модель была названа непродуктивной [5]. Тем не менее это утверждение спорно, поскольку лексических ограничений на использование такой модели в русском языке в явном виде не накладывается, и любое существительное можно встроить в анализируемую модель, предварительно поставив его в нужную форму.

Мы провели корпусное исследование, задав в лексико-грамматическом поиске Национального корпуса русского языка и Генеральном Интернет-корпусе Русского Языка описанную выше формулу. В результате мы получили больше пяти тысяч вхождений, среди которых есть, однако, те, что на самом деле не удовлетворяют запросу. Причиной этому стала омонимия форм имени существительного (напр., именительного падежа множественного числа и родительного падежа единственного) в НКРЯ и особенности синтаксиса и пунктуации в текстах, размещенных в социальных сетях, в ГИКРЯ. Затем мы проверили работу систем машинного перевода (Яндекс.Переводчик, Гугл.Переводчик, DeepL, Prompt, ЭТАП-3) с этой конструкцией, и тоже получили неудовлетворительный результат.

2 Описание синтаксической фраземы “Всем X-ам X”

Исходя из этого, мы предлагаем описать правила, по которым можно отличить синтаксическую фразему, строящуюся по указанной выше схеме, от схожих по морфологическому и/или лексическому составу словосочетаний. Для этого мы будем опираться на план, предложенный Л. Л. Иомдиным и др. в [1], а именно рассмотрим: 1) лексико-морфологическую идентификацию входящих в состав фраземы элементов, 2) характеристику синтаксических отношений, 3) синтаксические особенности лексем в составе фраземы.

Рассматриваемая синтаксическая фразема состоит из определительного местоимения «весь» в форме множественного числа дательного падежа и лексического повтора двух существительных (о пограничном случае, в котором это правило нарушается, будет сказано ниже), первое из которых совпадает по признакам с местоимением, а второе стоит в форме именительного падежа единственного или множественного числа.

Ядро фраземы - стоящее на второй позиции существительное, которое является частью определительного синтаксического отношения с местоимением и объектного - с существительным в третьей позиции.

Что касается синтаксических особенностей, данная фразема может вступать в предикативные отношения:

(2) *А трактор, видать, всем машинам машина.*

В них фразема является частью составного именного сказуемого. Кроме того, ее употребление распространено в неполных предложениях (нераспространенных и односоставных):

(3) *Всем книгам книга,*

(4) *Да уж, конфуз - всем конфузам конфуз.*

Соответственно, то, что не является микросинтаксической конструкцией, не будет выступать единым членом предложения. Формально такая ситуация, с одной стороны, возникает, если у какого-то из элементов рассматриваемой последовательности появляется зависимое слово, в ее состав не входящее. Таким случаям соответствуют, например, фразы:

(5) *сочувствующий всем людям человек*

(6) *угрожающая всем странам страна*

Тем не менее в некоторых случаях появление посторонних слов все-таки возможно. Во-первых, в инвертированных предложениях, когда подлежащее перемещается в середину нашего предиката. Во-вторых, когда $X > 1$ слова. В-третьих, если не относящееся к конструкции слово является вводным или междометием. Наконец, когда у X появляется атрибут. Ниже приведены примеры, включающие эти явления:

(7) *всем Вадимам он Вадим*

(8) *всем Бредам Питтам Бред Питт*

(9) *всем трендам, сцуко, тренд*

(10) *всем эпатажам самый классный эпатаж*

Кроме того, здесь можно говорить об эллиптических предложениях с опущенным предикатом (вместо которого стоит тире), причем существительное во второй позиции становится непрямым дополнением, местоимение - его атрибутом, а существительное, находящееся в третьей позиции - подлежащим. Такая синтаксическая схема представляется распространенной, к примеру, для лозунгов (*Всем компаниям - компания!*). Стоит отметить, что всецело полагаться на наличие тире между двумя существительными при формулировке правил для описания этой микросинтаксической конструкции не стоит, потому что в ходе анализа результатов нашего корпусного исследования были обнаружены употребления с тире между существительными, удовлетворяющие запросу

(11) *Вот какой Ваня столетний, всем Креням – Крень.*

Кроме тире “вклиниваться” между существительными, согласно проведенному исследованию, могут кавычки

(12) *Всем “языкам” языки.*

Получается, на данном этапе определить, является ли набор слов искомой фраземой, можно, если а) проверить, есть ли в предложении глаголы-связки составного именного сказуемого, б) если в предложении есть тире, то, вероятнее всего, мы имеем дело с “ложной” микросинтаксической конструкцией, хотя наверняка утверждать едва ли возможно, поскольку тире в предложении с “Кренями” не мотивировано правилами пунктуации русского языка.

Использование таких правил может пригодиться при подготовке обучающей выборки для нейросетевых моделей. Ее можно подготовить как в ручную, учитывая выявленные

признаки, так и в автоматическом режиме, например, с помощью регулярных выражений. Кроме того, перспективой исследования может стать преобразование описанных правил в функции для использования их при создании архитектуры нейронной сети.

3 О промежуточном случае данной микросинтаксической модели

В ходе анализа результатов, полученных при лексико-грамматическом поиске в Национальном корпусе русского языка, мы обнаружили несколько десятков словосочетаний, не попадающих под установленную в начале статьи модель, но также обладающую как способностью сочетаться с разными словами, так и приобретением нового значения. Такие конструкции построены по модели “всем X-ам мать (мати)/отец” и употребляются в двух значениях. Самое часто встречающееся и наиболее очевидное - использование существительных в третьей позиции для обозначения родства (“...всем детям мать”). В эту же группу попадают пословицы и поговорки, такие как, например, библейское “эта трава - всем трава мати” и историческое “Москва - всем городам мать”. Тем не менее возможны и такие случаи, как “водка - всем лекарствам мать” или “камень - всем камням отец”. Здесь, например, как и в предыдущей модели, выражение употреблено для того, чтобы передать превосходство.

Несмотря на это, современные машины справляются обработкой такого рода сочетаний достаточно хорошо. Например, машинный перевод на английский язык звучит естественно, в отличие от конструкции, описанной в начале статьи. Сравнение этих двух выражений в корпусе английского языка News of the Web подтвердило нашу гипотезу о том, что носитель языка с большей вероятностью скажет: «vodka is the mother of all medicines» (водка - мать всех лекарств), чем “this Thursday is Thursday for all Thursdays” (*этот четверг - четверг для всех четвергов), лексико-грамматический поиск в англоязычных корпусах British National Corpus и News of the Web показал, что в английском языке фразема-эквивалент “Всем X-ам X” “the X for all X” употребляется только с одушевленными существительными).

4 Заключение

Таким образом, отличить микросинтаксическую конструкцию, образованную по модели “Всем X-ам X” можно, используя следующие правила. 1) X - две словоформы одного слова; 2) вся конструкция является частью составного именного сказуемого, в котором глагольная часть может быть опущена; 3) у элементов фраземы нет зависимых слов (за исключением атрибутов X); 4) если между “X-ам” и “X” стоит тире, то можно сделать предположение о том, что последовательность слов идиоматичностью не обладает.

Сформулированные правила можно использовать в компьютерной лингвистике при проведении мультиязычных корпусных исследований для сравнения микросинтаксических структур. Кроме того, внедрение формальных описаний микросинтаксических конструкций улучшит обработку текста на синтаксическом и семантическом уровне, что, в свою очередь, скажется на таких направлениях дисциплины, как машинный перевод и информационный

поиск. В перспективе можно описать правила математическим языком для создания гибридных моделей, работающих и с правилами, и с нейросетями.

Библиографический список

- [1] Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л., Санников В.З. Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря. - М.: Языки славянских культур, 2010. - 408 с.
- [2] Иншакова Е. С., Иомдин Л. Л., Митюшин Л. Г., Сизов В. Г., Фролова Т. И., Цинман Л. Л. СИНТАГРУС СЕГОДНЯ // Труды Института русского языка им. В. В. Виноградова. - 2019. - №3. - С. 14-41.
- [3] Иомдин Л. Л. Конструкции микросинтаксиса, образованные русской лексемой раз // SLAVIA časopis pro slovanskou filologii. - 2015. - №3. - С. 291-306.
- [4] Рахилина Е. В. Лингвистика конструкций. - М.: Издательский центр "Азбуковник", 2010. - 584 с.
- [5] Ушакова А. П. Синтаксические фразеологизмы с опорным компонентом - местоименным словом // Верхневолжский филологический вестник. - 2016. - №4. - С. 61-65.
- [6] Шведова Н. Ю. Очерки по синтаксису русской разговорной речи. - М.: Издательство Академии наук СССР, 1960. - 373 с.
- [7] Шмелев Д. Н. Избранные труды по русскому языку. - М.: Языки славянской культуры, 2002. - 438 с.