

**ИНФОРМАЦИОННО ОПТИМАЛЬНОЕ ФИКСИРОВАНИЕ МЕТАСЕМАНТИЧЕСКИХ
ЗНАЧЕНИЙ ЗНАКОВ АЛФАВИТА СИНТАКСИСА В ДРЕВНЕКИТАЙСКОМ ЯЗЫКЕ**

**Крицук-Тарасов Никита Иванович, аспирант
Белорусский государственный университет**

**INFORMATION-OPTIMAL RECORDING OF METASEMANTIC MEANINGS OF SYNTAX
ALPHABET SIGNS IN THE ANCIENT CHINESE LANGUAGE**

**Krytsuk-Tarasau Mikita, postgraduate, nikita_krytsuk@mail.ru
Belarusian State University**

Аннотация. Предложенный способ фиксации метасемантических значений знаков алфавита синтаксиса в древнекитайском языке обеспечивает совместимость с ИТ. Обозначены условия фиксации. Результаты применимы для формализованного синтаксического описания древних текстов.

Ключевые слова: информационная оптимальность, синтаксический анализ, древнекитайский язык, метасемантическое значение.

Активное развитие информационных технологий (далее: «ИТ») в общем и искусственного интеллекта (далее: «ИИ») в частности оказывает коренное влияние на научно-исследовательскую деятельность, в том числе и в лингвистическом аспекте. С одной стороны, применение данных разработок позволяет существенно увеличивать охваты и скорость исследования. С другой стороны, широкое применение ИТ, а в особенности ИИ в определенной степени представляет собой своеобразного (не всегда полноценного) «конкурента» для деятелей научной сферы.

В наши дни существует множество примеров плодотворного и эффективного применения ИТ и ИИ учеными, исследователями и разработчиками. И несмотря на то, что в настоящий момент алгоритм ИИ по-прежнему нельзя назвать «интеллектом» в полном смысле этого слова, его дальнейшее применение во всех сферах жизнедеятельности человека – это скорее вопрос времени, нежели чем возможности. Насколько глубоким будет подобное применение, в значительной степени зависит от качества проработки ИИ и развития его в полноценный «интеллект».

Вместе с тем ИТ уже активно применяются исследователями и учеными. Если говорить о лингвистической области, то даже поиск подходящего исследования в интернете, а не в «традиционной» библиотеке, и применение компьютера вместо печатной машинки – это уже ИТ. Принимая во внимание повсеместность и глубину применения ИТ в исследовательской деятельности, необходимо оформлять результаты научных исследований в таком виде, который будет удобен не только для чтения человеком, но и для использования в ИТ: поиск по документу, фиксирование в памяти системы и т.д. В данной статье мы рассмотрим подобную оптимизацию оформления результатов научного исследования на примере метасемантических значений (далее в том числе «МЗ») *знаков алфавита синтаксиса* (сокращенно ЗАС) [1, с.228] в древнекитайском языке. Большинство из них представлены в наших предыдущих исследованиях [2, с.10, 57-58], но не оформлены оптимально в аспекте ИТ.

Предварительно отметим условия, которым должна соответствовать подобная фиксация метасемантических значений ЗАС:

- 1) отсутствие пересечения с другими элементами, являющимися объектами и инструментами исследования;
- 2) удобство восприятия текстовой формы для человека;
- 3) простота создания текстовой формы с помощью ИТ для человека;
- 4) целенаправленная однозначность результатов, выдаваемых при поисковом запросе;
- 5) полная или существенно полная передача в строке запроса в инструментах ИТ.

Далее продемонстрируем, какой вид фиксации был нами подобран, и рассмотрим его на соответствие требованиям, приведенным выше.

В тексте на разных уровнях возможно выделение различных единиц: предложений, топиковых [3, с.64, 82] структур, синтаксических элементов. К последним, помимо ЗАС, относятся разнообразные члены предложения. В нашей языковой традиции нет требования производить описание от большего к меньшему, в отличие от, например, китайского языка, где описание даты обяза-

тельно осуществляется в последовательности год-месяц-день. Это на несколько столетий или даже тысячелетий опередило стандарт ISO 8601 Международной организации по стандартизации. Несмотря на то, что именно данный стандарт касается форматов дат и времени, мы уверены, что не станем «пионерами», развивая подход «от большего к меньшему» в другие сферы, включая фиксацию метасемантических значений ЗАС.

Поскольку нами рассматривается именно кириллический формат, описание МЗ начинается с аббревиатуры ЗАС, которая не пересекается с прочими сокращениями. Рассмотрим фрагмент разбора предложения:

其幾千里也 qí jǐ qiān lǐ yě

其 /+ / [幾 + 千 + 里] + 也 qí /+ / [jǐ + qiān + lǐ] + yě

ЗАС5.1.a其 {П0} /+ / |Ск0| {непер. кач.} + [Оп0.1 + Оп0.2 + Оп-e0] {Компл_ст} + ЗАС_КЧ3.3.a也 <констатация, синтаксическое и стилистическое требования отсутствуют>

Он {ЗАС5.1.a其} <‘Кунь’> /+ / |большой| + [<на> неопределенное количество + тысяч + ли (древняя мера длины)] [4, с.959] {Компл_ст} + <же> {ЗАС_КЧ3.3.a也}

Очевидно, что во фрагменте присутствует два ЗАС – ЗАС5.1.a其 и ЗАС_КЧ3.3.a也.

Следующий шаг – при наличии такого выделения – подкатегория ЗАС. Мы предлагаем отделять категорию от подкатегории с помощью нижнего подчеркивания, как в примере выше. Исходя из подобного подхода, мы видим, что ЗАС_КЧ3.3.a也, относится к подкатегории КЧ (конечных части), а к примеру, ЗАС_не2.1.a不 – к подкатегории негаторов. Безусловно, в тексте может быть множество сочетаний «КЧ», а тем более сочетаний «не», но данный вопрос снимается, если мы будем искать в документе любое сочетание от «ЗАС_КЧ» и «ЗАС_не» до «_КЧ» и «_не» соответственно. С точки зрения написания нижнее подчеркивание легко вводится с клавиатуры и так же легко распознается системами ИТ. Также подобный вариант записи позволяет осуществлять поиск как любых ЗАС, так и их отдельных подкатегорий.

На следующем этапе мы вводим цифирное обозначение подкатегории ЗАС, номер конкретного ЗАС в номенклатуре соответствующей подкатегории, а затем – то метасемантическое значение, в котором данный ЗАС использовался в тексте. Подкатегорию, номер ЗАС и номер значения отделяем точками «.» с целью избежания пересечения с другими рассматриваемыми явлениями, требующими конкретной фиксации: для нумерации предложений и отдельных элементов в их составе мы используем знак двоеточие «:», для спецификации синтаксического анализа членов предложения пользуемся только цифирными обозначениями, без буквенных.

Изначально мы применяли надстрочные и подстрочные знаки для обозначения метасемантических значений ЗАС и разграничения членов предложения (разные простые предложения в составе одного сложного, различные синтаксические уровни анализа и т.д.) соответственно. Тем не менее, мы отказались от данного подхода по следующим причинам: во-первых, для ЗАС не требуются синтаксические разграничения, как для членов предложения, а для членов предложения – разграничения МЗ, как для ЗАС, поэтому в разграничении регистров нет необходимости; во-вторых, сочетание буквенных и цифирных обозначений позволяет избежать недопонимания касательно того, что именно подразумевают отдельные цифры; в-третьих, во многих строках запроса систем ИТ и ИИ указание верхнего и нижнего регистров физически невозможно, что делает бесполезным использование данных регистров при оперировании информацией в системах ИТ и ИИ.

Помимо этого, именно подобная последовательность перечисления «координат» МЗ позволяет четко понимать смысл каждого элемента: подкатегория – конкретный знак – конкретное МЗ знака. Вместе с тем такая фиксация позволяет легко находить определенный ЗАС, используемый в конкретных МЗ, через номенклатуру, поскольку не требует введения непосредственно иероглифа, тем самым экономя время запроса. Следует учитывать, что в связи с омофонией китайского языка, набор отдельных иероглифов с компьютера с помощью пиньиня – процедура, требующая времени, особенно если мы говорим об иероглифах, редких для современного языка.

Наконец, для простоты восприятия текста человеком описание ЗАС завершается непосредственно его иероглифической формой. Мы полагаем, что писать только пиньинь нецелесообразно: в связи с вышеуказанной омофонией китайского языка пиньинь во многих случаях будет вызывать путаницу, о каком именно иероглифе идет речь. Применение же иероглифа и пиньиня одновременно кажется слишком громоздким, к тому же, введение с клавиатуры диакритических знаков,

обозначающих тон, будет требовать излишних временных затрат. Более того, современное и древнее чтение иероглифов различается, что снимает острую необходимость фокусироваться на чтении в древних памятниках. Вместе с этим специалист, знакомящийся с текстом исследования, наверняка будет знать современное чтение иероглифа либо же сможет легко его найти в соответствующей справочной литературе.

Итак, ЗАС_КЧЗ.3.a也 – это иероглиф 也, относящийся под номером 3 к подкатегории конечных частиц категории ЗАС, и примененный в данном контексте в своем МЗ «а». Также для большего удобства восприятия информации человеком в прямоугольных скобках непосредственно после ЗАС возможно уточнение МЗ или его отдельной спецификации, как мы это видим из примера выше.

Таким образом, мы имеем полноценный способ фиксирования метасемантических значений ЗАС в древнекитайском языке, соответствующий всем вышеперечисленным условиям, который возможен для различных классификаций (основанных, например, не на комбинаторных, а на этимологических принципах).

Единственный вопрос, в настоящий момент остающийся неразрешенным – это обозначение «границ» функционального влияния ЗАС. Для наглядности продемонстрируем пример ниже:

其 + [遠 + 而 + 無 + [所 + [至 + 極]]] + 邪?

qí + [yuan + ér + wú + [suǒ + [zhì + jí]]] + yé?

[П2] + ЗАС5.1.в.其 + ^т[> Ск_всп2 {непер. кач.} + ЗАС4.2.a.而 <связь состояния и результата> + ЗАС_не2.2.a.無 {Ск_осн2} + ^с[> ЗАС5.4.a.所 <объективатор ПД_осн2> + ПД_осн2 <]^{5.4.a} <]^{4.2.a} + ЗАС_КЧЗ.4.a.邪 <альтернативный вопрос> ?

[Небо] + или {ЗАС5.1.в.其} + ^т[> далёкое + и при этом {ЗАС4.2.a.而} + не имеет {ЗАС_не2.2.a.無} + ^с[> того, где {ЗАС5.4.a.所} + достигается предел {ПД_осн2} <]^{5.4.a} <]^{4.2.a} + разве {ЗАС_КЧЗ.4.a.邪} ?

Для обозначения границ в настоящий момент используется сочетание квадратных скобок (обозначение функционирования участка как единого целого [2, с.10]) с вывернутыми треугольными скобками внутри (используются как стрелки, указывающие направление функционального влияния ЗАС) «>...<». Для обозначения непосредственно того ЗАС, который оказывает синтаксическое влияние на отрезок текста, взятый в скобки, используются иероглифический вид иероглифа слева от скобок и номер метасемантического значения ЗАС – справа, выделенные верхним регистром. Как уже отмечалось выше, применение различных регистров не является оптимальным для использования в ИТ. В то же самое время отсутствие регистра для иероглифической записи ЗАС, в особенности при использовании по отношению к оригинальному тексту, будет сбивать человека, знакомящегося с памятником, так как эта запись может казаться частью текста, а не обозначением синтаксического разбора. Полностью же убирать иероглифическую запись и номер метасемантического значения ЗАС нецелесообразно, поскольку «границы» функционального влияния различных ЗАС могут взаимно пересекаться, как это видно в отношении ЗАС4.2.a.而 и ЗАС5.4.a.所 из примера выше. Вероятно, целесообразно сохранить только номер метасемантического значения ЗАС слева и справа от скобок без верхнего регистра. Но окончательное решение данного вопроса мы осветим в будущих публикациях.

Таким образом, для оптимизации фиксирования результатов исследований с целью простоты оперирования ими в системах ИТ и ИИ при сохранении удобства восприятия человек можно сделать следующие выводы:

1. Не допускается использование визуально одинаковых способов и приемов фиксирования информации для разнотипных рассматриваемых явлений. Например, если мы в номерах МЗ используем знак точка «.», то он не должен применяться в номерах предложений.

2. Допускается некоторая степень визуальной однотипности в случае, если есть другие показатели, позволяющие понять, что обозначенные понятия являются разнородными. К примеру, знак точка «.» применяется как в номерах метасемантических значений ЗАС, так и в обозначении синтаксических показателей членов предложения, однако в таком случае отсутствуют буквенные указатели, наличествующие в номерах МЗ.

3. Способ фиксирования информации должен допускать возможность дополнительного комментирования рассматриваемых единиц (треугольные скобки «<...>»).

4. По возможности следует избегать выделения текста (подчеркивание, жирный шрифт, курсив и т.д.), применения различных (верхний и нижний) регистров и прочих средств оформления, недоступных во многих поисковых строках. Подобное оформление допустимо для повышения качества восприятия текста человеком, но не должно быть принципиально значимым для фиксации информации. Например, использование верхнего и нижнего регистров для обозначения метасемантических значений ЗАС и синтаксического порядка и уровня членов предложения.

Список использованных источников

1. Гордей, А. Н. Лингвистическая пропедевтика / А. Н. Гордей // Беларусь в современном мире : материалы IV Респ. науч. конф., 28 сентября 2005 г. / РИВШ ; редкол.: А. В. Шарапо [и др.]. – Минск, 2005. – С. 226–229.

2. Крыцук-Тарасаў, М. І. Даследаванне экспліцытнасці і імпліцытнасці моўных катэгорый у старажытнакітайскай мове (на прыкладзе тэксту 莊子 «Джуандзы») : дыс. ... маг. усх. : 1-23 80 10 / Мікіта Іванавіч Крыцук-Тарасаў ; Беларус. дзярж. ун-т. – Мінск, 2022. – 116 с.

3. Курдюмов, В. А. Идея и форма. Основы предикационной концепции языка / В. А. Курдюмов. – Москва : Воен. ун-т, 1999. – 194 с.

4. 古代汉语词典 : 大字本 / 主编: 陈复华 ; 副主编: 楚永安 ; 编写: 《古代汉语词典》编写组. – 北京 : 商务印书馆, 2002. – 2087页.