

АКАДЕМИЯ НАУК СССР
ГЕОГРАФИЧЕСКОЕ ОБЩЕСТВО СССР. ВОСТОЧНАЯ КОМИССИЯ

СТРАНЫ И НАРОДЫ ВОСТОКА

Под общей редакцией
члена-корреспондента АН СССР
Д. А. ОЛЬДЕРОГГЕ

ВЫПУСК XI

СТРАНЫ И НАРОДЫ ЦЕНТРАЛЬНОЙ,
ВОСТОЧНОЙ И ЮГО-ВОСТОЧНОЙ АЗИИ



ИЗДАТЕЛЬСТВО «НАУКА»
Главная редакция восточной литературы
Москва 1971

А. А. Торопов

СЕГМЕНТАЦИЯ КИТАЙСКОГО СРЕДНЕВЕКОВОГО ТЕКСТА

(на материалах текста «Цзин бэнь тунсу сяошо»)

Известно, с какими трудностями сталкивается любой лексикограф, составляющий словарную картотеку путем сплошной расписки текста. Особую трудность представляет работа с иероглифическим текстом.

Старый (в том числе средневековый) китайский текст состоит из сплошной цепочки иероглифов, стоящих один подле другого без каких-либо знаков пунктуации и без пробелов. В силу этого в отношении китайского текста не может быть использовано широко распространенное в работах по машинному переводу формальное определение слова как цепочки букв от одной до n , ограниченной слева и справа пробелами. Следовательно, задача лексикографа заключается в обнаружении соответствующих дифференциальных признаков (ДП), отличающих одну единицу от другой, и в нахождении методов выделения единиц языка (ЕЯ), зафиксированных в тексте. Без обращения к смыслу, без предварительного прочтения невозможно членение текста на слова или на какие-то другие ЕЯ или единицы речи (ЕР). Поэтому при выделении соответствующих единиц текста чаще всего приходится прибегать к чисто смысловым или к формально-смысловым ДП.

В статье С. Е. Яхонтова [7] дается анализ ряда ДП и формальных способов выделения ЕЯ. Вот основные из них:

- 1) цельнооформленность;
- 2) способность единиц к употреблению в качестве самостоятельного предложения;
- 3) возможность вставки отрезков речи между частями сложных образований;
- 4) способность единиц к перекрестной замене;
- 5) возможность опущения части предложения без нарушения его грамматической правильности;
- 6) метод отрицательной выделяемости (невыделяемости).

Как отмечается в указанной статье, универсальных признаков и способов выделения ЕЯ нет, разные авторы пользуются небольшим числом по-разному комбинируемых приемов. Ни один из методов не может быть применен ко всем языкам, причем каждый из них подвергается определенным ограничениям. Неприменимы они — ни каждый в отдельности, ни в комбинациях — и к китайскому языку, особенно к языку исследуемого нами памятника («Цзин бэнь тунсу сяошо», XII—XIV вв. н. э. В современном издании текст размечен) [6], ограниченного узкими временными рамками и имеющего сравнительно небольшой объем (7 рассказов, 56 тысяч знаков).

В своем докладе «Лексико-грамматические принципы составления словаря средневекового китайского языка (эпсха Сун)» [5, 101] автор данного сообщения уже приводил основные ДП, применяемые для выделения ЕЯ. Основные из них: синтаксическая цельноформленность (Ц_с) и раздельноформленность (Р_с), идиоматичность (И), воспроизводимость (В), устойчивость (У) и цельность значения (Ц_з). Побочные: морфологическая цельноформленность (Ц) и раздельноформленность (Р_м), наличие-отсутствие словообразовательного аффикса (А_с) и ряд других, внешних признаков (контекст, лексикографические традиции, параллелизм, словообразовательные модели и т. п.). Побочные ДП позволяют выделить ЕЯ более формальным способом, но они применимы к незначительному числу единиц. Основные ДП позволяют выделять ЕЯ менее формализованным способом, но охватывают основную массу единиц текста.

Содержание большинства из названных здесь ДП несколько отличается от традиционного, поэтому представляется необходимым дать более подробные пояснения к некоторым из них:

1. Ц_с — это способность ЕЯ к самостоятельному употреблению в какой-либо синтаксической функции (предложения, члена предложения, служебного показателя). Процедура применения Ц_с в основном сводится к выделению главных членов предложения или их групп, что позволяет обнаружить многие односложные слова¹ и наметить «кандидатов» в сложные (главным образом двусложные) слова.

2. Р_с — то же, но в отношении компонентов сложной единицы. При этом подразумевается не потенциальная способность компонента сложного образования к самостоятельному употреблению², а его реальное употребление в какой-либо синтаксической функции в данном конкретном случае. Этот ДП применим ко всем единицам, но иногда не дает однозначного решения, особенно в случае атрибутивных сочетаний. Например, по одному этому признаку трудно решить, чем (словом или словосочетанием) является двусложное образование *бянь*⁴ *жэнь*². Не помогает здесь и признак Ц_з. Но в другом месте компонент *жэнь*² стоит в начале, а компонент *бянь*⁴ отделен от него двумя самостоятельными элементами. Таким образом, налицо признаки У и И, что говорит в пользу отнесения данного образования к словосочетанию («подходящий человек»), а не к слову. Авторы ряда работ по проблеме слова [например, см.: 2, 12], использующие этот ДП для выделения ЕЯ, под самостоятельным употреблением компонентов понимают только употребление последних в качестве предложения. Некоторые авторы [3, 55], применяя признак Р_с, делят на части и сложные слова, так как фактически признают достаточным наличие всего лишь одного компонента сложного образования, способного к самостоятельному употреблению.

3. В — неоднократная повторяемость готовых единиц в конкретных случаях их употребления как следствие узуральности языкового знака. Различается несколько ступеней В — от единичных до регулярного по-

¹ Здесь используется некоторая фонетическая информация, заключенная в письменном тексте и вытекающая из известного положения, что иероглиф как письменное выражение языкового знака равен морфеме в системе звукового языка. А каждая такая морфема, обладающая синтаксической оформленностью, и есть односложное слово. Таким образом, процедура выделения морфем и основной массы односложных слов из китайского текста, как правило, не вызывает особых затруднений.

² В китайском языке (в том числе в нашем тексте) часто встречаются сложные слова, каждый из компонентов которых может быть употреблен в тексте и самостоятельно. Например, компоненты *тин*⁴ 'слушать' и *шо*¹ 'говорить' из двусложного слова *тин*⁴ *шо*¹ 'услышать'.

вторения в одинаковых или различных контекстах. Этот ДП применим ко всем единицам, но требует статистики.

4. У — невозможность без изменения значения замены, изъятия, взаимозамены или разъединения компонентов единицы. Как и В, У требует статистики. Из всех перечисленных элементов У для выделения ЕЯ чаще всего используется метод разъединения компонентов сложного образования вставкой [например, см.: 1, 90].

5. Ц_з — обязательная соотнесенность значения со всей единицей. С компонентами сложной единицы значение целого может соотноситься через посредство внутренней формы. Сложные единицы могут быть как со значимыми компонентами, так и без них. Это ДП, внутренне присущий каждой классической ЕЯ, но иногда трудно обнаруживаемый, особенно в сложных единицах с атрибутивным сочетанием компонентов.

6. Ц_м — совпадает с тем, что А. И. Смирницкий [4, 190, 197] называет просто «цельнооформленностью», подразумевая под этим грамматическую изменяемость всего сложного образования. Этот (как и два последующих) ДП дает всегда однозначное решение, но применим к крайне малому кругу единиц³ и далеко не в каждом конкретном случае их употребления.

7. Р_м — то же, но в отношении компонентов сложной единицы.

8. А_с — наличие словообразовательных суффиксов, таких, например, как суффиксы существительных *-цзы* и *-эр*.

9. И — различная по степени трудность мотивированности значения единицы значениями компонентов (или компонента) с точки зрения современных языковых норм. Данное определение нуждается в более подробном пояснении.

И трактуется здесь гораздо шире ставшего уже традиционным понятия И, введенного А. И. Смирницким [4, 199]. Согласно А. И. Смирницкому, под И подразумевается «невыводимость значения целого языкового образования из совокупности значений входящих в него частей...». Различие этих двух определений И заключается главным образом в признании или отрицании самой возможности выведения цельного значения из значений компонентов. Дело в том, что значение сложной единицы в конечном счете может быть выведено только из значений компонентов, но мотивировка этого значения с точки зрения норм современного языка может быть осложнена различными причинами. В зависимости от характера этих причин различается несколько видов И, которые можно условно назвать следующим образом:

а) лексико-семантическая И, когда причиной, осложняющей толкование цельного значения, являются лексические значения компонентов единицы. Например *ба²бу⁴* 'убегать, отрываться, освобождаться' (досл. 'выдергивать, извлекать' + 'шаг'). Именно этот вид И имеется в виду, когда И определяется как несводимость цельного значения к сумме значений компонентов. Степень этой несводимости может быть различной;

б) лексико-грамматическая И, когда мотивировка осложняется различной категориальной принадлежностью всей единицы, с одной стороны, и ее компонентов — с другой. Например: *бу⁴хао³* наречие 'неловко, неудобно' (досл. отрицание 'не' + прилагательное 'хороший, удобный');

в) структурно-синтаксическая И, являющаяся след-

³ Главным образом к некоторым глаголам, оформляющимся временными суффиксами, и к личным местоимениям, принимающим во множественном числе суффикс *-мэнь*.

ствием условности порядка компонентов единицы. Действительно, с точки зрения чисто синтаксических отношений трудно мотивировать твердый порядок компонентов в таком сложном слове с копулятивной связью компонентов, как *бао¹го³*: 1) существительное 'узел' (досл. 'узел' + 'узел'); 2) глагол 'завернуть (в узел)' (досл. 'завертывать' + 'завертывать').

Здесь приводится лишь самое общее деление И на виды. Оно позволяет отличать ЕЯ от ЕР. Дальнейшее подразделение ЕЯ на слова и фразеологизмы производится по более дробному делению И на подвиды, например по степени ощущения носителем языка образности: фразеологизмы, как правило, более образны, слова — нет.

Несомненно, что настоящая деление И на виды является условной абстракцией от реальных единиц. Каждая ЕЯ фактически может быть идиоматичной во всех планах, по всем направлениям.

Трудность мотивировки цельного значения значениями компонентов объясняется условным характером связи плана выражения с планом содержания, присущим языковому знаку, произвольностью ЕЯ как знака.

Таким образом, мы видим, что И является наиболее важным ДП, позволяющим отличать один тип единиц от другого. Это — свойство всех ЕЯ, внутренне присущее им и являющееся ярким выражением самой сущности отношений между языковым знаком и значением. Применение И для выделения ЕЯ не требует статистики.

Несмотря на содержательный характер понятия И, представляется возможным найти некоторые формальные признаки ее обнаружения. Так, например, при трансформации переменного атрибутивного словосочетания *бянь⁴жэнь²* 'подходящий человек' в предикативное *жэнь²бянь⁴* 'человек подходящий' не происходит нарушения ни грамматической правильности, ни общего смысла высказывания. Соответствующая трансформация номинативной ЕЯ *гань¹синь¹* (глагол 'примириться', досл. 'сладкий, приятный' + 'сердце') не может быть произведена без изменения значения.

В силу центрального положения И среди других ДП, отличающих ЕР от ЕЯ, это понятие нуждается в дальнейшей, более тщательной разработке как на материале китайского языка, так и языков других типов. Также требуется дальнейшая разработка в смысле уточнения и других ДП (*Ц₃*, *В*, *У*), по которым ЕР стлчаются от ЕЯ, а внутри последних — слова от устойчивых словосочетаний (по специальному приращению этих же ДП и по признакам *Р_с*, *Ц_м*, *Р_м* и *А_с*).

По причине явной неприложимости к исследуемому материалу здесь не рассматриваются такие не менее важные ДП, как фонетический и связанность значения компонентов.

Процедура выделения ЕЯ начинается с применения *Ц_с*. Затем, после выделения некоторых односложных слов, для определения характера оставшихся сложных образований последовательно применяются И, *Ц₃*, *Р_с*, *В* и *У*. Некоторые ЕЯ выделяются сразу же после применения *Ц_с* по наличию у них одного из побочных ДП (*Ц_м*, *Р_м*, *А_с*).

В результате соответствующих процедур выделяются морфемы, определяемые как самостоятельные или несамостоятельные, знаменательные или служебные. В исследуемом нами тексте встретились морфемы, характеризующиеся любой из комбинаций указанных здесь оппозиций: самостоятельные + знаменательные, самостоятельные + служебные и т. д. Имеются переходные случаи. Каждый тип морфемы характеризуется определенным набором ДП, классические сложные слова состоят из несамостоятельных, знаменательных (реже — служебных)

морфем и обладают такими ДП, как Ц_с, И, Ц_з, В и У. Кроме того, они могут характеризоваться наличием Ц_м и А, а также отсутствием Р_с и Р_м.

ЛИТЕРАТУРА

1. Гринберг Дж., Квантитативный подход к морфологической типологии языков, — сб. «Новое в лингвистике», вып. III, М., 1963.
2. Лу Чжи-вэй, Бэйцзин хуа даньиньцы цыхуэй (Словарь односложных слов пекинского диалекта), Пекин, 1956.
3. Ревзин И. И., Модели языка, М., 1962.
4. Смирницкий А. И., К вопросу о слове (Проблема «отдельности слова»), — сб. «Вопросы теории и истории языка в свете трудов И. В. Сталина по языкознанию», М., 1952.
5. Торопов А. А., Лексико-грамматические принципы составления словаря средневекового китайского языка (эпоха Сун), — сб. «Письменные памятники и проблемы истории культуры народов Востока. Тезисы докладов III годичной научной сессии ЛО ИНА», Л., 1967.
6. «Цзин бэнь тунсу сяошо» («Популярные рассказы, изданные в столице»), Шанхай, 1954.
7. Яхонтов С. Е., Методы выделения грамматических единиц, — сб. «Языковые универсалии и лингвистическая типология», М., 1969.