

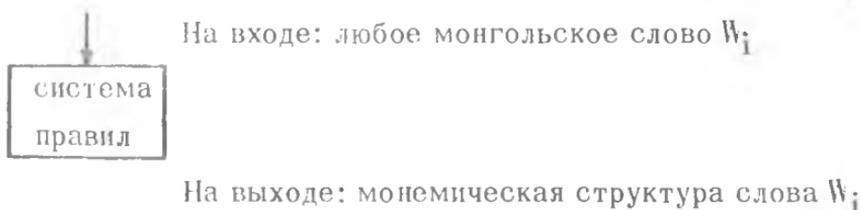
Х.-П. Фице

## ФОРМАЛЬНАЯ МОРФОЛОГИЧЕСКАЯ СТРУКТУРА МОНГОЛЬСКОГО СЛОВА

### 1. Монемы и морфемы

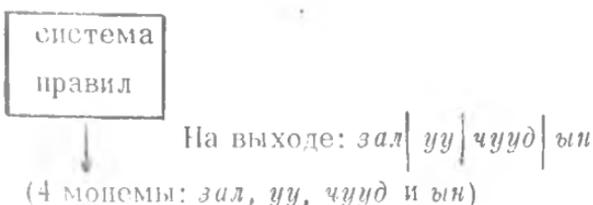
Мы употребляем название "монема" вместо "морфема" (по А. Мартине [1, 366-566], школе Фридриха Мейера и другим), потому что термин "морфема" нам нужен в другом месте.

Постулат: Имеется система правил, которая разбивает каждое написанное монгольское слово [5] на монемы [4]:



Пример:

На входе: *залуучуудын* 'молодежи'



Таким образом мы получаем в монгольском языке конечное множество  $A$  всех монем монгольских слов. (Для повышения математической проблематики достаточно элементарного введения в теорию множеств.) Элементы этого мно-

жества, т.е. отдельные монемы, мы обозначаем символами  $a_1, a_2, \dots, a_n$ :

$$A = a_1, a_2, \dots, a_n$$

Разумеется, эти монемы отличаются друг от друга, и их можно разделять по определенным признакам. Если такое разделение делается точно (в математическом смысле), оно может служить базой программы автомата, который добавляет к расчленению монгольского слова на монемы характеристику этих монем.

Из множества  $A$  можно выделить подмножество  $B: A \supseteq B$ .  $B$  является множеством всех монем  $a_i$ , так что существует  $a_j$  ( $i \neq j$ ), причем  $a_i = a_j$  и  $a_i \approx a_j$ . В действительности  $A \supset B$ .

Соотношение эквивалентности  $\approx$  определяет в множестве  $B$  разбиение на классы, причем каждый класс содержит по меньшей мере два элемента. Классы попарно дизъюнкты, т.е. пересечение двух любых классов всегда пустое. Объединение всех классов равно  $B$ . Мы обозначаем эти классы символами  $C_i$  ( $1 \leq i \leq k$ ); т.е.:  $C_i \subseteq B$ ;  $C_1 \cup C_2 \cup \dots \cup C_k = B$ ;  $C_i \cap C_j = \emptyset$ .

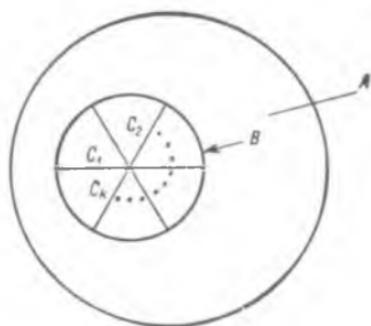
Класс  $C_i$  определяется следующим образом:  $C_i$  является множеством семантически равноценных монем  $\{a_1, a_2, \dots, a_k\}$  ( $1 < i \leq k$ ), которые находятся в комплементарной дистрибуции (в смысле дистрибуционного анализа Хэрри-са), т.е. могут быть заменены

между собой без влияния на семантику любого слова, к которому

они могут принадлежать. Элементы класса  $C_i$ , как правило, графически похожи друг на друга.

Пример одного подмножества  $C_i$ :

- (1) мал 'скот' - мал|ын' скота'
- (2) мод 'дерево' - мод|ны' дерева'
- (3) гэр 'юрта' - гэр|ийн' юрты'



(4) *цай* 'чай' – *цай|н* 'чая'

(5) *хөдөө* 'худон' – *хөдөө|ний* 'худона'

(6) *хөдөө|гийн* 'худона'

Шесть подчеркнутых монем образуют одно подмножество  $S_i$ . Они семантически равноценны (образуют родительный падеж). Их выбор определяется в каждом случае составом гласных и конечными согласными тех монем, к которым они добавляются. Примеры (5) и (6) являются одновременно примерами возможной замены двух монем без последствий в области семантики.

Класс  $S_i$  может содержать только два элемента, например: *чи* 'ты' и *чам* (в *чам шиг* 'как ты').

Каждый класс  $S_i$  называется "морфема", каждый элемент класса  $S_i$  – "алломорф". Морфемы являются множествами по меньшей мере двух монем. Алломорфы являются монемами. Подмножество  $B$  из  $A$  – множество всех морфем монгольского языка.

Чтобы упростить нотацию, не пишут "морфема *-ын, -ны, -ийн, -ний, -н, -гийн*", а выбирают один алломорф, который заключают в фигурные скобки { ... } вследствие чего он будет представителем морфемы. В данной статье мы воспользуемся для нотации морфем наклонными скобками /.../, чтобы не сталкиваться с символикой теории множеств (иначе морфемы могли бы быть поняты как множества с одним элементом).

Предлагается выбрать представителем морфемы алломорф, который (1) содержит графемы *а* или *ы*, является исходной формой (2) или встречается чаще всего (3), например:

/-лаа/ =-лаа, -лоо, -лээ, -лөө (прошедшее время);

/-чи/ =чи, чам 'ты';

/-жээ/ =-жээ, -чээ (прошедшее время).

Нотация морфем сложна в отдельных случаях, о чем речь пойдет ниже.

1) Рассмотрим следующие примеры:

*цагаан* 'белый' – *цав цагаан* 'совсем белый',

*хар* 'черный' – *хав хар* 'совсем черный',

*xox* 'синий' – *xov xox* 'совсем синий'.

Монемы *цав*, *хав* и *хов* семантически равноценны, они усиливают значение последующих монем. Кроме того они находятся в отношении комплементарной дистрибуции. Внутриязыковым критерием выбора является первая и вторая графема слова, значение которого эти монемы усиливают. Таким образом, мы, несомненно, имеем дело с морфемой.

Но невозможно выбрать один алломорф (например *цав*) представителем морфемы, так как это могло бы вести к сегментации слов *хов хox* на */цав/ + хox*. Поэтому предлагается критерием выбора представителя морфемы сделать критерий выбора алломорфов. Так, мы пишем эту морфему следующим образом:

$$/\#x_1, x_2, \dots, x_n\# \implies \#x_1, x_2, \text{ в } \#x_1, x_2 \dots x_n\# /$$

с возможным упрощением  $/x_1, x_2, \text{ в}/$ , где  $\#$  – пробел,  $x_1$  – буква,  $\implies$  – заменяй следующим,  $/\dots/$  – обозначение морфемы.

2) То же самое применимо и для таких случаев, как:

*ном* 'книга' – *ном ном* 'книги и т.д.',

*цаас* 'бумага' – *цаас маас* 'бумага и т.п.',

*мах* 'мясо' – *мах зах* 'мясо и т.п.',

чтобы писать морфему следующим образом:

$$/\#x_1, x_2, \dots, x_n\# = x_1, x_2, \dots, x_n\# \begin{matrix} [m] \\ [3] \end{matrix}, x_2, \dots, x_n\#$$

с упрощением  $[m \text{ или } 3], x_2, \dots, x_n /$ .

## 2. Лексемы и тагемы

Кроме выделения подмножества В можно разбить множество А (совокупность всех монем) на два класса D и E с условиями  $D, E \subseteq A$ ;  $D \cup E = A$ ;  $D \cap E = \emptyset$ .

Сначала рассмотрим класс D. Его элементы мы называем "лексемами". Лексемой является каждая монема, которую можно изолировать, т.е. такая монема, которая потенциально встречается между двумя пробелами, например:

(1) *мал* 'скот',

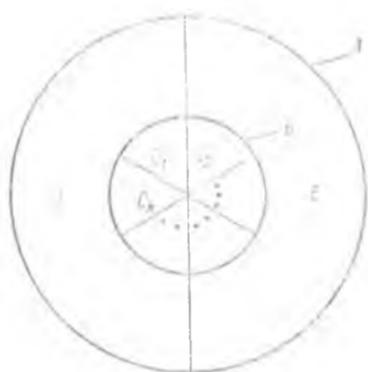
(2) *сур | и | ууль* 'школа' (только первую монему можно изолировать),

(3) *ав|ч|ир* 'принеси!' (первую и последнюю морфему можно изолировать, поэтому это слово содержит две лексемы).

Классе E легко можно определить как дифференцию  $A \setminus B$ , т.е. каждая морфема, которая не является элементом класса B (лексемой), есть элемент класса E;  $a \in E$  тогда и только тогда, когда  $a \notin B$ .

Элементы класса E мы называем "тагмемами". Примеры тагмем: все неподчеркнутые морфемы примеров (1) – (3).

Так мы получили следующую систему множеств:



- A – множество всех морфем,
- B – множество всех лексем,
- C – морфема,
- D – множество всех лексем,
- E – множество всех тагмем.

Разбиение множества A на классы D и E по является размельчением разбиения множества B на классы  $C_i$ , так как, если любой элемент одной морфемы является тагмемой, то и все остальные элементы этой морфемы есть тагмемы.

Множества  $C_i$ , которые находятся в пересечении  $D \cap B$ , называются "лексемы-морфемы", а те, которые в пересечении  $E \cap B$  "тагмемы-морфемы". Пример лексемы-морфемы:

*/руу/ = руу, рүү, луу, лүү* 'в направлении на'.

Пример тагмемы-морфемы:

*/-аарай/ = -аарай, -оорой, -ээрэй, -нарэй* (императив).

Все элементы пересечения  $B \cap D$  можно назвать "лексемы-алломорфы", все элементы пересечения  $B \cap E$  — "тагмемы-алломорфы". Лексемы-алломорфы и тагмемы-алломорфы являются монемами.

### 3. F-тагмемы и G-тагмемы

Все обозначения в этой статье мы выбрали с точки зрения системы множеств. В нашем распоряжении имеются еще такие термины, как суффикс, окончание, аффикс, постфикс, формант; грамматема, таксема; словообразующий, формообразующий (суффикс, окончание и т.д.) и др.

Множество E всей тагмем можно опять разбить на два класса F и G. Элементы класса F мы называем "F-тагмемы", а класса G — "G-тагмемы". На вопрос, принадлежит ли та или иная тагмема к F-классу отвечает следующая проверка: тагмема принадлежит к классу F, если она потенциально воздействует только на то слово, к которому она добавлена. Примеры:

(1) *тэмээ|ч|ин, адуу|ч|ин хоёр* 'верблюжий пастух', 'табунщик'; тагмемы *ч* и *ин* нужно добавить и к слову *тэмээ* 'верблюды' и к слову *адуу* 'табун', чтобы получить слова со значением 'верблюжий пастух', 'табунщик'.

(2) *өндөр шинэ|хэн байшин* 'высокое, довольно новое здание'; *хэн* имеет уменьшительное значение. Эта монема (как и монемы *ч* и *ин* первого примера) воздействует только на слово, к которому она добавлена, но не на слово *өндөр* 'высокий'. Таким образом, она является F-тагмемой.

При глагольных тагмемах проверка реализуется с помощью деепричастия:

(3) *Би түүн|ийг сан|уул|аад яв|уул|сан* 'Я ему напомнил (о чем-то) и отослал'; *уул* — тагмема с побудительным значением. Чтобы получить побудительные залоги лексем *сана-* 'вспоминать' и *яв-* 'идти', надо ее добавить к обоим лексемам. Поэтому она — F-тагмема.

(4) *Бидэц|оод тани|лц|сан* 'Мы пошли и познакомились';

-*ли*- — тагмема с кооперативным значением. Она добавлена только к лексеме *тани*- 'знать кого-то' и воздействует только на семантику этой лексемы, но не на семантику лексемы *оч*- 'идти'. Поэтому она тоже является F-тагмемой. Множество F всех F-тагмем можно опять разбить по синтаксическим, семантическим, позиционным, фонетическим и другим критериям на большое число классов. Разбиение по синтаксическим и семантическим признакам на классы отглагольных, отноминальных, глаголообразующих, побудительных, страдательных и многих других F-тагмем уже хорошо разработано [2; 3]. Мы здесь, как пример, кратко обрисовываем возможность разбить множество на классы по положению F-тагмем в слове. Классы мы обозначаем позиционными индексами (i):

Класс  $FT_{(1)}$ : F-тагмемы этого класса занимают в последовательности F-тагмем любого слова первое место, например, *-ли*- в слове *тани|ли|ах* 'познакомиться'.

Класс  $FT_{(2)}$ : F-тагмемы этого класса добавляются к лексемам и F-тагмемам классов  $FT_{(1)}$ ,  $FT_{(2)}$ , ...,  $FT_{(i-1)}$ , например *-уул*- в слове *тани|ли|уул|ах* 'познакомить'.

Класс  $FT_{(n)}$ : F-тагмемы этого класса занимают в последовательности F-тагмем любого слова последнее место, например, *агч* в слове *тани|ли|уул|агч* 'тот, кто знакомит'.

Мы уверены в том, что с помощью обратного словаря современного монгольского языка было бы возможно выработать целую систему позиционного поведения монгольских тагмем. Но это выходит за рамки нашей статьи.

Класс G; элементы которого мы называем "G-тагмемы", определяется следующим образом: тагмема принадлежит к классу G, если сфера ее влияния распространяется потенциально не только на то слово, к которому она добавлена, но и на предыдущее слово, например:

(1) *эх үр|ийн жарга л* 'счастье матери и ребенка'. Тагмема *-ийн* (родительный падеж) добавлена только к слову *үр* 'ребенок' (литер., но она распространяет свое влияние и на

предыдущее слово, как и тагмама *-оос* (аблатив) в следующем примере:

*аос үлэ хэср | оос* 'от отца и матери'. Так *-ийн* и *-оос* являются G-тагмамами.

При глагольных тагмамах мы опять воспользуемся для проверки деепричастием:

*Дорж хоооо амар | на*. 'Дорж поедет в худон и отдохнет'. Тагмама *-на* (будущее время) распространяется семантически и на предыдущее слово *явж*, и таким образом, она является G-тагмой.

И множество G всех G-тагмам можно разбить на классы  $GT_{(i)}$  ( $1 \leq i \leq n$ ) по позиционным признакам G-тагмам, например:

$GT_{(1)}$ : G-тагмамы этого класса стоят непосредственно после лексем и F-тагмам, например, *д* в слове *нөхө(р) д* 'товарищи'.

$GT_{(2)}$ : после лексем, F-тагмам и G-тагмам класса  $GT_{(1)}$ , например, *д* в *нөхө(р) | д | үүд* 'товарищи'.

$GT_{(1-1)}$ : как  $GT_{(2)}$ , но после G-тагмам классов  $GT_{(2)}, \dots, GT_{(1-1)}$ , например, *-ийн* в *нөхө(р) | д | үүд | ийн* 'товарищей'.

$GT_{(k)}$ : как  $GT_{(2)}$ , но после G-тагмам классов  $GT_{(2)}, \dots, GT_{(i)}, \dots, GT_{(k-1)}$ , например, последний *-д* в *нөхө(р) | д | үүд | ийн | д* 'в квартире (юрте, на месте работы) товарищей'.

$GT_{(n)}$ : G-тагмамы этого класса встречаются всегда в конце слова, например, *-оо* в *нөхө(р) | д | үүд | ийн | д | оо* 'в квартире (юрте, на месте работы) своих товарищей'.

Если не слишком далеко уходить от конкретных данных языка, то установить дизъюнктивность между множествами F и G не является возможным:  $F \cup G = E$ ,  $F \cap G = \emptyset$ .

Элементы в пересечении  $F \cap G$  мы называем "FG-тагмамы". Пример FG-тагмамы: *-о* в слове *сай(н) д* 'министр' (от *сайн* 'хороший') как F-тагмама. Проверка:

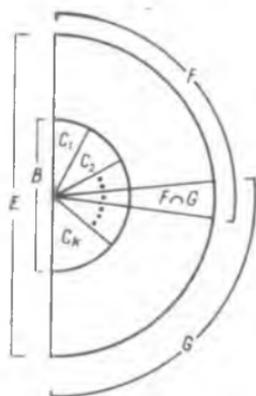
*нөхөр сайн | д* 'товарищ министр',

*-о* в слове *позий | о* 'свиавы и собаки' как G-тагмама. Проверка:

*позий | о* 'свиавы и собаки'.

В заключение мы рассмотрим опять соотношения между обоими не дизъюнктивными классами F и G и подмножеством B ( $B \subset A$ ,

$B = \{C_1, C_2, \dots, C_K\}$ ). Параллельно с определением лексем-морфем и тагмем-морфем мы обозначаем все множества  $G$  в пересечении  $F \cap B$  как – "F-тагмемы-морфемы", все  $C_j$  в пересечении  $G \cap B$  как "G-тагмемы-морфемы" и все множества  $C_n$  в пересечении  $(F \cap G) \cap B$  как – "FG-тагмемы-морфемы"



Пример F-тагмемы-морфемы: /-втар/ –-втар, -втор, -втэр, -втөр; значение: ослабление обозначений цветов, например, *хар* 'черный', *харавттар* 'черноватый'.

Пример G-тагмемы-морфемы: /-сан/ =-сан, -сон, -сэн, -сөн; значение: причастие, прошедшее время.

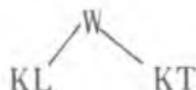
Пример FG-тагмемы-морфемы: /-нар/ =-нар, -нэр (если условиться, что *-нар* является монемой, а не словом). Значение: как F-тагмема /-нар/ обозначает множественное число, как G-тагмема оно изменяет значение слова *эх* 'мать' в *эхнэр* 'супруга'.

Все элементы в пересечении  $F \cap B$  называются "F-тагмемы-алломорфы", все элементы в  $G \cap B$  – "G-тагмемы-алломорфы" и все элементы в  $(F \cap G) \cap B$  – "FG-тагмемы-алломорфы". Все эти элементы есть монемы.

#### 4. Формальная морфологическая структура монгольского слова.

Каждое монгольское слово  $W$  можно разбить на конфигурацию лексем  $KL$  и конфигурацию тагмем  $KT$ , причем присутствие  $KL$  обязательно и присутствие  $KT$  факультативно:

$W \longrightarrow KL, KT$



Сначала рассмотрим конфигурацию лексем KL. Она начинается после пробела и кончается последней справа лексемой L.

Следующие структуры KL возможны:

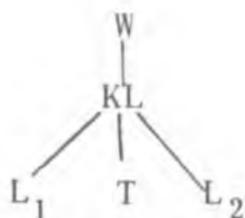
KL  $\longrightarrow$  L

KL  $\longrightarrow$  L<sub>1</sub>, L<sub>2</sub>

KL  $\longrightarrow$  L<sub>1</sub>, T, L<sub>2</sub>

Например: мал 'скот', *виѠ(Ѡ) ѡѡр* 'сегодня', *ав|ч|ир* 'принеси!', (ч – тагмема).

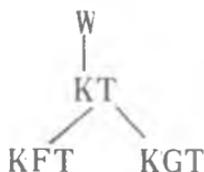
Таким образом, мы получаем для потенциальной конфигурации лексем KL и ее отношения к слову следующее дерево:



Примечание: По нашему определению KL кончается последней справа лексемой. Поэтому тагмема имеется только тогда в KL, когда лексема L следует за ней.

Конфигурацию тагмем KT можно разбить на конфигурацию F-тагмем – KFT и на конфигурацию G-тагмем – KGT:

KT  $\longrightarrow$  KFT, KGT



KFT начинается непосредственно после конфигурации лексем KL и кончается последней справа F-тагмемой слова. KFT может содержать n F-тагмем, которые расположены по их позиционным индексам:

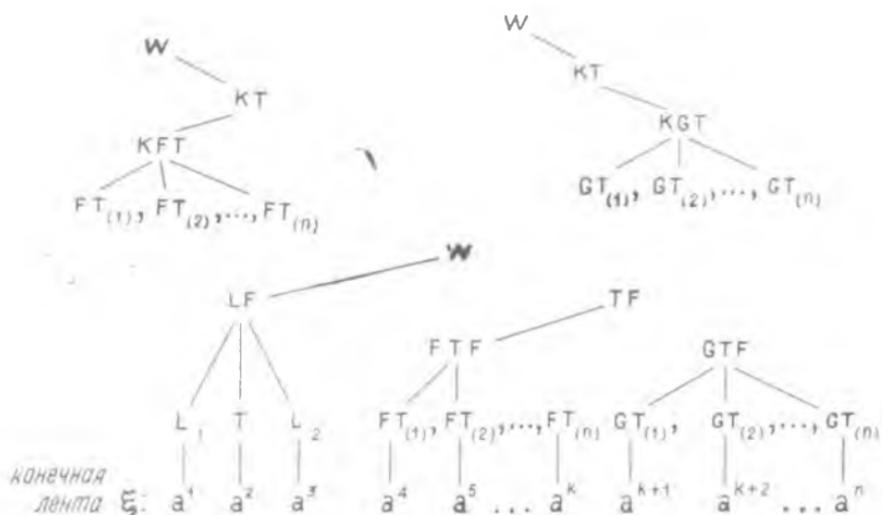
KFT  $\longrightarrow$  FT<sub>(1)</sub>, FT<sub>(2)</sub>, ..., FT<sub>(n)</sub> (пример см. раздел 3).

Конфигурация G-тагмем KGT начинается непосредственно после KFT или после конфигурации лексем KL, если KFT отсутствует. KGT содержит n G-тагмем:

$KGT \longrightarrow GT_{(1)}, GT_{(2)}, \dots, GT_{(n)}$  (Пример см. раздел 3).

Положение FG-тагмем нейтрально, т.е. они могут занимать как место F-тагмем в KFT, так и место G-тагмем в KGT.

Итак, мы получили потенциальную формальную морфологическую структуру монгольского слова:

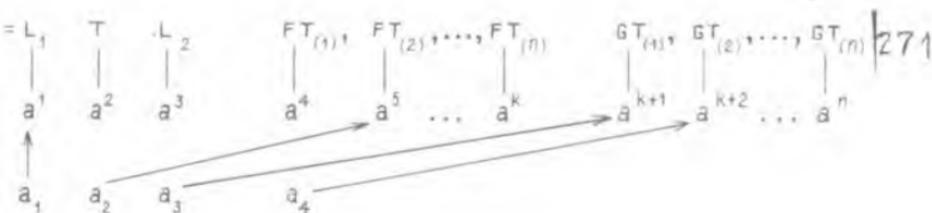


Конечная лента:  $a^1, a^2, a^3, a^4, a^5, \dots, a^k, a^{k+1}, a^{k+2}, \dots, a^n$ .

Каждое монгольское слово можно однозначно привязать к конечной ленте  $\xi$ . Взаимно-однозначная привязь на практике не встречается. Привязываются монемы  $a_1, a_2, \dots, a_n$  любого монгольского слова к элементам конечной ленты  $a^1, a^2, \dots, a^n$ . Для этой привязи действительны следующие правила:

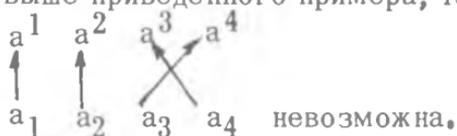
(1)  $a_1 \longrightarrow a^1$ ; (2)  $a_i (1 < i \leq n) \longrightarrow a^j (j \geq i, 1 < j \leq n)$ ,

т.е. каждая первая монема  $a_1$  ( $a_1 \in w$ ), которая должна быть лексемой, привязывается к  $a^1$  ( $a^1 \in \xi$ ). Каждая другая монема  $a_i$  ( $i > 1$ ) привязывается к элементу  $a^j$  из  $\xi$ , у которого тот же самый или более высокий индекс, чем у  $a_i$ . Пример:



$W = яв|уул|агч|аас$  'от отправителя'.

Но, в отличие от выше приведенного примера, такая привязь как, например,



Итак, мы получили в основном функцию, которая придает каждому монгольскому слову его формально-морфологическую структуру (здесь в абстрактном плане). Конкретный запас значений этой функции можно получить путем классификации монем монгольских слов по их позиционному поведению. Но это выходит за рамки этой теоретической статьи.

### Л и т е р а т у р а

1. Мартине А., Основы общей лингвистики, — «Новое в лингвистике», вып. III, М., 1963.
2. Санжеев Г.Д., Грамматика бурят-монгольского языка, М.-Л., 1941.
3. Санжеев Г.Д., Сравнительная грамматика монгольских языков, т. I. Глагол, М., 1953.
4. Vietze, Das Problem der Semantikstrukturierung bei einer automatischen monematischen Analyse, — «Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung» (в печати).
5. Vietze, Eine Definition des mongolischen Wortes, — «Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung», Bd 22/I, 1969.