

HANS-PETER VIETZE (Berlin)

Zur Struktur des Wortes in altaischen Sprachen

Summary

In continuation of two previous articles in which the word and its semantic structure in Altaic languages were discussed the word is regarded as a set of monemes. This set is dissected in different ways, especially under the aspect of the positional distribution of the monemes. The result is a model in the form of a tree ending in a string. This string reflects the potential morphological structure of any word of an Altaic language.

1. Corpus

Unter dem Begriff „altaische Sprachen“ werden die Turksprachen sowie die mongolischen und mandschurisch-tungusischen Sprachen zusammengefaßt, worüber unter den Fachwissenschaftlern im wesentlichen Einmütigkeit besteht. Umstritten ist die Zugehörigkeit des Koreanischen und Japanischen zu dieser Sprachfamilie. Das Problem der genetischen Verwandtschaft der altaischen Sprachen, die sogenannte altaische Hypothese, bleibt unbewiesen und vielleicht unbeweisbar, wenn man bedenkt, daß die ältesten alttürkischen Inschriften, die aus dem 8. Jh. stammen, fast 3000 Jahre jünger sind als die ältesten bekannten indoeuropäischen Quellen.

Typologisch vereinen die altaischen Sprachen in erster Linie die in der Morphologie stark dominierende Agglutination sowie in der Syntax das „altaische Grundprinzip“ (das Bestimmende steht vor dem zu Bestimmenden). Es gibt nur wenig eindeutige intersprachliche Lautgesetze. Die typologischen Gemeinsamkeiten sind jedoch groß genug, um Erkenntnisse zu einer Sprache in vielen Fällen auf eine andere transponieren zu können, wie auch im folgenden gezeigt werden soll, so daß die mit diesen Sprachen befaßte orientalistische Teildisziplin, die Altaistik, zumindest pragmatisch gerechtfertigt ist.

2. Moneme und Morpheme

Wir hatten in früheren Aufsätzen versucht, das mongolische Wort über einen Algorithmus zu definieren¹ und ein Regelsystem zu erarbeiten, das jedes geschriebene Wort einer altaischen Sprache in Moneme zerlegt,² das türkische Wort *dostlarım* 'meinen Freunden' also z. B. in die Monemfolge *dost|lar|ım|a*. Heuristisch seien die Moneme als kleinste bedeutungstragende Einheiten umrissen, definiert sei eine endliche Menge A aller Moneme a_i ($1 \leq i \leq n$) über das o. g. Regelsystem oder über eine Liste.

Aus A wird eine echte Teilmenge B ausgesondert: $A \supset B$. B ist die Menge aller Moneme a_i , für die ein a_j existiert, wobei gilt: $a_i \neq a_j$ und $a_i \approx a_j$. Die Äquivalenzrelation \approx definiert in B eine Zerlegung in paarweise disjunkte Klassen C_i ($1 \leq i \leq k$), deren Vereinigung wieder gleich der Menge B ist. Eine Klasse C_i ist eine aus mindestens zwei Elementen bestehende Menge semantisch gleichwertiger Moneme in komplementärer Distribution.

Beispiel für eine Klasse C_1 (mong.): malyn 'des Viehs', modny 'des Baumes', gerijn 'der Jurte', cajn 'des Tees', chödöönij oder chödöögijn 'der ländlichen Gegend'. Die sechs unterstrichenen Moneme sind semantisch gleichwertig (sie bilden den Genitiv). Ihre Auswahl wird vom Vokalbestand und vom Auslaut des jeweiligen Wortstammes bestimmt. Eine Klasse C_1 kann auch nur zwei Elemente enthalten, z. B. (mong.) či 'du' und čam 'du' (in čam šig 'wie du').

Jede Klasse C_1 heißt Morphem, jedes Element einer Klasse C_1 heißt Allomorph. Morpheme sind Mengen von mindestens zwei Monemen. Allomorphe sind Moneme. Die Teilmenge $B \subset A$ ist die Menge aller Morpheme der untersuchten Sprache.

Ein Morphem wird im folgenden als für C_1 repräsentatives Allomorph zwischen Strichen |...| notiert, da die übliche Notation (geschweifte Klammern) zu Verwechslungen mit der mengentheoretischen Symbolik führen könnte. Bei Schwierigkeiten in Ausnahmefällen, wie z. B. türk. eski 'alt' – epeski 'ganz alt', irak 'fern' – ıpirak 'ganz fern', uzun 'lang' – ıpuzun 'ganz lang' (ep, ıp und ıp sind per definitionem Allomorphe), kann das Morphem über eine Konvention als $|x_1p|$ notiert werden (x_1 : erster Buchstabe).

3. Lexeme und Suffixe

Neben der Aussonderung der Teilmenge B kann die Menge A (die Menge aller Moneme) in zwei Klassen D und E mit den Bedingungen $D, E \subset A$, $D \cup E = A$ und $D \cap E = \emptyset$ zerlegt werden. Die Elemente von D heißen Lexeme. Ein Lexem ist ein isolierbares, autosemantisches Monem, z. B. (1) türk. ev 'Haus', (2) mandsch. monggo|so 'Mongolen', (3) mong. av|čir 'bringe herbei!'. Das Beispiel (3) ist der seltene Ausnahmefall eines Wortes mit zwei getrennten Lexemen (av 'nimm!', ir 'komme!', č: Konverbkennzeichnung).

Die Klasse E kann leicht als die Differenz $A \setminus D$ definiert werden, d. h. jedes Monem, das nicht Element der Klasse D (kein Lexem) ist, ist ein Element der Klasse E : $a \in E$ genau dann, wenn $a \notin D$ (a : Monem). Die Elemente der Klasse E heißen Suffixe. Beispiele für Suffixe sind die nicht unterstrichenen Moneme in den o. a. Beispielen (2) und (3).

Lemma: Die Zerlegung der Menge A in die Klassen D und E ist keine Verfeinerung der Zerlegung der Menge B in die Klassen C_i , denn wenn ein beliebiges Element eines Morphems ein Suffix ist, dann sind auch alle übrigen Elemente dieses Morphems Suffixe.

$C_i \subset (D \cap B)$ (eine Menge, die sich im Durchschnitt $(D \cap B)$ befindet) heißt Lexem-Morphem, $C_i \subset (E \cap B)$ heißt Suffix-Morphem. Beispiel für ein Lexem-Morphem: mong. $|ruu| = ruu, rüü, luu, lüü$ 'in Richtung auf...'. Beispiel für ein Suffix-Morphem: mandsch. $|-na| = -na, -ne, -no$ (Illativ).

4. Wortbildende und formbildende Suffixe

Die Menge *E* aller Suffixe kann in zwei Klassen *F* und *G* zerlegt werden. Die Elemente der Klasse *F* heißen wortbildende Suffixe, die Elemente der Klasse *G* heißen formbildende Suffixe. Ein Suffix gehört zur Klasse *F*, d. h. es ist wortbildend, wenn es potentiell nur genau die Bedeutung desjenigen Wortstammes, an den es angefügt ist, zu einer neuen Wortbedeutung transformiert, z. B. (mong.):
 temee|č|in aduu|č|in chojor 'der Kamelhirt und der Pferdehirt'



Die Suffixe -č und -in müssen sowohl an das Wort temee 'Kamel' als auch an das Wort aduu 'Pferde(-herde)' angefügt werden, um neue Wörter mit der Bedeutung 'Kamelhirt' bzw. 'Pferdehirt' zu erhalten. Bei verbalen Suffixen wird dieser Test mit Hilfe von Konverben durchgeführt, z. B.:

Bi tüün|ijg san|uu|aad jav|uu| san. 'Ich habe ihn (daran) erinnert und weggeschickt.'



Das Kausativ-Suffix -uu muß sowohl an das Lexem sana- 'sich erinnern, als auch an das Lexem jav- 'gehen' angefügt werden, um die Bedeutungen 'erinnern' bzw. 'wegschicken' zu erhalten.

Die Menge *F* aller wortbildenden Suffixe kann nach morphologischen, syntaktischen semantischen, phonetischen u. a. Kriterien in eine große Zahl von Klassen, z. B. in Klassen deverbaler, verbbildender, hintervokalischer u. a. Suffixe zerlegt werden, was in jeder Grammatik getan wird. Im folgenden wird die Menge *F* nach der Position der wortbildenden Suffixe in Klassen *FS* zerlegt. Diese Klassen verstehen wir mit dem Positionsindex (*i*).

FS₍₁₎: Die wortbildenden Suffixe dieser Klasse nehmen in der Suffixkette eines beliebigen Wortes potentiell den ersten Platz ein, d. h. sie stehen nach einem Lexem, z. B. mandsch. -la in aisi|la- 'helfen'.

FS_(i): Die wortbildenden Suffixe dieser Klassen ($1 \leq i \leq n$) können an Lexeme oder an wortbildende Suffixe der Klassen *FS₍₁₎*, *FS₍₂₎*, ..., *FS_(i-1)* angefügt werden, z. B. mandsch. -ndu in aisi|la|ndu- 'einander helfen'.

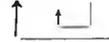
FS_(n): Die wortbildenden Suffixe dieser Klasse stehen in der Kette der wortbildenden Suffixe eines beliebigen Wortes an letzter Position. Das gilt z. B. für die Nomina actoris bildenden mongolischen Suffixe |-agč| oder für die türkische Negation |-ma|.

SWIFT war unseres Wissens der erste, der die türkischen Suffixe zunächst nach traditionellen Gesichtspunkten, dann aber auch nach ihrem o. a. positionellen Verhalten klassifizierte.³ Hier sei sein Beispiel koş|uş|tur|ul|ma- 'nicht miteinander herumrennen lassen' angeführt:

	1	2	3	4
Stamm	reflexiv / reziprok	kausativ	passiv	Negation
koş-	-uş	-tur	-ul	-ma
'rennen'				

Die Klasse *G*, deren Elemente formbildende Suffixe heißen, wird wie folgt abgegrenzt: Ein Suffix gehört zur Klasse *G*, wenn sein semantischer Wirkungsbereich über das Wort, dessen Bestandteil es ist, hinausgehen kann und ein oder mehrere gleichgeord-

nete vorhergehende Wörter mit umfaßt, z. B. (mong.): ech ür|*ijn* žargal 'das Glück der Mutter und des Kindes'



Das Suffix *-ijn* (Genitiv) ist nur an das Lexem ür 'Kind (lit.)' angefügt, erstreckt sich aber in seiner grammatischen Bedeutung auch auf das vorangegangene gleichgeordnete Wort ech 'Mutter'. Bei verbalen Suffixen wird der Test wieder mit Hilfe von Konverben durchgeführt, z. B.:

Dorz chödöo jav|ž amar|na. 'Dorž wird auf das Land fahren und sich erholen.'



Das Suffix *-na* (hier: Futur) erstreckt sich in seiner grammatischen Bedeutung sowohl auf das Lexem amar- 'sich erholen' als auch auf das Lexem jav- '(hier): fahren'.

Die Menge G der formbildenden Suffixe kann unter dem Gesichtspunkt ihres positionellen Verhaltens wieder in Klassen $GS_{(i)}$ ($1 \leq i \leq n$) zerlegt werden.

$GS_{(1)}$: Die formbildenden Suffixe dieser Klasse stehen immer unmittelbar nach Lexemen oder wortbildenden Suffixen, z. B. mong. Pl. *-d* in nöchö(r)|*d* 'Genossen'.

$GS_{(2)}$: Nach Lexemen, wortbildenden Suffixen und formbildenden Suffixen der Klasse $GS_{(1)}$, z. B. Pl. *-iiüd* in nöchö(r)|*d* *üüd* 'Genossen'.

$GS_{(i)}$: Wie $GS_{(2)}$, aber auch nach formbildenden Suffixen der Klassen $GS_{(2)}$, ..., $GS_{(i-1)}$, z. B. Genitiv *-ijn* in nöchö(r) |*d* |*iiüd* |*ijn* 'der Genossen', Lokativ *-d* in nöchö(r)|*d* |*iiüd* |*ijn* |*d* 'bei den Genossen'.

$GS_{(n)}$: Am Ende eines Wortes, z. B. Possessiv *-ee* in nöchö(r)|*d* |*iiüd* |*ijn* |*d* |*ee* 'bei seinen Genossen'.

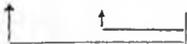
Die Mengen F und G können zumindest für die mongolische Sprache nicht als disjunkt betrachtet werden. Die Elemente (Suffixe) im Durchschnitt $F \cap G$, deren Zahl sehr gering ist, können sowohl wort- als auch formbildend sein, z. B.:

nöchör saj(n)|*d* 'Genosse Minister' (nöchör 'Genosse', sajn 'gut', sajd 'Minister', -*d* wortbildend, hier: Pluralis majestatis).



gachaj nocho(j)|*d* 'Schweine und Hunde' (gachaj 'Schwein', nochoj 'Hund', -*d* Plural,

formbildend).



Im Türkischen ist diese Erscheinung unseres Wissens nicht bekannt, wohl aber im Mandschurischen: bei| se 'Beise' (Titel). *-se* ist analog zu mong. *-d* an sajn(n) ein hier wortbildendes Suffix.

Im weiteren betrachten wir wieder die Relationen zwischen den nicht disjunkten Klassen F und G und der Teilmenge B ($B \subset A$, $B = C_1, C_2, \dots, C_k$). Parallel zur Definition der Lexem-Morpheme und der Suffix-Morpheme bezeichnen wir alle Mengen C_1 im Durchschnitt $F \cap B$ als wortbildende Suffix-Morpheme, alle C_j im Durchschnitt $G \cap B$ als formbildende Suffix-Morpheme und alle Mengen C_h im Durchschnitt $(F \cap G) \cap B$ als Suffix-Morpheme, die gleichzeitig wort- und formbildend sind. Beispiel für ein wortbildende Suffix-Morphem: mandsch. $|-la| = -la, -le, -lo, -ra, -re, -ro$ (denominales Verbbildungssuffix). Beispiel für ein formbildendes Suffix-Morphem: mandsch. $|-sa| = -sa, -se$ (Pluralsuffix). Beispiel für ein sowohl wort- als auch formbildendes Suffix-Morphem: mong. $|-nar| = -nar, -ner$ (formbildend als Pluralsuffix, wortbildend an ech 'Mutter': echner 'Gattin').

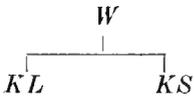
Alle Elemente im Durchschnitt $F \cap B$ heißen wortbildende Suffix-Allomorphe, alle Elemente im Durchschnitt $G \cap B$ formbildende Suffix-Allomorphe und alle Elemente im Durchschnitt $(F \cap G) \cap B$ Suffix-Allomorphe, die gleichzeitig wort- und

formbildend sind. Alle diese Elemente sind Moneme. Als Allomorphe von Suffix-Morphemen werden sie nur dann besonders ausgewiesen, wenn ihre Relation zur Teilmenge *B* zu diskutieren ist.

5. Darstellung der Wortstruktur

Jedes Wort einer altaischen Sprache kann in eine Lexemkonfiguration *KL* und in eine Suffixkonfiguration *KS* zerlegt werden, wobei nur *KS* leer sein darf:

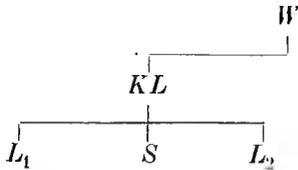
$$W = == \rangle KL, KS$$



Betrachten wir zunächst die Lexemkonfiguration *KL*. Sie beginnt nach einem Zwischenraum und endet mit dem im Wort am weitesten rechts stehenden Lexem *L*. Folgende Strukturen von *KL* sind möglich:

$$KL = == \rangle L \quad KL = == \rangle L_1, L_2 \quad KL = == \rangle L_1, S, L_2 \quad (S: \text{Suffix})$$

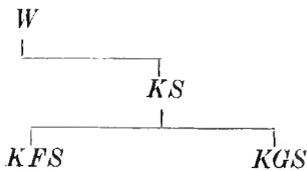
Somit erhalten wir für die potentielle Lexemkonfiguration *KL* und ihre Relation zum Wort *W* folgenden Baum:



Lemma: Da per definitionem *KL* mit dem im Wort am weitesten rechts stehenden Lexem endet, kann ein Suffix in *KL* nur dann auftreten, wenn ein weiteres Lexem noch nach diesem Suffix folgt.

Die Suffixkonfiguration *KS* kann in die Konfiguration der wortbildenden Suffixe *KFS* und in die Konfiguration der formbildenden Suffixe *KGS* zerlegt werden:

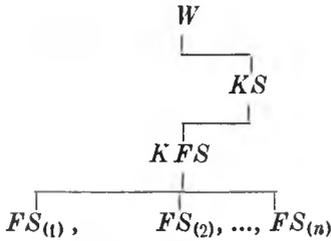
$$KS = == \rangle KFS, KGS.$$



KFS beginnt unmittelbar nach der Lexemkonfiguration *KL* und endet mit dem im Wort am weitesten rechts stehenden wortbildenden Suffix.

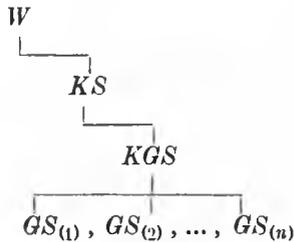
KFS kann *n* wortbildende Suffixe enthalten, die nach den Indices der Positionsklassen, zu denen sie gehören, geordnet werden:

$$KFS = == \rangle FS_{(1)}, FS_{(2)}, \dots, FS_{(n)} \quad (\text{Bsp. s. Abschnitt 4}).$$



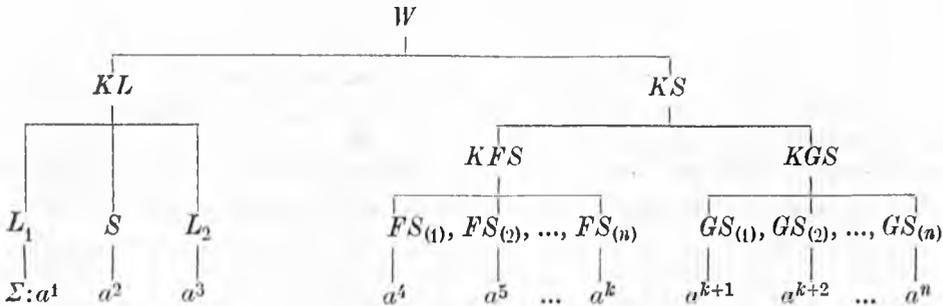
Die Konfiguration der formbildenden Suffixe *KGS* beginnt unmittelbar nach *KFS* oder nach der Lexemkonfiguration *KL*, wenn im konkreten Fall *KFS* leer ist. *KGS* enthält n formbildende Suffixe, die wieder ihrem positionellen Verhalten gemäß geordnet sind:

$KGS = = \Rightarrow GS_{(1)}, GS_{(2)}, \dots, GS_{(n)}$ (Bsp. s. Abschnitt 4).



Diejenigen Suffixe, die sowohl wort- als auch formbildend sein können, verhalten sich neutral, d. h. sie können sowohl den Platz von wortbildenden Suffixen in *KFS* als auch den Platz von formbildenden Suffixen in *KGS* einnehmen.

Das formale Gesamtmodell für das morphologische Grundinventar des Wortes und sein positionelles Verhalten stellt sich wie folgt dar:



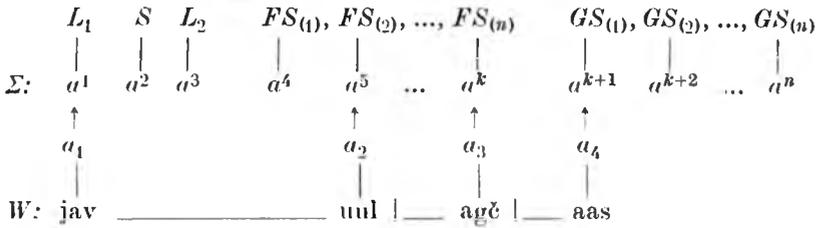
Jedes Wort einer altaischen Sprache kann eindeutig in die Endkette Σ abgebildet werden, wobei im Spezialfall, d. h. bei einer altaischen Sprache, die uns nicht bekannt ist, eine Modifizierung des Modells erfolgen kann. Eine eindeutige Abbildung kommt in praxi nicht vor. Die Urbilder der Abbildung sind die Moneme a_1, a_2, \dots, a_n eines beliebigen Wortes einer altaischen Sprache, ihre Bilder sind die Elemente a^1, a^2, \dots, a^n der Endkette Σ .

Für die Abbildung gelten die folgenden Regeln (1) und (2):

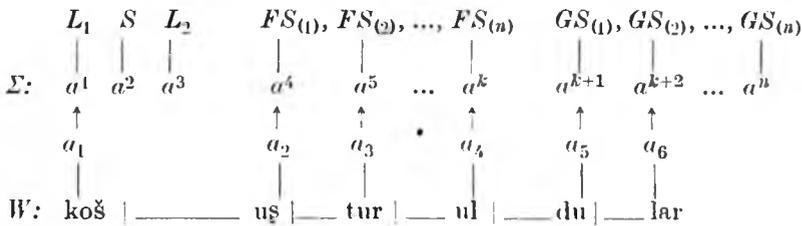
- (1) $a_1 \rightarrow a^1$ (2) $a_i (1 \leq i \leq n) \rightarrow a^j (j \geq i, 1 < j \leq n)$

d. h. jedes erste Monem $a_1 (a_1 \in W)$, das per definitionem ein Lexem sein muß, ist

a^1 ($a^1 \in \Sigma$) zuzuordnen. Jedes folgende Monem a_i ($i > 1$) wird einem Element a^j aus Σ zugeordnet, wobei a^j denselben oder einen höheren Index hat als a_i , z. B.:

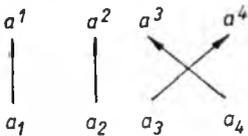


‘vom Absender’ (mong.), — : Überbrückung der grafisch bedingten Zwischenräume.



Sie haben herumrennen lassen’ (türk.)

Demgegenüber ist eine Abbildung wie z. B.



nicht zugelassen.

Damit haben wir die Grundzüge eines Modells erarbeitet, das das morphologische Grundinventar des Wortes formal und positionell im wesentlichen erfaßt und zumindest für die wichtigsten altaischen Sprachen gültig ist. Für andere altaische Sprachen, die hier nicht berücksichtigt werden konnten, kann es entsprechend modifiziert werden. Die aufwendig erscheinende Strukturierung des Wortes ist durch die ausgeprägte Agglutination in den altaischen Sprachen gerechtfertigt.

Literatur

VIETZE, H.-P., *Eine Definition des mongolischen Wortes*, in: ZPSK, Bd. 22 (1969), Heft 1, S. 77–82.
 —, *Das Problem der Semantikstrukturierung bei einer automatischen monematischen Analyse*, in: ZPSK, Bd. 23 (1971), Heft 6, S. 590–615.
 SWIFT, L. B., *A Reference Grammar of Modern Turkish*, Indiana University Publications, Uralic and Altaic Series, Vol. 19, Den Haag 1963, S. 114.