

Strukturalismus: Bloomfield, Harris, Chomsky

Im 19. Jahrhundert untersuchten die Sprachwissenschaftler hauptsächlich den Sprachwandel (*diachrone* Sprachwissenschaft).

Im 20. Jahrhunderts herrschte die *synchrone* Sprachwissenschaft vor: die Untersuchung einer (in engem Zeitraum) gebräuchlichen Sprache als ein System von Beziehungen zwischen sprachlichen Elementen, der *Strukturalismus*.

Ansatzpunkt der Strukturalisten (de Saussure, Bloomfield, Harris, Chomsky):

- ▶ Sprachliche Elemente (Wörter, Phoneme usw.) haben einen Sinn nur innerhalb eines Systems, durch Äquivalenzen und Kontraste untereinander.
de Saussure: wie die Figuren im Schach
- ▶ Das System wird durch Kombinationen der Elemente und Beschränkungen der Kombinationsmöglichkeiten beherrscht.

Sprachliche Ausdrücke werden i.a. aus kleineren Einheiten zusammengesetzt. Wenn man solche Einheiten identifiziert hat, kann man untersuchen,

1. welche ihrer Kombinationen werden als größere Einheiten erkannt/akzeptiert,
2. welche Kombinationsoperationen unterliegen welchen Einschränkungen?

Einschränkungen an die Kombination von Sätzen zu größeren Einheiten sind relativ schwach ausgeprägt, deshalb sind diese Fragen eher für Einheiten unterhalb der Satzebene interessant, z.B.

- ▶ Wörter innerhalb von Syntagmen
- ▶ Syntagmen innerhalb von Sätzen
- ▶ Phoneme innerhalb von Wörtern

Harris:

- ▶ Mangel an expliziten Methoden in der Syntax führt zu
 - ▶ intuitivem Vorgehen,
 - ▶ ungewisser Relevanz von Beobachtungen,
 - ▶ unklaren Beziehungen zwischen Beobachtungen
- ▶ Untersuchung von Distributionsverhältnissen
 - ▶ mit expliziten Methoden möglich (Substitution)
 - ▶ nicht, um daraus Aussagen über die Bedeutungen zu erhalten,
 - ▶ weil sie unabhängig von Bedeutungen untersuchbar sind.

Endliche Folgen kleinster Einheiten

Da sprachliche Äußerungen φ in der Zeit erfolgen und i.a. nicht atomar sind, lassen sie sich in endliche *Folgen* w kleinster Einheiten a zerlegen, die nacheinander geäußert werden.

Die kleinsten Einheiten sollten wenige sein, und von einander unterscheidbar. Nehmen wir an, eine endliche Menge $\Sigma = \{a, b, \dots\}$ solcher kleinsten Einheiten sei bestimmt, z.B. die Phoneme.

Sei Σ^* die Menge *aller* endlichen Folgen von Elementen aus Σ . Den *sprachlichen* Äußerungen entspricht eine Teilmenge $L \subseteq \Sigma^*$.

- ▶ Es gibt unter den $w \in L$ gewisse Regularitäten, z.B. treten manche $a, b \in \Sigma$ nicht am Ende eines w auf, manche nicht direkt nebeneinander, manche nur gemeinsam, usw.
- ▶ Da L sehr groß oder unendlich ist, ist es schwierig, die Regularitäten zu kennen.

Distribution von $w \in \Sigma^*$ in L

Um die Regularitäten aufzufinden, versucht man, Folgen $w \in \Sigma^*$ zu klassifizieren und Regularitäten auf der Ebene der Klassen zu finden.

Definition

Die *Distribution* $D_L(w)$ von $w \in \Sigma^*$ in L ist die Menge der *Kontexte* $u_v := (u, v)$, in denen w in L vorkommt:

$$D(w) := \{(u, v) \mid u, v \in \Sigma^*, uwv \in L\}.$$

Zwei Folgen $w_1, w_2 \in \Sigma^*$ sind *distributionsäquivalent bzgl. L* , kurz: $w_1 \equiv_L w_2$, wenn sie dieselbe Distribution haben, $D(w_1) = D(w_2)$.

Zwei Folgen $w_1, w_2 \in \Sigma^*$ sind *von komplementärer Distribution*, wenn sie keine Kontexte gemeinsam haben: $D(w_1) \cap D(w_2) = \emptyset$.

Seien Σ die Wortformen und L die Sätze des Deutschen.

- ▶ *Emil* und *Willi* sind äquivalent.
- ▶ *mich* und *dich* sind nicht äquivalent: *Ich schäme dich.* $\notin L$
- ▶ *wir* und *sprecht* sind komplementär.
- ▶ $D(\text{der Hund}) \subseteq D(\text{er})$. Aber: *er ißt* $\in L$, *der Hund ißt* $\in L$?

Beachte: die Begriffe sind nur bedingt operationalisierbar

- ▶ ist L endlich, so kann man $w_1 \equiv w_2$ feststellen, da man nur die endlich vielen u_v berücksichtigen muss, die man aus den $urv \in L$ bekommt. z.B. $\Sigma =$ Phoneme, $L =$ einf.Wortformen
- ▶ wenn L unendlich ist, müßte man unendlich viele Kontexte u_v durchlaufen, was man nur im Geiste kann. z.B. $L =$ Sätze
- ▶ die Definition ist relativ dazu, was $\in L$ bedeutet: manchmal ist die *Akzeptanz* von uvw gemeint, da man L nur indirekt kennt.

Besseres Verfahren zur Bestimmung der Morpheme einer Sprache:

1. Teile jeden Ausdruck in die kleinsten Phonemfolgen mit gleicher Bedeutung in *Morphemalternativen*: /knaif/ \neq /knaiv/.
2. Bilde *Morphemeinheiten*, d.h. Mengen der Morphemalternativen gleicher Bedeutung und komplementärer Distribution. Berücksichtige eine mehrelementige Einheit E nur, wenn

$$\bigcup \{D(w) \mid w \in E\} = D(e) \text{ für eine Einheit } E' = \{e\},$$

z.B. $E = \{am, are\}$ wegen $D(am) \cup D(are) = D(walk)$.

3. Unterscheiden sich die Alternativen zweier Einheiten E_1 und E_2 auf dieselbe Weise, so repräsentiere die Einheiten durch eines der Elemente und den Unterschied, z.B. $\{knife, kniv-\}$ und $\{wife, wiv-\}$ durch (/naif/, /f/ + /-z/ \mapsto /v/)

Ermittlung der Unterschiede zwischen den Alternativen:

1. Worin unterscheiden sich die Alternativen einer Einheit?
2. In welchen Kontexten kommt eine Alternative vor?
3. Welche Ähnlichkeit besteht zwischen Alternative und Kontext?
4. Welche Einheiten haben diese Unterschiede zwischen den Alternativen?

Z.B. die Kontexte B_C , in denen eine Alternative $a \in A$ vorkommt, sind selber Einheiten, die in ihre Alternativen $b \in B$, $c \in C$ zerfallen, und nur in manchen b_c tritt a auf.

Klassifizierung der Kontexte nach ähnlicher Bildung von Alternativen.

Substitutionsklassen

Die zu einem $w \in \Sigma^*$ bzgl. L distributionsäquivalenten w' bilden die *Distributions-* oder *Substitutionsklasse* von w ,

$$S(w) := \{w' \in \Sigma^* \mid w' \equiv_L w\} =: [w]_{\equiv_L}.$$

Die Elemente einer Substitutionsklasse kann man in beliebigen L -Kontexten durch einander ersetzen, ohne aus L herauszukommen:

$$\begin{aligned} w' \in S(w) &\iff w \equiv w' \iff D(w) = D(w'). \\ &\iff \forall u \in \Sigma^* \forall v \in \Sigma^* (uwv \in L \iff uw'v \in L). \end{aligned}$$

Wortarten als Distributionsklassen?

Man kann nicht erwarten, daß die Wortarten Substitutionsklassen sind: verschiedene Formen desselben Worts können oft gerade *nicht* in denselben Kontexten vorkommen, z.B. *Baum* und *Bäumen*.

Manchmal bilden alle Wörter einer Wortart (bzw. Unterart) *in derselben Form* eine Substitutionsklasse: z.B.

$$N_{neut}^{gen,sg} = \{Hauses, Kindes, \dots\}$$

Aber i.a. ist das nicht so: z.B. ist

$$N_{mask}^{nom,sg} = \{Baum, Hase, Hund, \dots\}$$

keine Substitutionsklasse (bzgl. L = deutsche Sätze), da

Ich sehe den Baum $\in L$, *Ich sehe den Hase* $\notin L$

Die Distribution der konkreten Formen kann allein nicht darüber entscheiden, was ein Satz ist: es kommt u.a. auf die abstrakten Formen an (z.B. auf den an der Wortform *Baum* nicht ablesbaren Kasus).

Erstmal braucht man die abstrakten Formen, die Wörter und die Wortarten. (Vgl. Hausaufgaben)

- ▶ Erhält man *Wörter* durch Abstraktion von ihren Formen? Die Distribution liefert etwa $D(\text{Baum}) = D(\text{Hund})$, unterscheidet aber nicht zwischen Nominativ- und Akkusativkontexten. Man muß also erstmal die Distributionen überladener Formen nach den abstrakten Formen trennen und ggf. diese geeignet zusammensetzen.
- ▶ Gibt es für abstrakte Formen, Wörter oder Wortarten *charakteristische Kontexte*? Z.B. Kontexte, in denen genau die Nomen im Nominativ, oder genau die Verben auftreten?

Literatur



L. Bloomfield.

Language.

Holt, 1933.



Z. Harris.

A Theory of Language and Information.

Clarendon Press, 1991.



Z. S. Harris.

Papers on Syntax.

D. Reidel, 1981.



J. Lyons.

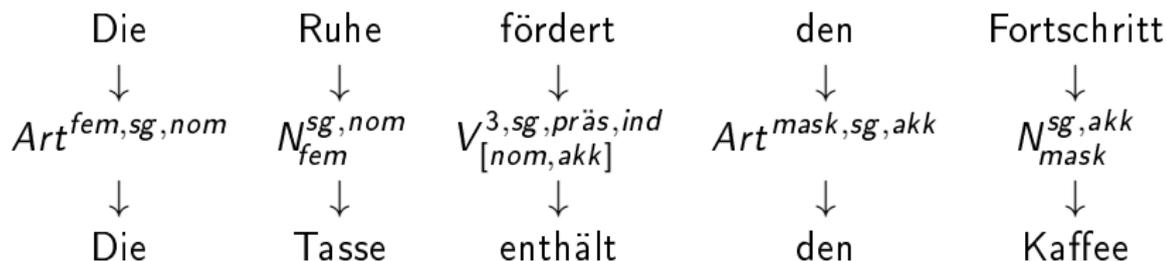
Introduction to Theoretical Linguistics.

Cambridge University Press, 1968.

Dt.: Einführung in die moderne Linguistik. C.H.Beck, 1989.

Satzschemata

Teilt man die Wortarten hinreichend fein in Unterarten (z.B. $V_{[nom,akk]}$: transitive Verben), so lassen sich in einem Satz Wörter derselben Unterart durch andere *derselben Form* ersetzen:



Aus jedem Satz erhält man ein *Satzschema*, eine Folge der Konkretisierungen abstrakter Formen. Aber:

Solche Schemata beschreiben die Regelmäßigkeiten im Satz zu eng. z.B. erlauben sie nicht, *Die Tasse* durch *der Becher* zu ersetzen.

Verschiedene konkrete Formen „eines“ Worts unterscheiden sich i.a. morphologisch höchstens durch

- ▶ durch kurze Affixe (am Anfang, im Innern oder am Ende),
- ▶ durch Lautveränderungen (Umlaute, Konsonantenwechsel),

so daß wir mit Endungstabellen, Umlautregeln u.ä. hoffentlich feststellen können, ob zwei Wortformen zum „selben“ Wort gehören.

Verschiedene Formen eines Worts sind i.a. komplementär:

Die Tasse enthält den Kaffee
* *Die Tassen enthält den Kaffee*

In diesem Fall müssen Wortformen des Kontexts „angepaßt“ werden, um wieder einen Satz zu haben: *enthält* \mapsto *enthalten*.

Bemerkung:

Wo verschiedene Formen desselben Worts nicht komplementär sind, spielen z.B. Bedeutungsänderungen mit:

<i>Die</i>	<i>Tasse</i>	<i>enthält</i>	<i>den</i>	<i>Kaffee</i>	<i>Gegenwart</i>
<i>Die</i>	<i>Tasse</i>	<i>enth<u>i</u>elt</i>	<i>den</i>	<i>Kaffee</i>	<i>Vergangenheit</i>
	<i>Wir</i>	<i>trinken</i>	<i>den</i>	<i>Kaffee</i>	<i>über Sprecher u.a.</i>
	<i><u>S</u>ie</i>	<i>trinken</i>	<i>den</i>	<i>Kaffee</i>	<i>nur über Andere</i>

Dann trenne man die Bezugssprache L in Teilsprachen $L_1 \dot{\cup} L_2$ und berücksichtige nur die Kontexte der Aussagen aus *einem* der L_i , z.B. die Distribution bzgl. $L_i = L^{präs}$ betrachten.

Wenn der Bezugsbereich L klein genug ist, sollten unterschiedliche Formen desselben Worts komplementär sein.

„Überladene“ Formen wie $Baum \in N_{mask}^{sg,nom} \cap N_{mask}^{sg,akk}$ lassen sich daran erkennen, daß *andere* Wörter in entsprechenden Kontexten verschiedene Formen verwenden:

Der	Baum	ist	alt		Der	Bär	ist	alt
Den	Baum	sehe	ich		Den	Bären	sehe	ich
Der	Bäume	gedenke	ich		Der	Bären	gedenke	ich

Wir können dann *abstrakte* Formen von Wort(unter)arten durch Kontexte aus konkreten Wortformen definieren:

Der	$N_{mask}^{sg,nom}$	ist	alt
Den	$N_{mask}^{sg,akk}$	sehe	ich
Der	$N_{mask}^{pl,gen}$	gedenke	ich

Dann bestimmt man das Paradigma einzelner Wörter wie *Baum*.

Man erhalte so zuerst nur verschiedene *Flexionsklassen*, und muß anschließend daraus Wort(unter)arten durch Vereinigung bilden:

$$N_{mask} = N_{mask, Flexionsklasse\ 1} \cup \dots \cup N_{mask, Flexionsklasse\ k} \quad (1)$$

Nehmen wir also an:

- ▶ man kann (über Kontexte konkreter Wortformen) Wortarten A mit zugehörigen abstrakten Formbereichen $Form_A$ festlegen,
- ▶ für jedes A und $F \in Form_A$ sei $A^F := \{W^F \mid W \in A\}$ die Menge der konkreten F -Formen W^F der Wörter W aus A .
- ▶ in den A^F definierenden Kontexten führen die W^F zu Sätzen.

Notation: **Formmerkmale** werden (hier) im oberen Index angegeben, **Artmerkmale** für Angabe von Unterarten durch den unteren Index:

$$N_{mask}^{sg, dat} := \text{maskuline Nomina im Singular und Dativ} \quad (2)$$

Wortarten und Flexionsdimensionen im Deutschen

Als *Flexionsarten* sind im Deutschen zu unterscheiden:

1. die Konjugation bei Verben,
2. die Deklination bei Artikeln, Nomina, Pronomina, Adjektiven,
3. die Komparation bei Adjektiven.

Die Gesamtheit der (abstrakten) Formen einer Flexionsart stellt man in mehrdimensionalen (Form-) Merkmalen dar. Wir verwenden für das Deutsche folgende *Merkmalsdimensionen*:

- ▶ bei Nomina: Numerus, Kasus,
- ▶ bei Pronomen: Person, Genus, Numerus, Kasus
- ▶ bei Artikeln: Genus, Numerus, Kasus,
- ▶ bei Adjektiven: Genus, Numerus, Kasus, Deklinationstyp, Komparation
- ▶ bei Verben: Numerus, Person, Tempus, Modus, Infinitheit

Flexionsdimensionen im Deutschen

Jede solche Merkmalsdimension der Wortarten hat einen endlichen Bereich von Werten, die das Merkmal annehmen kann:

Dimension	Werte	Kürzel
Kasus	Nominativ, Genitiv, Dativ, Akkusativ	nom, gen, dat, akk
Numerus	Singular, Plural	sg, pl
Person	erste, zweite, dritte	1, 2, 3
Genus	Femininum, Maskulinum, Neutrum	fem, mask, neut
Komparation	Positiv, Komparativ, Superlativ	pos, komp, sup
Deklin.typ	starke, schwache, gemischte Deklin.	st, sw, ge
Tempus	Präsens, Präteritum	präs, prät
Infinitheit	Infinitiv, Partizip I, Partizip II	inf, part1, part2
Modus	Indikativ, Konjunktiv, Imperativ	ind, konj, imp

Bem. Für Ausdrücke gibt es i.a. andere Dimensionen oder Werte, z.B. bei Sätzen: Diathese als Dimension, Perfekt als Tempuswert.

Satzschemata aus Mengen konkreter Wortformen?

Geht man von einem konkreten Satz

$$s = w_1^{F_1} \dots w_n^{F_n}$$

aus und ersetzt die konkreten Wortformen $w_i^{F_i}$ durch zu ihnen „passende“ Wortarten und abstrakte Formen $A_i^{F_i} \ni w_i^{F_i}$, so erhält man *potentiell* Satz­schemata $A_1^{F_1} \dots A_n^{F_n}$:

S	$Art^{fem,sg,nom}$	$N_{fem}^{sg,nom}$	$V_{[nom,akk]}^{3,sg,präs,ind}$	$Art^{mask,sg,akk}$	$N_{mask}^{sg,akk}$
	↑	↑	↑	↑	↑
	Die	Ruhe	fördert	den	Schlaf
	↓	↓	↓	↓	↓
*S	$Art^{mask,pl,nom}$	$N_{fem}^{sg,akk}$	$V_{[nom,akk]}^{3,sg,präs,ind}$	$Art^{fem,pl,dat}$	$N_{mask}^{sg,akk}$

Das zweite sollte kein Satzschema sein, obwohl es nur Sätze ergibt!

Daß alle Einsetzungen konkrete Sätze sind, ist nur ein notwendiges Kriterium:

*S	<i>Art</i> ^{mask,pl,nom}	$N_{fem}^{sg,akk}$	$V^{3,sg,präs,ind}$ [nom,akk]	<i>Art</i> ^{fem,pl,dat}	$N_{mask}^{sg,akk}$
S	↓ Die	↓ Tasse	↓ enthält	↓ den	↓ Kaffee

Das Schema besagt mehr, als sich in der Menge der Einsetzungen ausdrückt: eine bestimmte (Wortart,Form)-*Analyse* sei ein Satz.

Das Problem kommt daher, daß bei manchen Flexionsklassen die Formüberladung systematisch ist, z.B. $N_{fem}^{sg,nom} = N_{fem}^{sg,akk}$.

Also: Eine Folge $A_1^{F_1} \dots A_n^{F_n}$ von Wortarten A und abstrakten Formen $F_i \in \text{Form}_{A_i}$ ist höchstens dann ein *Satzschema* S , wenn beliebige Einsetzungen

$$w_1^{F_1} \dots w_n^{F_n}$$

der konkreten Formen $w_i^{F_i}$ von Wörtern w_i der Wortart A_i konkrete Sätze sind. Für diese Bedingung schreiben wir:

$$S \supseteq A_1^{F_1} \dots A_n^{F_n} \quad \text{oder} \quad S \rightarrow A_1^{F_1} \dots A_n^{F_n}.$$

Beachte: die A^F sind wie Mengenvariablen

- ▶ die Wahlen der $w_i \in A_i$ erfolgen unabhängig von einander
- ▶ kommt A^F mehrfach vor, so kann für jedes Vorkommen ein anderes w^F aus A^F gewählt werden.
- ▶ $A_i^{F_i}$ ist obligatorisch (muß konkretisiert werden) gdw. $\epsilon \notin A_i^{F_i}$.