Strukturalismus: Bloomfield, Harris, Chomsky

Im 19. Jahrhundert untersuchten die Sprachwissenschaftler hauptsächlich den Sprach*wandel* (*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 $\Sigma = \{a, b, \ldots\}$ 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.
- $D(\textit{der Hund}) \subseteq D(\textit{er})$. 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 *uwv* 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.

Substituionsklassen

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:

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

Wortarten als Distributionsklassen?

Man kann nicht erwarten, daß die Wortarten Substitutionsklassen sind: verschiendene 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.

$$N^{gen,sg}_{neut} = \{Hauses, Kindes, \ldots\}$$

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

$$N^{nom,sg}_{mask} = \{Baum, Hase, Hund, \ldots\}$$

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

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

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 $D(Baum) = D(Hund)$, 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?

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:

Die	Ruhe	fördert	den	Fortschritt
↓	↓	↓	↓	↓
$Art^{fem,sg,nom}$	$N^{sg,nom}_{fem}$	$V^{3,sg,präs,ind}_{[nom,akk]}$	$Art^{mask,sg,akk}$	$N^{sg,akk}_{mask}$
↓	↓	↓	↓	↓
Die	Tasse	enthält	den	Kaffee

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. erlaubt es nicht, *Die Tasse* durch *der Becher* zu ersetzen.

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.
Cambride University Press, 1968.
Dt.: Einführung in die moderne Linguistik. C.H.Beck, 1989.