



## Zirkuläre Korpusanalysen – Problemfälle aus der Referenzforschung

GUG-Jahrestagung und internationale Konferenz *Kognition und Interkulturalität*, Debrecen, 30.5.2009

### 1. Komplexanapher oder Anapher mit NP-Antezedent?<sup>1</sup>

#### 1.1 Begriffserklärung

Anaphern nehmen gewöhnlich einen Referenten wieder auf, der mit einem **NP-Antezedenten** eingeführt wurde (Consten/Schwarz-Friesel 2008):

(1) *Hier ist letzte Woche **ein Unfall** passiert. Meine Kollegin hat ihn / den Crash gesehen.*

Komplexanaphern etablieren dagegen Referenten wie Ereignisse, Prozesse und Zustände, die mit **satzwertigen Strukturen** eingeführt wurden (Consten/Knees/Schwarz-Friesel 2007, 2009):

(2) *Hier passieren immer wieder Unfälle. Dies / Diese Tatsache zeigt, dass die Stadt hier dringend etwas unternehmen müsste.*

(3) *Hier ist letzte Woche ein Unfall passiert. Meine Kollegin hat das gesehen.*

#### 1.2 Kodierprobleme in einem Zeitungskorpus

(4) *Etwa 25 Kilogramm Natur-Uralkonzentrat sind bei **einem Unfall** im thüringischen Wismut-Zwischenlager Seelingstädt freigesetzt worden. Der Zwischenfall ereignete sich bereits am 1. Juli [...] (TiGerKorpus<sup>2</sup> 2275f)*

Die Anapher *der Zwischenfall* ist hier sowohl als Komplexanapher mit dem ganzen vorhergehenden Satz als Antezedent analysierbar, als auch als einfache Anapher mit dem Antezedent *einem Unfall*.

(5) ***Israelischer Angriff im Gaza-Streifen** fordert mindestens 13 Tote, darunter viele Zivilisten. General spricht von „unabdingbarer Operation“ gegen Hamas, für palästinensischen Unterhändler zielt dieses „Massaker“ auf die Friedensmission von Solana. („Massaker“ im „Terrornest“, taz, 08.10.2002)*

*Dieses „Massaker“* ist analysierbar als Komplexanapher mit dem gekennzeichneten Satz als Antezedenten – wofür die grammatische Form der demonstrativen NP spricht – aber auch als Anapher mit der NP *Israelischer Angriff im Gaza-Streifen*. Auch hier ist der referenzielle Sachverhalt, auf den die Anapher Bezug nimmt, derselbe; allenfalls ist festzustellen, dass *dieses „Massaker“* als Komplexanapher eher als die elliptische Prädikation *„unabdingbarer Operation“* das Ergebnis des Angriffs und nicht den Angriff als solchen fokussiert.

<sup>1</sup> Das Projekt „KomplexTex – zur textuellen Funktion von Komplexanaphern“ wurde von 2003 bis 2008 von der Deutschen Forschungsgemeinschaft finanziert.

<sup>2</sup> Das TiGer-Korpus besteht aus ca. 50.000 Sätzen/ 900.000 Tokens aus der Tageszeitung *Frankfurter Rundschau* und ist morphosyntaktisch kodiert.

Die Mehrdeutigkeit liegt ausschließlich in der linguistischen Strukturanalyse; referenziell bestehen keine unterschiedlichen Lesarten. Experimentdaten z.B. aus Lesartentests wären daher nicht vereindeutigend.

### 1.3 Heuristik als Lösung

Wenn eine Lesart als Komplexanapher plausibel ist, wird Beleg als Komplexanapher kodiert. Zusätzlich wird ein Merkmal „analytisch ambig“ (versus „referenziell ambig“) kodiert.

## 2. Neueinführung, indirekte Anapher oder unikale Referenz?

### 2.1 Kodierprobleme in einem Spontansprachkorpus

Loll (2007) und Consten/Loll (2009/im Druck) beschäftigen sich mit Artikelsetzung in einem Längsschnittkorpus zum Zweitspracherwerb des Deutschen.<sup>3</sup>

Während überprüft werden soll, inwieweit der definite Artikel zielsprachlich zur Kennzeichnung identifizierbarer Referenten verwendet wird, zeigt sich, dass bei sog. indirekten Anaphern (Schwarz 2000, Schwarz-Friesel 2007) eine Charakterisierung der Referenteneigenschaften nur in zirkulärer Abhängigkeit vom sprachlichen Ausdruck möglich ist.

#### (3) @Bg: nacherzählung aus der erinnerung

\*NAS: also es ist über ein mädchen, **das [/] das war plötzlich irgendwie krank** .  
%mor: PP|über ART:INDEF|ein N|mädchen:EINF:NIREF .

\*NAS: **hatte irgendwas irgendwelches [\*] blutvergiftung oder so** .

\*NAS: und da gab es keine mitteln [\*] [//] mittel noch dagegen, weil eh@fp das war im jahr achtundfünfzig .

\*NAS: <und &de> [//] und **der arzt**, der sagten [\*] [//] sagte .

%mor:ART:DEF|der N|arzt:WAUF:IREF . (Loll 2007: 20)

#### 2.1.1 Indirekte Anapher oder Neueinführung?

*Der Arzt* ist als indirekte Anapher (**WiederAUFnahme**) konzeptuell verankert durch Vorerwähnung eines Krankheitsfalles interpretierbar. Dies rechtfertigt, die Setzung des definiten Artikels als zielsprachlich zu bewerten. Andererseits wäre ebenso als zielsprachlich interpretierbar (vgl. auch *einem Elektriker* in (4)):

(3') \*NAS: <und &de> [//] und **ein arzt**, der sagten [\*] [//] sagte .

%mor:ART:INDEF|ein N|arzt:EINF:NIREF . (konstruiert)

<sup>3</sup> Das Korpus wurde erstellt im DFG-Projekt „Der Altersfaktor im Erwerb des Deutschen als Zweitsprache“ (DaZ-AF), geleitet von Ursula Stephany, Köln, und Christine Dimroth, Nijmegen, 2000-2002. Hier zitiert: Spontansprach-Interviews mit Nastja, 9 Jahre 10 Mon. alt, Muttersprache Russisch, seit 1 Jahr 3 Mon. in Deutschland.

Die hier auszugsweise zitierte Kodierung ist von Loll (2007). Die hier relevanten Kodiermerkmale sind: IREF = Referent identifizierbar / NIREF = Referent nicht identifizierbar; EINF = Neueinführung des Referenten in das Textweltmodell / WAUF = anaphorische Wiederaufnahme des Referenten, einschließlich indirekter Anaphorik. Transkriptionszeichen (CHILDES-Standard): [\*] deviant, eh@fp gefüllte Pause, [/] Wiederholung, [//] Selbstreparatur. Vgl. Stephany/ Bast/ Lehmann 2001.

Trotz der daraus resultierenden anderen Kodierung von Referenteneigenschaften evoziert (3') keine andere Lesart.

„In einer recht großen Grenzzone zwischen indirekter Anaphorik und Neueinführung von Referenten muss die Artikelwahl somit als eine Art stilistischer Variation angesehen werden, die sich einer präzisen Erklärung durch die hier untersuchten Variablen entzieht.“ (Loll 2007: 21)

Forts. von (3): Das kranke Mädchen ist durchgehender Diskursreferent und somit guter Anker für die folgenden (potenziellen) indirekten Anaphern.

- (4) \*NAS: und danach <haben sie> [//] hat [=? hatte] der [/] <der vater eine idee> [/] der vater eine idee .  
%mor:ART:DEF|der N|vater:WAUF:IREF .
- \*NAS: der hat ihr gesagt, <das der> [//] das weihnachten eh@fp also früher ist .
- \*NAS: dieses jahr <als &and> [/] <als eh@fp andere> [//] als in anderen jahren .
- \*NAS: also nicht am vierundzwanzigsten, sondern am [/] am zweiten .
- \*NAS: und dann hat der dem lehrer gesagt, das die sollten in der schule schon mal basteln anfangen .  
%mor:ART:DEF|der N|lehrer:EINF:IREF .
- \*NAS: und &de dem bäcker hat er gesagt, also der bäcker hat versprochen .  
%mor:ART:DEF|der N|bäcker:EINF:IREF ART:DEF|der:N|bäcker:WAUF:IREF .
- \*NAS: sein schokoladenschiff früher auf (da)s fenster zu stellen .  
%mor:PP:DEF|auf N|fenster:WAUF:IREF .
- \*NAS: <und so> [/] und so hat er das, und das ganze dorf hatte [?] sowas nur für diese Janine .  
[...]
- \*NAS: sie [/] eh@fp sie hatten eine [//] einem elektriker gesagt .  
%mor:ART:INDEF|ein N|elektriker:EINF:NIREF. (Loll 2007: 18)

### 2.1.2 Interaktion von Unikalität und indirekter Anaphorik?

Die Definitheit von *dem Lehrer* und *dem Bäcker* erklärt Loll (2007: 18) mit Unikalität der Referenten innerhalb der Referenzdomäne. Daraus resultiert Kodierung als „EINFührung“ eines neuen, aber „identifizierbaren“ (weil unikal) Referenten.

Zumindest *der Lehrer* könnte aber auch indirekte Anapher mit dem Anker *ein Mädchen* (3) sein, oder Definitheit ist durch Kombination von anaphorischer Verankerung und Unikalität erklärbar. Dies könnte auch gelten für *das Fenster*. Kodiert als indirekte Anapher WAUF (*Bäcker* → Bäckerei → (Schau-)Fenster der Bäckerei), aber ohne Annahme von Unikalität wäre Indefinitheit *ein(s von mehreren) Fenstern* zielsprachlich. Dies lässt sich mit den klar abgrenzbaren, dichotomischen Kodierungsmerkmalen, die zur statistischen Analyse erforderlich sind, nicht ausdrücken.

## 2.2 Heuristik als Lösung

Wenn die Artikelsetzung im Korpus sinnvoll interpretierbar ist, wird sie als zielsprachlich gewertet. „Es kann jedoch nicht ausgeschlossen werden, daß die Sprecherin eigentlich etwas anderes intendierte und der Artikel nicht zielsprachlich gewählt wurde“ (Loll 2007: 19). Das Problem in 2.1.2 lässt sich hierdurch nicht lösen. Da beide Fälle aber den definiten Artikel fordern, ist dies nicht relevant für die Analyse in Loll (2007).

## 3. Fazit:

- Durch Heuristiken können inkonsistente Kodierungen in Korpora – gerade bei mehreren Kodierern – vermieden werden.
- Es muss aber transparent gemacht werden, dass Kodierungen nicht immer in einem 1:1-Verhältnis zur Realität der Sprachrezeption stehen, sondern Artefakte der linguistischen Analyse sein können – so ist eine Ambiguität bezüglich der Kodierung nicht unbedingt eine Ambiguität für den natürlichen Hörer/Leser.
- Kodier-Heuristiken regeln Probleme nur innerhalb des Erkenntnisinteresses der jeweiligen Studie (vgl. 2.2), sie bilden daher die Möglichkeiten theoretischer Kategorisierungen nicht immer ab.

## 4. Literatur

- Consten, Manfred/ Knees, Mareile/ Schwarz-Friesel, Monika, 2007. The Function of Complex Anaphors in Texts. In: Schwarz-Friesel, Monika / Consten, Manfred / Knees, Mareile (Hg.), *Anaphors in Texts*. Amsterdam: Benjamins. 81-102.
- Consten, Manfred/ Knees, Mareile/Schwarz-Friesel, Monika, 2009. Complex Anaphors. The Impact of Ontology, Cotext and Conceptual Knowledge. In: Zlatev, J., Johansson Falck, M., Lundmark, C., and Andrén, M (Hg.), *Studies in Language and Cognition*. Cambridge Scholars Publishing, Newcastle. 285-302.
- Consten, Manfred/ Loll, Annegret, 2009 (im Druck). Indirekte Anaphern – ein Zirkularitätsproblem zwischen Grammatik und Pragmatik. In: Péteri, Attila (Hg.), *An der Grenze zwischen Grammatik und Pragmatik*. Frankfurt/M: Peter Lang, 323-330.
- Consten, Manfred./ Schwarz-Friesel, M., 2008. Anapher. In: Hoffmann, Ludger (Hg.), *Deutsche Wortarten*. Berlin / New York: de Gruyter. 265-292.
- Loll, Annegret, 2007. *Determinierer im Erwerb des Deutschen als Zweitsprache - eine Fallstudie*. (Arbeitspapier N.F. 52). Köln: Institut für Linguistik.  
[www.uni-koeln.de/phil-fak/ifl/asw/forschung/forschung\\_frames\\_d.html](http://www.uni-koeln.de/phil-fak/ifl/asw/forschung/forschung_frames_d.html)
- Schwarz, Monika, 2000. *Indirekte Anaphern in Texten*. Tübingen: Niemeyer (LA 413).
- Schwarz-Friesel, Monika, 2007. Indirect anaphora in text: a cognitive account. In: Schwarz-Friesel, Monika/ Consten, Manfred/ Knees, Mareile (Hg.), *Anaphors in Text*. Amsterdam: Benjamins, 3-20.
- Stephany, Ursula/ Bast, Conny/ Lehmann, Katrin, 2001. *Computer-Assisted Transcription and Analysis of Speech*. (Arbeitspapier N.F. 41). Köln: Institut für Linguistik.  
[www.uni-koeln.de/phil-fak/ifl/asw/forschung/ap/childes.pdf](http://www.uni-koeln.de/phil-fak/ifl/asw/forschung/ap/childes.pdf)