



So Typen wie du
Eine syntaktische Fundierung
von so als Artikel

Hagen Hirschmann

Forschungskolloquium Korpuslinguistik
14.05.2014

Plan



- Einführung: Empirische Daten zu so+N
- Der syntaktische Status von so+N
(→ Es IST so: Es gibt eine Verwendung von so als Artikel)
- Erklärung: Warum so+N trotz mehrerer grammatikalisierterer Varianten?

so, los geht's

- So bekannt als Form mit verschiedenen Funktionen und Bedeutungen:
 - Subjunktion (konditional) *So es der radikalen Linken gelingt (36502452)*
 - Satzadverb (kausal, circumstance) *Nach den Bergen lockt das Meer und so zieht dann die gesamte Gruppe in der zweiten Woche an den Strand (7031467)*
 - Satzadverb/anaphor. Korrelat (konditional/temporal, konzessiv) *Wurde die neu begonnene Fremdsprache in der Vorstufe mit sechs Jahreswochenstunden begonnen, so wird in der Qualifikationsphase durchgängig mit drei Stunden unterrichtet (137499165)*
 - VP-Adverb/Korrelat (manner) *Um sich keine Blöße zu geben tut er so, als wäre das Bike für ihn der Schlüssel zur großen Unabhängigkeit (29809943)*
 - VP-Adverb (manner) *So klang er wirklich nicht (47797743)*
 - Intensivierungspartikel (graduierend) *Warum fällt man barfuß immer gleich so negativ auf? (56735309)*
 - Vergleichspartikel/Korrelat *Der mit einem Hauch von Wahnsinn umwobene , muskelbepackte Frontmann schüttelte sein schütterere Mähne fast so oft wie der Autor dieser News beim letzten RWE-Spiel (64301667)*
 - Artikel/Adjektiv? *Von so Leuten kann eine große Gefahr ausgehen (3543234)*

(Alle Daten aus DeWaC I)

Datengrundlage



- Alle angeführten Korpusdaten stammen aus dem DeWaC-Korpus (Baroni et al. 2009)
 - "Opportunistisch" erhobene Internetdaten: Forentexte, Blogs, online-Magazine, ...
 - Warum?
 - breites Spektrum an Textsorten, Registern, Stilen (Variation über diese Varietäten ist nicht Gegenstand der Analyse)
 - großer Anteil an konzeptionell mündlichen Daten
 - medial schriftliche Kontexte schließen ausschließlich mündliche Faktoren (Häsitationen, Kiezdeutsch, ...) aus

Datengrundlage: DeWaC

- Partitionierung: 6 Subkorpora à ca. 270.000.000 Token
- Eignet sich sehr gut für die Untersuchung seltener Phänomene
- Meistens genügt Auswertung eines Teils
- Frei verfügbar im CQP-Interface der HU Berlin (www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/korpora/cqp/)
- Annotationen: automatisch vergebene pos-Tags und Lemmata (Treetagger; Schmid 1994)
- Suche: N-determinierende so können nicht an allen syntaktischen Positionen automatisch gefunden werden
→ zusätzliche manuelle Korpussuche nötig

DeWaC 1	268.849.871
DeWaC 2	268.857.448
DeWaC 3	268.885.815
DeWaC 4	268.932.235
DeWaC 5	268.909.724
DeWaC 6	282.734.464
Σ	1.627.169.557

Forschungskontext: so als adnominales Element



- so meistens im Zusammenhang mit *ein* als Determinierer eingestuft
 - Zifonun et al. 1997, S. 1938: *so* und *solch* als "Determinativmodifikatoren (Prädeterminative)" eingestuft, also kein eigenständiges Element
 - Wiese 2011: *so* als Fokusmarker (*hab so Bier geholt so*)
 - Diese Verwendung wird hier nicht untersucht
 - Hole/Klumpp 2000: *so'n* als dritter Standardartikel im Gesprochenen (später)
 - Umbach/Gust 2014: *so ein* als deiktischer Vergleichsausdruck (später)
 - Umbach/Ebert 2009: *so* als Heckenausdruck und Intensivierer

Erste Daten



- (I) *Ich hab da auf den Fotos **so** schwarze Würstchen gesehen (4875614)*
- (II) *Wir haben mit **stattbau** zusammengearbeitet. Die machen **so** Wohnprojekte (170798978)*
- (III) *Ich meine klar handeln ne Menge Texte von Gothicliedern von **so** Themen wie Satanismus, aber das ist doch eigentlich das einzige, was mir da einfällt (77609014)*
- (IV) *Wie müsste denn idealerweise ein Becken für **so** Tierchen aussehen (43338533)*

(Alle Daten aus DeWaC I)



Fragestellungen

- Wie ist der grammatische Status von $so+N$ zu beschreiben?
 - Was kann man daraus für eine deskriptive Syntax der DP ableiten, die so und äquivalente Elemente berücksichtigen will?
- Wie kommt die Struktur $so+N$ zustande?
- Wie beschreibt man die Relation mit den intuitiv vergleichbaren Elementen $solch+$ und $so'n$?

Theoretische Voraussetzungen

- Einfaches NP/DP-Modell (beliebig komplex denkbar; vgl. gestrigen Vortrag Said Sahel):
 - $DP[D_{NP}[AP[A]N]]$
- (Kommt eine deskriptive Syntax, die so berücksichtigt, damit aus?)
- Achtung: keine Behandlung von *so* bei deverbalen oder deadjektivischen Nomina oder Adjektiven innerhalb der NP (!*ein so Lieber*; !*so große Blumen*; !*so Befreite*)

so+N

Syntaktischer Status – Hypothese 1



- so ist Determinierer
- DP[**SO** NP[AP[A]N]]
- *Die machen* DP[**so** NP[Wohnprojekte]]
solche
diese
- *Die machen* DP[**so** NP[AP[schöne] Wohnprojekte]]
solch
manch
solche NP[AP[schönen] Wohnprojekte]
diese

so+N

Syntaktischer Status – Hypothese 1



- so ist Determinierer
- Folgerung 1: so gehört zur Klasse der unflektierten/unflektierbaren Indefinitartikel: *manch, solch, kaum, ...*
- Folgerung 2: so existiert parallel mit *so'n* und *solch* (flektiert und unflektiert)
 - Verwendungspräferenzen sind
 - a) situativer Natur (Registerunterschiede) oder
 - b) abhängig von bestimmten bevorzugten Konstruktionen (dieser Aspekt wird überprüft)

so+N

Syntaktischer Status – Hypothese 2



- so ist Adjektiv
- DP[D NP[AP[so]N]]
- Gegenevidenz:
*Die machen *diese/*jene/*meine so Wohnprojekte*
- Aber:
- *Er soll dem Kanzler ausrichten, dass er gefälligst kein so Gesicht mehr machen soll (DeWaC 2, 213430974)*
- *Ich habe an jede Leine ein paar so Frischhaltetüten gebunden (DeWaC 1, 91226414)*
- *Das ist zum Beispiel Jason , der Schlagzeuger von den Weakerthans , oder Haakon Gebhardt von Motorpsycho . Das sind halt zwei so Typen , die sind (...)* (DeWaC 1, 91308105)
- *Und wenn jemand ein so Riesen-Macho ist (...)* (DeWaC 1, 9657351)
- *wenn Papi auch lauter so langweilige Sachen macht wie Spiegelquälen (DeWaC 1, 51277626)*

so+N

Syntaktischer Status – Hypothese 2



- so ist Adjektiv
- Hypothese 2.1: Die Fälle, in denen so inkompatibel mit Determinierern ist, beruhen auf semantischen Faktoren
- Vgl. *viel*+:
- DP[NP[AP[*viele*] *Menschen*]
- DP[NP[AP[*nette*] *Menschen*]
- DP[*die/jene/diese/deine* NP[AP[*vielen/netten*] *Menschen*]
- DP[*kaum NP[AP[*viele*] *Menschen*]
- aber:
- *Sehr **viele und schöne** Bilderserien hast Du* (DeWaC I, 615633; 190 akzeptable Belege)
- Keine Koordination von so mit Adjektiven belegbar

So als Determinierer



- Beobachtung: *so*+N kann immer durch *solch*+N oder *so'n*+N ausgetauscht werden
- Frage:
 - Warum existiert diese Konstruktion neben den Alternativen
 - (Was legitimiert sie)?

So als Determinierer

- Vgl. Hole/Klumpp 2000 zu *so'n*:
 - vollständig ausgeprägtes Paradigma
 - schließt Lücke zw. *ein* und *der* durch Referenz auf indefinites Token definiten Typs
- *so*: Kein Paradigma
- Im Gegensatz zu *so'n*: Nur Nomina mit fakultativem Artikel können determiniert werden
- Analog zu *so'n*: Beide können rechts von gewissen quantifizierenden Ausdrücken stehen: *kein*, *ein*, *ein paar*, Kardinalzahlen

Semantik von so+N

- **Umbach&Gust 2014** (*so ein*):
- *so* als AP-Adjunkt, VP-Adjunkt und in *so ein* als NP-Determinierer
Komparationsausdruck
- *so ein* (wie andere Demonstrativa) → keine Identitätsrelation zw.
deiktischem Ausdruck und Referenten
- Frage in **Umbach&Gust 2014**: Wie kann die Ähnlichkeitsrelation
zwischen Referenten wie in "*so ein Auto ~ ein Auto wie dieses dort*"
semantisch erfasst werden?
- Mit Blick auf die vorliegenden Daten:
 - In medial schriftlichen Texten meistens keine Relation zwischen
Referenten, sondern zw. Referent und Kategorie
 - Meistens keine Ähnlichkeit, sondern Übereinstimmung aller
Merkmale
- Im Folgenden: Welche klar abgrenzbaren Lesartunterschiede besitzt
so ein Auto bzw. die Pendants ohne *ein*?

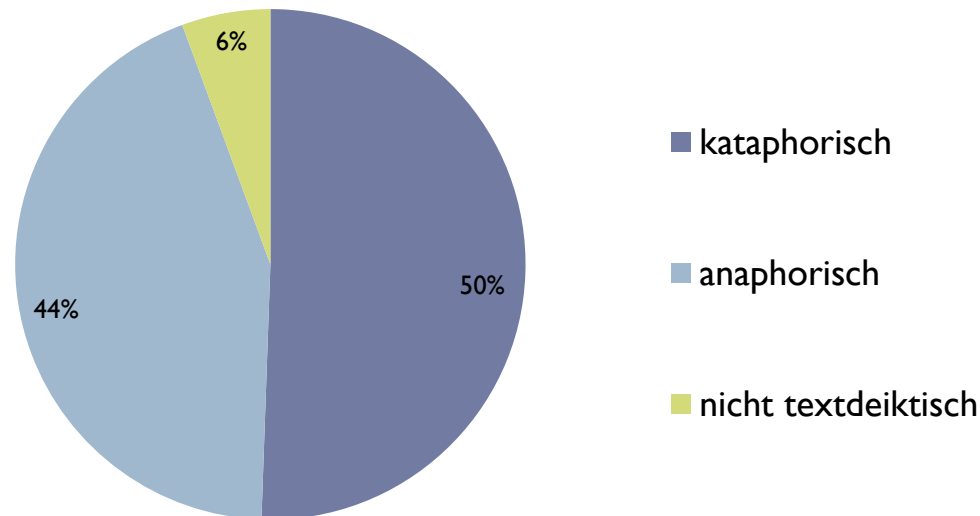
Semantik von so+N

- Durch Sichtung von vorerst 300 Fällen
→ 3 Klassen bei originären Nomina (es gibt bestimmt mehr)
 - 1. anaphorisch (klassenzuweisend)
 - *Ich kann mit **so** Fantasyzeug eh nicht viel anfangen* (4470822)
(Bezug: Tolkien-Werke)
 - 2. kataphorisch (klassenzuweisend)
 - *(...) die kann man doch überhaupt nicht mit **so** Künstlern wie herbert vergleichen* (49125639)
 - 3. nicht textdeiktisch (Ähnlichkeit zw. Referent u. Klasse)
 - *Am besten zieht man die Samen einzeln auf der Fensterbank vor (in **so** Joghurtbechern oder kleinen Blumentöpfen)*
(127494646)
- (Alle Daten aus DeWaC I)

Verteilung der 3 semantischen Klassen



- Manuelle Filterung von 300 positiven Treffern aus zufällig extrahierten Sätzen mit so-N-Verbindungen aus DeWaC I
- Annotation der drei Klassen



Häufigste Klasse: kataphorisches *so*

- Meistens *so-N-wie*-Konstruktion
- Aber auch andere Strukturen: *à la*; z. B.; ...
- Hypothese: Genau diese Vergleichsstrukturen grammatikalisieren das Auftreten von *so* in NP-Kontexten (*so-A-wie*-APs werden in [Umbach&Gust 2014](#) als Prototyp des vergleichenden *so* behandelt)
- Datensichtung: In 200 Sätzen mit *so+N* vs. *solch+N* tritt *wie*+Komplement bei *so* 16,02 Mal häufiger auf als bei *solch*
- Ziel: Analyse größerer Datenmengen, mit Einbindung von *so'n*

Häufigste Klasse: kataphorisches *so*

- Kolligationsanalyse: Ziehen sich *so*+N und *wie*+Vergleichsphrase signifikant stärker an als *so'n*+N und *solch*+N?
- Problem in großen Datenmengen: Vermeidung false hits
→ Einschränkung auf Vorkommen innerhalb von PPn
- Berechnung von Mutual Information je Typ

	Strukt.1: Prep-X-N	Strukt.2: <i>wie</i> +Komp.	Kook.	Größe	MI
<i>so</i>	154	267431	44	13334000	3,4696
<i>so'n</i>	90	267431	3	13334000	0,6856
<i>solch</i> +	19571	267431	349	13334000	-0,197

Zusammenfassung

- so vor Nomina hat Determiniererstatus
- Kann quantifizierende Ausdrücke vorangestellt haben, aber keine anderen Elemente
- Beschreibung von NP_n/DP_n: Eine Syntax, die das berücksichtigt, muss links von einer D-Position eine Q-Position zulassen
 - (Kardinalzahlen können an dieser Position oder an der Adjektivposition zwischen D und N stehen)
- Drei informationsstrukturell-semantische Verwendungen von so+N unterschieden/annotiert (viele andere Klassifikationen denkbar):

Zusammenfassung



- Das Vorkommen von *so* neben *solch* und *solch+* lässt sich morphosyntaktisch und gebrauchsbasiert-semantisch erklären:
 - *so* steht nur in fakultativen D-Positionen (N-Restriktion)
 - *so*, *solch+* und *so'n* besitzen vergleichbare Funktionen
 - *so* ist gegenüber den anderen stark restringiert
 - Aber: *so* ist anscheinend durch die Vergleichskonstruktion "*so ... wie*" (Prototyp von *so* in Positionen unterhalb von VP) als Determinierer lizenziert, was die Kollokationsstärke zw. *so+N* und *wie+Komplement* belegt

Vielen Dank!



Kontakt:


hirschhx@hu-berlin.de

Vielen Dank!



BEANTWORTETE FRAGEN

so und solch?



What's the difference between them?
I see 'so' more often, so etwas, so weit wie, so dass...
But I see this sentence:
Solch ein Mensch begegnet einem selten.
Can solch here be replaced by so? If not, why?
Thanks

Beste Antwort - Gewählt vom Fragenden
There's no difference. (...)

<http://www.italki.com/question/82147>

Kontakt:

hirschhx@hu-berlin.de

Literatur



- Baroni, Marco; Bernardini, Silvia; Ferraresi, Adriano; Zanchetta, Eros (2009) The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43(3): S. 209-226. DeWaC-Korpus: <http://wacky.sslmit.unibo.it/doku.php?id=corpora>
- Hole, Daniel; Klumpp, Gerson (2000) Definite type and indefinite token: the article *son* in colloquial German. *Linguistische Berichte* (182), S. 231-244.
- Schmid, Helmut (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees, *Proceedings of International Conference on New Methods in Language Processing*, Manchester.
- Umbach, Carla; Gust, Helmar (2014) Similarity Demonstratives. *Lingua*.
- Umbach, Carla; Ebert, Cornelia (2009) German demonstrative *so* – intensifying and hedging effects. *Sprache und Datenverarbeitung. International Journal for Language Data Processing* 33 (1-2), S. 153-168.
- Wiese, Heike (2011) *So* as a focus marker in German. *Linguistics* 49/5, S. 991-1039.
- Zifonun, Gisela; Hoffmann, Ludger; Strecker, Bruno (1997) Grammatik der deutschen Sprache. Band 2. Berlin/New York; de Gruyter.