Syntactic annotation of non-standard language

Guest lecture series of the CRC 732
„Incremental Specification in Context“
University of Stuttgart
Mo, 8.12.2014 | 14:00–15:30

Burkhard Dietterle
(Humboldt University Berlin)

Outline

1. Preliminaries
2. Non-standard language
3. Challenges for annotation
4. Annotation solutions
5. Use in corpus research
6. References

Preliminaries

- The following considerations on the problematics of non-standard language in corpus linguistics are based on experience gained within the „**No**n-**sta**ndard **D**eutsch“ (NoSta-D) project:

  CLARIN-D | F-AG 7 | KP 2

  Humboldt University Berlin

  Ruhr-University Bochum

  funded period: 2012 – 2013;

  prolonged period (HU only): 2013 – 2014

  homepage @ CLARIN: http://clarin-d.de/de/fachspezifische-arbeitsgruppen/f-ag-7-computerlinguistik/kurationsprojekt-2.html

  homepage @ HU: http://u.hu-berlin.de/nosta-d

Preliminaries

- NoSta-D reference paper:
  - Dipper, Stefanie; Lüdeling, Anke; Reznicek, Marc (2013): NoSta-D: A Corpus of German Non-Standard Varieties.

    In: Zampieri, Marcos & Sascha Diwersy (eds.): Non-Standard Data Sources in Corpus-Based Research 5. Köln: Shaker (ZSM Studien), 69–76.

    Online: http://www.linguistics.ruhr-uni-bochum.de/~dipper/papers/nosdac13.pdf

Preliminaries

- NoSta-D based variational study:
  - Dietterle, Burkhard; Lüdeling, Anke; Reznicek, Marc (submitted): Zur Syntax von Plauderchats.

    In: Beißwenger, Michael (ed.): Empirische Erforschung internetbasierter Kommunikation. Berlin: Akademie Verlag.

    Online: http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/nosta-d/empirikom-paperpackage

Preliminaries

- previous talk, which happened to become the reason for today's invitation:
  - Dietterle, Burkhard; Dipper, Stefanie; Lüdeling, Anke; Reznicek, Marc (2014): CLARIN-D-Kurationsprojekt: Linguistische Annotation von Nichtstandardvarietäten. Guidelines und „Best Practises“ (F-AG 7). Heidelberg University

    Online: http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/nosta-d/dietterle_et_al-2014

Non-standard language

- for this talk defined as a data driven notion:
  - Non-standard language is language with a distribution of structures different from the corresponding distribution in standard language.

Non-standard language

- example (1), adopted from Storrer 2013:
  - Inflectives are often found in chat.
  - Inflectives are rarely found in newspapers.

- example (2), adopted from Dietterle et al. submitted:
  - Parentheses are rarely found in chat.
  - Parentheses are often found in newspapers.

  - Taking newspaper language as standard results in positing chat language to be non-standard.

‘Inflektiv’ von Teuber (1998)
‘Aktionswort’ bei Storrer (2013)

Non-standard language

- delineation from theory driven definitions:
  - Not the mere presence of 'non-canonical' structures is used here as a defining property of non-standard language, but their wider distribution therein in contrast to their sparse distribution in standard language.
  - (yet, even before regarding non-canonical structures a comparatively sparse distribution of canonical structures is as well used as a defining property of non-standard language)

'nicht-kanonisch' i.S.v. Hirschmann et al. (2007)

Non-standard language

- example (3), adopted from Dietterle et al. submitted:
  - Repairs are found in spoken language.
  - Repairs are not found in newspapers.
  - Taking newspaper language as standard results in positing spoken language as non-standard.
  - We don't need to say that repairs cannot appear in newspapers.
  - We can say that repairs just happen not to appear in newspapers, but possibly could.

‘Reparatur’ hier im weitesten Sinne inklusive Neustarts und Wiederholungen, vgl. Belz (2013)

Non-standard language

- yet another instantiation of grammar:

  Non-standard language instantiates the same grammar as standard language.

  This single grammar needs to allow for both canonical and non-canonical structures.

  (Interestingly, the lexicon is readily accepted to contain standardly uninstantiated items: cf. pejorative vocabulary which, being part of the lexicon (especially of native speakers!), is not instantiated in newspapers, but in chat and spoken language. Asterisked pejoratives in newspapers illustrate this even more.)

Vergleiche ‘Mindergebrauch’ bzw. ‘Übergebrauch’ in Zeldes et al. (2008) und Hirschmann et al. (2013).

Hier eben nur nicht zwischen Lernern und Muttersprachlern, sondern zwischen Nichtstandard und Standard.

vgl. auch Sandig (1973)

Non-standard language

- example (4), adopted from Hirschmann et al. 2007:
  - Genitive-marked (or genitive-homophonous) accusative prepositional complements are found in learner language: e. g. *„für das Haushaltsjahrs“*
  - Genitive-marked (or genitive-homophonous) accusative prepositional complements are not found in newspapers.
  - Taking newspaper language as standard results in positing the use of standardly uninstantiated genitive-marked (genitive-homophonous) accusative prepositional complements as indicator of the non-standardness of learner language.

Challenges for annotation I

- standardly uninstantiated structures:
  - not covered by commonly used annotation schemas
  - ('non-canonical')

- How to code presence in non-standard language?

Challenges for annotation I

- ‘non-canonical’, standardly uninstantiated structures in German:
  - @-postings
  - emoticons
  - unspaced writings of syntactically distinguishable, but phonetically not cliticised words
  - spaced compounds
  - inflectives
  - repairs
  - dangling (‘Chinese-style’) topics
  - fragments (annotated via their opposite—ellipted structures)

‘Dangling topic’, ‘Chinesisches Topik’ im Sinne von Wallace Chafe (1976), zit. nach Xu (2006)

‘Ellipsen’ hier im weitesten Sinne; vgl. Reich (2011)

Challenges for annotation II

- ambiguous structures:
  - For structures with multiple, but systematic meanings default annotation guidelines can be formulated, which to use if disambiguation by context fails.
- example (5) of default annotation guidelines:
  - *„Annotate modifiers as high as possible, but avoid annotating modifiers to auxiliaries.“*

    (cf. TIGER 2003)

‘Ambiguität’ hier im Sinne von Fries (1980)

TIGER 2003 = Albert et al. 2003

Challenges for annotation II

- vague structures:
  - For structures with multiple, but unsystematic meanings default annotation guidelines can be formulated as well, but these won't reflect a choice of one certain interpretation out of a clear-cut set of interpretations, rather a choice of some uncertain interpretation out of a diffuse set of interpretations.

‘Vagheit’ hier im Sinne von Fries (1980)

Challenges for annotation Ⅱ

- examples of a vague modifier:
  - (6) *„ich mache quasi“* (NoSta-D-bematac_2012-10-30-B, Segment 132)
  - (7) *„gehe quasi rüber“* (NoSta-D-bematac_2012-10-30 - B, Seg. 133)
  - (8) *„senkrecht nach oben bis du an der Laborantin so oberhalb der da Laborantin bist quasi“* (NoSta-D-bematac_2012-01-19-A, Seg. 278)
- For (6) and (7) one can default to annotating *quasi* as modifier of a finite verb.
- For (8) one can default to annotating *quasi* as modifier of the highest finite verb (here the missing matrix verb) or the nearest finite verb (here the embedded auxiliary).
- Yet, neither annotation of *quasi* does resolve the vagueness of the utterances containing it.

Nämliches für „sozusagen“

Challenges for annotation III

- standardly low-frequent structures
  - more or less satisfactorily covered by commonly used annotation schemas
  - (accordingly, more or less ‘canonical’ depending on applied annotation schema)

- Which annotation schemas to take?
  - TIGER 2003, TüBa-D/Z 2005, Foth 2006, …
- recent tendency towards dependency-based annotation schemas

Skut et al. (1997): noch allgemein formuliert. Tektogrammatisch besser als Phenogrammatisch; am besten aber nicht Tektogrammatik in Reinform – Dependenzgrammatik – sondern Kombination aus Tekto- und Phenogrammatik. Also letztlich TIGER.
Dependenzgrammatik könne keine nichtlexikalischen Köpfe haben. In NoSta-D-Dependenzen aber schon, und zwar die Segmentwurzel!

Kübler & Prokić (2006): Dependenzgrammatik ausnahmsweise für das Deutsche besser als für andere Sprachen wegen besserer Abbildung weiter syntaktischer Relationen und von Koordination.

Rehbein (2010): intuitiv verständliche direkte Kodierung der Prädikat-Argument-Struktur in Dependenzgrammatik; geeignet für automatische Bedeutungsextraktion; Parsingeffizienz mit wenig komplexen Algorithmen.

Seeker & Kuhn (2012): Dependenzkonversion aus Konstituenten (TIGER 2003) mit explizierten Verbellipsen (nur Verbellipsen?). In NoSta-D Ellipsen (nicht nur Verbellipsen) in NORM expliziert, in ORIG nicht annotiert (Fragmente annotiert!). In ORIG und NORM je ein nichtlexikalischer Knoten – die Segmentwurzel.

Frank (2013:76 u. 88): lange Tradition der Dependenzgrammatik von Antike bis Tesnière (1959), gegenwärtig Revival dank effizienter Parsingalgorithmen. Besonders sprachübergreifend anwendbar mit geringem Anpassungsaufwand. Also auch varietätenübergreifend!

TüBa-D/Z = Telljohann et al. 2012

Challenges for annotation III

- cases of standardly low-frequent, almost '(non-)canonical' structures in German:
  - complex prefield
  - subjunction + V1/V2
  - complex subjunction / sentential prepositional complement
  - comparative correlatives
  - non-integrable direct speech
  - parentheses

'mehrfache Vorfeldbesetzung' Müller (2003)
'komplexes Vorfeld' (Breckle & Zinsmeister 2011)

als-V1, dass-V2, weil-V2

als dass, als ob, als wenn, wie wenn

auf dass, bis dass, ohne dass

comparative correlatives, s. Zeldes (2012)

'Nicht integrierbare direkte Rede' eher in Belletristik

'Parenthesen' hier im engen Sinne von nicht integrierbaren Einschüben

Challenges for annotation Ⅳ

- standardly instantiated structures
  - uncontroversially covered by commonly used annotation schemas
  - (‘canonical’)

- But, how to annotate in case of absence in non-standard language, i. e.

  how to annotate presence of ‘ellipted structure’ in non-standard language?

Challenges for annotation Ⅳ

- cases of standardly instantiated, but non-standardly uninstantiated ellipted structure
  - verbs
  - verbal complements
    - subjects
    - objects
  - prepositional complements
- These ellipted structures are ambiguous and disambiguable by context. If not the lexical item itself can be resolved, than at least its category: missing verb or (pro)noun.
- Also ambiguous and disambiguable is ellipsis of noun and adjective complements and ellipsis of modifiers in informationally non-neutral context.
- However, in informationally neutral contexts ellipsis of noun and adjective complements and ellipsis of modifiers remains vague.

21

Wenn nur Kategorie disambiguierbar, werden semantisch schwache oder Dummy-Verben bzw. Dummy-Nomen normalisiert.

– „Wie war dein Urlaub?“
– „Super! Am ersten Tag [meines Urlaubs] ging's nach …“

Dass *erster Tag* einen nicht realisierten Genitivmodifikator – und zwar *mein Urlaub* – haben kann, ist aus dem informationell nicht neutralen Kontext schließbar.

– „Wie geht's?“
– „(Äußerst, immer noch, leider, sehr, ziemlich) schlecht“

Was für weggelassene Modifikatoren *schlecht* in diesem informationell neutralen Kontext haben könnte, bleibt ungewiss (bis auf dass wahrheitswertändernde Modifikatoren ausgeschlossen sind)

Annotation solutions

- assumption:
  - For each sequence or original language data there exist one sequence of normalised language data such that
    - the normalised sequence is a possibly false superset of the original sequence
    - the intersecting elements of normalised and original sequences follow the same order
    - the normalised sequence does not cross borders with another normalised sequence
    - the normalised sequence contains no unorthographic or alternative writings
    - the normalised sequence contains no fragments which are complements or modifiers of ellipted structure
    - the normalised sequence is a contextually plausible paraphrase of the original sequence
    - there is no other normalised sequence involving less additional assumptions, i. e.
      - being shorter
      - being orthographically closer to the original
      - requiring the least amount of structures to be posited as standardly uninstantiated
      - being plausible for a larger subset of the set of all considerable contexts—from one-word context up to full-text context

22

vgl. aber ‘Zielhypothese’ von Lüdeling (2008)

Alle Elemente von ORIG sind notwendigerweise auch Elemente von NORM, aber
nicht alle Elemente von NORM sind notwendigerweise auch Elemente von ORIG.

Normalisierungen können Parenthesen enthalten, diese überschreiten aber nicht die Grenzen der Normalisierung, sondern sind vollständig innerhalb ihrer Grenzen enthalten.

Normalisierungen können ausgelassene Komplemente und Modifikatoren haben, wenn deren Komplemente und Modifikatoren ebenso ausgelassen sind.

Wir müssen nur sieben im (Zeitungs)standard uninstanziierte Strukturen annehmen, s. oben.

Im Idealfall gilt ORIG = NORM, wenn ORIG im Standard instanziiert werden kann.

Annotation solutions

- normalisation:
  - not itself a variety
  - not ‘correct’ in any traditional sense
  - annotatable

vgl. ‘Zielhypothese Eins’ und ‘Zielhypothese Zwei’ im Falko-Korpus, für die der Anspruch, selbst eine Varietät zu sein, gilt. Es sollen muttersprachlich äquivalente Varietäten zu den Lernerdaten sein.

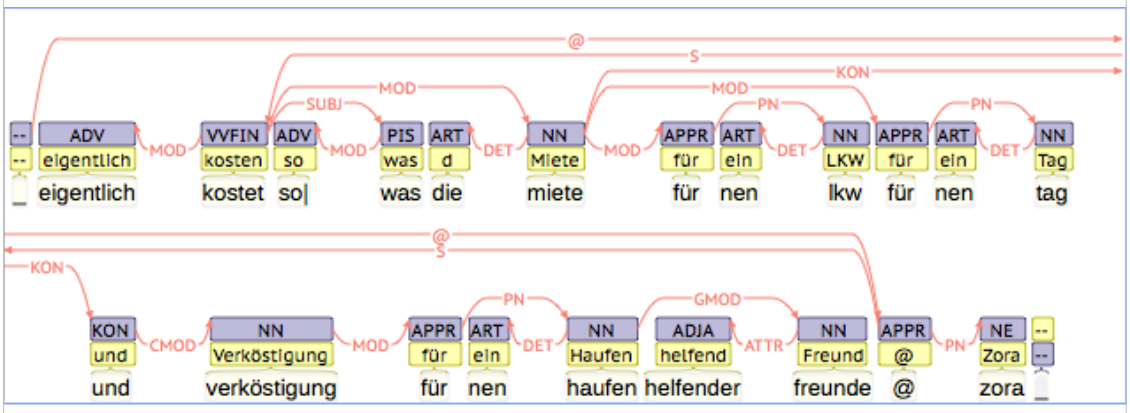
Der NoSta-D-Normalisierung am nächsten kommt in Falko die ‘Zielhypothese Null’, die der Zielhypothese Eins abzüglich Wortstellungsänderungen entspricht. Die Zielhypothese Null hat daher auch nicht den Anspruch, eine muttersprachliche Varietät zu sein (wird im Gegensatz zur NoSta-D-Normalisierung allerdings nicht unmittelbar annotiert, sondern mittelbar über die Zielhypothese Eins)

s. Reznicek et al. 2012, 2013

Annotation solutions

- concept of workflow:
  - normalisation of original data
  - annotation of normalisation
  - transferring annotations of normalisation to original data
    - parallel transfer, where possible
    - unparallel transfer, where parallel transfer not possible

@-postings



NoSta-D-unicum, Segment 221

25

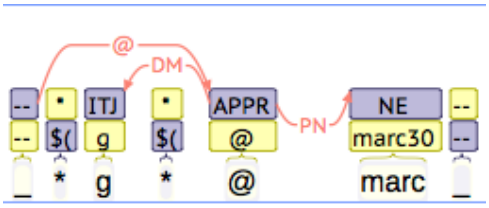
vgl. Discourse-Level-Konstituente in TIGER 2003
auch für vorangestelltes @+Adressat

@-postings

part	pos-tag	govenor	dependency label
@	APPR	root or next highest root-like node	@
content	as if independent	@	as if independent
addressee	N…, P…	@	PN

POS-Tags nach STTS 1999 (=Schiller et al. 1999)
Das „@“ in @-Adressierungen allerdings nicht XY, sondern APPR.

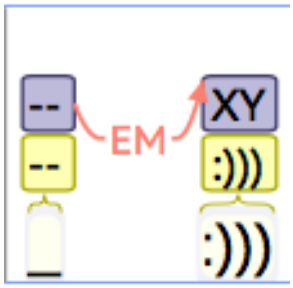
Asterisk-expressions



NoSta-D-unicum, Segment 262

Asteriske wie auch sonstige Interpunktion nicht dependenzfähig

Emoticons



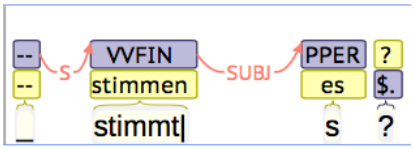
NoSta-D-unicum, Segment 32

pos-tag	govenor	dependency label
XY	segment node (root)	EM

28

keine Interpunktion, daher dependenzfähig.
aber nicht-kombinatorisch (zumindest in unserem Sample), daher allein segmentfüllend und mit alternativlosem Dependenzlabel

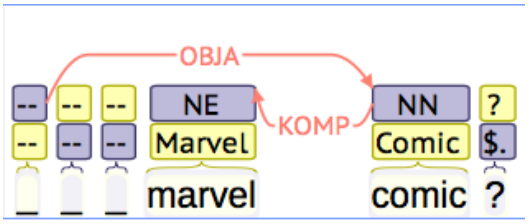
Unspaced writings



NoSta-D-unicum, Segment 100

token$_i$	**token$_{i+1}$**	**pos-tag, govenor, dependency label**
…|	…	as if spaced

Spaced compounds



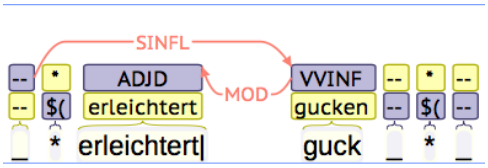
NoSta-D-unicum, Segment 737

dependent	govenor	dependency label
morphological non-head	morphological head	KOMP

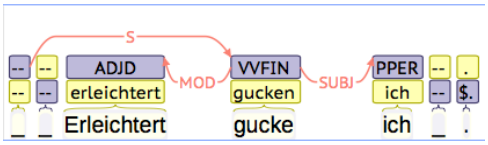
bei mehr als zweigliedrigen Komposita rekursive Regelanwendung

Inflectives

orig



norm

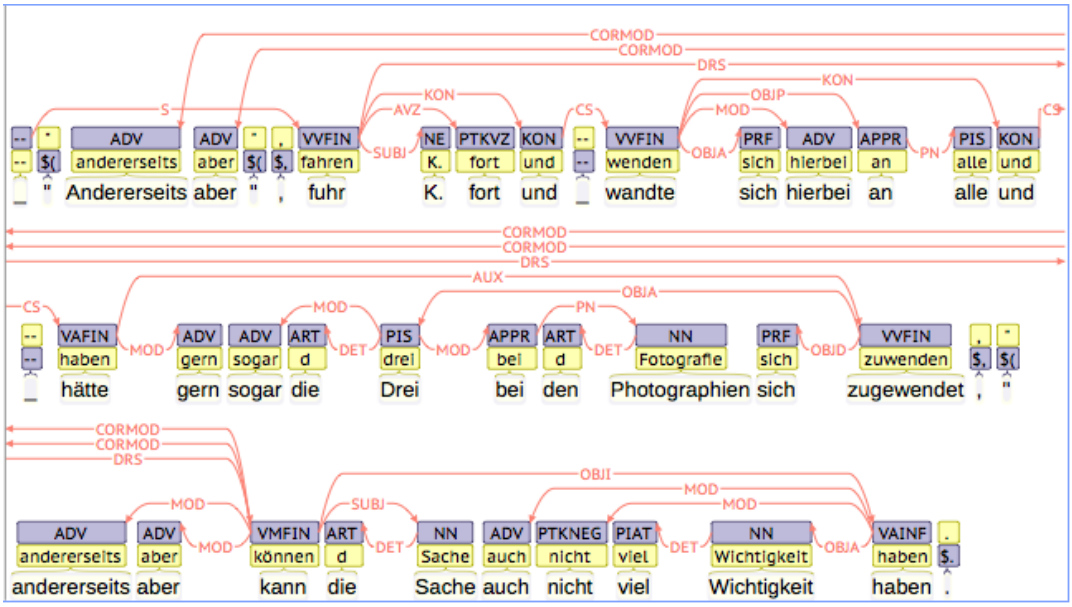


NoSta-D-unicum, Segment 15

	text layer	pos-tag	govenor	dependency label
	orig	VVINF	root	SINFL
	norm	VVFIN	root	S

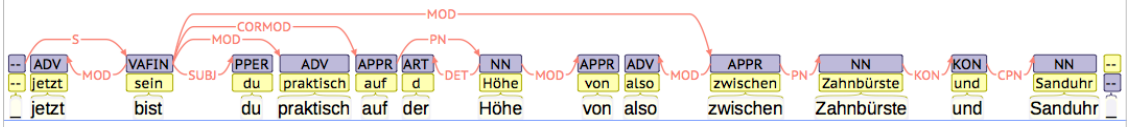
s. auch Schlobinski 2001 zu ‘Inflektivkonstruktion’

Repairs



NoSta-D-kafka, Segment 162

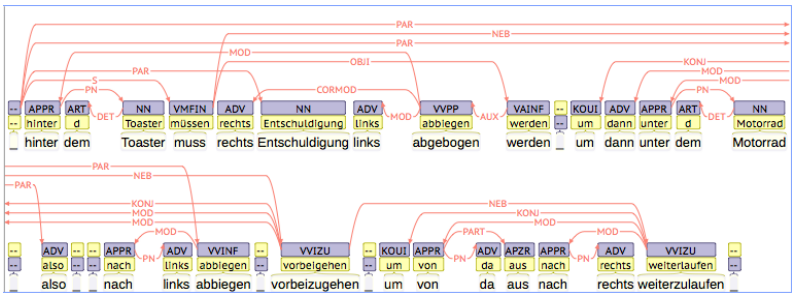
Repairs



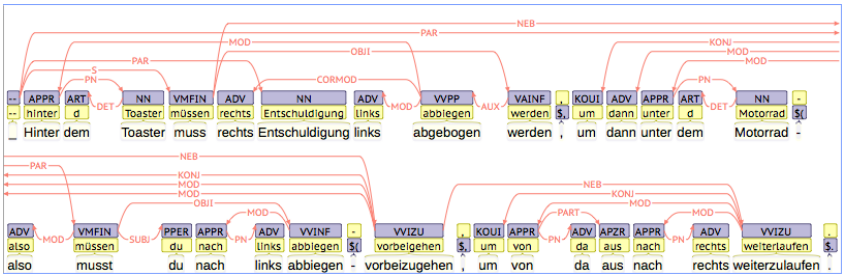
NoSta-D-bematac_2012-11-02-B, Segment 204

Repairs

orig

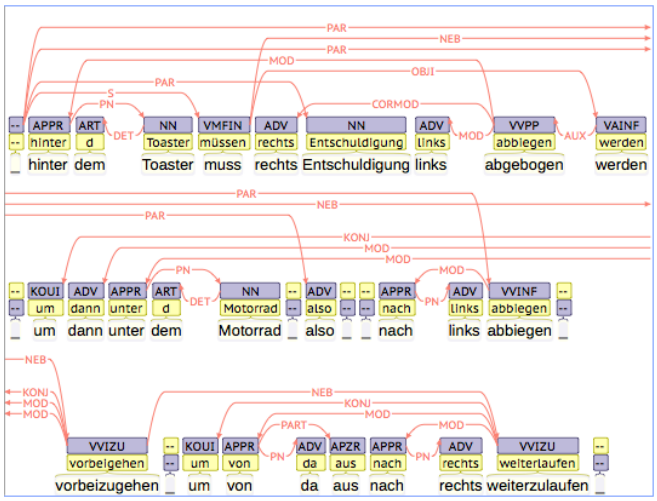


norm



NoSta-D-bematac_2012-10-31-A, Segment 25

Repairs



NoSta-D-bematac_2012-11-02-B, Segment 204

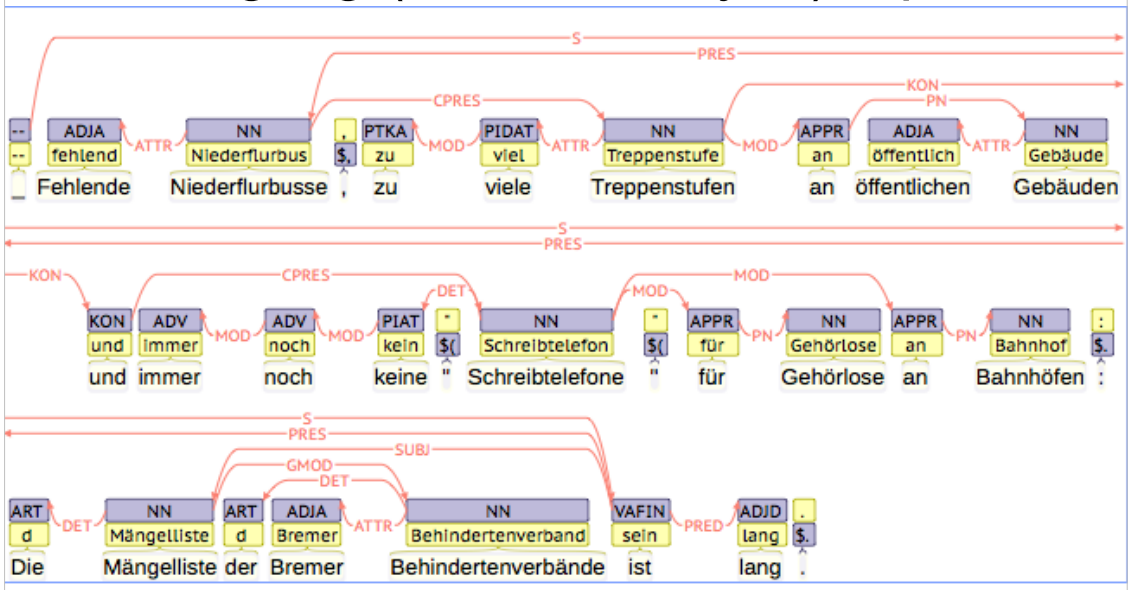
Repairs

part	govenor	dependency lable
reparandum	as if independent	COR × as if independent
interregnum	root	PAR
reparans	as if independent	as if independent
reparans introducing modifier	reparans	MOD

Wenn Govenor des Reparandums ambig, gilt der tiefere.

Reparans einleitende Modifikatoren („also“ und „so“), könnten auch zu den Interregna gezählt werden. Sie scheinen allerdings eher den Beginn des Reparans zu markieren, während Interregna eher das Ende des Reparandums markieren.

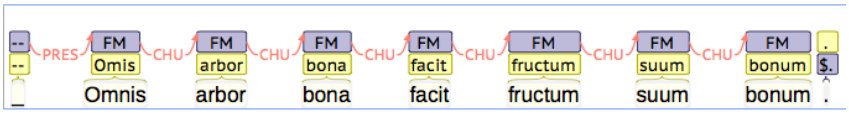
Dangling (‘Chinese-style’) topics



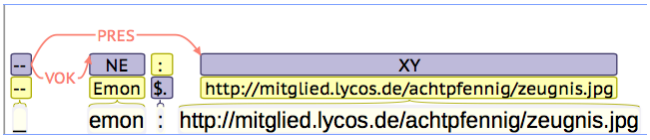
NoSta-D-tuebadz, Segment 49

Dangling (‘Chinese-style’) topics

type	govenor	dependency label
dangling topic	clause node	PRES
title, subtitle, headline, caption, signature etc. not normalisable to sentences	root	PRES



NoSta-D-anselm, Segment 1



NoSta-D-unicum, Segment 184

Fragments

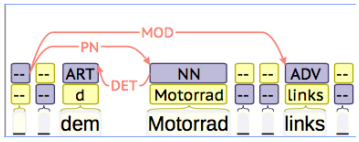
- Fragmentary structure is annotated via its opposite—ellipted structures.
- Ellipted structure is made explicit in normalisation.
- Govenor of ellipted structure in normalisation becomes govenor of fragment in original data.
- Dependency label of fragment in original data is chosen according to the type of node the govenor belongs to.

root and root-like nodes

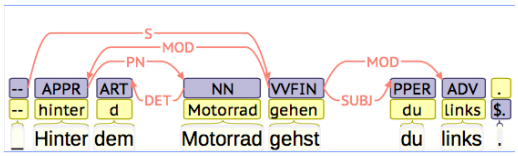
type	govenor	crosslabelling	label of governed fragments
segment node (root)	„_“	none	lowest label of ellipted dependency chain, but in case of governed parenthetical fragments: PAR
@-node	„@“	none	lowest label of ellipted dependency chain
clause node	finite verb	only in case of governed fragments: X × …	highest label of ellipted dependency chain
conjunction node	KON	C × (DR ×) …	lowest label of ellipted dependency chain
repair node	govenor of repaired and repairing structure	if reparandum: COR × (C|X ×) … if reparans: as if independent	as if reparans or as if independent, resp.
discourse-level head node	govenor of direct speech non-integrable due to ellipsis or discourse-level head (cf. TIGER 2003)	DR × …	lowest label of ellipted dependency chain or label, the direct speech part would have independently from discourse-level head, resp.
other governed nodes		only in case of governed fragments: X × …	highest label of ellipted dependency chain

Fragment govenor = root

orig

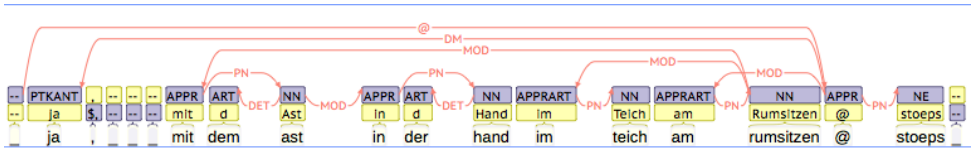


norm

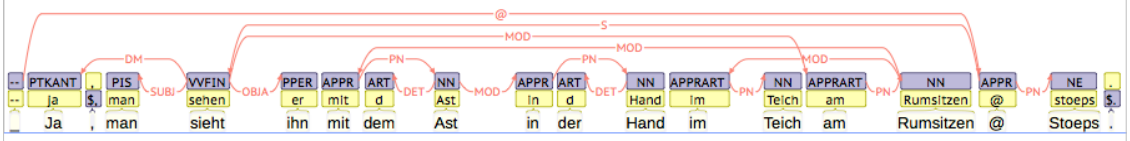


Fragment govenor = @-node

orig



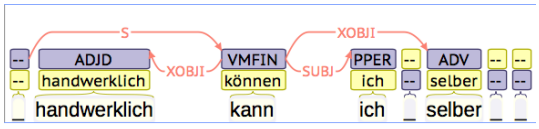
norm



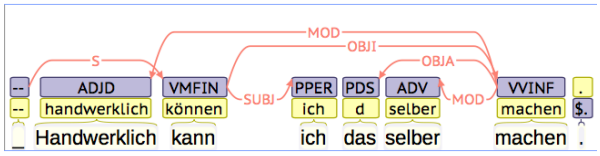
NoSta-D-unicum, Segment 518

Fragment govenor = clause node

orig



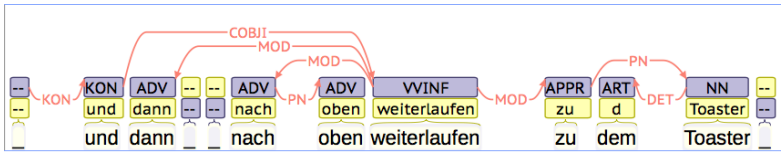
norm



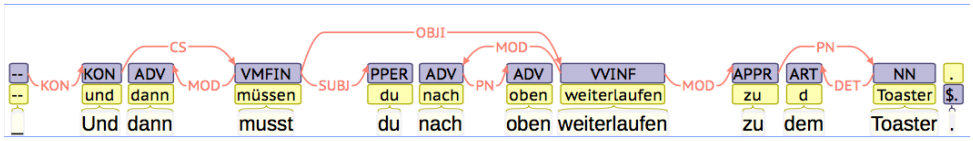
NoSta-D-unicum, Segment 260

Fragment node = conjunction node

orig



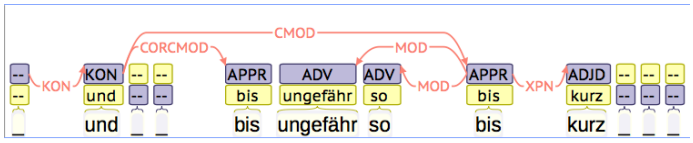
norm



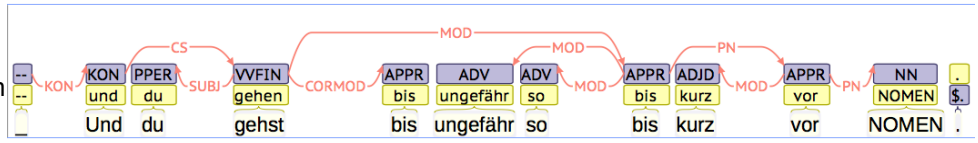
NoSta-D-bematac_2012-10-31-A, Segment 24

Fragment node = repair node (CORC)

orig



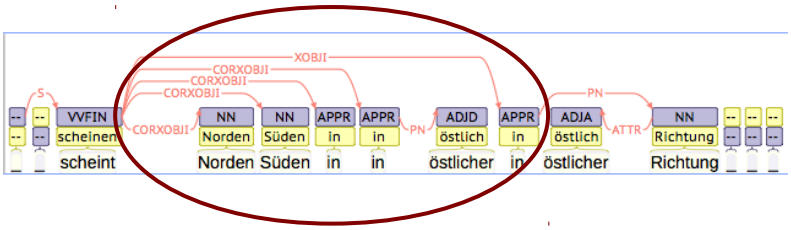
norm



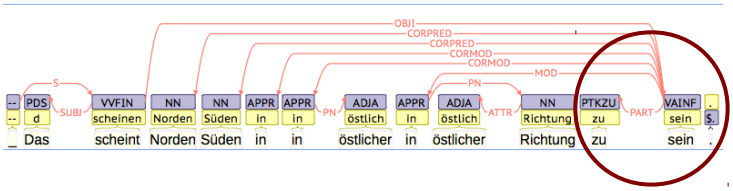
NoSta-D-bematac_2012-01-19-A, Segment 227

Fragment node = repair node (CORX)

orig



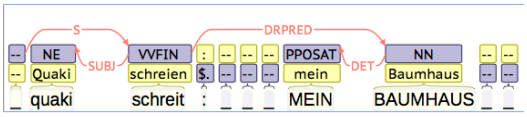
norm



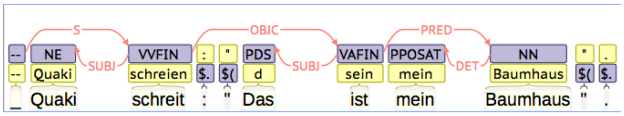
NoSta-D-bematac_2012-11-08-A, Segment 11

Fragment node = discourse-level head node

orig



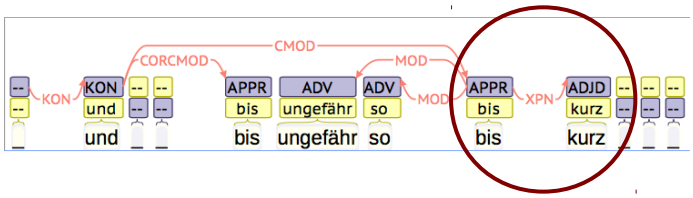
norm



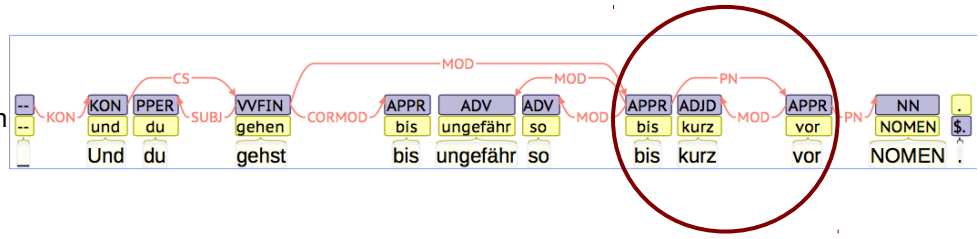
NoSta-D-unicum, Segment 583

Fragment node = other governed node

orig



norm



NoSta-D-bematac_2012-01-19-A, Segment 227

Use in corpus research

- enables sophisticated variational research on non-standard varieties in a two-step way:
  - Step 1: Intravariety comparison:
    What is the difference between original data of variety A and its normalisation?
    What is the difference between original data of variety B and its normalisation?
  - Step 2: Intervariety comparison:
    What is the differecence between the [differences found in Step1]?

- example (9) from Dietterle et al., submitted:

  Comparative analysis of fragments (in original data) and ellipted structures (explicitly coded in normalised data) proved chat language to be quite different from spoken language, contrary to common claims about the orality of chat.

  In chat, there are less often cases of ellipted structures being themselves complements or modifiers and leaving fragmented complements and modifiers.

  In spoken, there are more often cases of ellipted structures being themselves complements or modifiers and leaving fragmented complements and modifiers.

  Chat is less fragmentary than spoken language.

vgl. z.B. Storrer (2013)

Use in corpus research

- should facilitate building of non-standard corpora:
  - manual normalisation should prove faster and more reliable than manual annotation
  - normalised data should be parsable automatically
  - transfer of normalisation annotation to original data is straightforward

References

(including those in slide comments)

Albert, Stefanie; Anderssen, Jan; Bader, Regine; Becker, Stephanie; Bracht, Tobias; Brants, Sabine; Brants, Thorsten; Demberg, Vera; Dipper, Stefanie; Eisenberg, Peter; Hansen, Silvia; Hirschmann, Hagen; Janitzek, Juliane; Kirstein, Carolin; Langner, Robert; Michelbacher, Lukas; Plaehn, Oliver; Preis, Cordula; Pußel, Marcus; Rower, Marco; Schrader, Bettina; Schwartz, Anne; Smith, George; Uszkoreit, Hans (2003): TIGER Annotationsschema. Saarland/Stuttgart/Potsdam: Universität des Saarlandes/Universität Stuttgart/Universität Potsdam.

Beißwenger, Michael (2013): Das Dortmunder Chat-Korpus. In: Zeitschrift für germanistische Linguistik 41(1), 161–164.

Belz, Malte (2013) Disfluencies und Reparaturen bei Muttersprachlern und Lernern - eine kontrastive Analyse. Master thesis at Humboldt University Berlin.

Breckle, Margit & Zinsmeister, Heike (2011): AleSKo-Korpus. Annotationsrichtlinien “Funktion des Vorfeld”. Erweitert für die funktionale Annotation von nicht-Vorfeld Kategorien.

Dietterle, Burkhard; Lüdeling, Anke; Reznicek, Marc (submitted): Zur Syntax von Plauderchats. In: Beißwenger, Michael (ed.): EMPIRIKOM.

References

(including those in slide comments)

Dietterle, Burkhard; Dipper, Stefanie; Lüdeling, Anke; Reznicek, Marc (2014): CLARIN-D-Kurationsprojekt: Linguistische Annotation von Nichtstandardvarietäten. Guidelines und „Best Practises“ (F-AG 7). Talk at Heidelberg University.

Dipper, Stefanie; Lüdeling, Anke; Reznicek, Marc (2013): NoSta-D: A Corpus of German Non-Standard Varieties. In: Zampieri, Marcos & Sascha Diwersy (eds.): Non-Standard Data Sources in Corpus-Based Research 5. Köln: Shaker (ZSM Studien), 69–76.

Foth, Kilian A. (2006): Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Technischer Report. Hamburg: Universität Hamburg.

Frank, Anette (2013): A Tour of Grammar Formalisms. In: King, Tracy Holloway & de Paiva, Valeria (eds.): From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen. CSLI Publications.

References

(including those in slide comments)

Fries, Norbert (1980): Ambiguität und Vagheit. Einführung und kommentierte Bibliographie. Tübingen: Niemeyer.

Hirschmann, Hagen; Doolittle, Seanna; Lüdeling, Anke (2007): Syntactic Annotation of Non-canonical Linguistic Structures. In: Proceedings of Corpus Linguistics 2007, Birmingham.

Hirschmann, Hagen; Lüdeling, Anke; Rehbein, Ines; Reznicek, Marc und Zeldes, Amir (2013): Underuse of Syntactic Categories in Falko – A Case Study on Modification. In: 20 Years of Learner Corpus Research. Looking Back, Moving Ahead.

Kübler, Sandra & Prokić, Jelena (2006): Why is German Dependency Parsing More Reliable than Constituent Parsing? In: Proceedings of TLT 5. 7–18.

Lüdeling, Anke (2008): Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In: Walter, Maik & Grommes, Patrick (eds.): Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung. Tübingen: Max Niemeyer. S. 119–140.

References

(including those in slide comments)

Müller, Stefan (2003): Mehrfache Vorfeldbesetzung. In: Deutsche Sprache 31(1). S. 29–62.

Rehbein, Ines (2010): Der Einfluss der Dependenzgrammatik auf die Computerlinguistik. In: Zeitschrift für Germanistische Linguistik 38(2). 224 - 228.

Reich, Ingo (2011): Ellipsis In: Maienborn, Claudia, Klaus von Heusinger und Paul Portner (eds.): Semantics: An International Handbook of Natural Language Meaning. Berlin/New York: de Gruyter, 1849–1874.

Reznicek, Marc; Lüdeling, Anke; Krummes, Cedric; Schwantuschke, Franziska; Walter, Maik; Schmidt, Karin; Hirschmann, Hagen; Thorsten, Andreas (2012): Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 2.0.1. Technical report: Humboldt University Berlin.

Reznicek, Marc; Lüdeling, Anke; Hirschmann, Hagen (2013): Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: Díaz-Negrillo, Ana; Ballier, Nicolas und Thompson, Paul (eds.): Automatic Treatment and Analysis of Learner Corpus Data. Amsterdam: John Benjamins, 101–124.

Sandig, Barbara (1973): Zur historischen Kontinuität normativ diskriminierter syntaktischer Muster in spontaner Sprechsprache. In: Deutsche Sprache 3. S. 37–57.

References

(including those in slide comments)

Schiller, Anne; Teufel, Simone; Stöckert, Christine; Thielen, Christine (1999): Guidelines für das Tagging deutscher Textkorpora mit STTS. Technischer Bericht. Stuttgart/Tübingen: Universität Stuttgart/ Universität Tübingen.

Schlobinski, Peter (2001): *knuddel zurueckknuddel dich ganzdollknuddel*. Inflektive und Inflektivkonstruktionen im Deutschen. In: Zeitschrift für Germanistische Linguistik 29(2), 192–218.

Seeker, Wolfgang & Kuhn, Jonas (2012): Making Ellipses Explicit in Dependency Conversion for a German Treebank. In: Proceedings of the 8th International Conference on Language Resources and Evaluation, p. 3132–3139. Istanbul.

Skut, Wojciech; Brants, Thorsten; Krenn, Brigitte; Uszkoreit (1997): Annotating Unrestricted German Text. In: Proceedings der 6. Fachtagung der Sektion Computerlinguistik der Deutschen Gesellschaft für Sprachwissenschaft. Heidelberg.

Storrer, Angelika (2013): Sprachstil und Sprachvariation in sozialen Netzwerken. In: Frank-Job, Barbara; Mehler, Alexander; Sutter, Tilmann (eds.): Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW. Wiesbaden: VS Verlag für Sozialwissenschaften, 331-366.

References

(including those in slide comments)

Telljohann, Heike; Hinrichs, Erhard W.; Kübler, Sandra; Zinsmeister, Heike; Beck, Kathrin (2012): Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z).

Teuber, Oliver (1998): fasel beschreib erwähn — Der Inflektiv als Wortform des Deutschen. In: Germanistische Linguistik 141/142, 7–26.

Xu, Liejiong (2006): Topicalization in Asian Languages. In: Everaert, Martin & van Riemsdijk, Henk (eds.): The Blackwell Companion to Syntax. Volume 5. p. 137–174.

Zeldes, Amir; Lüdeling, Anke; Hirschmann, Hagen (2008): What's Hard? Quantitative Evidence for Difficult Constructions in German Learner Data. In: Arppe, Antti: Sinnemäki, Kaius; Urpo, Nikanne (eds.): Proceedings of Quantitative Investigations in Theoretical Linguistics 3 (QITL-3). Helsinki, 74-77.

Zeldes, Amir (2012): Productivity in Argument Selection: A Usage-Based Approach to Lexical Choice in Syntactic Slots. Dissertation at Humboldt University Berlin.