



# Searching with POS tags (2)

Query: "search for a form of werden and a predicative adjective in the same sentence"

- ▶ Result: 1318 sentences
  - ▶ inspection of the first 20: 5 inchoatives, 3 passives, 9 adverbs (mostly passives), 3 within different clauses
  - ▶ second inchoative sentence: ... daß dieser asketische Kunstradfahrer endlich mal gegen Ausgrenzung aktiv wird ...
    - ▶ (that this ascetic trick cyclist finally becomes active against marginalization)

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

8 / 64

# Searching with lemma information

Query: [word = "w[e]u[rde]([n]t)?" | word = "wirst" | word = "wird" | word="geworden"] [\* [tpos = "ADJD"] within s

- ▶ Still rather awkward, would be easier if we could state that all these are forms of the lemma "werden"
  - ▶ this can only be used in a query if lemma information is possible
- ▶ Resulting query: [lemma = "werden"] [\* [tpos = "ADJD"] within s
  - ▶ should not change number of retrieved sentence, but mistakes are possible in lemmatization

Query could not be tested because Tüba-D/Z does not include lemma information at present

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

9 / 64

# A basic setup for corpus searches

## Meurers 2005

- Corpora
- ▶ 39,5 million words *Frankfurter Rundschau* (FR)
  - ▶ 8,5 million words *Donaukurier* (DK)

Corpus preparation (Helmut Feldweg, Oliver Christ)

- ▶ tokenization
- ▶ tagging with ELWIS tagset (46 tags, → STTS)
- ▶ sentence segmentation

Search tool: cqp/Xkwc (Oliver Christ, Bruno Schulze)

To search for theoretically relevant examples, a phenomenon has to be characterized in terms of

- ▶ occurrences of word forms and part of speech tags in
- ▶ direct linear sequence or
- ▶ linear sequences within a search window.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

10 / 64

# Word occurrences in domains

Exploration of a phenomenon:  
What hypotactic chains of modal verbs in what interpretations are possible in German?

Search in the DK

How does one find a hypotactic chain of modals?

A) Restrict problem to: two occurrences in a sentence

```
[tpos="V.*" & (word="(ge)?k[aöo]nn.*" | word="(ge)?w[oi]ll.*" | word="(ge)?d[äü]rf.*" | word="(ge)?soll.*" | word="(ge)?m[üu][sß]s.*" | word="m[a][g].*" | word="(ge)?m[oo][gc].*")] [*]
[tpos="V.*" & (word="(ge)?k[aöo]nn.*" | word="(ge)?w[oi]ll.*" | word="(ge)?d[äü]rf.*" | word="(ge)?soll.*" | word="(ge)?m[üu][sß]s.*" | word="m[a][g].*" | word="(ge)?m[oo][gc].*")]
within s
```

2053 matches

Expressing queries in terms of regular expressions on word forms is error prone and cumbersome  
⇒ lemmatization of corpora very useful

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

11 / 64

# Word occurrences in domains (cont.)

B) Additionally eliminate the following material in-between two occurrences of a modal verb in a sentence:

- ▶ commas, begin/end of direct speech
- ▶ coordinating elements

87 matches (70 actual examples)

(1) Und irgendwann **will** ich auch ein Löschfahrzeug and at one point want I also a fire.truck steuern **können**.

steer be able to

'At one point I want to be able to steer a fire truck.'

(2) Ich **möchte** dies nicht entscheiden **müssen**.

I want this not decide must

'I do not want to have to decide this.'

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

12 / 64

(3) Montags und mittwochs **sollen** sich die Mitarbeiter voll auf die Sachbearbeitung Mondays and Wednesdays shall self the employees fully on the paperwork konzentrieren **können**.

concentrate be.able.to  
'On Mondays and Wednesdays, the employees are supposed to be able to concentrate entirely on their paperwork.'

(4) In Eichstätt **sollen** die Kinder mehr mitbestimmen **dürfen**, etwa in Eichstätt shall the children more decide be.allowed.to for.example beim Straßenbau.  
concerning building.of.streets

'In Eichstätt the children are supposed to be allowed to decide more, for example concerning which streets are built.'

(5) "Wenn wir schon das führende Wirtschaftsland sind, **sollte** kein Deutscher mehr mit if we already the leading economic.force are should no German only with 850 DM nach Hause gehen **müssen**", erklärte der bayerische Landesvorsitzende.  
850 DM to home go must declared the Bavarian party.leader

'If we are one of the leading economic powers, no German should have earned only 850 DM, declared the Bavarian party leader.'

(6) Das sei eine ganz wichtige Funktion des Treffens, erklärte die gastgebende that is a very important function of.the meeting declared the hosting Äbtissin, denn die Frauen **müssen** gegenseitig ihre Erfahrungen austauschen abbot since the women must among.each.other their experiences exchange **können**.

be.able.to

'This is a very important function of the meeting, declared the hosting abbot, since women must be able to exchange their experiences.'

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller 2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

13 / 64

# Topological Fields

Generalization (den Besten and Edmondson 1983):  
Speakers of Middle-Bavarian, South-Bavarian and Franconian use an otherwise inexistent verbal complex order when they "attempt to sound non-dialect like".

(7) daß er singen<sub>3</sub> hat<sub>1</sub> müssen<sub>2</sub>  
that he sing has must  
'that he has had to sing'

(8) damit unser Lager von einer Lawine nicht  
so.that our camp of an avalanche not  
getroffen<sub>4</sub> hätte<sub>1</sub> werden<sub>3</sub> können<sub>2</sub>  
hit had been be.possible  
'so that our camp had not been possible to be hit by an avalanche'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Topological Fields

Checking the generalization with a search in FR  
A sequence of three immediately adjacent verbs outside of the verbal complex is rare, so we try the query:

[tpos = "V.\*"] [tpos = "VFIN"] ([tpos = "V.\*" | ([tpos = "PTKZU"] [tpos = "VINF"])])

189 matches (10 actual examples)

(9) Zu dem Zeitpunkt, an dem ich mich entscheiden<sub>3</sub>  
at the time at which I me decide  
hätte<sub>1</sub> müssen<sub>2</sub>, war das Gesangsbuch  
had have was the hymn.book  
wichtiger.  
more.important

'At the time at which I would have had to decide, the hymn book was more important to me.'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

(10) der Glaube, daß jener Clan, der als nächster Mogadischu kontrolliert,  
the belief that the clan that as next Mogadischu controls  
sich nach dem Vorbild der Marehan von Siad Barre genauso  
self after the model of.the Marehan of Siad Barre equally  
bereichern<sub>3</sub> wird<sub>1</sub> können<sub>2</sub>  
enrich will be.able  
'the belief that the clan which controls Mogadischu next will be able to enrich following the model of Siad Barre'

(11) laut der der Landeszuschuß nicht bei den Betriebskosten  
according to.which the subsidy not for the operating.costs  
berücksichtigt<sub>4</sub> hätte<sub>1</sub> werden<sub>3</sub> sollen<sub>2</sub>  
considered have be should  
'according to which the subsidy should not have been considered for the operating costs'

(12) die Ortskernsanierung in Steinkirchen, die sicher 1993  
the sanitation of Steinkirchen which surely 1993  
abgeschlossen<sub>4</sub> werden<sub>3</sub> hätte<sub>1</sub> können<sub>2</sub>  
completed be have could  
'the sanitation of Steinkirchen, which surely could have been completed by 1993'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Direct reference to topological fields

Leaving out the topological field information results in low precision, e.g., due to

1. other example patterns matching the query:

- ▶ topicalized [V V] followed by finite verb-second
- ▶ finite verb-last followed by extraposed [V V]
- ▶ [V V] followed by extraposed intransitive V
- ▶ special constructions, e.g.,

(13) Doch gelernt ist gelernt.  
but learned is learned  
'Things you really learned, you remember.'

(14) Von der Sowjetunion lernen heißt siegen lernen  
of the Soviet Union learn means win learn  
'To learn from the Soviet Union means to learn how to win.'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Direct reference to topological fields (cont.)

Other topological field notions (*Mittelfeld*, *Nachfeld*) are impossible to translate into words and POS-tags.

⇒ Annotating corpora with topological information is highly useful for linguistically motivated corpus searches.  
(→ structural tags, treebanks)

2. erroneous corpus annotation, e.g., tagging errors

⇒ Annotation tools (taggers, shallow parsers) usually geared towards disambiguation at any cost. For linguistic searches preferable to preserve certain ambiguities.

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Constituents

Exploration of a phenomenon:

Müller (1999) mentions an example with a topicalized constituent consisting of a past participle and an agentive "von [by]"-PP

(15) [Von Grammatikern angeführt] werden auch Fälle  
of grammarians mentioned are also cases  
mit dem Partizip intransitiver Verben.  
with the participle intransitive verbs  
'Grammarians also mention cases with the participle of intransitive verbs'

Is a "[von-PP passive-participle]" constituent generally available with passives?

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields**
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Constituents (cont.)

Search in DK corpus

requires approximation of

- ▶ the structure of a *von*-PP
- ▶ the *Vorfeld* as topological field before the finite verb

```
<s> "Von" [tpos != "VFIN"]* [tpos = "NN"]
  [tpos = "VPP"] [tpos = "VFIN"] within s
```

35 examples

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Constituents

Search Results: Agentive passive (*Vorgangspassiv*)

(16) [Von ihrer 21 Monate alten Enkelin ausgesperrt] wurde by her 21 months old granddaughter lock.out was Montag mittag eine 58jährige Hausfrau aus der Monday noon a 58-year-old housewife from the Mercystraße. Mercystreet

'On Monday at noon, a 58 year old house wife living on Mercystreet was locked out by her 21 month old granddaughter.'

(17) [Von den Bürgern angeregt] wurde, an der Straße in by the townsmen suggested was at the road in Richtung Friedhof eine weitere Straßenlampe anzubringen. direction cemetery a further street-lamp attach.

'It was suggested by the townsfolk to add another street lamp at the road towards the cemetery.'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Constituents

Search Results: Stative passive (*Zustandspassiv*)

(18) [Von den Entwicklungen auf dem Arbeitsmarkt besonders by the developments at the job-market particularly betroffen] sind laut Arbeitsamt Ingolstadt affected are according-to labor-exchange Ingolstadt Männer und ausländische Arbeitnehmer. men and foreign employees

'Labor exchange at Ingolstadt reports that the current development of the work market particularly affected men and foreign workers.'

(19) [Von Baggern umklammert] ist derzeit Riedenburg. by excavators embraced is currently Riedenburg

'Riedenburg is currently embraced by excavators.'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Constituents

Search Results: Other passives

Stative passive embedded under raising verb (or idiomatic)

(20) Von Pech verfolgt scheint in dieser Saison Abwehrspieler By bad.luck followed seems in this season defense.player Dietmar Habermeier zu sein . . . Dieter Habermeier to be

'This season, the defense player Dietmar Habermeier is followed by his bad luck.'

Other passive

(21) [Von einem Unbekannten verfolgt] fühlt sich ein by a person unknown followed feels himself a Imker aus Bad Abbach. bee-keeper from Bad Abbach

'A bee-keeper from Bad Abbach feels followed by a person unknown.'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Grammatical relations

Generalization by Pafel (1995):

"[A]rguments of the noun can be extracted, but modifiers cannot:

(22) \* Mit rotem Einband habe ich ein Buch gelesen. with red cover have I a book read

[. . .] Unextractability of noun modifiers is attested at least for English (Huang 1982:488; Chomsky 1986:80), Italian (Giorgi & Longobardi 1991: 62), and French (Godard 1992: 238)."

Checking the generalization with a search in FR

A) Restrict search to specific prepositions followed by a simple NP structure at the beginning of a sentence, i.e., before a finite verb:

```
<s> "Aus" [tpos="ART"]? []? [tpos="N.*"]
  [tpos="VFIN"]
```

1469 matches

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Grammatical relations

Search Results

(23) Aus dem English Theater stehen zwei Modelle in den Vitrinen. from the English Theater stand two models in the display.cases 'Two models from the English Theater are shown in the display cases.'

(24) Aus dem 17. Jahrhundert erklangen in dynamisch differenziertem Spiel und mit weich gestaltendem Ansatz Tanzsätze von Johann play and with soft shaped lipping dances by Johann Christoph Pezelius und Michael Praetorius Christoph Pezelius and Michael Praetorius 'Dances from the 17th century by Johann Christoph Pezelius and Michael Praetorius were played.'

(25) Aus der A-Jugend stoßen Jens Schneider, Thomas Gölzenleuchter From the A-youth come Jens Schneider, Thomas Gölzenleuchter und Achim Nau zu den Aktiven and Achim Nau to the actives 'Jens Schneider, Thomas Gölzenleuchter and Achim Nau from the A-youth join the grown-up team'

- Corpus Linguistics
- Syntactic Searching
- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents**
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Direct reference to grammatical relations

- B) Use a treebank with special query tools:
- ▶ VERBMOBIL treebanks (Tübingen, OSU)
  - ▶ NEGRA treebanks & search tools (Saarbrücken)
- Tree description language to specify linear order, dominance, grammatical functions, ...

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

26 / 64

# Direct reference to grammatical relations

Returning to our example

Search a PP modifier, a finite verb to the right of it and a NP modified by the PP to the right of the verb, so that the PP is only dominated by constituents which also dominate the finite verb.

⇒ Also finds examples with richer internal constituent structure, e.g., coordinated NPs

(26) In Cockpit und Kabine wurden neue Gehaltsstrukturen in cockpit and cabin were new salary structures mit "marktkonformen" Anfangsgehältern vereinbart. with market adequate starting.salaries agreed on 'New salary structures in cockpit and cabin with starting salaries in line with real marked conditions were agreed on.'

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

27 / 64

# Searching in tree structures

- A problem for our first example (of "werden"):
- ▶ too many examples of adverbial use
    - ▶ STTS does not distinguish between adverbs and predicative adjectives
  - ▶ But this information is present in syntactic annotation:
    - ▶ predicative adjectives project to an adjectival phrase with grammatical function "PRED" (predicate)

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

**Searching w/ syntax**  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

28 / 64

# TIGERSearch

graphical search interface:

- ▶ allows to search in two-dimensional tree structures
- ▶ 2 forms of query design:
  - ▶ text mode (logical query language):  
#n1: [cat="SIMPX"] >\* [word=("werden"|"wird")] & #n1 >\* [pos="ADJD"]
  - ▶ graphical mode (construct partial trees)
- ▶ searchable relations: direct dominance, dominance (transitive closure), direct linear precedence, linear precedence (transitive closure)
- ▶ limited negations

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

**Searching w/ syntax**  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

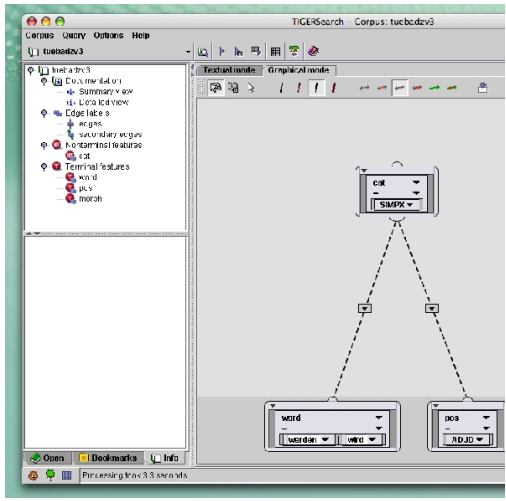
Summary

LSE  
Comparative correlatives

References

29 / 64

# TIGERSearch graphical interface



Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

**Searching w/ syntax**  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

30 / 64

# Searching in tree structures

As with POS annotation, you need to know how phenomena are annotated

- ▶ Goal: search for forms of "werden" + adjectival phrase with grammatical function "PRED" (predicate)
  - ▶ Result: 504 sentences
  - ▶ Inspection of the first 20 sentences: 10 correct sentences, 9 with 2 clauses, 1 passive sentence
- ▶ Restrict search to both occurrences in the same clause
  - ▶ Result: 244 sentences
  - ▶ Inspection of the first 20 sentences: 18 correct, 2 passives (verb complex not restricted to only head)

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents

**Grammatical relations**

**Searching w/ syntax**  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

31 / 64



## Refining the Treebank Query

The query `[pos="PTKVZ"] . [pos=finite]` wrongly matches relative and interrogative clauses such as:

- (28) dem Anfang der neunziger Jahre Hohn und Spott  
 which beginning of nineties years mock and sneer  
**zuteil wurde**  
 PART become  
 'which were mocked at the beginning of the nineties'

A query with more explicit reference to specifics of the syntactic annotation is required to exclude such examples.

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

38 / 64

## A More Explicit Treebank Query

```
#s:[cat="S"] > #part:[pos="PTKVZ"] &
#part . [pos=finite] &
#s >@! #leftcorner &
#leftcorner:[pos= ! (prorel | printer | conjunction)]
```

1. Search for a sentence node (#s) that dominates a particle (#part)
2. which is adjacent to a finite verb.
3. Search for the left edge of the #s.
4. The part-of-speech of the left edge may not be a relative pronoun, interrogative pronoun or conjunction.

Caveat: The more elaborate a query, the stronger its dependence on the specifics and quality of the syntactic annotation.

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

39 / 64

## Extrapolation and Subjacency: The Issue

- (29) `[NP Many books [PP with [stories t]] t']` were sold  
 [that I wanted to read].

- (30) `[NP Many proofs [PP of [the theorem t]] t']` appeared  
 [that I wanted to think about].

Baltin (1981) and Chomsky (1986, p. 40):

- Relative clause cannot be related to *t*, since Subjacency excludes crossing of more than one barrier.

This view is very common, including recent textbooks (Haegeman 1994, p. 422, Klenk 2003, p. 96, Baltin To Appear).

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

40 / 64

## Extrapolation is a Non-Local Dependency Adjuncts

Extrapolation may cross arbitrarily many NP boundaries (cf. Müller 1999, 2004):

- (31) Karl hat mir [ein Bild [einer Frau -i]] gegeben, [die schon Karl has me a picture of a woman given who already lange tot ist].  
 long dead is

- (32) Karl hat mir [eine Fälschung [des Bildes [einer Frau -i]]] gegeben, [die schon lange tot ist].  
 given who already long dead is

- (33) Karl hat mir [eine Kopie [einer Fälschung [des Bildes [einer Karl has me a copy of a forgery of a picture of a Frau -i]]]] gegeben, [die schon lange tot ist].  
 woman given who already long dead is  
 'Karl gave me a copy of a forgery of the picture of a woman who has been dead for a long time.'

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

41 / 64

## Extrapolation is a Non-Local Dependency Complements

- Some proposals assume that extraposed adjuncts, such as those in the previous example, are base-generated and that coreference/coindexation is established via special mechanisms (Kiss 2005).
- But it is also possible to extrapose sentential complements:

- (34) Ich habe [von [der Vermutung -i]] gehört, [daß es I have of the assumption heard that it Zahlen gibt, die die folgenden Bedingungen numbers exist which the following conditions erfüllen].  
 satisfy

and such extraposition is equally non-local.

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

42 / 64

## Extrapolation is a Non-Local Dependency Complements

- (35) Ich habe [von [einem Beweis [der Vermutung -i]]] gehört, I have of the proof of the assumption heard [daß es Zahlen gibt, die die folgenden Bedingungen that it numbers exist which the following conditions erfüllen].  
 satisfy

- (36) Ich habe [von [dem Versuch [eines Beweises [der I have of the attempt of a proof of the Vermutung -i]]]] gehört, [daß es Zahlen gibt, die die assumption heard that it numbers gives that the folgenden Bedingungen erfüllen].  
 following conditions satisfy

'I have heard of the attempt to prove the assumption that there are numbers for which the following conditions hold.'

Selectional restrictions of the matrix head have to be ensured so that one cannot avoid establishing a relation between the governing noun and the extraposed element

Corpus Linguistics  
 Syntactic Searching

Introduction

Searching w/o  
 syntax

Basic searching

Meurers 2005

Word occurrences in domains

Topological Fields

Constituents

Grammatical relations

Searching w/ syntax

TIGERSearch

Meurers & Müller

2007

Fronted Particles

Extrapolation and Subjacency

Multiple Fronting

Summary

LSE

Comparative correlatives

References

43 / 64

## Can We Find Such Examples in a Corpus?

- ▶ A basic query using an unannotated corpus:  
Search for nouns that take sentential complements + *daß* (or use a reverse dictionary to search for *ung*, i.e., *gnu*).
  - ▶ Reasoning:
    - ▶ We know many verbs that govern an object clause.
    - ▶ Heavy clauses tend to be extraposed, so looking for nominalizations that actually appear with the complement clause should provide some results.
  - ▶ Pro: Query doesn't require syntactically annotated corpus.
  - ▶ Con: False positives, requiring extensive manual filtering.
- ▶ Can we improve on this with a query based on a treebank?

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

44 / 64

## A Query Based on the TIGER Treebank

```
#xp:[cat="NP"] >OC [ ] &
[cat="NP|PP"] > #xp &
discontinuous(#xp)
```

1. Search for an NP node (#xp), that immediately dominates an object clause (OC).
2. #xp is immediately dominated by a NP or PP node.
3. #xp is discontinuous, that is, the object clause is usually extraposed.

We get 12 hits in the 2003 version of the Tiger corpus (40018 sentences).

- ▶ We get this result in 1.1 seconds + time for typing the query in comparison to several hours that are needed for the manual approach.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

45 / 64

## Multiple Fronting: The Issue

- ▶ German is a V2 Language: Usually only one constituent can be placed in front of the finite verb.
 

(37) a. Maria stellt Max Peter vor.  
Maria introduces Max Peter PART  
'Maria introduces Max to Peter.'

b. \* Maria Max stellt Peter vor.  
Maria Max introduces Peter PART
- ▶ Sometimes this property of German is used as a constituent test: Those elements that can be fronted together form a constituent.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

46 / 64

However, sentences with multiple fronted constituents exist (Müller 2003):

- (38) [Trocken] [durch die Stadt] kommt man am  
dry through the city comes one on  
Wochenende auch mit der BVG. (taz berlin, 10.07.1998)  
weekends also with the BVG
- (39) [Gezielt] [Mitglieder] [im Seniorenbereich] wollen die  
targeted members in.the senior.group wants the  
Kendoka allerdings nicht werben. (taz, 07.07.1999)  
Kendoka however not advertise
- (40) [Wenig] [mit Sprachgeschichte] hat der dritte Beitrag in  
little with language.history has the third article in  
dieser Rubrik zu tun,  
this group to do  
(Zeitschrift für Dialektologie und Linguistik, 3/2002)

This data can be analyzed with an empty verbal head in the *Vorfeld*, which preserves the V2 property (Müller 2005), but nevertheless such data is problematic for constituency tests.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

47 / 64

## Can We Find Such Examples in a Corpus?

```
#s:[cat="S"] >HD #fin:[pos=finit] &
#s >@l #sleftedge &
#s > #vf1 & #vf1 >@l #sleftedge &
#s > #vf2 & #vf1 . #vf2 & #vf2 . #fin
```

1. Search for a sentential node that immediately dominates the finite Verb.
2. #s has a left edge #sleftedge.
3. #s immediately dominates #vf1
4. and #vf1 has the same left edge #sleftedge, that is, #vf1 is the left-most node under #s.
5. #s immediately dominates another node #vf2.
6. #vf1 borders #vf2
7. and #vf2 borders the finite verb.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

48 / 64

## False Positives

- ▶ The query unintentionally returns relative and interrogative clauses and parts of coordinations.
- ▶ It also finds examples with adverbs such as *indes*, *jedoch*, *wiederum*:

(41) Hier wiederum mangelt es an  
here again lacks it of  
Opferbereitschaft.  
readiness to make sacrifice

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subadjacency

Multiple Fronting

Summary

LSE  
Comparative correlatives

References

49 / 64

# Making the Query More Precise

- ▶ Since the class of such adverbs is small, one can rule them out explicitly by listing them:

```
#vf2:[! (word= ("aber" | "also" | "auch" | "allerdings" |
"dagegen" | "freilich" | "hingegen" |
"jedenfalls" | "jedoch" | "nämlich" |
"schließlich" | "wiederum"))]
```

- ▶ Once we also exclude the relative clauses and coordination sentences, we get: nothing!

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Summary (so far)

- ▶ Electronic corpora can be used to search for examples of linguistically relevant phenomena in order to
  - ▶ verify generalizations or
  - ▶ obtain a wide empirical basis exemplifying the phenomenon.
- ▶ Corpus data are attractive since they
  - ▶ exhibit a wide variation of known and unknown parameters and
  - ▶ come with a context.
- ▶ The linguistic terminology used to single out the relevant phenomenon needs to be reconstructed in terms of the annotation of the corpus.
- ▶ Danger of selective results due to the translation of the characterization of the phenomenon to the query language.

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Summary (cont.)

- ▶ Depending on the task, the following levels of annotations are needed:
  - ▶ basic: word forms, part of speech tags, lemmata, simple structural tags (e.g., sentence boundaries)
  - ▶ topological fields and constituents: special structural tags or treebanks
  - ▶ grammatical relations: treebanks
- ▶ While many corpus annotations (lemmata, part of speech tags, sentence segmentation, shallow parsing chunks) can be obtained automatically, different from standard use, annotation tools for linguistic purposes should
  - ▶ allow for ambiguity preserving annotation (for ambiguities which cannot be resolved with high accuracy by the efficient algorithms),
  - ▶ possibly followed by more costly algorithms for ambiguity resolution.

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Linguist's Search Engine (LSE)

Resnik and Elkiss (2005)

## Motivation:

- ▶ Linguistic judgments are imperfect
- ▶ A need for new data sources, on the scale of the web (to deal with Zipf's Law)
  - ▶ But web searches do not allow for all the necessary linguistic detail
  - ▶ e.g., seeking instances of a particular verb used only intransitively

## The Linguist's Search Engine (LSE)

(<http://lse.umiaccs.umd.edu:8080/>) addresses these concerns by:

- ▶ providing a natural interface for linguists to use
- ▶ which allows for syntactic searching over the web

See the paper for some description of the system architecture

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# A motivating example

Resnik et al. (2005)

What if we search for implicit object uses for rare verbs (e.g., *titrate*)?

(42) ?They titrated

- ▶ We need linguistic annotation to know that this is an intransitive verb.
- ▶ But we need the scale of the web to find a range of uses of a verb

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Query by example

For the case of *titrate*, we can enter *John titrates the solution* into the query editor

- ▶ Then, click on "parse"
- ▶ Then, edit the tree (see next slide)
- ▶ And finally update the query to search
  - ▶ e.g., (S1 (S NP (VP (VBZ titrates) (! NP))))

Note that you don't have to agree (or completely understand) the annotation, so long as it's the same as what you're looking for

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary
- LSE
- Comparative correlatives
- References

# Tree Editor

You can right click on a node and do any of the following:

- ▶ Edit Node
- ▶ Undo
- ▶ Revert
- ▶ Add Node
- ▶ Remove Nodes (e.g., the children under NP)
- ▶ Change Relation (e.g., specify that the VP does not immediately dominate the NP)

Then, click on "update query" to change the query before searching

See [http://lse.umiacs.umd.edu:8080/lseuser/Tree\\_Editor.html](http://lse.umiacs.umd.edu:8080/lseuser/Tree_Editor.html)

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

# LSE interface concepts

- ▶ Minimize learning/ramp-up time
- ▶ Have a linguist-friendly look & feel
- ▶ Permit rapid interaction
- ▶ Permit large-scale searches
- ▶ Allow searches using linguistic criteria

These concepts come in conflict (e.g., searching linguistic criteria vs. minimizing ramp-up time)

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

# Search features

You have different options for searching, allowing one to:

- ▶ download results in a KWIC format
- ▶ negate relationships
- ▶ match nodes based on WordNet relations
- ▶ restrict to particular level of difficulty
- ▶ filter offensive content
- ▶ use specified collections of text

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

# Build custom collection

You can build your own collection of pages to search through, by having LSE:

- ▶ Query Altavista to find pages containing, e.g., any form of *titrate*
- ▶ Extract only sentences containing this word
- ▶ Annotate & index those sentences
- ▶ Augment the collection by iterating with different specifications

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

# Implicit object uses

Resnik et al. (2005)

Returning to the implicit object case before, the LSE reveals examples like the following from a custom collection:

- (43) a. You titrate with sulfuric acid using gloves and safety glasses.  
 b. We titrated till we reached the final endpoint of pH 4.5.  
 c. Conversely, if we titrate in the opposite direction...  
 d. I no longer titrate, but that is because...

There are also non-implicit object uses (needing linguistic judgment!)

- (44) a. Note that the detected response titrates linearly.  
 b. The catheter is inserted and the optimal dose titrated.  
 c. Although sulfite usually titrates as thiosulfate...

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

# Comparative correlatives

Resnik et al. (2005)

Much debate in the literature in how to handle comparative correlatives, e.g.:

- (45) The more pizza Romeo eats, the fatter he gets
- An important issue is when the main copular verb can be deleted, e.g.,:
- (46) The better an advisor (is), the more successful the student (is)
- (47) The more obnoxious Fred is/ $\emptyset$ , the less attention you should pay attention to him.

Corpus Linguistics  
Syntactic Searching

- Introduction
- Searching w/o syntax
- Basic searching
- Meurers 2005
- Word occurrences in domains
- Topological Fields
- Constituents
- Grammatical relations
- Searching w/ syntax
- TIGERSearch
- Meurers & Müller 2007
- Fronted Particles
- Extrapolation and Subjacency
- Multiple Fronting
- Summary

**LSE**  
Comparative correlatives  
References

## Comparative correlatives (cont.)

Previously theorized that copula deletion occurs with generic subjects

- ▶ LSE data showed that deletion occurs in the first clause when it occurs in both clauses *in all cases*

There also appears to be a factor of phonological weight

(48) a. \*The more obnoxious Fred, the less attention you should pay to him

b. ?The more obnoxious a child, the less attention you should pay to him

c. The more obnoxious Fred's younger brother, the less attention you should pay to him

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

63 / 64

## Comparative correlatives (cont.)

LSE also allowed for an investigation into the use of *then* in the second clause

(49) a. The darker the coffee bean, then the less caffeine

b. The more playing time in the past then the less regression to the mean needed

c. The more that you can arrange for everyday life to happen almost automatically then the less you have to concentrate on what is happening

- ▶ Corroborates the idea that these are conditionals

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

64 / 64

## References

Baltin, Mark (1981). Strict Bounding. In Carl Lee Baker and John J. McCarthy (eds.), *The Logical Problem of Language Acquisition*, Cambridge: Massachusetts, London: England: The MIT Press.

Baltin, Mark (To Appear). Extraposition, the Right Roof Constraint, Result Clauses, Relative Clause Extraposition, and PP Extraposition. Draft of 23.09.2001. In *The Syntax Companion* (<http://www.uilots.let.uu.nl/syncom/>), Utrecht Institute of Linguistics.

Bennis, Hans (1991). Theoretische Aspekten van Partikelvooropplaatsing. *TABU Bulletin voor Taalwetenschap* 21(3), 89–95.

Bierwisch, Manfred (1963). *Grammatik des deutschen Verbs*. studia grammatica 11. Berlin: Akademie Verlag.

Brants, Sabine, Stefanie Dipper et al. (2004). TIGER: Linguistic Interpretation of a German Corpus. *Research on Language and Computation* 2(4), 597–620.

B'uring, Daniel and Katharina Hartmann (2001). The Syntax and Semantics of Focus-Sensitive Particles in German. *Natural Language and Linguistic Theory* 19(2), 229–281. <http://www.linguistics.ucla.edu/people/buring/>.

Chomsky, Noam (1973). Conditions on Transformations. In Stephen R. Anderson and Paul Kiparski (eds.), *A Festschrift for Morris Halle*, New York: Holt, Rinehart & Winston, pp. 232–286.

Chomsky, Noam (1986). *Barriers*, vol. 13 of *Linguistic Inquiry Monographs*. Cambridge: Massachusetts, London: England: The MIT Press.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

64 / 64

Chomsky, Noam (1993). *Lectures on Government and Binding – The Pisa Lectures*. No. 9 in *Studies in Generative Grammar*. Berlin, New York: Mouton de Gruyter, 7th edn.

Cowart, Wayne (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.

den Besten, Hans and Jerold A. Edmondson (1983). The Verbal Complex in Continental West Germanic. In Werner Abraham (ed.), *On the Formal Syntax of the Westgermania*, Amsterdam/Philadelphia, pp. 155–216. (Linguistik Aktuell 3).

Duden (2005). *Grammatik der deutschen Gegenwartssprache*, vol. 4. Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 7th edn.

D'urscheid, Christa (1989). *Zur Vorfeldbesetzung in deutschen Verbzweit-Strukturen*. No. 1 in FOKUS. Trier: Wissenschaftlicher Verlag.

Eisenberg, Peter (1994). *Grundriß der deutschen Grammatik*. Stuttgart, Weimar: Verlag J. B. Metzler, third edn.

Eisenberg, Peter (1999). *Grundriß der deutschen Grammatik*, vol. 2. Der Satz. Stuttgart, Weimar: Verlag J. B. Metzler.

Engel, Ulrich (1977). *Syntax der deutschen Gegenwartssprache*, vol. 22 of *Grundlagen der Germanistik*. Berlin: Erich Schmidt Verlag.

Engel, Ulrich (1994). *Syntax der deutschen Gegenwartssprache*, vol. 22 of *Grundlagen der Germanistik*. Berlin: Erich Schmidt Verlag. 3., völlig überarbeitete Auflage.

Erdmann, Oskar (1886). *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung*, vol. 1. Stuttgart: Verlag der J. G. Cotta'schen Buchhandlung. Reprint: Hildesheim: Georg Olms Verlag, 1985.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

64 / 64

Fanselow, Gisbert (1987). *Konfigurationsalität*. No. 29 in *Studien zur deutschen Grammatik*. Tübingen: Gunter Narr Verlag.

Fanselow, Gisbert (1991). Minimale Syntax. *Groninger Arbeiten zur Germanistischen Linguistik* 32.

Fanselow, Gisbert (1993). Die R'uckkehr der Basisgenerierer. *Groninger Arbeiten zur Germanistischen Linguistik* 36, 1–74.

Grewendorf, Günther (1988). *Aspekte der deutschen Syntax. Eine Rektions-="Bindungs"-Analyse*. No. 33 in *Studien zur deutschen Grammatik*. Tübingen: Gunter Narr Verlag.

Grewendorf, Günther (1990). Verb-Bewegung und Negation im Deutschen. *Groninger Arbeiten zur Germanistischen Linguistik* 30, 57–125.

Grubačić, Emilija (1965). Untersuchungen zur Frage der Wortstellung in der deutschen Prosadichtung der letzten Jahrzehnte. Ph.D. thesis, Philosophische Fakultät, Zagreb.

Haegeman, Liliane (1994). *Introduction to Government and Binding Theory*. No. 1 in *Blackwell Textbooks in Linguistics*. Oxford, UK/Cambridge, USA: Blackwell Publishers, second edn.

Haider, Hubert (1982). Abhängigkeiten und Konfigurationen: Zur deutschen V"-Projektion. *Groninger Arbeiten zur Germanistischen Linguistik* 21, 1–60.

Haider, Hubert (1990). Topicalization and other Puzzles of German Syntax. In Günther Grewendorf and Wolfgang Sternefeld (eds.), *Scrambling and Barriers*, Amsterdam, Philadelphia: John Benjamins Publishing Company, no. 5 in *Linguistik Aktuell / Linguistics Today*, pp. 93–112.

Haider, Hubert (1991). *Fakultativ kohärente Infinitivkonstruktionen im Deutschen*. Arbeitspapiere des SFB 340 No. 17, IBM Deutschland GmbH, Heidelberg.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

64 / 64

Haider, Hubert (1993). *Deutsche Syntax – generativ. Vorstudien zur Theorie einer projektiven Grammatik*. No. 325 in *Tübinger Beiträge zur Linguistik*. Tübingen: Gunter Narr Verlag.

Haider, Hubert (1996). Downright Down to the Right. In Uli Lutz and Jürgen Pafel (eds.), *On Extraction and Extraposition in German*, Amsterdam: Benjamins, no. 11 in *Linguistik Aktuell / Linguistics Today*, pp. 245–271.

Haider, Hubert (1997a). Precedence among Predicates. *The Journal of Comparative Germanic Linguistics* 1, 2–41.

Haider, Hubert (1997b). Projective Economy. On the Minimal Functional Structure of the German Clause. In Werner Abraham and Ely van Gelderen (eds.), *German: Syntactic Problems—Problematic Syntax*, Tübingen: Max Niemeyer Verlag, no. 374 in *Linguistische Arbeiten*, pp. 83–103.

Haider, Hubert, Susan Olsen and Sten Vikner (1995). Introduction. In Hubert Haider, Susan Olsen and Sten Vikner (eds.), *Studies in Comparative Germanic Syntax*, Dordrecht, Boston, London: Kluwer Academic Publishers, vol. 31 of *Studies in Natural Language and Linguistic Theory*, pp. 1–45.

Hoberg, Ursula (1997). Die Linearstruktur des Satzes. In Hans-Werner Eroms, Gerhard Stickel and Gisela Zifonun (eds.), *Grammatik der deutschen Sprache*, Berlin, New York: Walter de Gruyter, vol. 7.2 of *Schriften des Instituts für deutsche Sprache*, pp. 1495–1680.

Hoeksema, Jack (1991). Theoretische Aspekten van Partikelvooropplaatsing. *TABU Bulletin voor Taalwetenschap* 21(1), 18–26.

Jacobs, Joachim (1986). The Syntax of Focus and Adverbials in German. In Werner Abraham and S. de Meij (eds.), *Topic, Focus, and Configurationality. Papers from the 6th Groningen Grammar Talks, Groningen, 1984*, Amsterdam, Philadelphia: John Benjamins Publishing Company, no. 4 in *Linguistik Aktuell / Linguistics Today*, pp. 103–127.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o  
syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller  
2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE

Comparative correlatives

References

64 / 64

Jacobson, Pauline (1987). Phrase Structure, Grammatical Relations, and Discontinuous Constituents. In Geoffrey J. Huck and Almerindo E. Ojeda (eds.), *Discontinuous Constituency*, New York: Academic Press, vol. 20 of *Syntax and Semantics*, pp. 27–69.

Kathol, Andreas (1996). Discontinuous Lexical Entries. In Philippe Blache (ed.), *Third International Conference on HPSG—Abstracts. 22-24 May 1996*. Marseille, pp. 43–44.

Kefer, Michel (1989). *Satzgliedstellung und Satzstruktur im Deutschen*. Studien zur deutschen Grammatik 36. Tübingen: Narr. (= Phil. Diss., Lüttich, Belgium, 1983).

Kiss, Tibor (1994). Obligatory Coherence: The Structure of German Modal Verb Constructions. In John Nerbonne, Klaus Netter and Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, Stanford: CSLI Publications, no. 46 in CSLI Lecture Notes, pp. 71–108.

Kiss, Tibor (2005). Semantic Constraints on Relative Clause Extraposition. *Natural Language and Linguistic Theory* 23(2), 281–334.

Klenk, Ursula (2003). *Generative Syntax*. Narr Studienbücher. Tübingen: Gunter Narr Verlag.

Lötscher, Andreas (1985). Syntaktische Bedingungen der Topikalisierung. *Deutsche Sprache* 13, 207–229.

Lüdeling, Anke (1997). Strange Resultatives in German: New Evidence for a Semantic Treatment. In Ralph C. Blight and Michelle J. Moosally (eds.), *Texas Linguistic Forum 38: The Syntax and Semantics of Predication. Proceedings of the 1997 Texas Linguistics Society Conference*. Austin, Texas: University of Texas Department of Linguistics, pp. 223–233.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

64 / 64

Meurers, Walt Detmar (1999). Lexical Generalizations in the Syntax of German Non-Finite Constructions. Phil. diss., Universität Tübingen, Tübingen. Available as No. 145 in Arbeitspapiere des SFB 340, 2000.

Müller, Stefan (1999). *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. No. 394 in Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag.

Müller, Stefan (1999). *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. No. 394 in Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag. <http://www.cl.uni-bremen.de/~stefan/Pub/hpsg.html>. 06.04.2009.

Müller, Stefan (2003). Mehrfache Vorfeldbesetzung. *Deutsche Sprache* 31(1), 29–62. <http://www.cl.uni-bremen.de/~stefan/Pub/mehr-vf-ds.html>. 06.04.2009.

Müller, Stefan (2004). Complex NPs, Subjacency, and Extraposition. *Snippets* 8, 10–11. <http://www.cl.uni-bremen.de/~stefan/Pub/subjazenz.html>. 06.04.2009.

Müller, Stefan (2005). Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung. *Linguistische Berichte* 203, 297–330. <http://www.cl.uni-bremen.de/~stefan/Pub/mehr-vf-lb.html>. 06.04.2009.

Neeleman, Ad and Fred Weermann (1993). The Balance between Syntax and Morphology: Dutch Particles and Resultatives. *Natural Language and Linguistic Theory* 11, 433–475.

Olsen, Susan (1997a). Pr<sup>a</sup>adikative Argumente syntaktischer und lexikalischer K<sup>o</sup>pfe: Zum Status der Partikelverben im Deutschen und Englischen. *Folia Linguistica* 31(3–4), 301–329.

Olsen, Susan (1997b). Zur Kategorie Verbpartikel. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 119, 1–32.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

64 / 64

Pafel, Jürgen (1995). Kinds of Extraction from Noun Phrases. In Uli Lutz and Jürgen Pafel (eds.), *On Extraction and Extraposition in German*, Amsterdam/Philadelphia: Benjamins, vol. 2 of *Linguistik aktuell*.

Paul, Hermann (1919). *Deutsche Grammatik. Teil IV: Syntax*, vol. 3. Halle an der Saale: Max Niemeyer Verlag. 2nd unchanged edition 1968, Tübingen: Max Niemeyer Verlag.

Reis, Marga (1976). Reflexivierungen in deutschen A.c.I.–Konstruktionen. Ein transformationsgrammatisches Dilemma. *Papiere zur Linguistik* 9, 5–82.

Resnik, Philip and Aaron Elkiss (2005). The Linguist's Search Engine: An Overview. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*. Ann Arbor, Michigan: Association for Computational Linguistics, pp. 33–36. <http://www.aclweb.org/anthology/P05-3009>.

Resnik, Philip, Aaron Elkiss, Ellen Lau and Heather Taylor (2005). The Web in Theoretical Linguistics Research: Two Case Studies Using the Linguist's Search Engine. In *31st Meeting of the Berkeley Linguistics Society*. pp. 265–276. <http://umiacs.umd.edu/~resnik/pubs/bls2005.pdf>.

Rohrer, Christian (1996). Fakultativ kohärente Infinitivkonstruktionen im Deutschen und deren Behandlung in der Lexikalisch Funktionalen Grammatik. In Gisela Harras and Manfred Bierwisch (eds.), *Wenn die Semantik arbeitet. Klaus Baumgärtner zum 65. Geburtstag*, Tübingen: Max Niemeyer Verlag, pp. 89–108.

Ross, John Robert (1967). Constraints on Variables in Syntax. Ph.D. thesis, MIT. Reproduced by the Indiana University Linguistics Club.

Schütze, Carson T. (1996). *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago: Univ. of Chicago Press.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

64 / 64

Stiebels, Barbara (1996). *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag verbaler Pr<sup>a</sup>fixe und Partikeln*. studia grammatica XXXIX. Berlin: Akademie Verlag.

Stiebels, Barbara and Dieter Wunderlich (1992). *A Lexical Account of Complex Verbs*. Arbeiten des SFB 282 No. 30, Seminar f<sup>u</sup>r Allgemeine Sprachwissenschaft. Universität at D<sup>u</sup>sseldorf.

Suchsland, Peter (1994). "Äußere" und "innere" Aspekte von Infinitivbetonungen im Deutschen. In Anita Steube and Gerhild Zybato (eds.), *Zur Satzwertigkeit von Infinitiven und Small clauses*, Tübingen: Max Niemeyer Verlag, no. 315 in Linguistische Arbeiten.

Uszkoreit, Hans, Thorsten Brants, Denys Duchier, Brigitte Krenn, Lars Konieczny, Stephan Oepen and Wojciech Skut (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextrapolation im Deutschen. *Kognitionswissenschaft. Themenheft Ressourcenbeschränkungen* 7(3), 129–133.

van de Velde, Marc (1978). Zur mehrfachen Vorfeldbesetzung im Deutschen. In Maria-Elisabeth Conte, Anna Giacalone Ramat and Paolo Ramat (eds.), *Wortstellung und Bedeutung: Akten des 12. Linguistischen Kolloquiums, Pavia 1977*, Tübingen: Max Niemeyer Verlag, no. 61 in Linguistische Arbeiten, pp. 131–141.

von Stechow, Arnim and Wolfgang Sternefeld (1988). *Bausteine syntaktischen Wissens. Ein Lehrbuch der Generativen Grammatik*. Opladen/Wiesbaden: Westdeutscher Verlag.

Webelhuth, Gert and Farrell Ackerman (1999). A Lexical-Functional Analysis of Predicate Topicalization in German. *American Journal of Germanic Linguistics and Literatures* 11(1), 1–61.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

64 / 64

Wunderlich, Dieter (1984). Zur Syntax der Pr<sup>a</sup>ositionalphrase im Deutschen. *Zeitschrift f<sup>u</sup>r Sprachwissenschaft* 3(1), 65–99.

Zifonun, Gisela (1999). Wenn mit alleine im Mittelfeld erscheint: Verbpartikeln und ihre Doppelgänger im Deutschen und Englischen. In Heide Wegener (ed.), *Deutsch kontrastiv. Typologisch=vergleichende Untersuchungen zur deutschen Grammatik*, Tübingen: Stauffenburg Verlag, pp. 211–234.

Corpus Linguistics  
Syntactic Searching

Introduction

Searching w/o syntax  
Basic searching

Meurers 2005  
Word occurrences in domains  
Topological Fields  
Constituents  
Grammatical relations

Searching w/ syntax  
TIGERSearch

Meurers & Müller 2007  
Fronted Particles  
Extrapolation and Subjacency  
Multiple Fronting

Summary

LSE  
Comparative correlatives

References

64 / 64