Tree Adjoining Grammars

XTAG-Analyses of Syntactic Phenomena

Laura Kallmeyer & Timm Lichte

HHU Düsseldorf

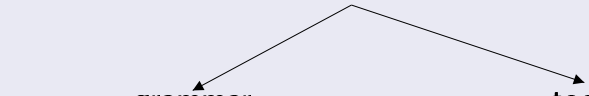
WS 2012

24.10.2012

Outline

The XTAG-project

... was located at the University of Pennsylvania (ca. 1988-2001)

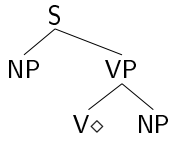


URL: `http://www.cis.upenn.edu/~xtag/`
Manual: [XTAG Research Group, 2001]

The architecture of the XTAG-grammar

Morph Database	inflected form → root form, POS, inflectional information
Syntactic Database	root form, POS → list of tree templates or tree families, list of feature equations
Tree Database	list of tree templates and tree families

Example: **Tree template** for the declarative transitive verb (αnx0Vnx1), where $\diamond$ marks the lexical insertion site:



The architecture of the XTAG-grammar

Morph Database	inflected form → root form, POS, inflectional information
Syntactic Database	root form, POS → list of tree templates or tree families, list of feature equations
Tree Database	list of tree templates and tree families

A tree family

- is a set of tree templates,
- represents a subcategorization frame, and
- unifies all syntactic configurations the subcategorization frame can be realized in.

Example: αnx0Vnx1 $\in$ Tnx0Vnx1

The architecture of the XTAG-grammar - Counts

subcategorization frame	# tree fam.	# tree temp.
intransitive	1	12
transitive	1	39
adjectival complement	1	11
ditransitive	1	46
prepositional complement	4	182
verb particle constructions	3	100
light verb constructions	2	53
sentential complement (full verb)	3	75
sentential subject (full verb)	4	14
idioms (full verb)	8	156
small clauses/predicative	20	187
equational 'be'	1	2
ergative	1	12
resultatives	4	101
it clefts	3	18
total	57	1008

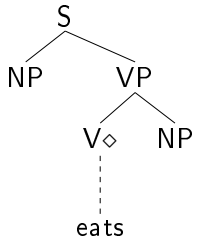
(from [Prolo, 2002])

Lexical insertion

Lexical insertion

Drawing an edge between the lexical anchor and the lexical insertion site

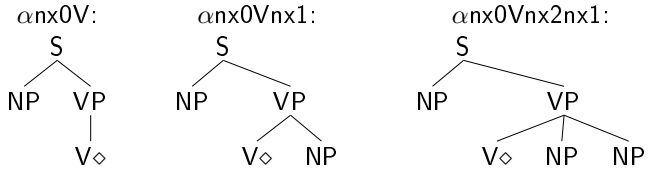
- prior to substitution and adjunction
- The feature structures of the **lexical anchor** and the **insertion site** unify.



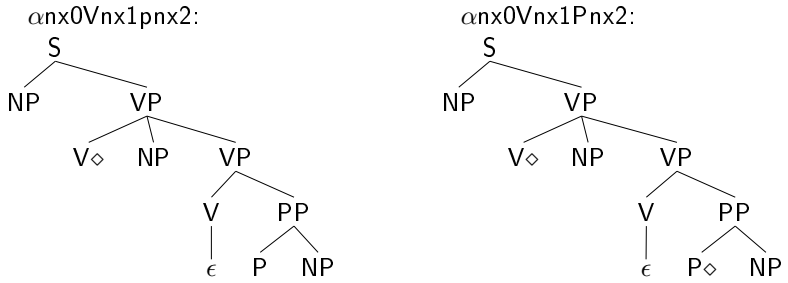
Outline

Complementation with NPs and PPs: The base cases

Complementation with NPs:



Complementation with PPs: substitution or co-anchor



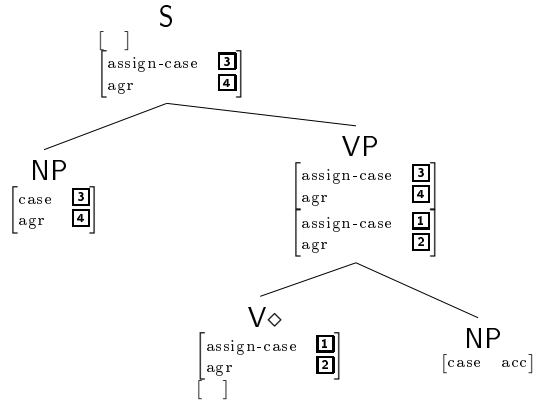
Case assignment and subject-verb agreement

Two modes of case assignment in tree templates:

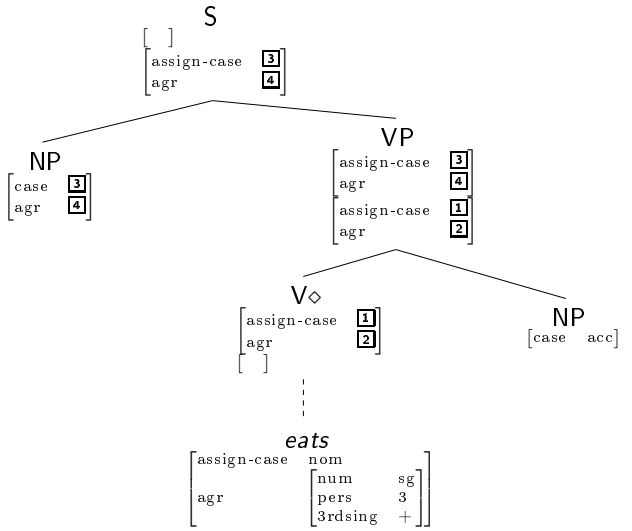
- Direct case assignment with case
- Indirect case assignment with assign-case

  ⇒ by the lexical anchor (during lexical insertion) or by adjoining trees

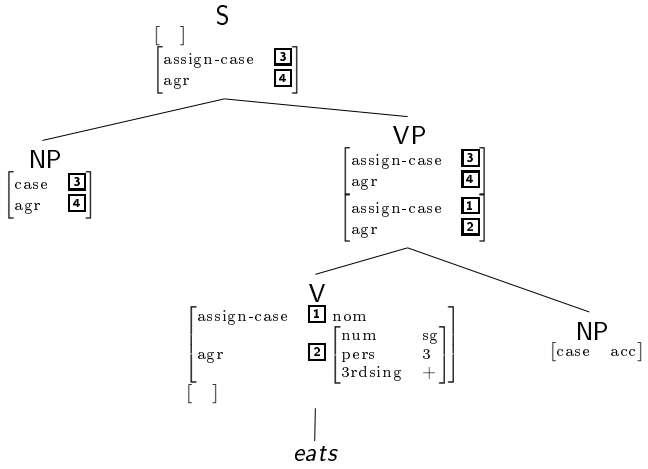
αnx0Vnx1:



Case assignment and subject-verb agreement



Case assignment and subject-verb agreement



Sentential complement structures

In XTAG, a distinction is drawn between sentential complements with **(i) finite verbs**, sentential complements with **(ii) to-infinitives**, and **(iii) small clauses**.

(1) a. Kim said [that Sandy left]. (finitive)
b. Dana preferred [for Pat to get the job]. (to-infinitive)
c. Leslie wanted [Chris to go].
d. René tried [PRO to win].
e. [Kim] seems [to be happy].
f. Tracy proved [the theorem false]. (small clauses)
g. Bo considered [Lou a friend].
h. Gerry expects [those children off the ship]
(from [Pollard and Sag, 1994])

To-infinitives: Controlling and Raising its subject

XTAG assumes different syntactic structures/derivations for superficially very similar sentences:

(2) a. John tries [PRO to leave].
b. [John] seems [to leave].

Why is that?

XTAG adopts the **projection principle** from GB [Chomsky, 1981], according to which "meaning maps transparently into syntactic structure" [Culicover and Jackendoff, 2005, 47], such that the following equivalence relation holds:

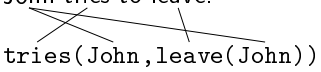
Complement of the verb $\Longleftrightarrow$ Argument of the predicate

$\Rightarrow$ θ-criterion for TAG from [Frank, 2002]

To-infinitives: Controlling and Raising its subject

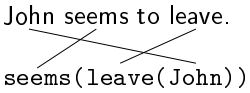
Complement of the verb $\Longleftrightarrow$ Argument of the predicate

(3) John tries to leave.



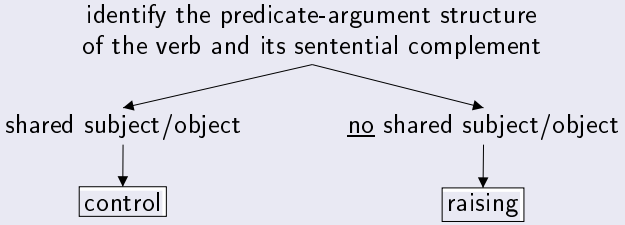
$\Rightarrow$ *John* is the complement of both *tries* and *to leave*.
$\Rightarrow$ Empty element (PRO) is used to avoid complement sharing.
$\Rightarrow$ PRO needs to be "controlled".
$\Rightarrow$ **Control**

(4) John seems to leave.



$\Rightarrow$ *John* is not the complement of *seems*.
$\Rightarrow$ Argumenthood is the primary syntactic factor, not agreement!
$\Rightarrow$ An alien complement looks like a regular complement.
$\Rightarrow$ **Raising**

Raise or control?



- Classfication game:

(5) a. They asked Jan to leave. (object control)
b. Bo turns out to be obnoxious. (subject raising)
c. Sandy is willing to go to the movies. (subject control)
d. Terry was expected to win the prize. (subject raising)
e. Kim believed a unicorn to be approaching. (object raising)

Raise or control?

identify the predicate-argument structure
of the verb and its sentential complement

shared subject/object → control

no shared subject/object → raising

- Pifalls and special cases:

(6) a. It is important for Bill to dance. (PP-raising?)
b. Christy left the party early to go to the airport. (modifier?)
c. Peter kept standing in the doorway. (no to-infinitive)

Control verbs

Control verbs establish the coreference between their subject/object and the unexpressed subject (PRO) of their sentential complement.

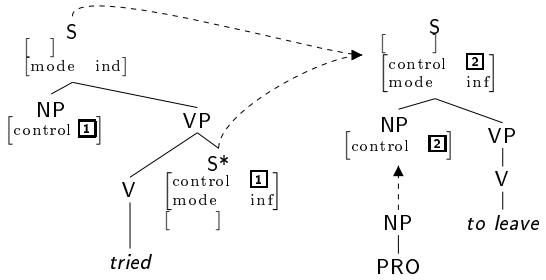
(7) a. John tried [PRO to leave]. (subject control)
b. John persuaded him [PRO to leave]. (object control)
c. *There tries [PRO to be disorder after a revolution].

⇒ **Control verbs assign semantic role to the controller!**

Control verbs - XTAG-Analysis

- control feature for coindexation
- PRO tree or PRO as coanchor of the verb

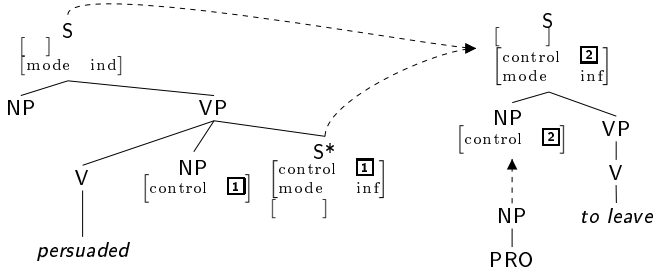
Example for subject control:



Control verbs - XTAG-Analysis

- control feature for coindexation
- PRO tree or PRO as coanchor of the verb

Example for object control:



Raising verbs

Raising verbs determine case and agreement properties of the subject complement of the (non-finite) sentential complement. Since the "raised" constituent is no immediate part of the argument structure of the raising verb, this is called **Exceptional Case Marking (ECM)**.

(8) a. [John] seems [to leave]. (subject raising)
b. Sue expects [him to leave]. (object raising)
c. [There] seems [to be disorder after a revolution].
d. John expected [it to rain].

⇒ **allow for expletive pronouns (*it*/*there*)**

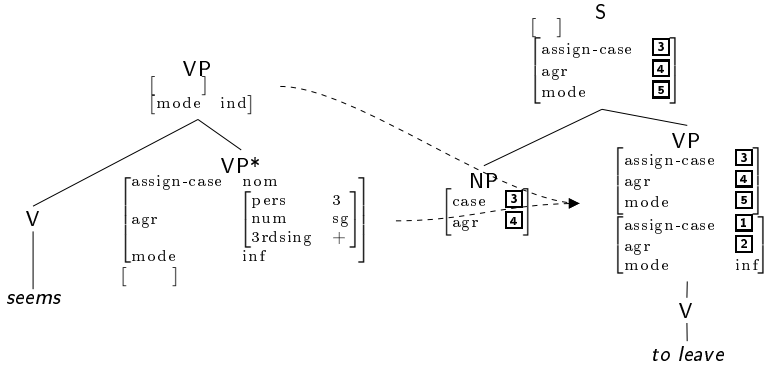
(9) John seems unhappy.
*John tries unhappy.

⇒ **allow for small clauses**

Raising verbs - XTAG-Analysis (1)

- no PRO
- The "raised" constituent is still part of the to-infinitive!
- ECM via assign-case feature

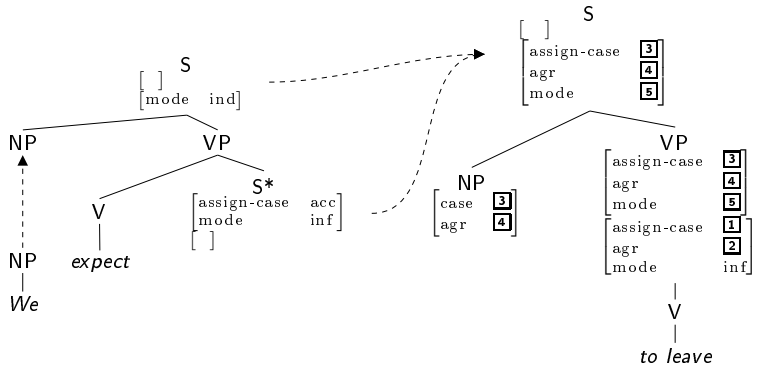
Example for subject raising:



Raising verbs - XTAG-Analysis (2)

Example for object raising:

(10) We expect him to leave.



"Ist's eins? Sind's zwei?" (Goethe, 1819)

Question:

What complements does the verb *consider* take?

(11) a. We consider [Kim to be an acceptable candidate].
b. We consider [Kim an acceptable candidate].
c. We consider [Kim quite acceptable].
d. We consider [Kim among the most acceptable candidates].
e. *We consider [Kim as an acceptable candidate].

Similar verbs: *prove, expect, rate, count, want*

1. **One sentential complement (small clause)**, where *to be* can be omitted
2. **A noun and a predicative phrase**

Small clauses - Pro and contra (1)

Pro:

- Homomorphism between argument structure and complement structure (in GB: Projection Principle, UTAH; in TAG: θ-Criterion)
- Uniformity of the subcategorized constituents:

  Instead of NP, AP, PP, IP/S, ... as possible categories of the complements, there is only one complement category.

Small clauses - Pro and contra (2)

Contra:

- Passivization (object-to-subject shift)

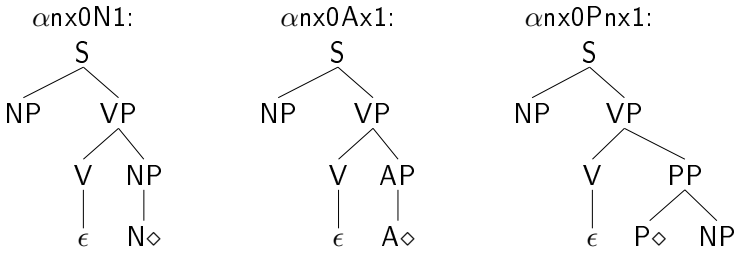
  (12) We considered [Kim quite acceptable].
  Kim was considered [__ quite acceptable].

- Idiosyncratic restrictions on the predicative phrase

  (13) a. I consider/*expect [this Island a good vacation spot].
  b. I consider/*expect [this man stupid].
  I expect [that man to be stupid].
  c. We rate/*consider [Kim as quite acceptable]

⇒ The verb should be indifferent to the categorial status of the small clause predicate!

Small clauses - XTAG-Analysis (1)

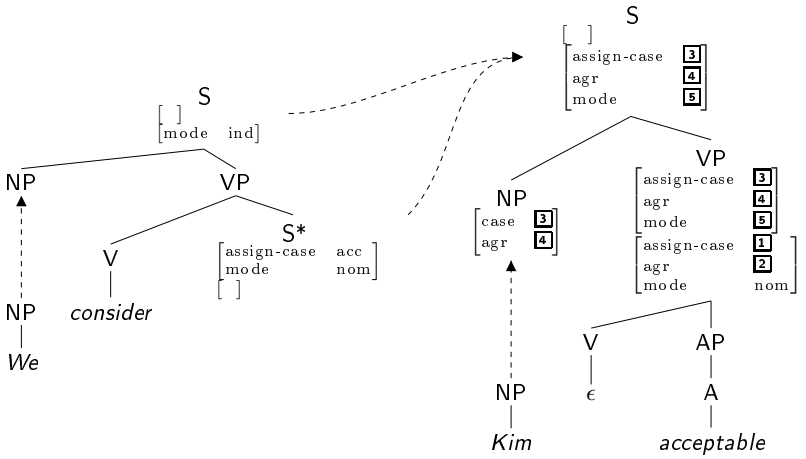


Small clauses have the structure of regular sentences , except that the verb is missing.

⇒ The superordinate verb is represented as auxiliary tree that adjoins at VP or S.

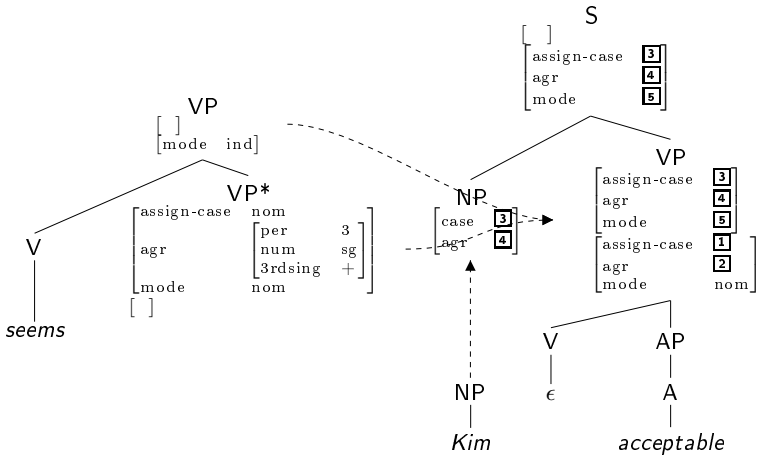
Small clauses - XTAG-Analysis (2)

(14) We consider Kim acceptable.



Small clauses - XTAG-Analysis (3)

(15) Kim seems acceptable.



Raise and control - Summary

control verbs	**raising verbs**
assign semantic role (to the controlled subject)	assign no semantic role (to the raised subject)
PRO (incomplete sent. complement)	no PRO (complete sent. complement)
assign no case (to the controlled subject)	assign case via ECM (to the raised subject)
no small clauses	small clauses
XTAG: adjoin to S	XTAG: adjoin to S or VP

Outline

Extraction - Basics

The movement metaphor:

- Relating syntactic configurations in a derivational hierarchy.
- **Traces** and **coindexation** are used to express derivational subordination.

Topicalization/Extraction:

Placing a post-verbal constituent into a sentence-initial position.

(16) a. Sandy loves Kim. (base configuration)

b. Kim_i, Sandy loves $\_\__i$. (NP-topicalization)

c. On Kim_i, Sandy depends $\_\__i$. (PP-topicalization)

Wh-extraction - Basics

Wh-Extraction:

Placing a constituent **as wh-phrase** into a clause-initial position.

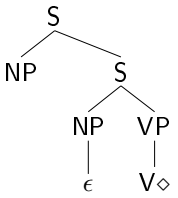
(17) a. I wonder [who$_i$ Sandy loves $\_\__i$] . (indirect question)

b. Who$_i$ does Sandy love $\_\__i$. (direct question)

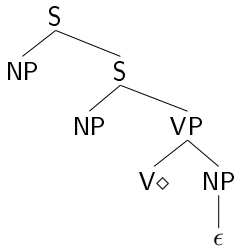
c. Sandy loves Kim$_i$ [who$_i$ Irmgard hates $\_\__i$]. (relative clause)

Extraction - Tree templates

subject extraction
(αW0nx0V)

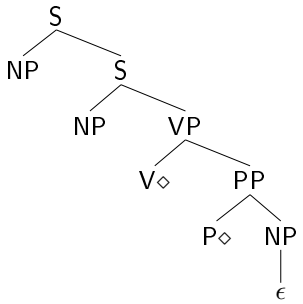


object extraction
(αW1nx0Vnx1)

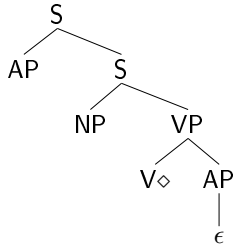


Extraction - Tree templates

preposition stranding
(αW1nx0VPnx1)



adjective complement extraction
(αWA1nx0Vax1)



Unbounded dependency

Unbounded dependency:

The dependency between an extracted constituent and its trace may extend **across arbitrarily many clause boundaries**.

(18) a. Kim$_i$, Sandy loves $\_\__i$.

b. Kim$_i$, Chris knows [Sandy loves $\_\__i$].

c. Kim$_i$, Dana believes [Chris knows [Sandy loves $\_\__i$]].

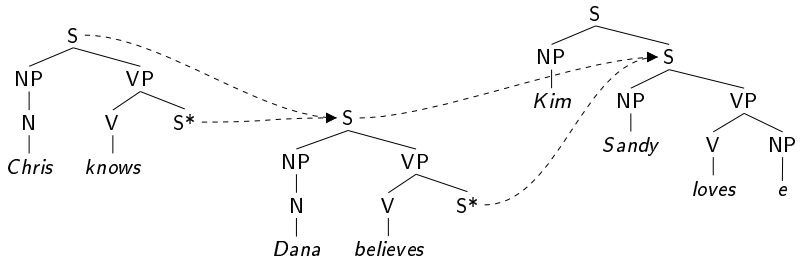
(19) a. I wonder [who$_i$ Sandy loves $\_\__i$].

b. I wonder [who$_i$ Chris knows [Sandy loves $\_\__i$]].

c. I wonder [who$_i$ Dana believes Chris knows [Sandy loves $\_\__i$]].

Unbounded dependency - XTAG-analysis (outline)

(20) Kim$_i$, Dana believes [Chris knows [Sandy loves __$_i$]].



⇒ extended domain of locality and factoring of recursion (recursive adjunction)

Islands for extraction

- **Adjuncts:**

  (21) *[Which movie]$_i$ did Gorgette fall asleep [after watching __$_i$].

  ⇒ No such elementary tree for the adjunct!

- **Coordination**

  (22) *Who$_i$ did Sandy love [__$_i$ and Kim].

  ⇒ No such elementary trees for the coordinated NP and for the governing verb!

Islands for extraction

- **Finite sentences with complementizer** (subject extraction)
  (In GB: Empty Category Principle/Subjacency):

  (23) *Who$_i$ did Alice say [that __$_i$ left].
  Who$_i$ did Alice say [__$_i$ left].

  ⇒ No such elementary trees!

- **Finite sentences with complementizer** (object extraction)

  (24) *Who$_i$ did the elephant whisper [that the emu saw __$_i$] ?
  Who$_i$ did the elephant say [that the emu saw __$_i$] ?

  ⇒ Filtering by features:
  comp = nil, where non-bridge verbs attach (*whisper*)
  comp = nil/that, where bridge verbs attach (*say*)

Subject-auxiliary inversion

Subject-auxiliary inversion

The auxiliary verb ('do', 'have', 'be', 'can', ...) precedes the subject.

- **No subject-auxiliary inversion** in embedded wh-questions:

  (25) a. I wonder [what$_i$ John reads __$_i$].
  b. *I wonder [what$_i$ **does** John read __$_i$].

- **Obligatory subject-auxiliary inversion** in direct questions with object extraction:

  (26) a. What$_i$ **does** John read __$_i$?
  b. *What$_i$ John **does** read __$_i$?
  c. *What$_i$ John reads __$_i$?

- **No subject-auxiliary inversion** in topicalization:

  (27) a. *This report$_i$ **does** John read __$_i$.
  b. This report$_i$ John **does** read __$_i$.

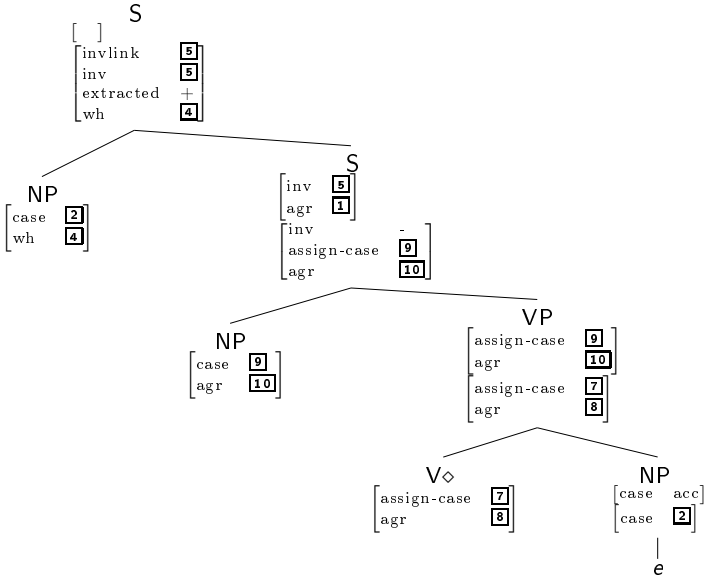
Subject-auxiliary inversion - XTAG-analysis (1)

Features for extraction:

- extracted := {+,-}
  - ⇒ to indicate extraction in the S-node
- wh := {+,-}
  - ⇒ to indicate the presence of a wh-pronoun
- inv := {+,-}
  - ⇒ to indicate inversion
- invlink := {+,-}
  - ⇒ to link wh und inv via the **root restriction**

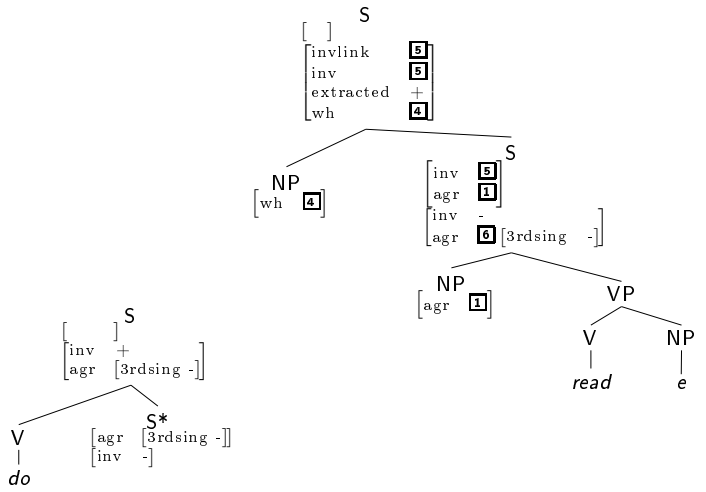
Subject-auxiliary inversion - XTAG-analysis (2)

Tree template for object extraction (simplified):



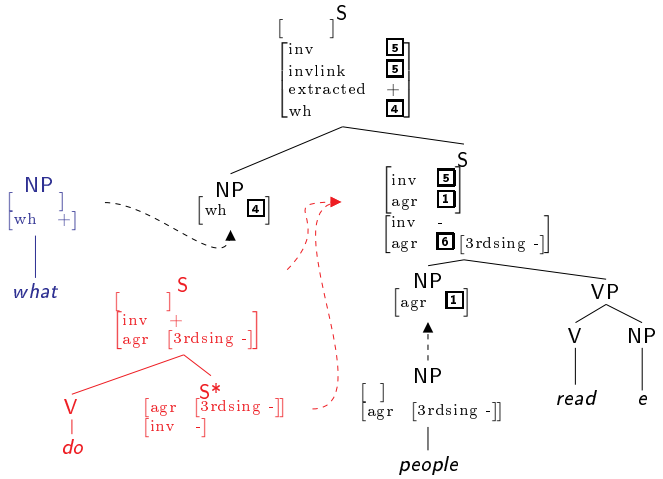
Subject-auxiliary inversion - XTAG-analysis (3)

Elementary tree with object extraction (even more simplified) and elementary tree for the inverting auxiliary *do*:



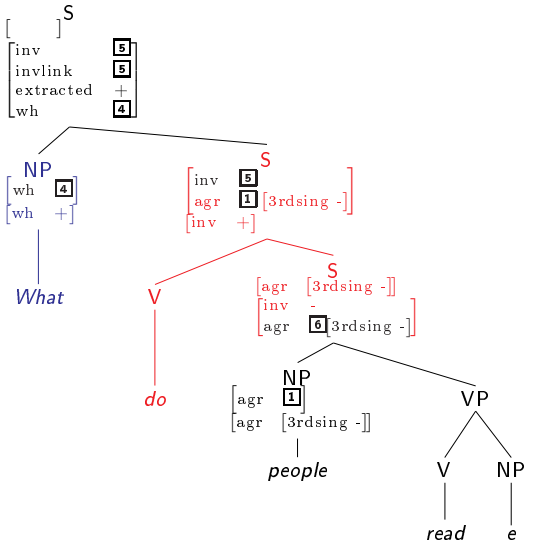
Subject-auxiliary inversion - XTAG-analysis (4)

Example derivation:



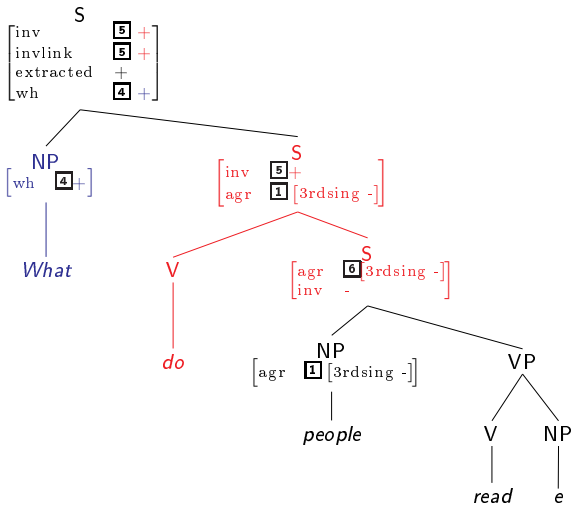
Subject-auxiliary inversion - XTAG-analysis (4)

Example derivation:



Subject-auxiliary inversion - XTAG-analysis (4)

Example derivation:



Subject-auxiliary inversion - XTAG-analysis (5)

- **No subject-auxiliary inversion** in embedded wh-questions:
  - ⇒ The governing verb selects a sentential complement with inv = - in the root node.
- **Obligatory subject-auxiliary inversion** in direct questions:
  - ⇒ In the root node: wh = +, inv = +
- **No subject-auxiliary inversion** in topicalization:
  - ⇒ In the root node: wh = -, inv = -

Problem

How to impose that wh = inv in non-embedded object extraction, without including embedded sentences or subject extraction?

Subject-auxiliary inversion - XTAG-analysis (6)

Root restriction

"A restriction is imposed on the **final root node** of any XTAG derivation of a tensed sentence which equates the wh feature and the invlink feature of the final root node." [XTAG Research Group, 2001, 296]

Crucial difference:

- The trees for object extraction have the invlink.
- The trees for subject extraction do not have the invlink.

Effects:

- Only in non-embedded object extractions the wh-pronoun depends on inversion and vice versa.
- The same tree can be used for embedded and non-embedded object extraction.

Relative clauses - Basics

"Relative clauses are NP modifiers involving extraction of an argument or an adjunct" (XTAG manual)

(28) a. the dog [which ate the cake] (wh-relatives)
b. the export exhibition [Muriel planned] (wh-less relatives)
c. [What$_i$ Sandy loves __$_i$] is Kim. (free wh-relatives)
d. the girl [reading the magazine] (gerunds ???)

(29) Somebody$_i$ lives nearby [who has a CD-burner]$_i$. (extraposition)

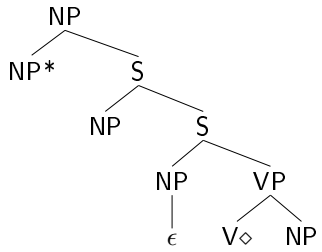
⇒ **internal vs. external syntax**

Relative clauses - XTAG-analysis (1) - Wh/that-relatives

(30) a. The dog$_i$ [that$_i$ ate the cake] (subject extraction)
b. The person$_i$ [who$_i$ I talked to __$_i$]. (preposition stranding)

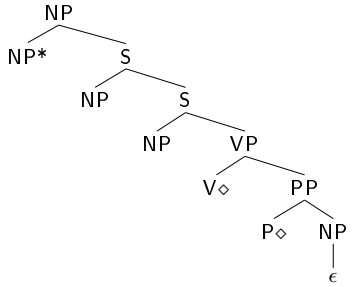
internal syntax: same as wh-extraction
external syntax: adjunction at a NP-node

βN0nx0Vnx1:



subject extraction

βN1nx0VPnx1:



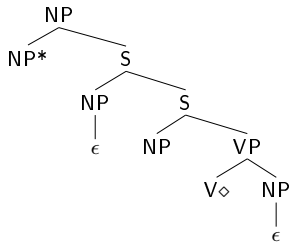
preposition stranding

Relative clauses - XTAG-analysis (2) - Wh-less relatives

(31) a. the export exhibition [Muriel planned/is planning]
b. the export exhibition [(being) planned by Muriel]

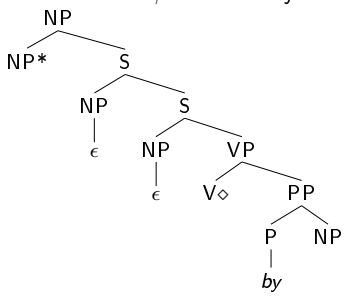
internal syntax: same as wh-extraction, but missing wh-pronoun
external syntax: adjunction at a NP-node

βNc1nx0Vnx1:



missing wh-object

βNc1nx1Vbynx0:



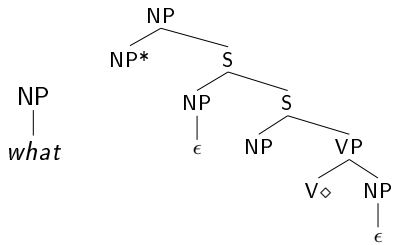
missing wh-subject in passive

Relative clauses - XTAG-analysis (3) - Free wh-relatives

Also known as **Pseudoclefts** !

(32) [What$_i$ Sandy loves __$_i$] is Kim$_i$.

internal syntax: same as wh-less relative clause
external syntax: adjunction at a wh-pronoun

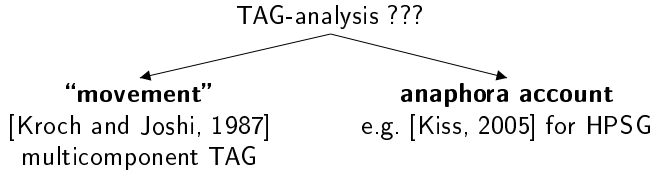


⇒ XTAG covers only free wh-relatives in object position!

Extraposed relative clauses

(33) a. Somebody$_i$ lives nearby [who$_i$ has a CD-burner].
b. Karl hat mir [von der Kopie [einer Fälschung [eines Bildes [einer Frau __$_i$]]]] erzählt, [die schon lange tot ist]$_i$.

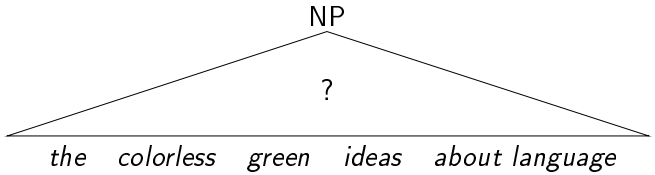
internal syntax: same as wh-extraction
external syntax: no adjunction at a NP-node, but to the right periphery of the sentence

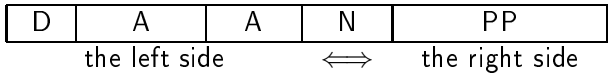


Extraction - Summary

- Topicalization and wh-extraction obtain a uniform analysis.
- Account for unbounded dependency via extended domain of locality + factoring of recursion
- Island constraints can be modelled rather naturally (wrt. TAG).
- Relative clauses are realized as auxiliary trees. Their internal structure is analysed as ordinary wh-extraction.

The inner structure of NPs





1. The left side of nouns
   - Determiners
   - Adjectives
2. The right side of nouns
   - PP-complements/-adjuncts of nouns
   - Relative clauses

The left side of nouns - Determiners

'Determiners' labels a rather heterogenous set of items:

- articles (*the*, *a*)
- demonstratives (*this*, *that*)
- genitives (*my*, *Bill's*, *that man's*)
- quantifiers (*all*, *some*, *every*, *most*, *many*)

Determiners can be stacked:

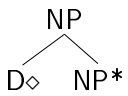
(34) all these many ideas

⇒ The pattern of determiner stacking is very complex!

The left side of nouns - Determiners - XTAG-analysis

XTAG uses β**Dnx** for all determiners:

βDnx:



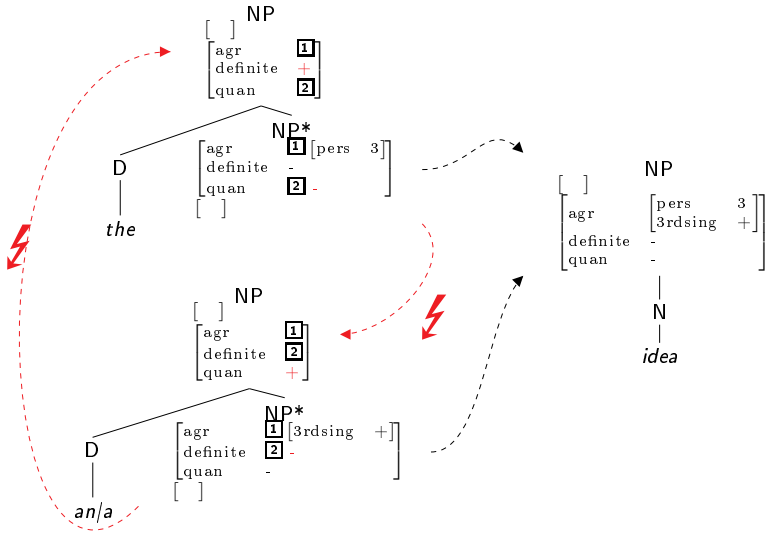
XTAG uses a set of **9 features** to handle determiner stacking:

- definite:= $\{+,-\}$ marks definite determiners (*the*, *this*, *that*, ...)
- quant:= $\{+,-\}$ marks quantifiers and non-definite articles (*a*, *all*, *some*, *every*, ...)
- plus: card(inality), gen(itive), wh, decreas(ing), const(ancy), compl(ement), and arg

⇒ We only consider definite and quan in what follows.

The left side of nouns - Determiners - XTAG-example

⇒ The feature structures are considerably simplified!



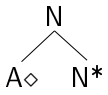
The left side of nouns - Adjectives

XTAG assumes that adjectives can appear in any order:

(35) a. the colorless green ideas
b. the green colorless ideas

In XTAG, adjective trees adjoin to N, where no special feature is required:

βAn:



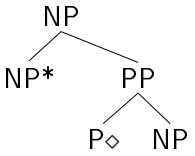
The right side of nouns - PP-complements/adjuncts

XTAG assumes that PP-complements/adjuncts can appear in any order.

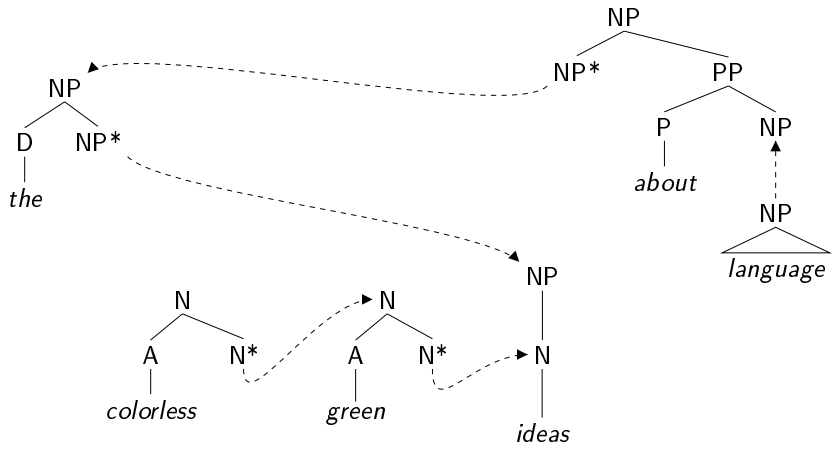
(36) a. the ideas about language from Germany
b. the ideas from Germany about language

In XTAG, PP-complements/adjuncts adjoin to NP, and no special feature is required:

βnxPnx:



The inner structure of NPs - Putting the pieces together



⇒ The order of adjunction of determiners and PPs is not fixed!

Gerund NPs

NPs made from gerunds basically fall into two groups:

1. The gerund verb is treated like a regular noun.
2. The gerund verb and its complements/adjuncts preserve a sentential structure, but are treated as regular NP.

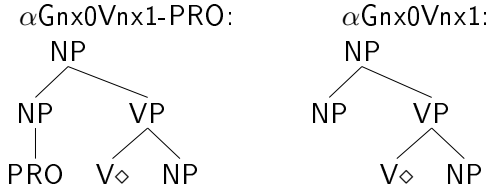
Determiner gerunds (aka derived nominalizations):

(37) a. [The proving of the theorem] succeeds.
b. *[The proving the theorem] succeeds.

NP gerunds (aka sentential gerunds):

(38) a. [Proving the theorem] succeeds.
b. [John proving the theorem] succeeds.
c. *[The Proving the theorem] succeeds.

Gerund NPs - XTAG-analysis of NP gerunds



Chomsky, N. (1981).
Lectures on Government and Binding.
Foris, Dordrecht.

Culicover, P. W. and Jackendoff, R. (2005).
Simpler Syntax.
Oxford University Press, Oxford.

Frank, R. (2002).
Phrase Structure Composition and Syntactic Dependencies.
MIT Press, Cambridge,MA.

Kiss, T. (2005).
Semantic constraints on relative clause extraposition.
Natural Language and Linguistic Theory, 23:281–334.

Kroch, A. S. and Joshi, A. K. (1987).
Analyzing extraposition in a tree adjoining grammar.
In Huck, G. J. and Ojeda, A. E., editors, Discontinous Constituency, number 20 in Syntax and Semantics, pages 107–149. Academic Press, Inc.

Pollard, C. and Sag, I. A. (1994).
Head-Driven Phrase Structure Grammar.
University of Chicago Press, Chicago and London.

Prolo, C. A. (2002).
Generating the XTAG English grammar using metarules.
In Proceedings of COLING-02, pages 814–820, Taipei, Taiwan.

XTAG Research Group (2001).
A Lexicalized Tree Adjoining Grammar for English.
Technical report, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA.