Scope below the word level

Markus Egg

Humboldt-Universität Berlin

Semantikkreis, Universität Göttingen

28. Oktober 2009

Structure of the talk

- topic: semantic scope relations below the level of syntactic atoms (words)
  - due to them, the syntactic and semantic structure of linguistic expressions can differ considerably
  - these expressions are very similar to structural ambiguities
- a challenge for the syntax-semantics interface
  - theoretical definition of the relation between syntactic and semantic structure
  - practical semantic construction in NLP systems: bridge the difference between syntactic and semantic structure
- several interface strategies that are proposed in the literature are compared
  - expressivity
  - coverage

The data 1

- well-know example: modification of indefinite pronouns

  (1) *everyone in this room*

- intuition: the modifier scopes between *every-* and *-one*:

  every- (in this room (-one))

- this cannot be modelled with standard semantic construction

  (2)(a) $\lambda P \forall x.\underline{\mathbf{person}'}(x) \rightarrow P(x)$

  (b) $\lambda P \lambda x.P(x) \wedge \mathbf{in}'(x, \mathbf{R})$

  (c) $\lambda P \forall x.\mathbf{person}'(x) \wedge \mathbf{in}'(x, \mathbf{R}) \rightarrow P(x)$

The data 2

- the phenomenon occurs across languages, cp. Icelandic DPs with enclitic determiners

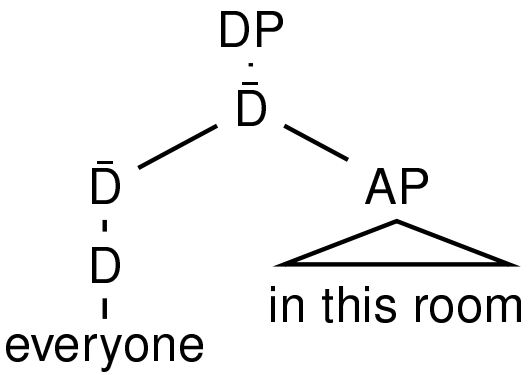
  (3) *rauða* *hús* *-ið*
  red house DEF
  'the red house'

- here it is not obvious how to model the relation between syntactic and semantic structure
- most important: semantic scope corresponds to the syntactic 'c-command'
- a node A c-commands a node B in a tree iff
  - A is the immediate daughter of a branching node that dominates B
  - A does not dominate B or vice versa

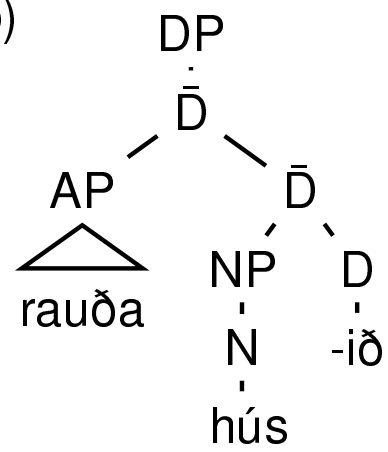
The data 3

- surface-oriented syntactic structures for *everyone in this room* and *rauða húsið*

(4) (a)



(b)



- parallel structures
- scope doesn't show up in the syntax as c-command

The data 4

- Turkish derivation affixes

  (5) *yağız* *at* *-lı*

  dark.brown horse provided.with

  'someone with a dark brown horse'

- Turkish inflection: the *-ip*-construction

  (6) *yi* *-y* *-ip* *iç* *-eceğ* *-im*

  eat -F -IP drink -FUT -1sg

  'I will eat and drink'

- this phenomenon does not depend on morphological transparency

  (7) *Amélie opened the door for two hours*

  (8) CAUSE($\mathbf{a}$, BECOME($\mathbf{for\_2h'}(\mathbf{open'}(\mathbf{D}))$))

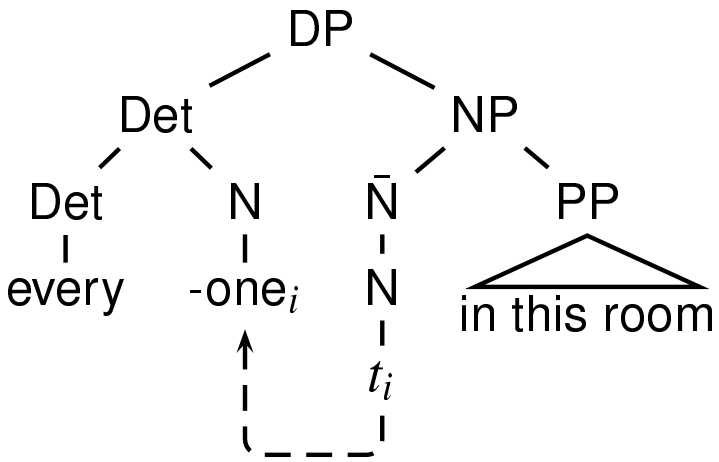
Proposed analyses

- how to bridge the difference between syntactic and semantic structure?
  - Generative Grammar: syntactic 'preprocessing' in the direction of the semantic structure
  - underspecified analysis: very expressive syntax-semantics interface
  - LTAG and LFG for Turkish: very expressive syntactic construction
- the approaches bring with them specific prerequisites
  - additional syntactic layer(s)
  - complex syntax-semantics interface
  - only partial parallel between syntactic and semantic structure
  - dependency on morphological structure
  - complex syntactic construction

Previous analyses: Generative Grammar 1

- 'preprocessing' of syntactic structures before interpreting them
- semantically relevant syntactic structures are not directly visible
- they are systematically related to the surface structure
- application to the modification of indefinite pronouns (Abney 1987)
  - they consist of a determiner and an enclitic noun
  - head-to-head movement of the noun and incorporation into the determiner

(9)



Previous analyses: Generative Grammar 2

- advantages of the approach
  - the parallel between syntax and semantics is maximised: semantic scope corresponds to syntactic c-command
  - very simple syntax-semantics interface (mainly functional application)
- disadvantages
  - additional syntactic layer(s)
  - the range of indefinite pronouns cannot be explained

    some/any/every/no + *one/body/thing*
  - lexical ambiguity for the second elements must be stipulated
    - free vs. bound version
    - each one with a different interpretation

Previous analyses: Generative Grammar 3

- disadvantages (ctd)
  - the analysis cannot explain specific restrictions (this is problematic for all approaches)

    (10) **everyone former* (no reading 'every former person')
  - the analysis (tacitly?) presupposes morphological transparency

    (11) *jeder/jemand in diesem Zimmer* 'everyone/someone in this room'

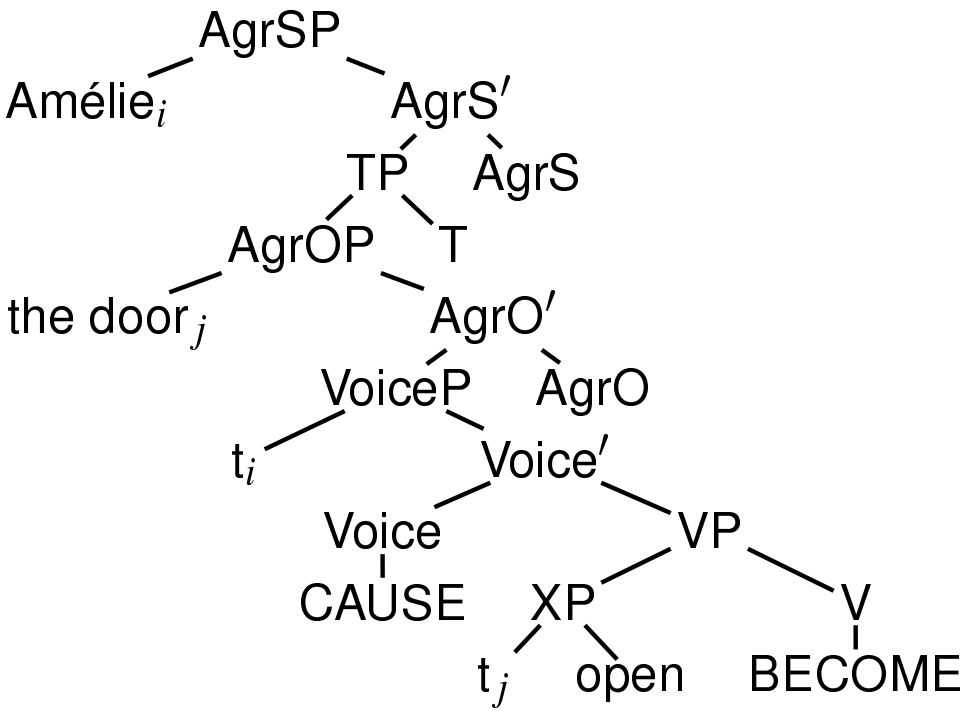
    (12) *etwas neues* 'something new'
  - few analyses (e.g., von Stechow 1996) suggest syntactic decomposition of cases like (13)

    (13) [= (7)] *Amélie opened the door for two hours*

Previous analyses: Generative Grammar 4

- Stechow's analysis: „CAUSE(BECOME(open))" is phonetically realised as *open*

(14)

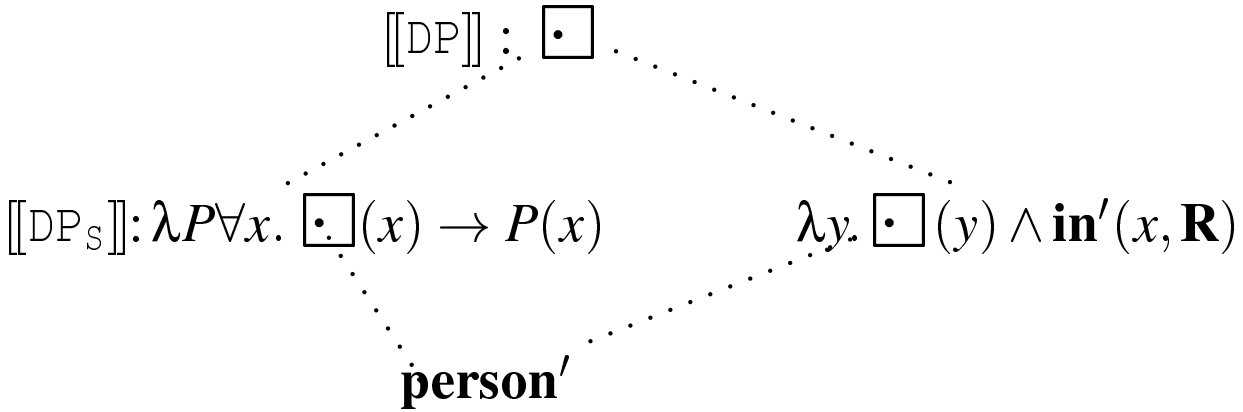


- poststate predicate is syntactically accessible for modification (as XP)

Previous analyses: underspecification 1

- underspecified representation ('constraint') of (16)

  (15)



  - sets of semantic representations (here, λ-terms) are described on a meta-level
  - ingredients: fragments of λ-terms, 'holes', and relations between them
  - the described semantic representations ('solutions' of the constraint) are derived by identifying fragments and holes (the 'plugging' of Bos 2004)

- the only solution of (15) is (16) [= (2c)]

  (16) $\lambda P \forall x. \mathbf{person}'(x) \wedge \mathbf{in}'(x, \mathbf{R}) \rightarrow P(x)$

Previous analyses: underspecification 2

- structural ambiguities may also show up below the word level

  (17) *Amélie opened the door again* (Dowty 1979)

  (18) CAUSE(**a**, BECOME(**again′**(**open′**(**D**))))

  (19) **again′**(CAUSE(**a**, BECOME(**open′**(**D**))))

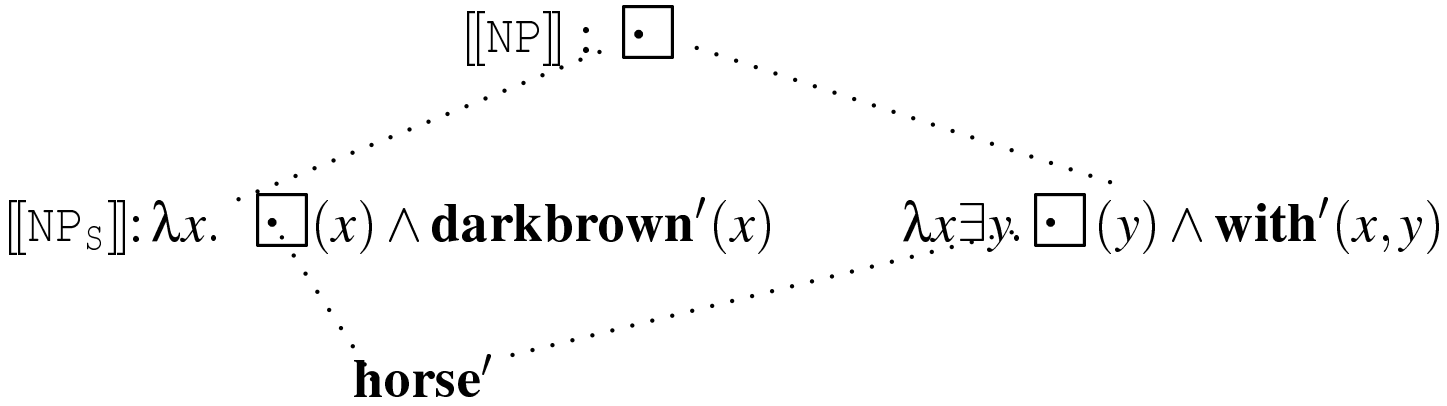
  (20) *genç* *at* *-lı*
  young horse provided.with
  'someone with a young horse/young rider'

- the other cases are different in that there a potential reading is ruled out
  - *open* is aspectually bounded, its poststate predicate 'be open' is not
  - durative adverbials select for unbounded predicates, *again* does not
  - *yağız* 'dark brown' is preferably used to refer to animals

Previous analyses: underspecification 3

- representation of *yağız atlı*

  (21)



- the sole solution of (21)

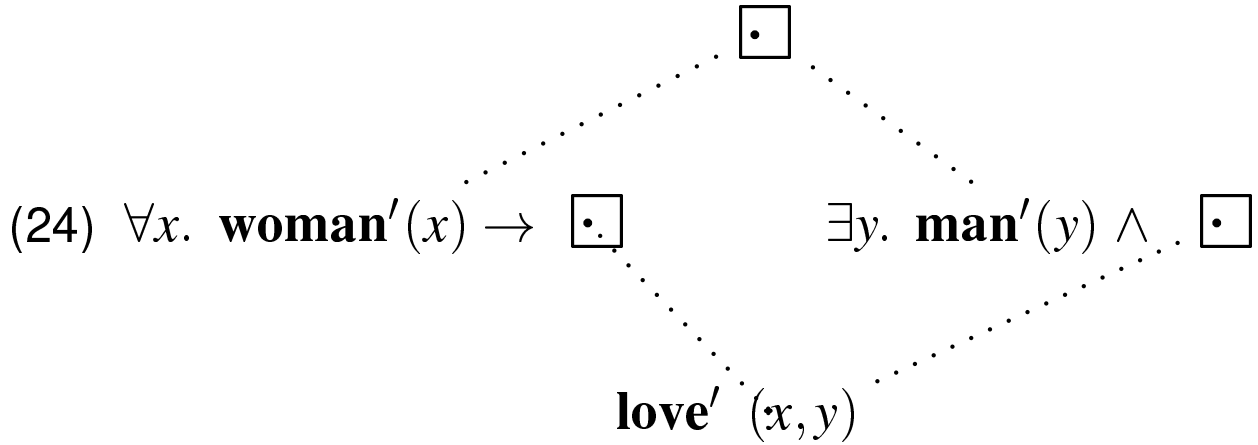
  (22) $\lambda x \exists y.\mathbf{horse}'(y) \wedge \mathbf{darkbrown}'(y) \wedge \mathbf{with}'(x,y)$

- if *yağız* ‘darkbrown’ is replaced by *genç* ‘young’, the constraint has two solutions

Previous analyses: underspecification 4

- this representation (and its construction) can also serve for the underspecified representation of scope ambiguity like in (23)

  (23) *Every woman loves a man*



- here the scopally ambiguous fragments can be arranged in two ways that correspond to the two readings

Previous analyses: underspecification 5

- advantages
  - surface-oriented syntactic representation (words are atoms)
  - immediate modelling of the similarity to structural ambiguities
  - the syntax-semantics interface for phrase level scope ambiguity can be reused
  - lexical licensing for scope below the word level: structured lexical entries, e.g., for indefinite pronouns

    (25) $[\![\mathrm{D}]\!] : \lambda P \forall x. \boxed{\bullet} (x) \rightarrow P(x)$

    $[\![\mathrm{D_S}]\!] : \mathbf{person}'$
  - the analysis does not depend on morphological transparency
- disadvantages
  - the relation between c-command and scope is not represented
  - the syntax-semantics interface is complicated

Previous analyses: LFG for Turkish 1

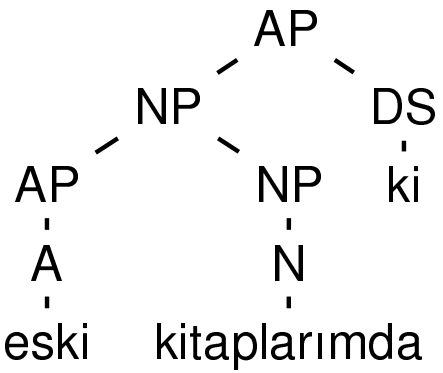
- part of the ParGram project (Çetinoğlu and Oflazer 2006)
- describes cases like (5) and (6) in terms of 'inflectional groups' (specific word stems)
- as constituents they are accessible syntactically for modification

(26) *eski* *kitap* *-lar* *-ım* *-da* *-ki*

old book PL my LOC KI

'in my old books'

(27)



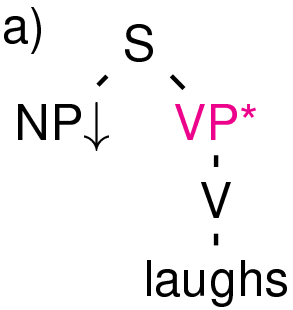
Previous analyses: LFG for Turkish 2

- *kitaplarımda* 'in my books' is an inflectional group
- the 'relative suffix' *-ki* is a constituent 'DS' (derivational suffix) of its own
- *-ki* is not an ordinary suffix (e.g., no vowel harmony; Kornflit 1997)
- only an ad-hoc solution, because *eski* 'old' pertains only to the root *kitap* 'book' (or to *kitaplar* 'books')
  - the adjective is in the scope of the locative *-da*, otherwise, the DP would mean 'old things in my books'
  - such a semantic case takes scope over the whole DP (see also Butt and King 2005)
- many more suffixes would have to be separated as constituents, which would weaken the boundary between morphology and syntax

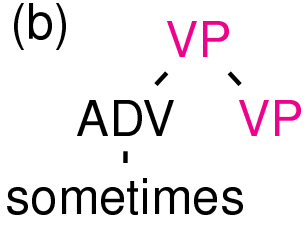
Previous analyses: LTAG 1

- in LTAG (Joshi, Kallmeyer, and Romero 2007), difficult cases like (1) can be modelled directly in terms of adjunction
- syntactic heads introduce the whole tree fragment for their projection (minus subcategorised elements) and determine its meaning
- in adjunction a specific internal node in a tree fragment is substituted by another fragment
- the first fragment is split in two, and the second fragment is inserted in between

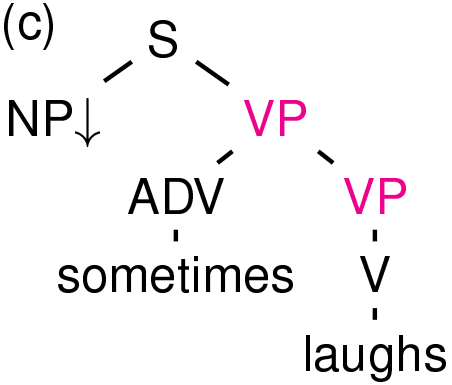
(28) (a)



(b)



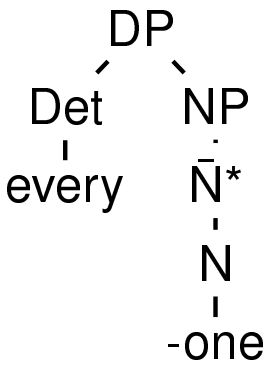
(c)



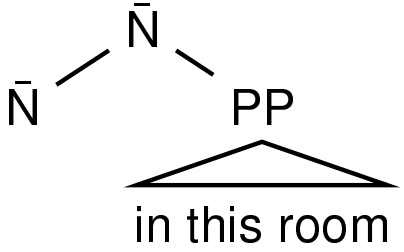
Previous analyses: LTAG 2

- adjunction for *everyone in this room* (Kallmeyer and Romero 2008)

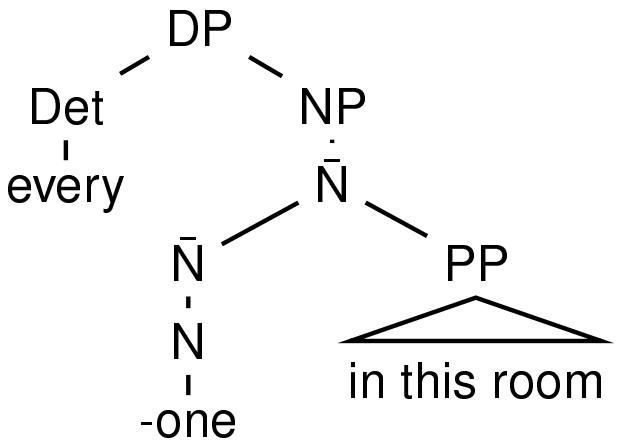
(29) (a)



(b)



(c)



- if modification is feasible, the potentially modified expression determines how the semantics of the modifier is integrated with the semantics of the whole projection

Previous analyses: LTAG 3

- semantic entries for *everyone* and *in this room* (strongly simplified; Kallmeyer and Romero 2008)

(30)
$$\begin{bmatrix} \text{DP} \mid \text{SEM} & \begin{bmatrix} \text{TOP} & \lambda P \forall x. \boxed{1}(x) \rightarrow P(x) \\ \text{BOTTOM} & \boxed{1} \end{bmatrix} \\ \bar{\text{N}} \mid \text{SEM} & \begin{bmatrix} \text{TOP} & \boxed{1} \\ \text{BOTTOM} & \mathbf{person}' \end{bmatrix} \end{bmatrix}$$

(31)
$$\begin{bmatrix} \bar{\text{N}} \mid \text{SEM} & \begin{bmatrix} \text{TOP} & \lambda x. \boxed{2}(x) \wedge \mathbf{in}'(x, \mathbf{R}) \\ \text{BOTTOM} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

Previous analyses: LTAG 4

- unification of the TOP BOTTOM values of modifier and modified expression
- result for *everyone in this room*

$$\left[\begin{array}{l|ll} \text{DP} & \text{SEM} & \left[\begin{array}{ll} \text{TOP} & \lambda P \forall x.\mathbf{person}'(x) \wedge \mathbf{in}'(x,\mathbf{R}) \rightarrow P(x) \\ \text{BOTTOM} & \lambda x.\mathbf{person}'(x) \wedge \mathbf{in}'(x,\mathbf{R}) \end{array}\right] \end{array}\right]$$

- in case there is no modifier, TOP and BOTTOM values of the potentially modified expression are unified
- lexemes introduce tree fragments anyway, i.e., lexical entries for indefinite pronouns are in this respect normal, e.g., (28a) and (29a)
- modification is anticipated in the modified expression just like in the underspecified analysis

Previous analyses: LTAG 5

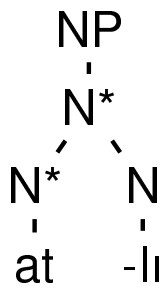
- advantages
  - the relation between c-command and scope is modelled
  - no additional syntactic levels
- problems
  - the analysis is motivated morphologically
    - then pairs of synonyms that only differ w.r.t. morphologic transparency would have to be treated differently
      sterben/*totgehen* 'die', *benetzen*/*naß machen* 'wet', . . .
    - this runs counter to intuitions
      (32) *die Tür fünf Minuten lang öffnen* 'open the door for five minutes'
      (33) *die Tür fünf Minuten lang aufmachen*

Previous analyses: LTAG 6

- problems (ctd)
  - scope below the word level must be modelled in the syntax, which is inadequate for (at least) semantically relevant inflection
    - semantic case
    - tense: scope over quantifying adverbs, *-ip*-construction

      (34) *Amélie always went to the library*
  - what to do in the case of structural ambiguity below the word level?
    - several potential adjunction sites?

      (35)

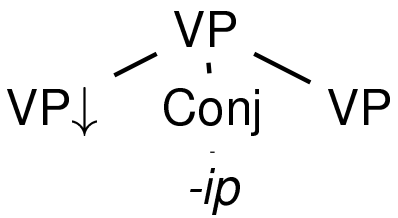
      

    - careful bookkeeping in the semantics would be needed

Previous analyses: LTAG 7

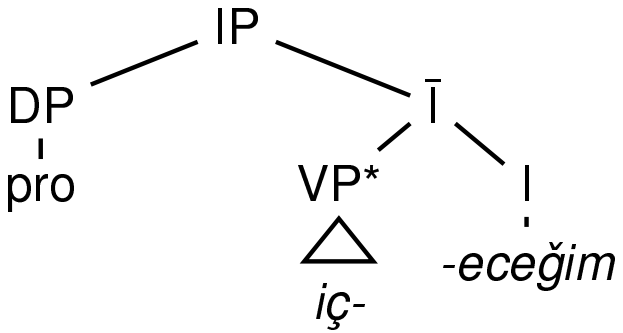
- the *-ip-* construction
- lexical entry for *-ip*:

(36)



- tree for *içeceğim* 'I will eat'

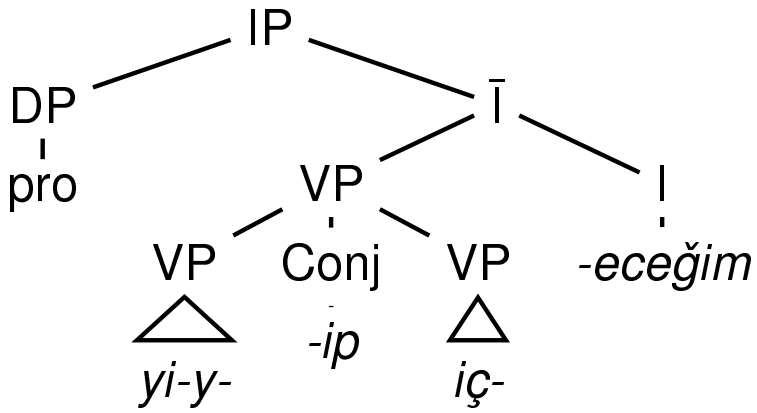
(37)



Previous analyses: LTAG 8

- analysis of *yiyip içeceğim* 'I will eat and drink'

(38)



Conclusion

- there is no optimal analysis that combines all advantages
  - the underspecified analysis cannot model the relation between scope and c-command
  - the LTAG analysis does not generalise to morphologically intransparent cases
  - the generative analysis cannot be based onto a surface-oriented syntactic analysis
- the weight of these shortcomings depends on the (theoretical or applied) approach to the syntax-semantics interface

References

Abney, S. (1987). *The English noun phrase in its sentential aspect*. Ph. D. thesis, MIT.

Bos, J. (2004). Computational semantics in discourse: Underspecification, resolution, and inference. *Journal of Logic, Language and Information 13*, 139–157.

Butt, M. and T. King (2005). The status of case. In V. Dayal and A. Mahajan (eds), *Clause Structure in South Asian Languages*, 153–198. Dordrecht: Kluwer.

Çetinoğlu, Ö. and K. Oflazer (2006). Morphology-syntax interface for Turkish LFG. In *Proceedings of the 21st COLING and the 44th ACL*, 153–160.

Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.

Joshi, A., L. Kallmeyer, M. Romero (2007). Flexible composition in LTAG: Quantifier scope and inverse linking. In H. Bunt and R. Muskens (eds), *Computing Meaning 3*, 233–256. Amsterdam: Springer.

Kallmeyer, L. and M. Romero (2008). Scope and situation binding in LTAG using semantic unification. *Research on language and computation 6*, 3–52.

Kornflit, J. (1997). *Turkish*. Descriptive Grammars. London: Routledge.

von Stechow, A. (1996). The different readings of 'Wieder': A structural account. *Journal of Semantics 13*, 87–138.

The interface rules 1

- the semantic contribution of every syntactic constituent C distinguishes a *main* fragment '$[\![\mathtt{C}]\!]$'and an embedded *secondary* fragment '$[\![\mathtt{C_S}]\!]$'
- interface rules address them and determine them for the constructed constituent

(39) $[\![\mathtt{D}]\!] : \lambda P \forall x.\ \boxed{\cdot}\ (x) \rightarrow P(x)$

$[\![\mathtt{D_S}]\!] : \mathbf{person}'$

- '$[\![\mathtt{C}]\!] : \mathtt{F}$' expresses that the main fragment of C is defined as fragment F

(40) $[_{\bar{\mathtt{X}}}\ \mathtt{X}]$ $\overset{(\mathrm{SSI})}{\Rightarrow}$ $[\![\bar{\mathtt{X}}]\!] : [\![\mathtt{X}]\!]; \quad [\![\bar{\mathtt{X}}_\mathtt{S}]\!] : [\![\mathtt{X}_\mathtt{S}]\!]$

The interface rules 2

- the rule for modification

(41) $[_{\bar{X}_1}\ \mathrm{Mod}\ \bar{X}_2] \overset{(SSI)}{\Rightarrow}$

$[\![\bar{X}_{1S}]\!] : [\![\mathrm{Mod}]\!](\ \boxed{\bullet}\ )$ $\quad [\![\mathrm{Mod}]\!]: [\![\mathrm{Mod}_S]\!]$ $\quad [\![\bar{X}_1]\!]:[\![\bar{X}_2]\!]$

$[\![\bullet\bar{X}_{2S}]\!]$

- the rule for projecting $\bar{X}$ constituents to XP

(42) $[_{XP}\ \bar{X}] \overset{(SSI)}{\Rightarrow}$

$[\![XP]\!] : \boxed{\bullet}$

$[\![\bullet XP_S]\!]:[\![\bar{X}]\!]$ $\quad [\![\bullet\bar{X}_S]\!]$

The interface rules 3

- semantic construction for *everyone in this room*
  - semantic representations (43) [= (39)] and (44) of pronoun (and $\bar{\text{D}}$) and PP

(43) $[\![\text{D}]\!] : \lambda P \forall x. \boxed{\cdot} (x) \rightarrow P(x)$

$[\![\text{D}_\text{S}]\!] : \mathbf{person}'$

(44) $[\![\text{PP}]\!], [\![\text{PP}_\text{S}]\!]: \lambda P \lambda x. P(x) \wedge \mathbf{in}'(x, \mathbf{R})$

  - result of the modification rule (41)

(45) $[\![\bar{\text{D}}]\!] : \lambda P \forall x. \boxed{\cdot}(x) \rightarrow P(x)$ $\qquad$ $[\![\bar{\text{D}}_\text{S}]\!] : \lambda x. \boxed{\cdot}(x) \wedge \mathbf{in}'(x, \mathbf{R})$

$\mathbf{person}'$

  - rule (42) adds the upper half of the dominance diamond (15)

The analysis 1: *yağız atlı*

- rule (47) describes the semantic effect of affixing *-lı* to a nominal base

(46) $[_{\mathtt{X}}\ \mathtt{Bs}\ \mathtt{Aff}] \overset{(\text{morph})}{\Rightarrow} [\![\mathtt{X}]\!] : [\![\mathtt{Aff}]\!](\boxed{\bullet})$

$[\![\mathtt{X_S}]\!] : \ [\![\mathtt{Bs}]\!]$

- (47) is the affix semantics

(47) $[\![\mathtt{Aff}]\!]$, $[\![\mathtt{Aff_S}]\!]$: $\lambda P \lambda x \exists y.P(y) \wedge \textbf{provided-with}'(x,y)$

- the semantics of *atlı* 'someone provided with a horse'

(48) $[\![\mathtt{N}]\!] : \lambda x \exists y.\boxed{\bullet}(y) \wedge \textbf{provided-with}'(x,y)$

$[\![\mathtt{N_S}]\!] : \ \textbf{horse}'$

The analysis 4: The *ip*-construction

- the interface rule

  (49) $[_{\mathrm{V2}}\ \mathrm{Bs}\text{-}ip\ \mathrm{V1}] \overset{(\mathrm{SSI})}{\Rightarrow}$

  $[\![\mathrm{V_2}]\!] : [\![\mathrm{V_1}]\!] \qquad [\![\mathrm{V_{2S}}]\!] : [\![\mathrm{Bs}]\!] \& [\![\mathrm{V_{1S}}]\!]$

- constraint for (6)

  (50) $[\![\mathrm{V_2}]\!] : \exists e.e_0 < e \wedge \boxed{\cdot} \vdots\ (\mathbf{speaker'})(e)$

  $[\![\mathrm{V_{2S}}]\!] : \mathbf{eat'} \ \& \ \mathbf{drink'}$

- solution of this constraint

  (51) $\exists e.e_0 < e \ \wedge \mathbf{eat'}(\mathbf{speaker'})(e) \wedge \mathbf{drink'}(\mathbf{speaker'})(e)$

Describing the nontrivial cases 5: The *ip*-construction

- simplified tense account
- constraint for (6)

(52) $[\![\mathrm{V}_2]\!] : \exists e.e_0 < e \wedge \boxed{\cdot}\ (\mathbf{speaker}')(e)$

$[\![\mathrm{V}_{2\mathrm{S}}]\!] : \mathbf{eat}' \ \& \ \mathbf{drink}'$

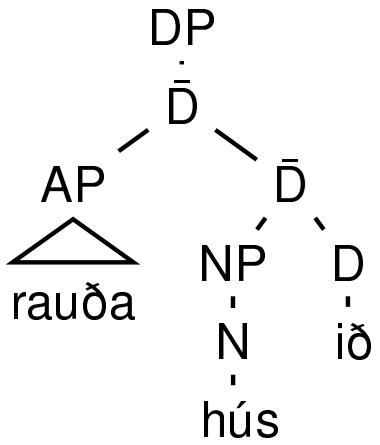
- solution of this constraint

(53) $\exists e.e_0 < e \ \wedge \mathbf{eat}'(\mathbf{speaker}')(e) \wedge \mathbf{drink}'(\mathbf{speaker}')(e)$

The analysis 5: *rauða húsið*

- syntactic structure of *rauða húsið*

(54)



- analogous semantic construction pattern as in (1)

(55) $\lambda P \exists !x.[\mathbf{red}'(x) \wedge \mathbf{house}'(x)] \wedge P(x)$

(56) $\lambda P \exists !x.[\underline{\mathbf{house}'(x)}] \wedge P(x)$

(57) $\lambda P \lambda x.\mathbf{red}'(x) \wedge P(x)$

- difference: the modified expression is syntactically complex