Which syntax is required by semantics?

Markus Egg

Rijksuniversiteit Groningen

Workshop 'Which syntax feeds semantics?'

14 August 2008

rijksuniversiteit groningen

Main argument of the talk

- do nontrivial cases of semantic construction yield clues to the workshop question?
  - their syntactic (surface) structure and semantic structure differ considerably
  - syntactic analyses with considerable 'preprocessing' of syntactic structures seem to have an advantage here over surface-oriented ones
- these nontrivial cases are closely related to structural ambiguities
- underspecification formalisms can represent these ambiguities as well as these nontrivial cases
- there are syntax-semantics interfaces to map from (surface-oriented) structures to underspecified semantic representations
- the interfaces can be reused for the nontrivial cases of semantic construction
- consequently, these nontrivial cases are no straightforward argument for or against specific syntactic analyses

Structure of the talk

- the nontrivial cases of semantic construction
- their relation to structural ambiguities
- their description in terms of underspecified semantic representations
- *only in the paper:* the interface to derive these representations
- related work

rijksuniversiteit groningen

Nontrivial cases of semantic construction 1

- Turkish derivational affixes (see also Bozsahin 2008)

  (1) *yağız* *at* *-lı*

  dark.brown horse provided.with

  'someone with a dark brown horse'

- Turkish inflectional morpology: the *-ip*-construction

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

  eat -F -IP drink -FUT -1sg

  'I will eat and drink'

- Islandic enclitic determiners

  (3) *rauða* *hús* *-ið*

  red house -the

  'the red house'

rijksuniversiteit groningen

Nontrivial cases of semantic construction 2

- in all these examples, morphemes within one word have scope over (an)other word(s)
- these other words can have scope over the rest of the first word, e.g., in the Icelandic example:

  (4) *-ið (rauða (hús))* ‘the (red (house))’
- possible way out: regard items like *-ið* as clitics, words of their own whose independent status is hidden by orthography and phonology
- but these examples pattern with definitely non-clitic cases

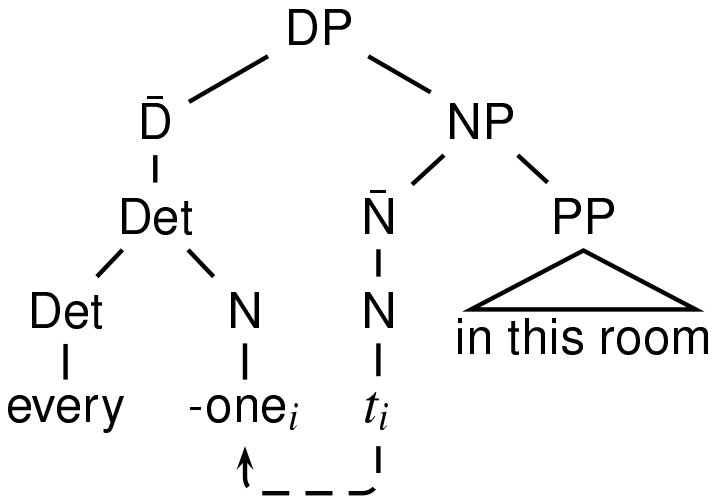
  (5) *everyone in this room*
- scope relations in (5):

  (6) *every- (in this room (-one))*

Nontrivial cases of semantic construction 3

- does this call for syntactic decomposition and movement (Abney, 1987)?
  - *everyone* consists of a determiner *every* and an enclitic noun *-one*
  - the noun is incorporated with the determiner after head-to-head movement

(7)



  - the semantic interpretation would follow directly from the underlying structure
  - this analysis looks like an argument in favour of generative grammar

Nontrivial cases of semantic construction 4

- syntactic decomposition and movement for *everyone*
  - problem 1: this pattern is highly restricted:
    some/any/every/no + *one/body/thing*
  - problem 2: one must stipulate lexical ambiguity for the feasible second elements: a bound and a free variant with different meanings
  - problem 3: this analysis presupposes morphological transparency
    (8) *jeder/jemand in diesem Zimmer* 'everyone/someone in this room'
- these examples pattern with other morphologically opaque cases
  (9) *Amélie left for two hours*
  (10) BECOME($\mathbf{for\_2hrs}'(\mathbf{be\_away}'(\mathbf{a}))$)

Nontrivial cases and structural ambiguities 1

- these examples are closely related to structural ambiguities

  (11) *Amélie left again*

  (12) BECOME($\mathbf{again'}(\mathbf{be\_away'}(\mathbf{a}))$)

  (13) $\mathbf{again'}$(BECOME($\mathbf{be\_away'}(\mathbf{a})$))

  (14) *genç* *at* *-lı*

  young horse provided.with

  'someone with a young horse/young rider'

- the difference seems to be that in the non-ambiguous cases one of the potential readings is ruled out (or, at least strongly dispreferred) by
  - *leave* is aspectually bounded, its aftermath (the state of being away) is not
  - *for two hours* selects for unbounded predicates, *again* does not
  - *yağız* 'dark brown' is preferably used for animals

rijksuniversiteit groningen

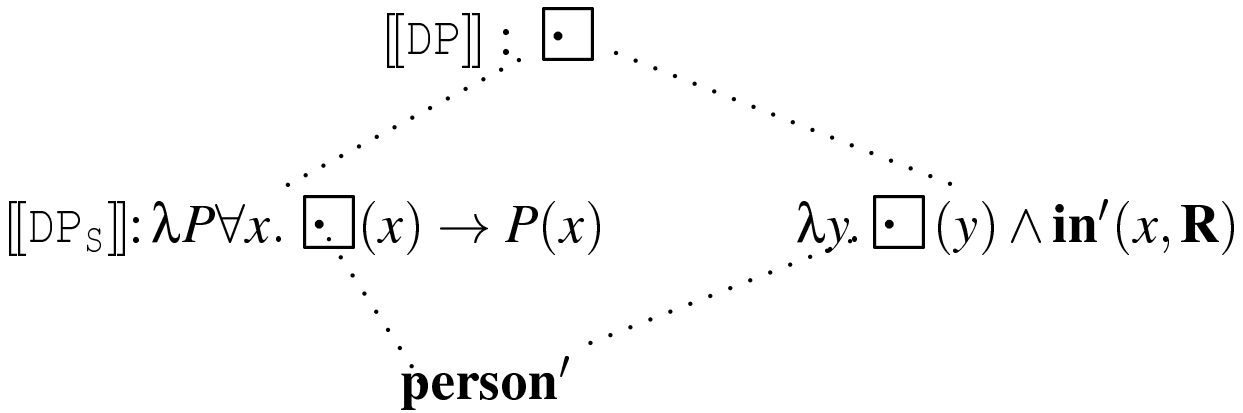
Nontrivial cases and structural ambiguities 2

- these structural ambiguities have already been successfully described in underspecification formalisms
- these formalisms allow the formulation of very powerful interfaces that can mediate between structural differences of syntactic and semantic structures
- these formalisms were designed to be used with surface-oriented syntactic analyses such as HPSG (Pollard and Sag, 1994) or LFG (Dalrymple, 2001)
- this suggests using them for the (closely related) unambiguous challenging cases of semantic construction, too
- they are needed anyway, so we can reuse them at no extra cost

Describing the nontrivial cases 1

- describing (16) in an underspecification formalism (basic intuition)

  (15)



  - three ingredients: fragments of λ-terms, 'holes', and relations between holes and fragments (depicted by dotted lines)
  - meta-level description of object-level semantic representations ('solutions')
  - deriving solutions by putting together the fragments ('jigsaw puzzle')

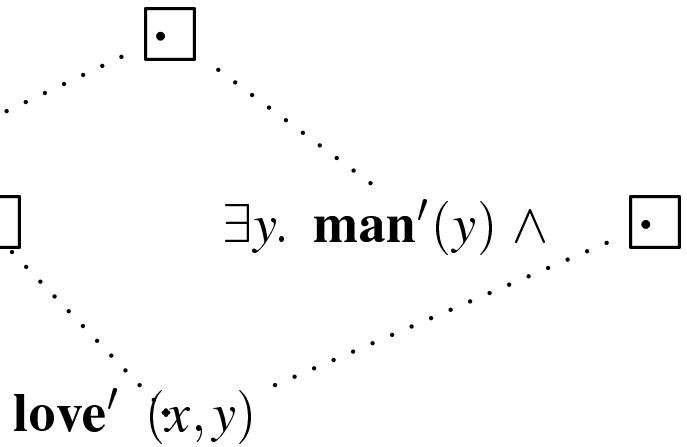
- (15) has only (16) as a solution

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

Describing the nontrivial cases 2

- compare this to the underspecified representation of (17)

  (17) *Every woman loves a man*

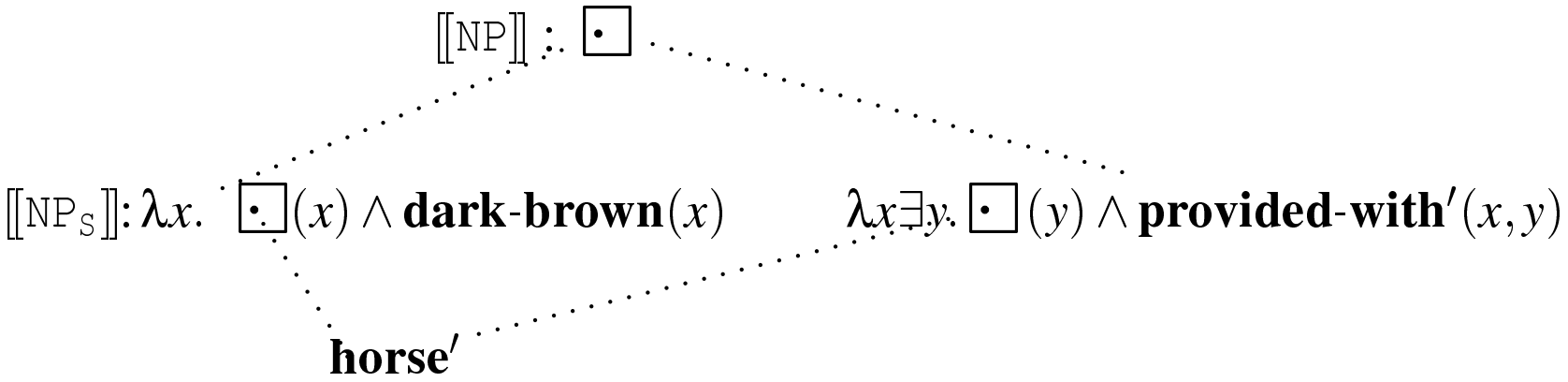
(18) 

- here, there are two readings because the scopally ambiguous material can be arranged in either way

Describing the nontrivial cases 3: *yağız atlı*

- semantics of *yağız atlı*

  (19)



- the sole solution of (19)

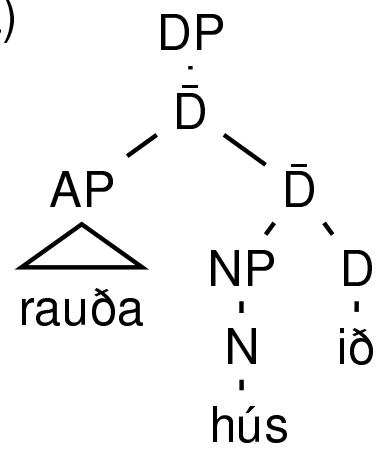
  (20) $\lambda x \exists y.\mathbf{horse}'(y) \wedge \mathbf{dark\text{-}brown}'(y) \wedge \mathbf{provided\text{-}with}'(x,y)$

- replacing *yağız* ‘dark brown’ by *genç* ‘young’ would give rise to ambiguity

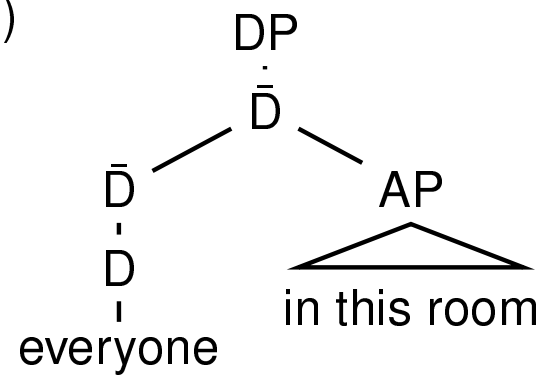
Describing the nontrivial cases 4: *rauða húsið*

- syntactic structure of *rauða húsið* and *everyone in this room*

(21) (a)



(b)

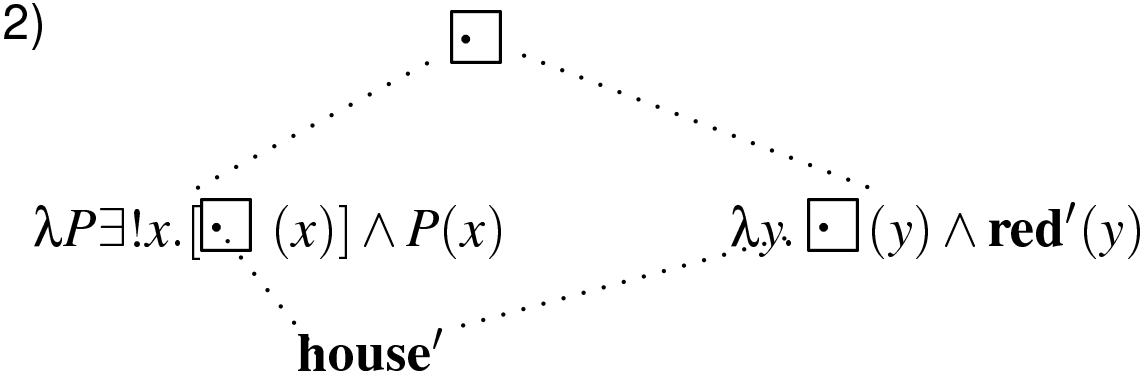


- almost identical syntactic structure, excepting
  - the ordering of D̄ and AP
  - the inner structure of the D̄ element

Describing the nontrivial cases 5: *rauða húsið*

- analogous semantic representation in (15) and (22)

(22)



- sole solution:
  - $\lambda P \exists !x.[\mathbf{house}'(x) \wedge \mathbf{red}'(x)] \wedge P(x)$

Summary

- challenging cases of semantic construction exhibit considerable mismatch between syntactic and semantic structure
- this looks like an argument in favour of syntactic analyses that do part of the mapping in the syntax
- their similarity to cases of structural ambiguity makes possible a treatment in terms of the (underspecified) approaches to structural ambiguity and the related interfaces
- these challenging cases cannot be used as straightforward arguments for or against specific syntactic analyses

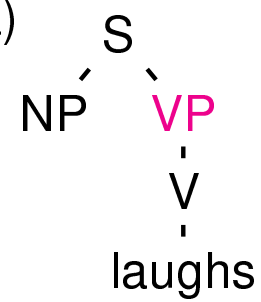
Related work 1: the Turkish LFG

- part of the ParGram project (Çetinoğlu and Oflazer, 2006)
- addresses semantic construction for cases like (14) and (2) in terms of 'inflectional groups'
- as nodes in the constituent structure they are accessible syntactically for processes like modification
- problem: they assume too few of these inflectional groups to cover semantic construction in Turkis fully, and assuming more of these groups would blur the boundary of morphology and syntax

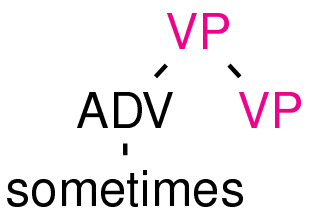
Related work 1: Lexical Tree-Adjoining Grammar

- the approach to scope underspecification in L-TAG Joshi et al. (2007) can be used directly for a representation of cases such as (3)
- heads introduce the whole subtree for their own projection and its semantics
- adjunction is modelled by replacing an internal node by a tree fragment, i.e., the original tree is split in two parts, and then the tree fragment is inserted between these two parts

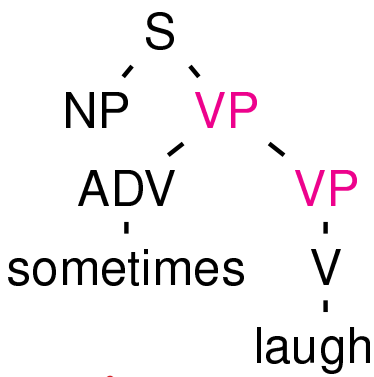
(23) (a)



(b)



(c)



Related work 2: Lexical Tree-Adjoining Grammar

- relations between syntactic nodes define semantic scope in the usual way (mostly, c-command)
- then the adjoined modifier gets intermediate scope, it outscopes only the tree part below the node replaced during adjunction (in (23a), VP)
- this intermediate scope is motivated syntactically, but to be specified in the semantics of the adjoinable tree: what is the semantics of the lower tree part?
- this parallels the anticipation of intermediate modifier scope in my interface rules

Related work 3: Lexical Tree-Adjoining Grammar

- L-TAG entry for *laugh* (Joshi et al., 2007; Kallmeyer and Romero, 2008):

(24)

$$l_1 : laugh\left(\boxed{2}, \boxed{1}\right): \left[\begin{array}{ll} \text{NP} & \left[\text{TOP} \mid \text{INDEX} \quad \boxed{1}\right] \\ \text{VP} & \left[\begin{array}{ll} \text{TOP} & \left[\begin{array}{ll} \text{PROPOSITION} & \boxed{4} \\ \text{SITUATION} & \boxed{3} \end{array}\right] \\ \text{BOTTOM} & \left[\begin{array}{ll} \text{PROPOSITION} & l_1 \\ \text{SITUATION} & \boxed{2} \end{array}\right] \end{array}\right] \end{array}\right]$$

Related work 4: Lexical Tree-Adjoining Grammar

- an analogous entry for *húsið*:

(25)

$$l_1{:}\mathit{the}\left(\boxed{1},l_2,\boxed{3}\right),l_3{:}\mathit{house}\left(\boxed{2}\right):\left[\begin{array}{ll}\text{D} & \left[\text{TOP}\ \left[\begin{array}{ll}\text{PROP} & l_1\\ \text{INDEX} & \boxed{1}\end{array}\right]\right]\\ \text{NP} & \left[\begin{array}{ll}\text{TOP} & \left[\begin{array}{ll}\text{PROP} & l_2\\ \text{INDEX} & \boxed{1}\end{array}\right]\\ \text{BOTTOM} & \left[\begin{array}{ll}\text{PROP} & l_3\\ \text{INDEX} & \boxed{2}\end{array}\right]\end{array}\right]\end{array}\right]$$

- difference: L-TAG approaches base the intermediate scope eventually on syntactic adjunction structures
- intuitive for *everyone* and *húsið*, much less so for change-of-state verbs

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

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

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

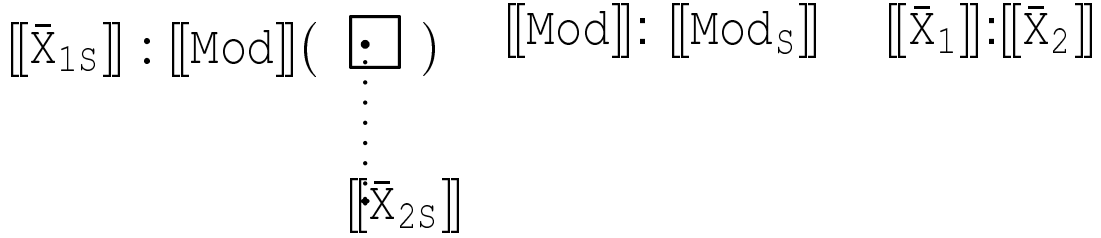
- '$[\![\mathtt{C}]\!] : \mathtt{F}$' expresses that the main fragment of C is defined as fragment F

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

The interface rules 2

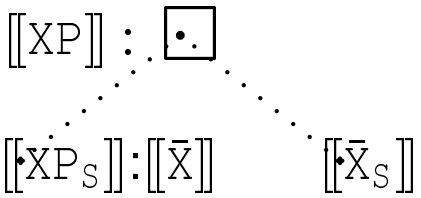
- the rule for modification

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



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

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



The interface rules 3

- semantic construction for *everyone in this room*
  - semantic representations (30) [= (26)] and (31) of pronoun (and $\bar{\text{D}}$) and PP

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

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

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

  - result of the modification rule (28)

(32) $[\![\bar{\text{D}}]\!] : \lambda P \forall x. \boxed{\bullet}(x) \rightarrow P(x)$ $\qquad$ $[\![\bar{\text{D}}_S]\!] : \lambda x. \boxed{\bullet}(x) \wedge \mathbf{in}'(x, \mathbf{R})$

$\mathbf{person}'$

  - rule (29) adds the upper half of the dominance diamond (15)

The analysis 1: *yağız atlı*

- rule (34) describes the semantic effect of affixing *-lı* to a nominal base

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

- (34) is the affix semantics

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

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

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

rijksuniversiteit groningen

The analysis 4: The *ip*-construction

- the interface rule

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

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

- constraint for (2)

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

  $\vdots$

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

- solution of this constraint

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

rijksuniversiteit groningen

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

- simplified tense account
- constraint for (2)

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

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

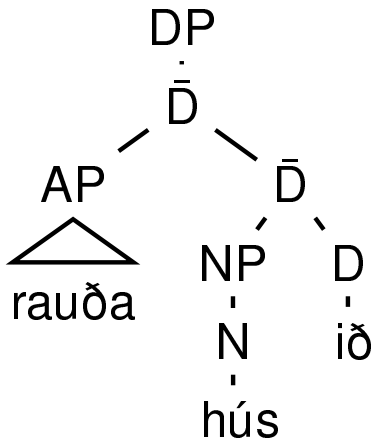
- solution of this constraint

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

The analysis 5: *rauða húsið*

- syntactic structure of *rauða húsið*

(41)



- analogous semantic construction pattern as in (47)

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

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

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

- difference: the modified expression is syntactically complex

Related work 1: the Turkish LFG

- part of the ParGram project (Çetinoğlu and Oflazer, 2006)
- addresses semantic construction for cases like (14) and (2)

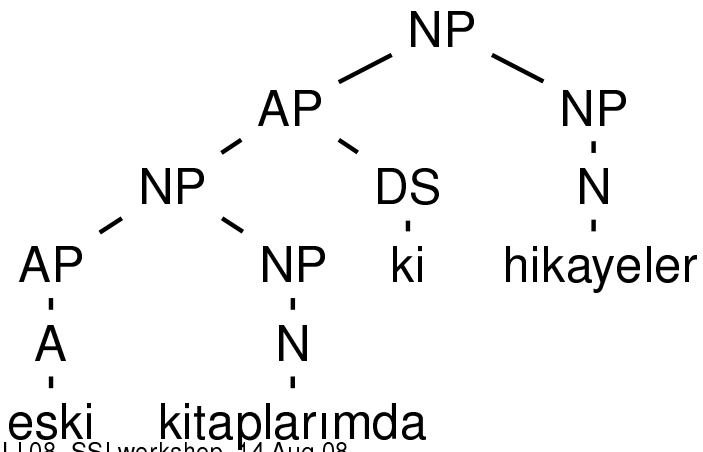
  (45) *eski* *kitap* *-lar* *-ım* *-da* *-ki* *hikaye* *-ler*

  old book PL my LOC KI story PL

  'the stories in my old books'

- strategy: 'inflectional groups' below the word level as nodes in the constituent structure (thus accessible for processes like modification already in the syntax)

(46)



Related work 2: the Turkish LFG

- *kitaplarımda* 'in my books' is an inflectional group
- the element *-ki* 'related to' is singled out in a constituent 'DS' (derivational suffix)
- *-ki* is no standard derivational suffix (e.g., no vowel harmony; Kornflit 1997)
- but: this is only ad hoc in that *eski* 'old' pertains to the root *kitap* 'book' exclusively
  - it is in the scope over the definiteness (from the possessive)
  - it is in the scope of the dative *-da*, otherwise the NP would mean 'old item(s) located in my books'
- i.e., there would have to be much more inflectional groups, which blurs the boundary of morphology and syntax
- affixes like *-li* undergo vowel harmony and precede inflectional affixes, they are thus less amenable to a syntactic treatment than *-ki*

rijksuniversiteit groningen

1 Conclusion and outlook

In this paper, I showed that underspecified approaches to semantics, which emerged as attempts to handle ambiguity in natural language, lend themselves for the derivation of difficult cases of semantic construction on the basis of surface-oriented syntactic analyses. The very powerful syntax-semantics interfaces in these approaches provide the necessary machinery to handle these cases of semantic construction. I have reviewed a number of such cases and outlined their semantic construction, comparing the proposed analysis to other competing approaches.

The increasing (re-)use of underspecified approaches to ambiguity for semantic construction is highly relevant for the question of what syntax is necessary from a semantic point of view, because powerful and flexible syntax-semantics interfaces can do a lot of the work of semantic construction themselves, and are less dependent on specific preprocessing of syntactic structures, as e.g. offered in the Logical Form of Generative Grammar. This development

is not restricted to scope-related issues as the ones presented in this paper, it is a general trend in the field, which is for instance also present in the analysis of negative concord in Richter and Sailer (2006), which makes use of techniques that allow the representation of specific ambiguities in Afrikaans tense marking (Sailer, 2004).

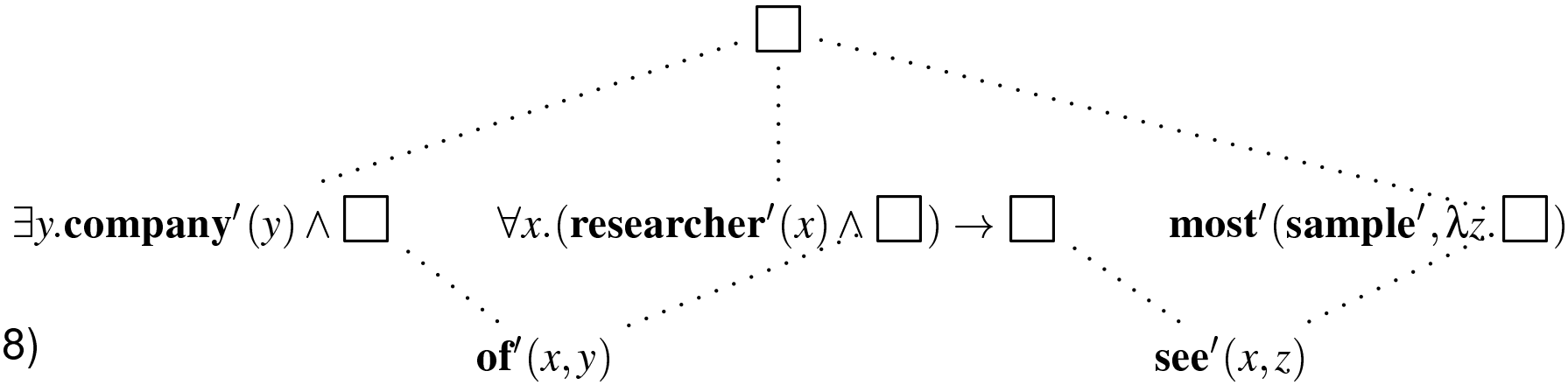
Underspecification in semantics 1

- deliberate omission of information in linguistic descriptions to capture alternative realisations of a linguistic phenomenon in one representation
- mostly, a meta-level is introduced to describe (rather than enumerate) object-level representations
- underspecification emerged in phonology, caught on in semantics in the 1980's
- typically used to model structural ambiguity, in particular, of scope relations
- a host of underspecification formalisms has been developed and coupled to various syntactic approaches

Underspecification in semantics 2: example

(47) *Every researcher of a company saw most samples*

(48)

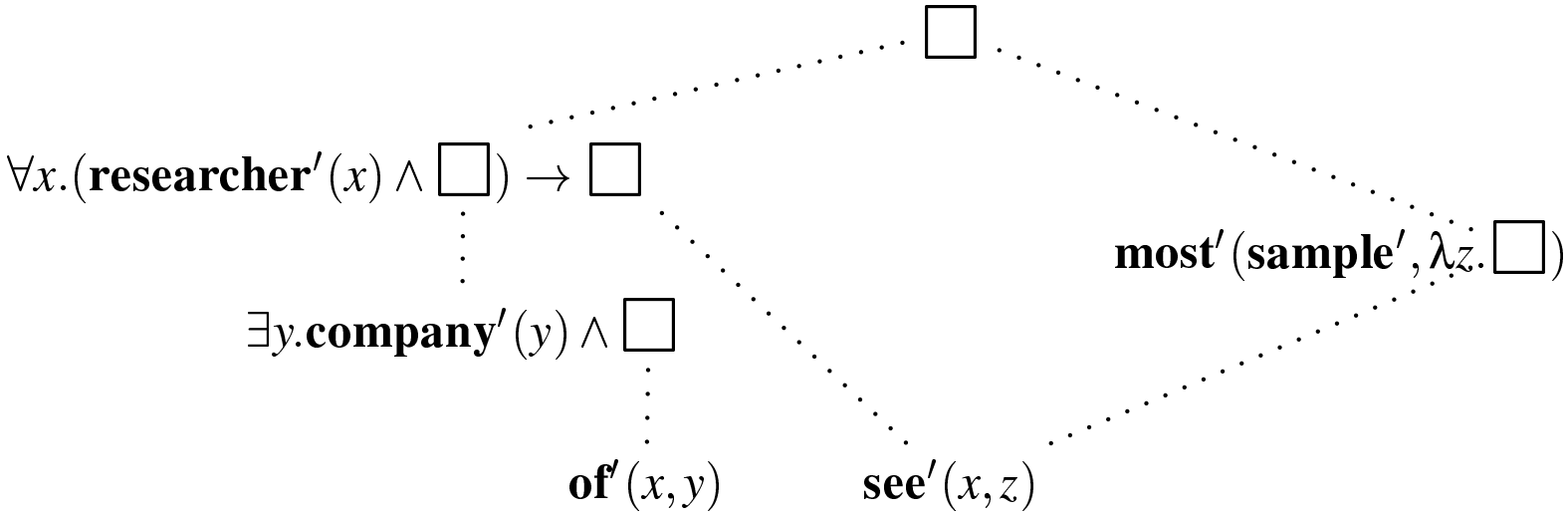


- (48) describes 5 semantic representations
- this configuration can be implemented in different ways

Underspecification in semantics 2a

- first partial disambiguation

(49)

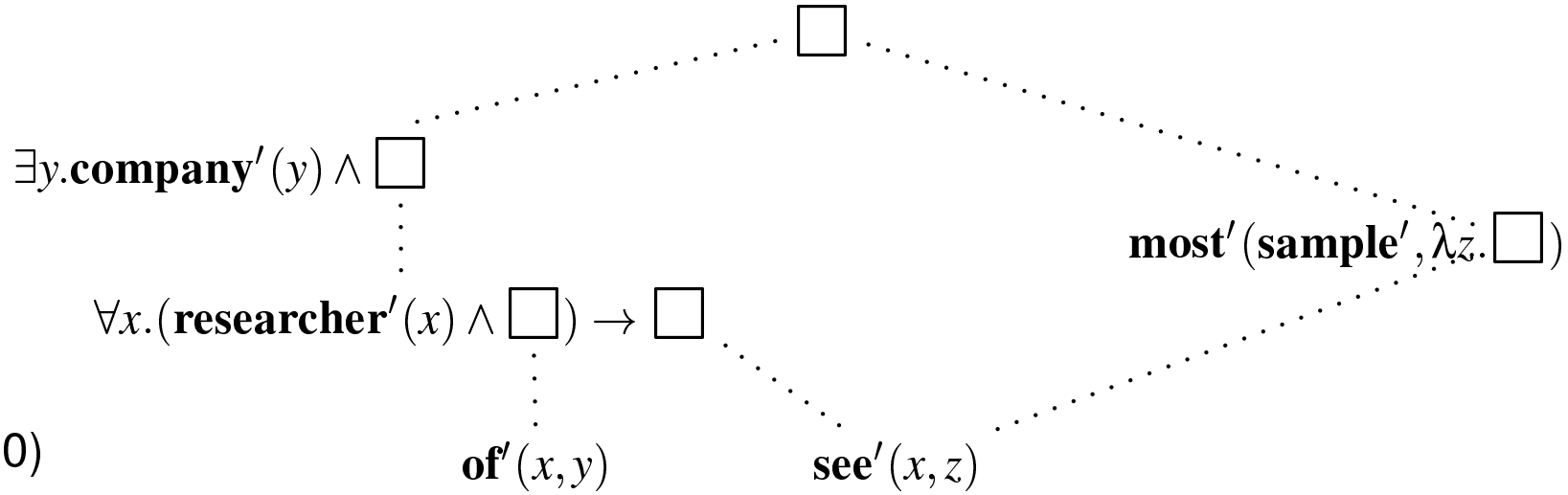


- two readings

Underspecification in semantics 2b

- second partial disambiguation

(50)



- three readings

Underspecification in semantics 3

- how many readings has (47)?
  - consensus: there is no reading $\forall > \mathbf{most}' > \exists$ (Hobbs and Shieber (1987); Egg et al. (2001); Bos (2004),)
  - but is there a reading $\exists > \mathbf{most}' > \forall$ (Park (1995); Kallmeyer and Romero (2008),)?
- this brings in the issue of expressivity König and Reyle (1999); Ebert (2005)
  - formalisms must be able to express any subset of readings of an ambiguous expression
  - but how to block the reading $\exists > \mathbf{most}' > \forall$ in terms of a configuration like (48)?

Underspecification in semantics 4

- theoretically, the number of readings seems tractable (low number of scope-bearing items in a sentence)
- but in real-life applications, scope ambiguities abound Koller and Thater (2006)
- the record-bearer from the Rondane Treebank (2.4×10^{12} readings)

  (51) *Myrdal is the mountain terminus of the Flåm rail line (or Flåmsbana) which makes its way down the lovely Flåm Valley (Flåmsdalen) to its sea-level terminus at Flåm.*

- advanced methods of resolving underspecified representations and/or redundancy elimination are called for (Alshawi (1992); Koller et al. (2008),)
- this is the problem of efficiency that will resurface in underspecified approaches to discourse structure

rijksuniversiteit groningen

References

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

Alshawi, H. (ed.) (1992). *The Core Language Engine*. Cambridge: MIT Press.

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

Butt, M. and T. H. 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 International Conference on Computational*

Linguistics and the 44th annual meeting of the ACL, 153–160. Association for Computational Linguistics.

Copestake, A., D. Flickinger, C. Pollard, and I. Sag (2005). Minimal Recursion Semantics. An introduction. *Research on Language and Computation 3*, 281–332.

Dalrymple, M. (2001). *Lexical Functional Grammar*. Number 34 in Syntax and Semantics. New York: Academic Press.

Dalrymple, M., J. Lamping, F. Pereira, and V. Saraswat (1997). Quantifiers, anaphora, and intensionality. *Journal of Logic, Language, and Information 6*, 219–273.

Ebert, C. (2005). *Formal investigations of underspecified representations*. Ph. D. thesis, King's College, London.

Egg, M. (2004). Mismatches at the syntax-semantics interface. In S. Müller

(ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 119–139. Stanford: CSLI Publications.

Egg, M. (2008). Reference to embedded eventualities. In J. Dölling, T. Heyde-Zybatow, and M. Schäfer (eds), *Event structures in linguistic form and interpretation*, 149–172. Berlin: de Gruyter.

Egg, M., A. Koller, and J. Niehren (2001). The Constraint Language for Lambda-Structures. *Journal of Logic, Language, and Information 10*, 457–485.

Hobbs, J. and S. Shieber (1987). An algorithm for generating quantifier scoping. *Computational Linguistics 13*, 47–63.

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

Kishimoto, H. (2000). Indefinite pronouns and overt N-raising. *Linguistic Inquiry 31*, 557–566.

Koller, A., M. Regneri, and S. Thater (2008). Regular tree grammars as a formalism for scope underspecification. In *Proceedings of ACL-08*. To appear.

Koller, A. and S. Thater (2006). An improved redundancy elimination algorithm for underspecified descriptions. In *Proceedings of COLING/ACL-2006*, Sydney.

König, E. and U. Reyle (1999). A general reasoning scheme for underspecified representations. In H. J. Ohlbach and U. Reyle (eds), *Logic, Language and Reasoning. Essays in Honour of Dov Gabbay*, 1–28. Dordrecht: Kluwer.

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

Larson, R. (1998). Events and modification in nominals. In D. Strolovitch and A. Lawson (eds), *Proceedings from SALT VIII*, Ithaca, 145–168. CLC Publications.

Park, J. (1995). Quantifier scope and constituency. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 05)*, 205–212.

Pollard, C. and I. Sag (1994). *Head-driven Phrase Structure Grammar*. CSLI and University of Chicago Press.

Reyle, U. (1993). Dealing with ambiguities by underspecification: construction, representation, and deduction. *Journal of Semantics 10*, 123–179.

Richter, F. and M. Sailer (2006). Modeling typological markedness in semantics: the case of negative concord. In S. Müller (ed.), *Proceedings of the HPSG 06 Conference*, Stanford, 305–325. CSLI Publications.

Rooth, M. and B. Partee (1982). Conjunction type ambiguity and wide scope "or". In *Proceedings of the First West Coast Conference on Formal Linguistics*, 353–362. Stanford University.

Sailer, M. (2004). Past tense marking in Afrikaans. In C. Meier and M. Weisgerber (eds), *Proceedings of Sinn und Bedeutung 8*, 233–248.

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