Comparison: range of potential solutions 1

- two ways of evaluating underspecified descriptions
  - assume that they explicitly mention all the material that may show up in solutions
  - assume that material not yet mentioned by them may show up in solutions
- in the first approach, deriving solutions can be regarded as turning a partial order of fragments into a total order
  - this is the strategy in UDRT (Reyle 1993)
  - it resembles the type-driven translation proposed in Klein and Sag (1985)
  - analogy: jigsaw puzzle
- in the second approach, solutions that make do with the material explicitly mentioned in the description are special ('constructive solutions')

Comparison: range of potential solutions 2

- the additional expressivity of the second approach is needed for metonymy and ellipsis
  - e.g., in metonymy, additional material shows up in the utterance meaning that is not contributed by any constituent of the utterance

    (93) Amélie labelled the wine

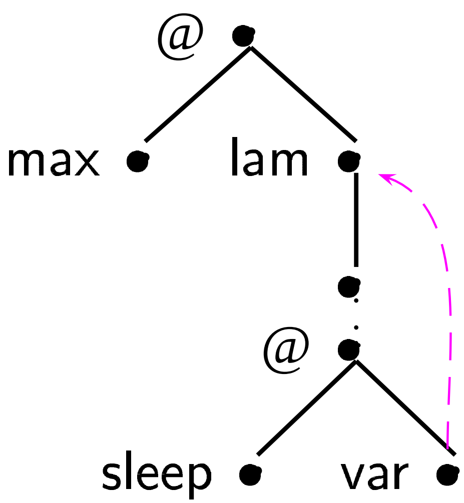
    (94) Der Wein steht auf dem Tisch (‘the wine is standing on the table’)
  - either sentence mentions containers for the wine (which are labelled and have a maximal axis, respectively)
  - not even in a ‘Generative Lexicon’ approach (Pustejovsky 1995) would a constituent of (93) or (94) introduce the notion of containers
- but for issues of scope underspecification, no additional material (beyond the one mentioned in the description) is needed

Comparison: integration of syntax and semantics

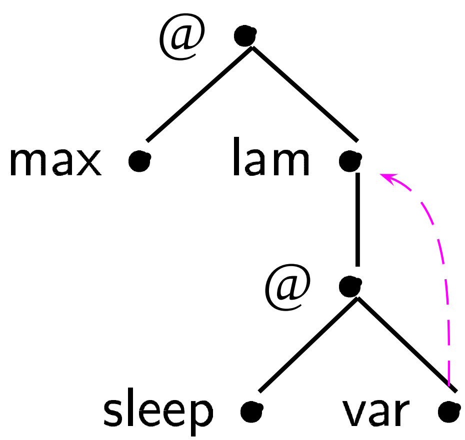
- semantic representations can be integrated to a different degree with the underlying syntactic representations
  - very strong integration in Lexical Resource Semantics (LRS) (Richter and Sailer 2003)
  - e.g., here λ-terms are rendered as typed feature structures, too
  - in CLLS, the integration is less strict
  - e.g., the resolution of ambiguous representations is handled by an external resolver

CLLS: comparison to MRS 1

- how to connect parts of representations
  - MRS: identify handle arguments in relations of MRS with handles of other relations
  - e.g., from $h_1 : \mathbf{max}'(h_n), h_2 : \mathbf{sleep}'$ to $h_1 : \mathbf{max}'(h_2), h_2 : \mathbf{sleep}'$
  - CLLS: identify unlabelled leaf nodes of fragments (maximal sets of nodes connected by immediate dominance) with roots of other fragments
  - e.g., from



to



CLLS: comparison to MRS 2

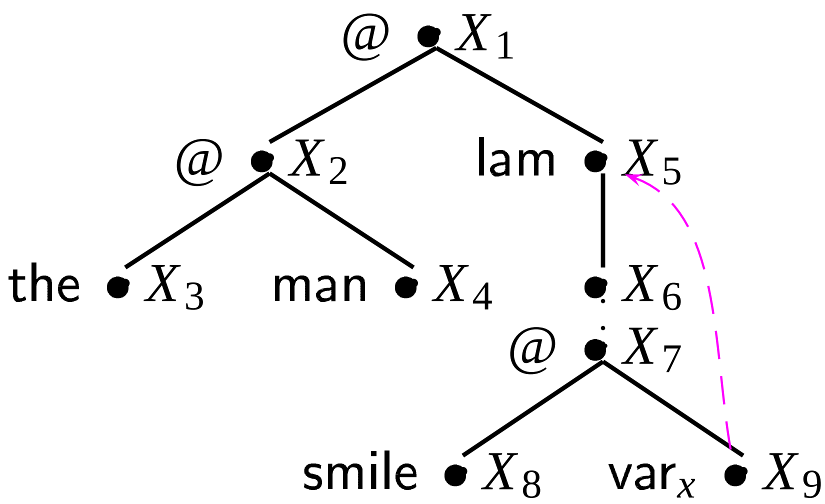
- how to rule out unwanted readings
  - MRS relies on the correctness of variable bindings and explicit dominance (qeq) constraints between handles
  - CLLS does not rely on correctness of variable binding: Every variable node is dominated by the node for its λ-binder
  - CLLS uses more dominance constraints to rule out unwanted readings than MRS, e.g., scope arguments of quantifiers are determined in CLLS but not MRS
  - (surfacy) generalisation of these observations
    - ∗ in MRS, there are several groups of constraints that block unwanted readings
    - ∗ in CLLS, the different constraints are more integrated

From CLLS to MRS: introduction

- intuitively, there is a lot of common ground between MRS and CLLS representations
- but some things are implemented differently (e.g., scope of quantifying expressions)
- Niehren and Thater (2003) specified a translation between MRS representations and CLLS-like dominance constraints (DCs)
- this translation only works for MRS representations that are nets, i.e., whose DC counterparts obey specific structural constraints (also called nets)
- first, a short sketch of the translation will be provided (including a definition of nets)
- second, the results of an evaluation of the translation will be discussed
- note the ambiguity of the expressions ‘MRS’ and ‘DC’

From CLLS to MRS: definition of nets 1

- fragments in a constraint are maximal sets of nodes that are pairwise connected by immediate dominance
  - e.g., there are two fragments in the following constraint:



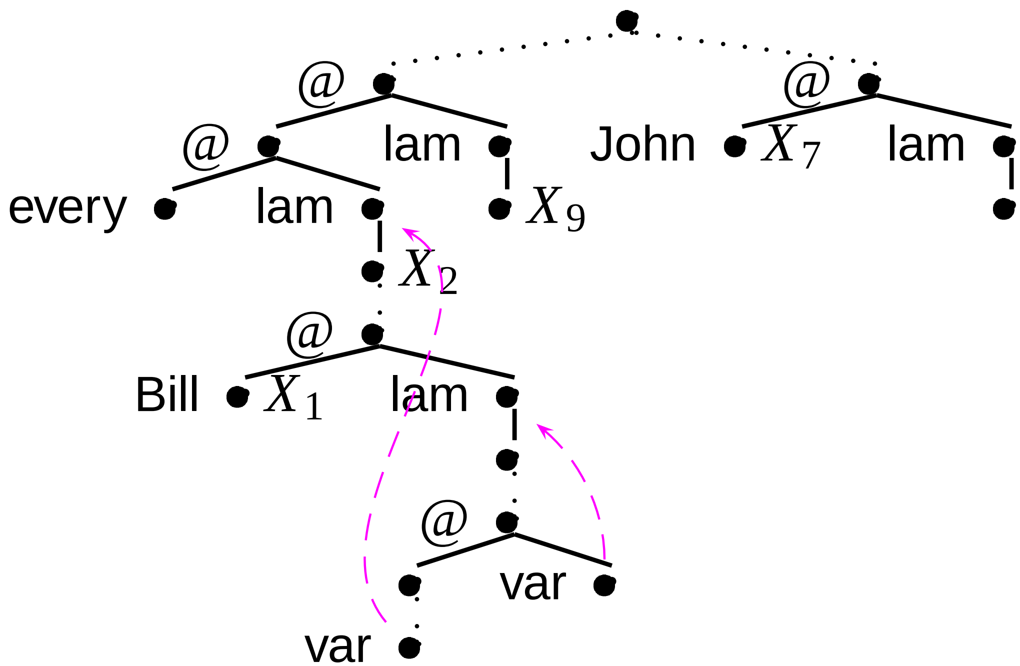
- the definition of nets characterises the fragments that may show up in them
- the definition is simplified in that compactness is left out of account (tree fragments would have to be condensed to have height 1)

From CLLS to MRS: definition of nets 2

- conditions on fragments in nets
  - all non-labelled leaves of a fragment ('holes') except the last one must dominate another fragment
  - the last hole dominates another fragment iff the root doesn't
    - ∗ dominance relations between roots do not occur in ordinary CLLS constraints
    - ∗ they only emerge as the result of rendering the effect of variable binding in MRS with the help of dominance constraints in the DC
  - if one hole dominates two different fragments, they are connected
    - ∗ roughly, the roots of the fragments are connected by a series of dominance relations (in either direction)
    - ∗ formally, they must be connected by a hypernormal path
- pictures . . .

From CLLS to MRS: definition of nets 3

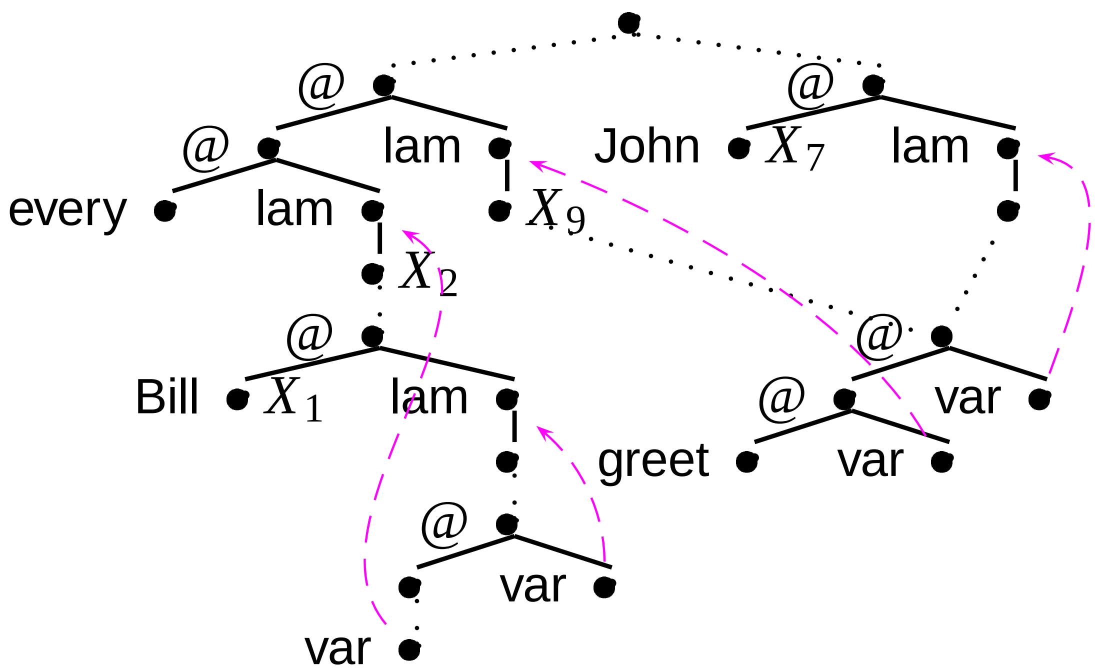
- this constraint is no net:



Egg & Kordoni, ESSLLI 2004, Nancy

From CLLS to MRS: definition of nets 4

- in contrast, the following constraint is a net:



From CLLS to MRS: preprocessing of MRSs

- to translate a MRS (structure) into a DC, it is preprocessed first
- preprocessing step 1 for relations with equated handles
  - introduce a new EP that makes the conjunction explicit
  - assign new handles to the EPs with the equated handles
  - formally: replace $h : EP_1, h : EP_2$ with $h : \&(h_1, h_2), h_1 : EP_1, h_2 : EP_2$, where h_1 and h_2 are fresh handles not yet used in the MRS that is to be translated
- preprocessing step 2: the MRS is then compactified
  - MRS fragments are maximal sets of EPs such that the address handle of one EP is the argument of another EP
  - these fragments would have to be condensed into a single EP

From CLLS to MRS: translation 1

- then the MRS literals are translated
  - MRS relations are translated into labelled nodes of CLLS:

$$h : P(x_1, \ldots, x_n, h_1, \ldots, h_m) \mapsto h : P_{x_1,\ldots,x_n}(h_1, \ldots, h_m)$$

  - in prose: the resulting node variable h is labelled by $P_{x_1,\ldots,x_n}$ and has the immediate daughter nodes $h_1, \ldots, h_m$
  - qeq relations between handles are translated into dominance constraints between node variables:

$$h =_q h' \mapsto h \triangleleft^* h'$$

- but recall that CLLS structures include more dominance constraints than there are qeq relations in MRSs

From CLLS to MRS: translation 2

- i.e., the effect of variable binding in MRS must be taken into account in the translation
  - variable binding enforces scope relations in MRS
  - if a quantifier EP_1 has the address handle h_1 and the bound variable v and another EP_2 has the address handle h_2 and v as an argument a dominance constraint is introduced:

    $h_1 \triangleleft^* h_2$
- normalisation: every $root_1$-to-$root_2$ dominance relation is replaced
- instead, introduce a dominance relation for the (necessarily existing) last hole on the fragment of $root_1$ that does not yet dominate anything
- the second argument of this dominance relation is the $root_2$ of the other fragment

From CLLS to MRS: example 1

(95) Every old dog barked

- MRS representation:

  (96) $\langle\{h_2 : \mathbf{every}'(x, h_3, h_4), h_5 : \mathbf{dog}'(x), h_5 : \mathbf{old}'(x), h_6 : \mathbf{bark}'(x)\}, \{h_0 =_q h_6, h_3 =_q h_5\}\rangle$

- after preprocessing:

  (97) $\langle\{h_2 : \mathbf{every}'(x, h_3, h_4), h_5 : \&(h_{11}, h_{12}), h_{11} : \mathbf{dog}'(x), h_{12} : \mathbf{old}'(x), h_6 : \mathbf{bark}'(x)\}, \{h_0 =_q h_6, h_3 =_q h_5\}\rangle$

From CLLS to MRS: example 2

- after the translation step:

(98) $\{h_2 : \mathbf{every}'_x(h_3, h_4),\ h_5 : \&(h_{11}, h_{12}),\ h_{11} : \mathbf{dog}'_x,\ h_{12} : \mathbf{old}'_x,\ h_6 : \mathbf{bark}'_x,$
$h_0 \lhd^* h_6, h_3 \lhd^* h_5, h_2 \lhd^* h_6\}$

- note that all the h_n are node variables and expressions like '$\mathbf{old}'_x$' are node labels

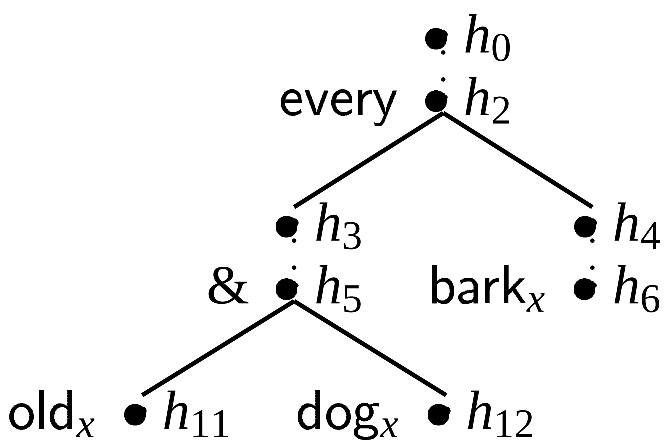
- the first two dominance constraints are direct translations of qeq relations, the last one, a result of MRS variable binding

From CLLS to MRS: example 3

- after normalisation:

(99) $\{h_2 : \mathbf{every'}_x(h_3, h_4),\ h_5 : \&(h_{11}, h_{12}),\ h_{11} : \mathbf{dog'}_x,\ h_{12} : \mathbf{old'}_x,\ h_6 : \mathbf{bark'}_x,$
$h_0 \triangleleft^* h_6, h_3 \triangleleft^* h_5, h_4 \triangleleft^* h_6\}\rangle$

- the resulting dominance constraint (assuming that h_0 must dominate any other node):



From CLLS to MRS: evaluation 1

- hypothesis of Niehren and Thater (2003): MRS structures are nets
- Fuchss et al. (2004) investigated this on the output of the English Resource Grammar (ERG; Copestake and Flickinger 2000) for the sentences of the Redwood Treebank (Oepen et al. 2002)
- 6230 of the 6612 sentences in the Redwoods Treebank could be parsed successfully by the ERG
- 83% of the resulting representations were nets
- tentative claim: the remaining 17% non-nets of the ERG evaluation are linguistically problematic

From CLLS to MRS: evaluation 2

- core problem for non-nets 1: numerical expressions

  (100) okay that would be one hundred and eighty-six euros

  – such sub-areas of languages (like expressions for times) tend to be problematic anyway
  – it might be best to relegate them to some special preprocessing step

- core problem for non-nets 2: coordination

  (101) a dog and a cat sleep

  – too many (sometimes also incorrect) readings, e.g., 16 for (101)
  – due to an attempt to cover the distributive/collective interpretation

CLLS and MRS: evaluation 3

- comparison of runtimes for solvers: the solver for CLLS is faster and exhibits a constant runtime per solution
- these comparison of runtimes should be taken to represent trends
- the evaluation was done with a meanwhile no longer current version of the ERG
- in the current version, much work has been done on the problematic issues, i.e., a repetition of the evaluation might lead to different results
- the practical use of this translation enterprise
  - make available to the HPSG community effective resolution techniques for MRS representations
  - make available to the CLLS community the large coverage of ERG