Scaling up a joint model of word meaning and sentence meaning: Situation Description Systems and the Visual Genome

Katrin Erk, University of Texas
joint work with Aurélie Herbelot

Logical form integrated with fine-grained word meaning representations

Variety of approaches, different in the problems they address.

Many use embeddings,
and many use probabilities or graded representations.

- Word meaning in context: Asher, Erk&Herbelot, Emerson
- Phrase meaning: Baroni et al, Sadrzadeh et al
- Quantifiers, negation, adjectives: Bernardy et al
- Word learning, vagueness: Larsson

Logical form integrated with fine-grained word meaning representations

Variety of approaches, different in the problems they address.

Many use embeddings,
and many use probabilities or graded representations.

- Word meaning in context: Asher, Erk&Herbelot, Emerson
- Phrase meaning: Baroni et al, Sadrzadeh et al
- Quantifiers, negation, adjectives: Bernar
- Word learning, vagueness: Larsson

This talk: taking the next step
in one of these approaches,
Situation Description Systems

Situation Description Systems: Influences on word meaning in context

Local semantic context: selectional constraints

> The player <u>caught</u> the **ball**.
>
> The committee <u>organized</u> a <u>**ball**</u>.

Also wider topical context, “scenario” (examples by Ray Mooney)

> The athlete <u>ran</u> to the **ball**.
>
> The débutante <u>ran</u> to the **ball**.

Same verb sense, so disambiguation must come from scenario

Situation Description Systems: Influences on word meaning on context

Selectional constraints and overall scenario can interact

The astronomer married the **star**.

Both factors must be active at the same time:
If the selectional constraint of "marry/Theme" were resolved first, there would not be a pun.

Situation Description Systems

- Content words evoke underlying lexicalized concepts that characterize their meaning in context
- Concepts tap into underlying scenarios
- Probabilistic model:
    - Concept has probability distribution over words that can express it: $P(w|c)$
    - Selectional constraint: Concept has probability distribution over concepts that can fill a role: $P(c_{arg} \mid c_{pred}, \text{role})$
    - Scenario has probability distribution over concepts that can appear in the scenario: $P(c \mid s)$
- Concept underlying "star" conditionally dependent on both selectional constraint and scenario

Scaling up Situation Description Systems

Previously:

- Toy-size system
- hand-constructed list of concepts and scenarios with their probabilities

Aims in scaling up:

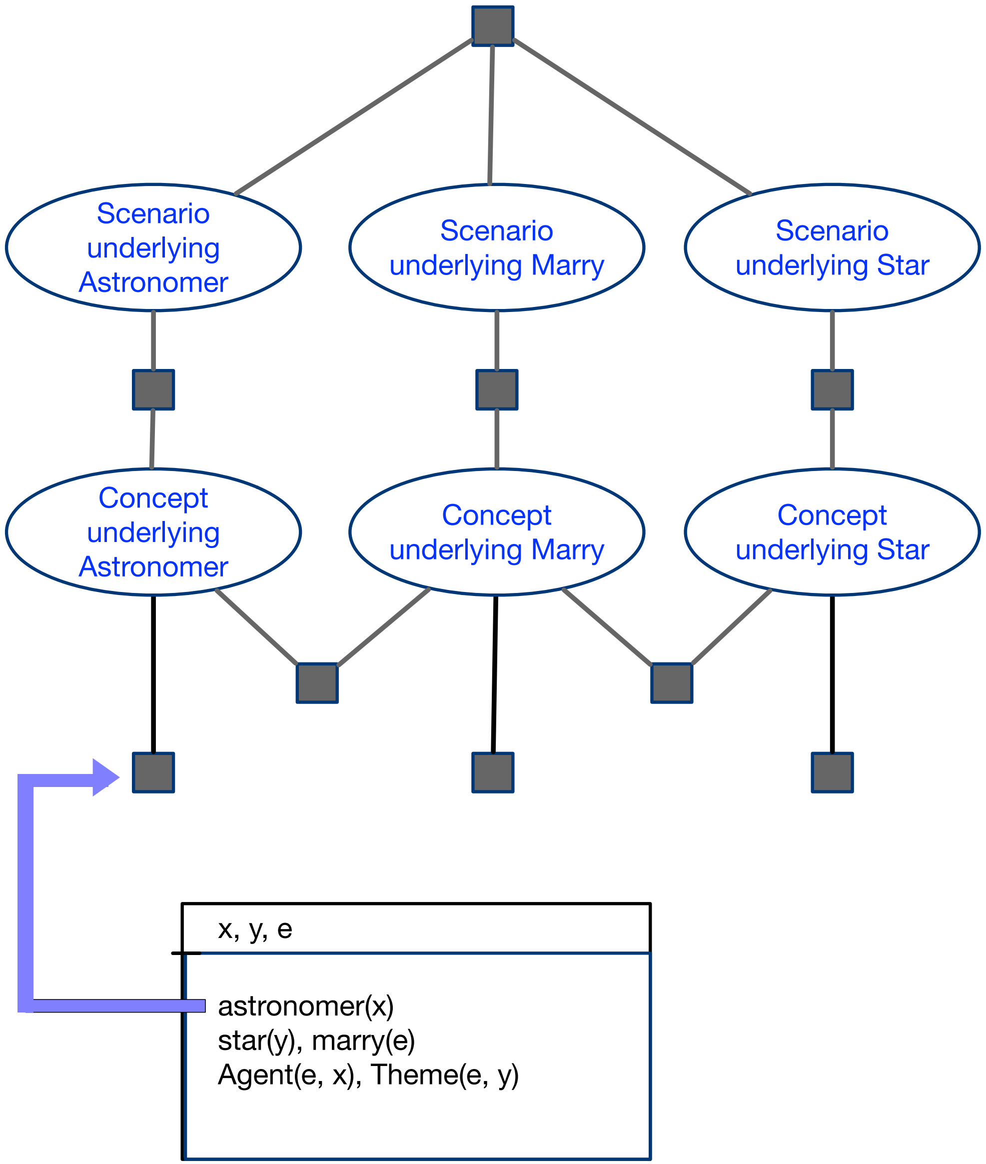
- Address the "vastness of the lexicon" (Baroni et al 2014): Learn concepts and scenarios from data
- For now, not end to end:
  - Implement specific algorithms for individual phenomena
  - Maybe an end-to-end variant in the future, for the question of learnability
- Extension: Interaction of reference and polysemy

Situation Description Systems as factor graphs

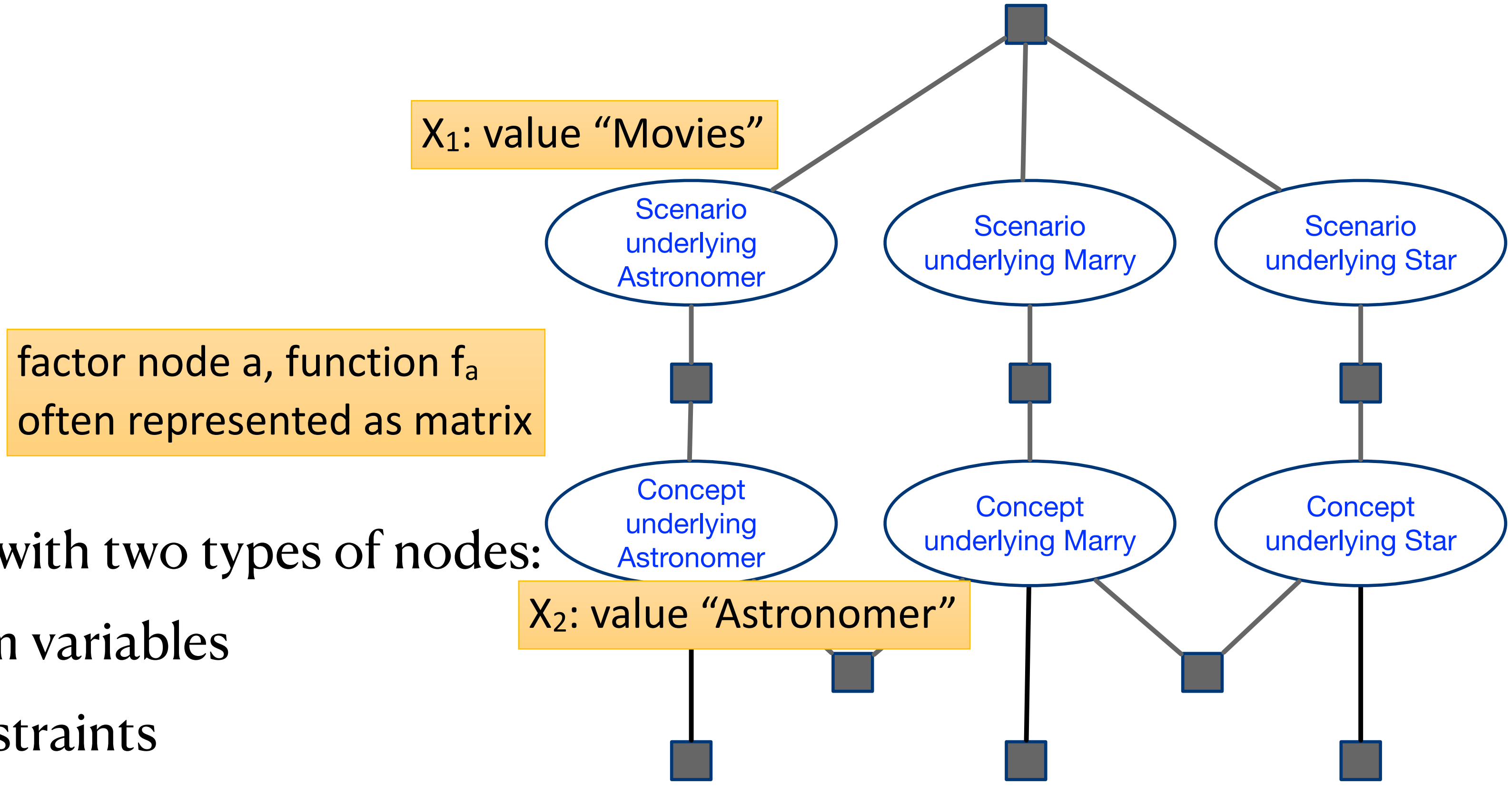
Situation descriptions: Now as factor graphs

Logical form plus conceptual graph

- Logical form: Discourse Representation Structure (DRS)
- Conceptual graph:
  For each lexical item in the utterance, a random variable for its underlying concept
- Concept value is constrained by factors, shown as black boxes in the graph



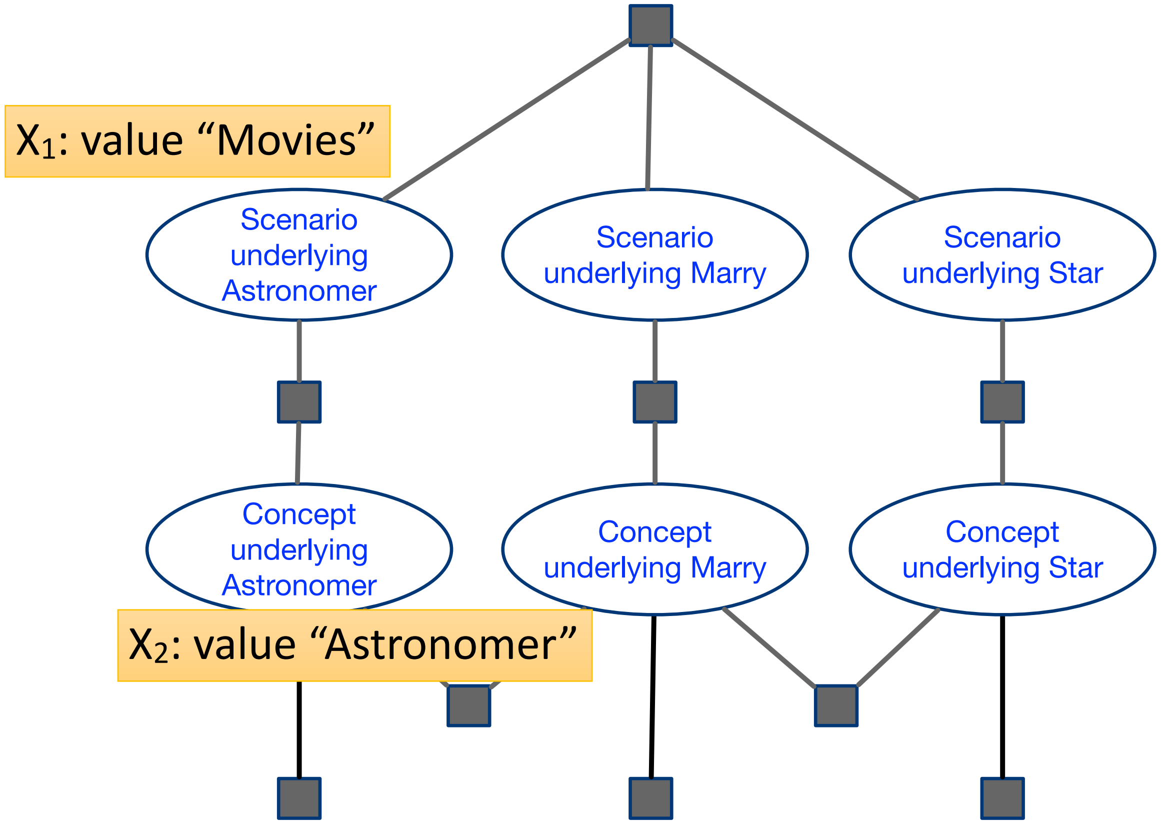
Factor graphs



- Probabilistic graphical model with two types of nodes:
  - Variable nodes are random variables
  - Factor nodes express constraints on adjacent variable nodes
- Factor node a has associated function f_a: values for all adjacent variables -> weight
  - Example for the factor node f_a from above:

f_a: P(concept|scenario)	**Movies**	**Space**
Astronomer	0	0.3
Marry	0.3	0.3
Star-person	0.3	0
Star-sun	0	0.3
Director	0.3	0

Factor graphs



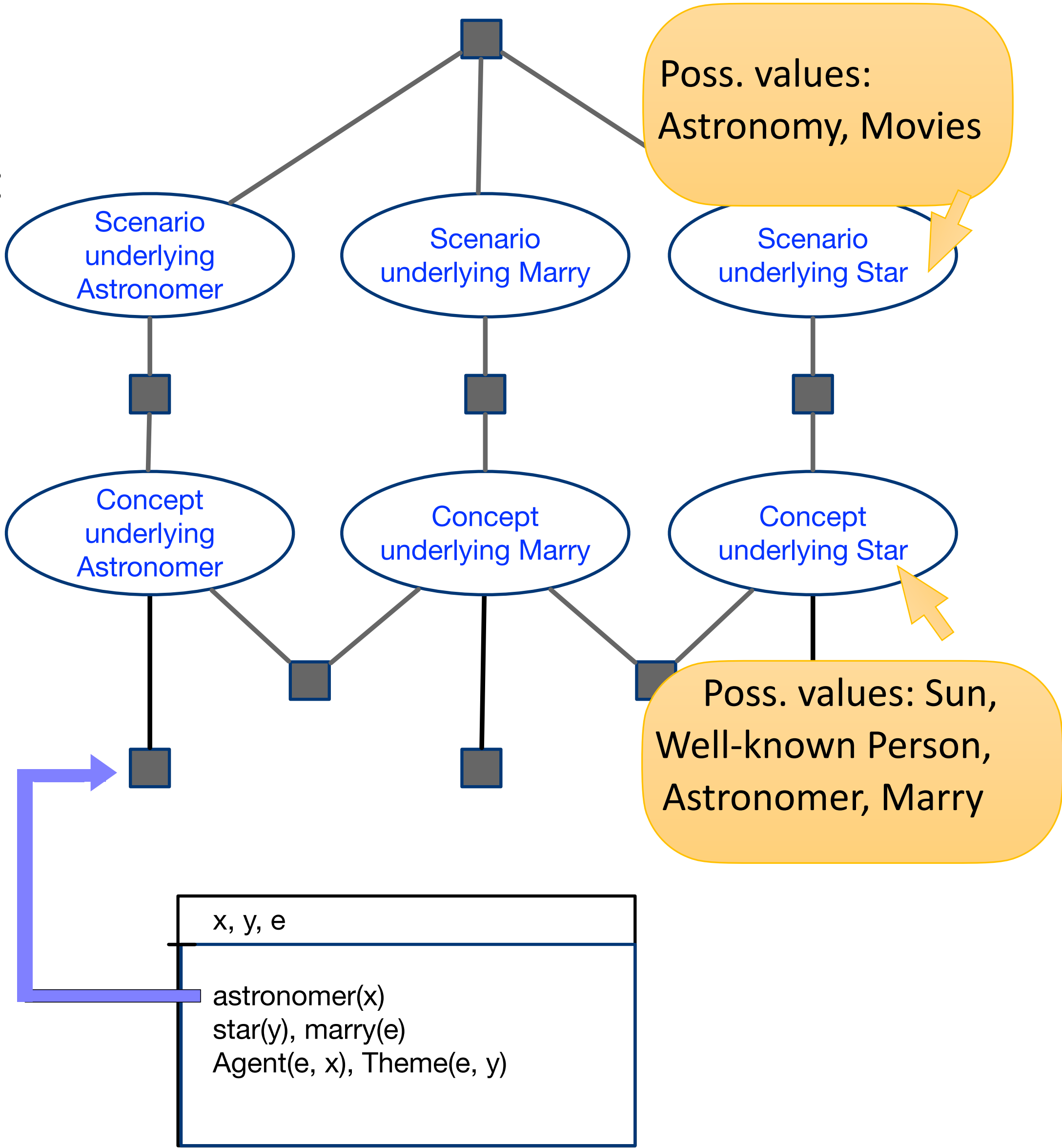
(Non-normalized) probability of assignment $\mathbf{x} = x_0..x_n$ to all variables $X_0..X_n$:

$$p(\mathbf{x}) = \prod_{a \in F} f_a(x_a)$$

where F set of factors, x_a assignments to variables adjacent to factor a

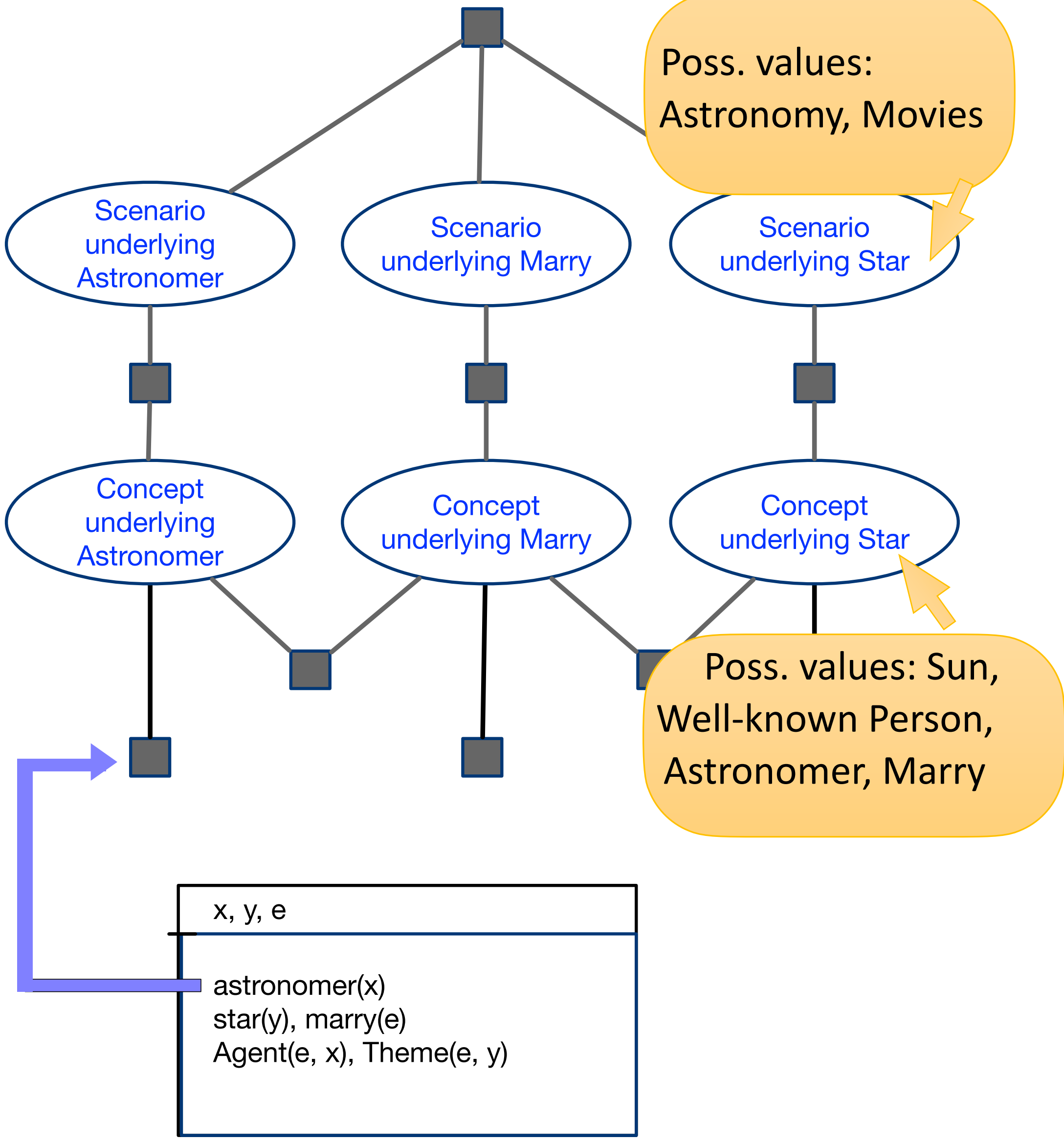
Random variables for concepts and scenarios

- Random variables that stand for concepts: Possible values are lexicalized concepts, like Sun, Well-known Person for “star”
- Random variables that stand for scenarios: Possible values are scenario types, like Astronomy, Movies
- Probabilistic assignment of values



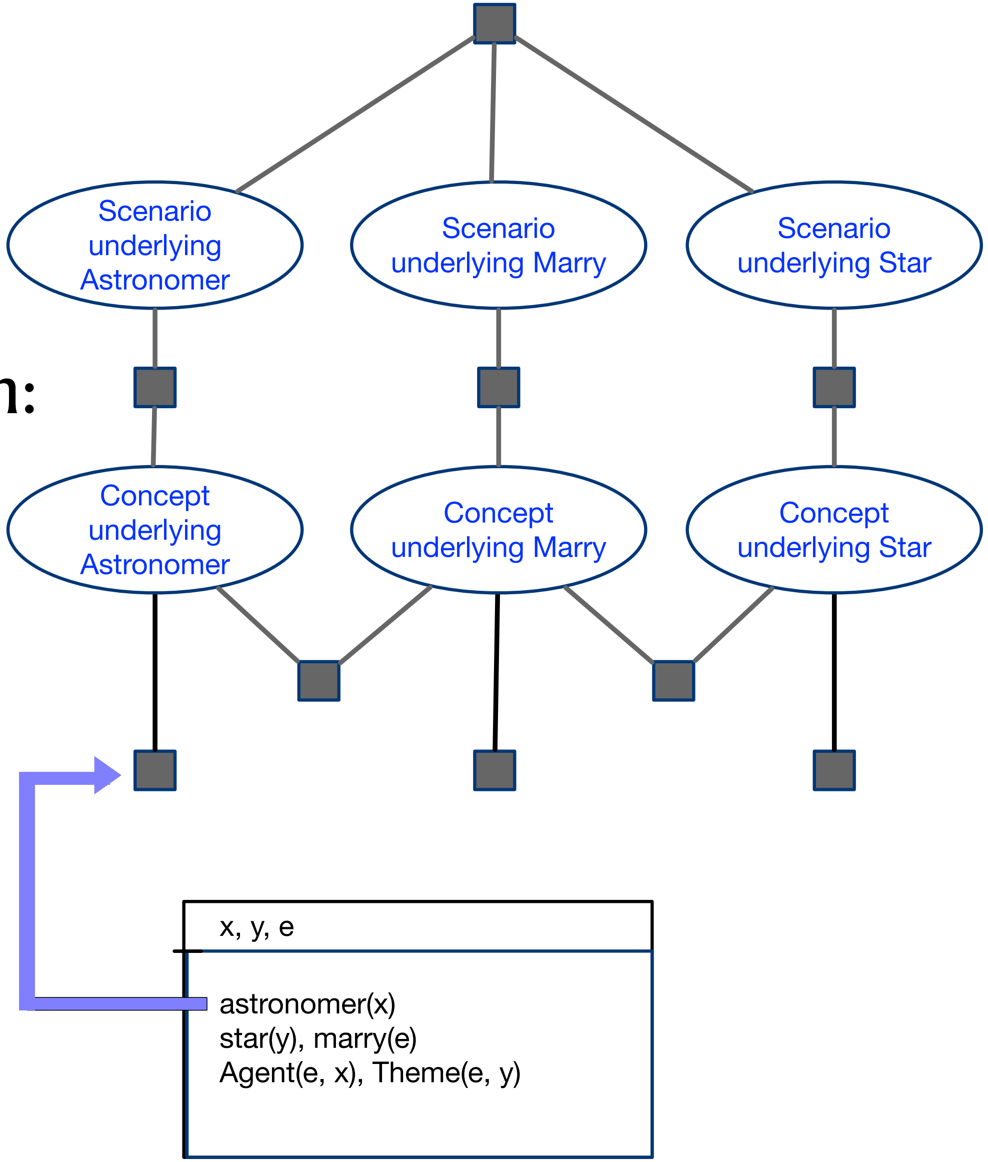
Probabilistic assignment of values

- Pun sentence "The astronomer married the star": Assign some probability mass to both Sun and Person
- "She revels in arguments, and loses no opportunity to declare her political principles":
  - Does she revel in quarrels, or in viewpoints pro and con, or in logical sequences of statements?
  - Maybe a bit of each: Assign some probability mass to each of the three senses



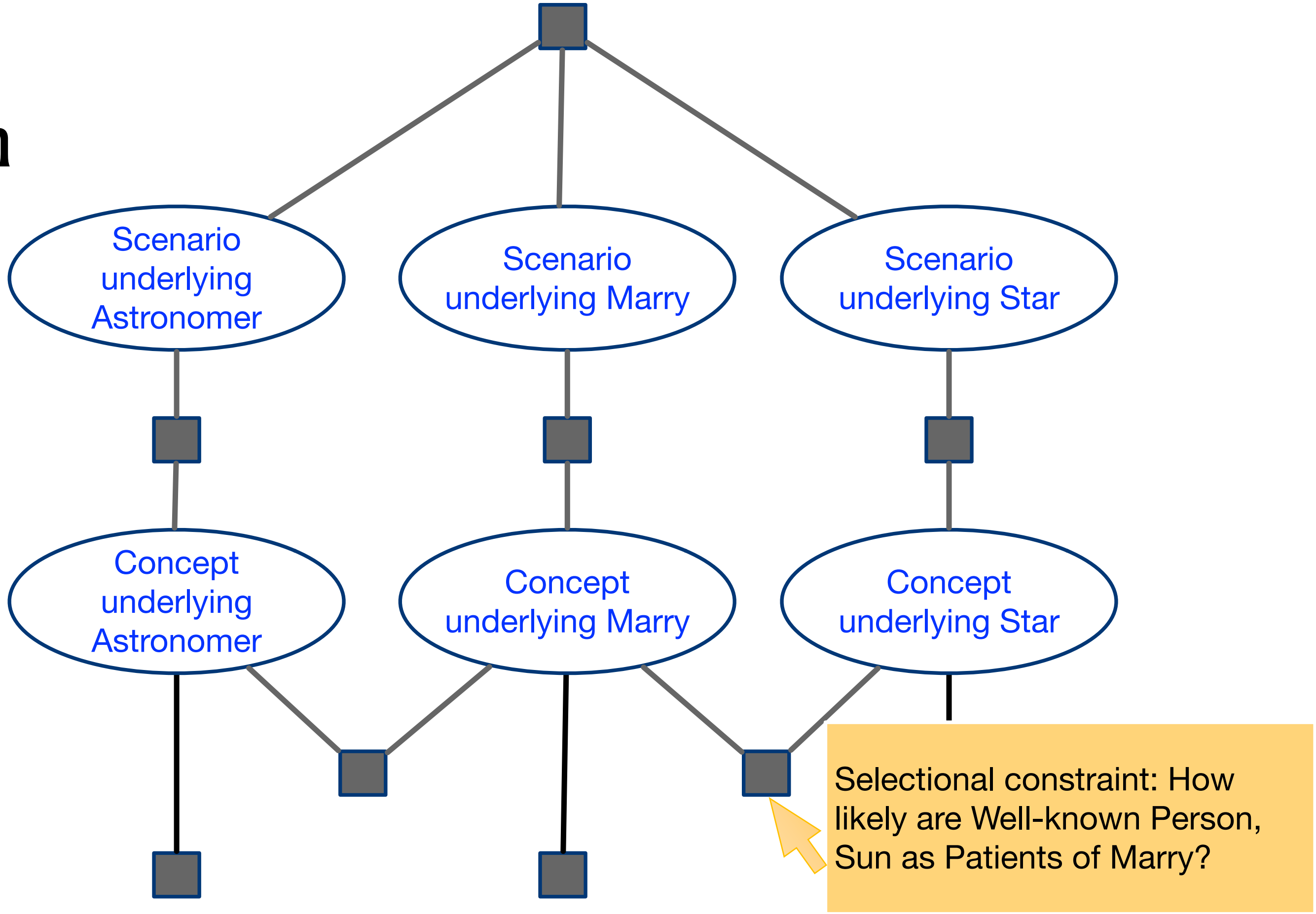
Constraints on word meaning in context as factors in the factor graph

- For every unary literal in the DRS, say astronomer(x):
  - Underlying concept for the lexical item: variable node
  - Factor constraining the concept node matching the observed predicate of the literal
- Observations here become unary factors



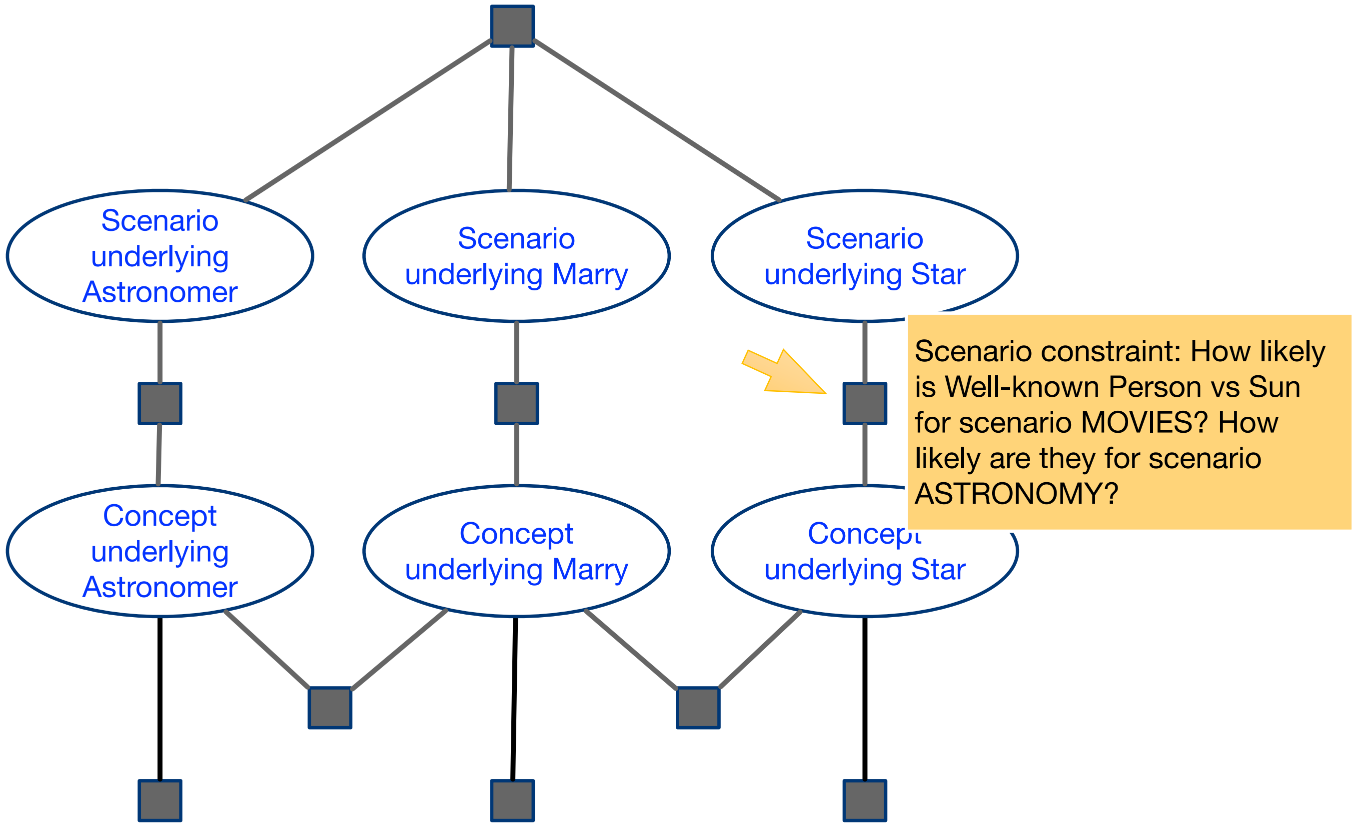
Constraints on word meaning in context as factors in the factor graph

- For every binary literal from the DRS, like Theme(e, y): binary factor implementing a selectional preference



Constraints on word meaning in context as factors in the factor graph

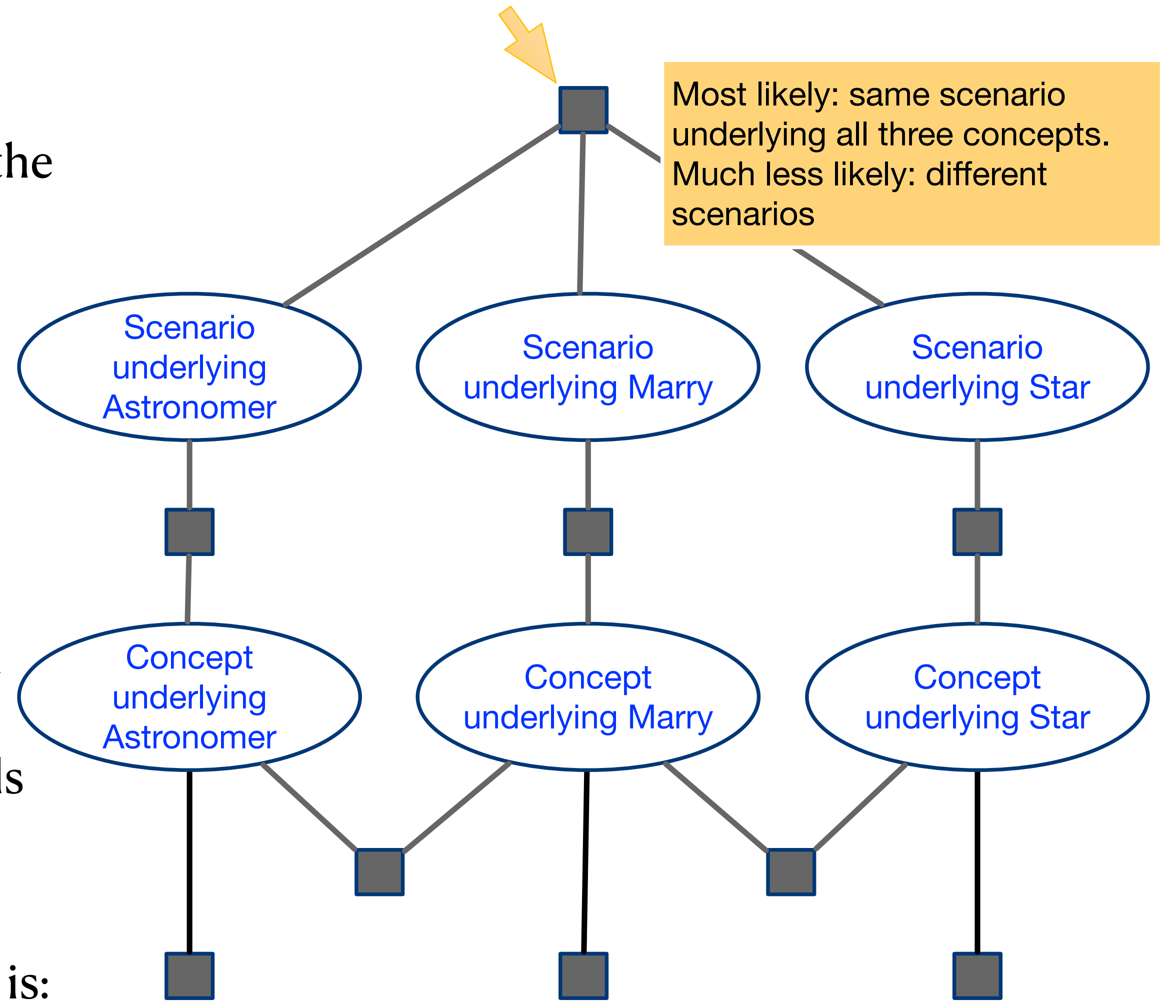
- For every variable node that stands for a concept, a variable node that stands for a scenario
- Factor:
  Sun concepts are more likely to appear in an Astronomy scenario,
  Well-known Person concepts are more likely to appear in a Movies scenario



Constraints on word meaning in context as factors in the factor graph

- One factor connecting all scenarios: More likely to have the same scenario underlying all three concepts, less likely to have different scenarios
- Modeling (adapting Latent Dirichlet Allocation, Blei 2003):
  - Multinomial distribution over scenarios for the sentence
  - Multinomial is drawn from a Dirichlet, a distribution over multinomials. Dirichlet parameter alpha: When alpha < 1, prefer to sample sparse multinomials
- Marginalizing over the multinomial distribution, all scenarios drawn from a Dirichlet-Multinomial: Probability of scenario vector **s**, with n scenarios overall, is:

$$P(\mathbf{s}, n, \alpha) = \int_p \mathrm{Mult}(\mathbf{s} \mid n, \mathbf{p})\ \mathrm{Dir}(\mathbf{p}|\alpha)\ d\mathbf{p}$$



Reference in Situation Description Systems

Definite descriptions disambiguated by entities in the context

- Copestake 1995

  Please sit in the apple juice chair
  attested by Downing 1977, "context where there was a table already set with a glass of orange juice by three places and apple juice by the fourth"

- Ambiguous: relation between the nouns
  $\mathrm{apple\_juice}(x) \wedge \mathrm{chair}(y) \wedge R(x, y)$

- Disambiguation comes from the (non-linguistic) context

Disambiguation by entities in the context

Latent relations:

- Copestake 1995: Please sit in the apple juice chair
  $\text{apple\_juice}(x) \wedge \text{chair}(y) \wedge R(x, y)$
- Similarly, Asher and Denis 2004:

  All the children were drawing fish. Suzie's salmon was blue.
  $\text{salmon}(y) \wedge R(x, \text{Suzie})$
- Similarly, McNally & Boleda 2015, "Conceptual versus referential affordance in concept composition":

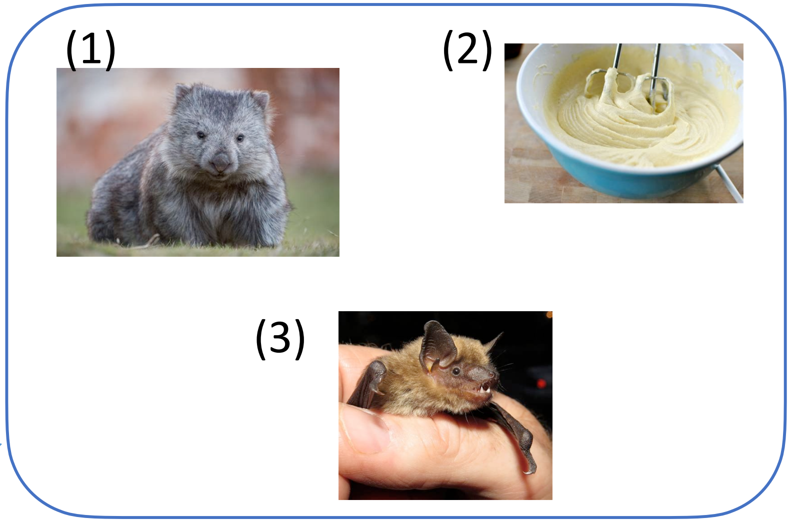
  Prince Edward and wife begin Canadian visit
  $\text{visit}(y) \wedge R(x, \text{Canada})$

Conclusion: Disambiguation may rely on reference, link to entity library / mental files.

So, do disambiguation and reference resolution concurrently

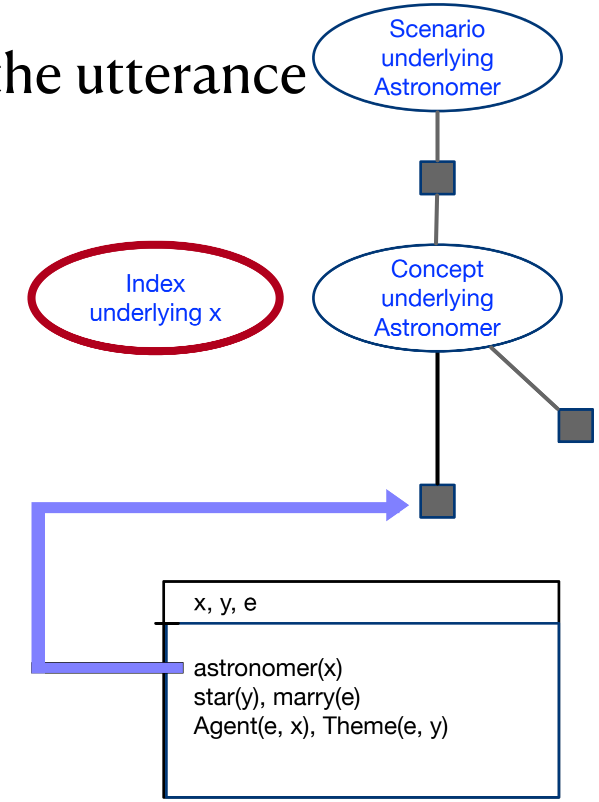
Disambiguation and reference in Situation Description Systems

- How disambiguation and reference should interact:
  - Utterance: “the bat”
  - Entity library salient entities:
  - Identify discourse referent for the bat with (3), disambiguating the utterance

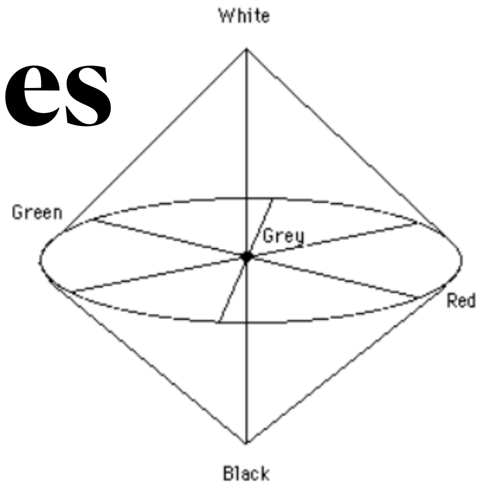


- Changes:
  A. To conceptual graph:
    - So far, underlying concepts for content words.
    - Now also: underlying index for discourse referents

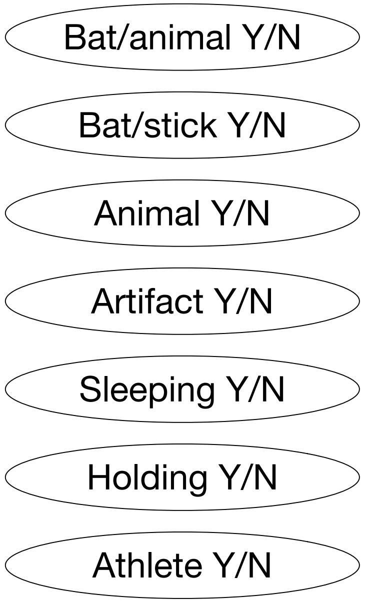
  B. Add entity library / mental files



Storing information about entities

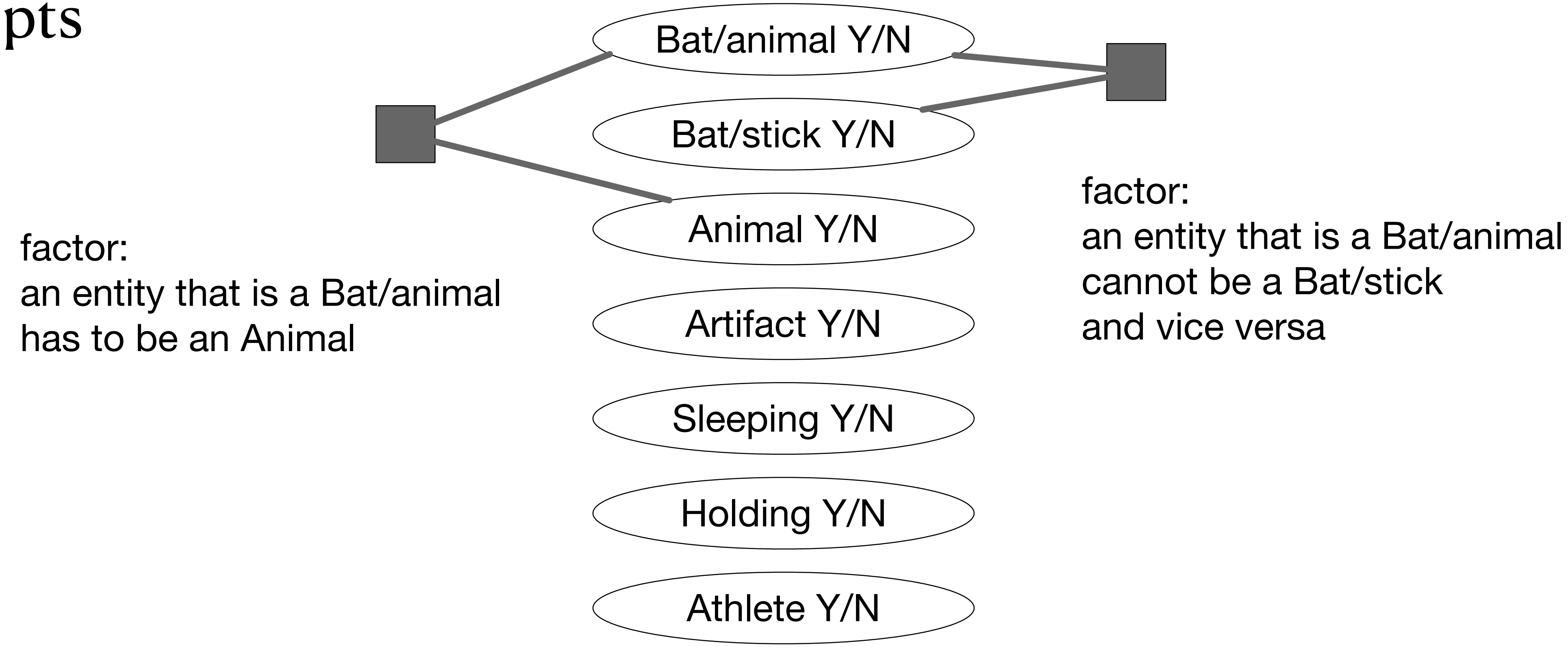


- Gärdenfors, Conceptual Spaces (2000):
  - Concepts as regions in spaces of quality dimensions
  - **Entity: point in conceptual space**
    - Concepts of which it is an instance: Concept regions in which it is situated
- **Re-represent in a different, simpler space** as practically working with Gärdenfors spaces is difficult (Bouraoui et al, 2022):
  - **One binary dimension per concept.**
  - Value of 1 = entity is instance of concept.
    Value of 0 = entity is not an instance.
  - Stored entity = vector of Bernoulli random variables



Storing information about entities

- Entity as a vector of Bernoulli variables
- Factors to express general constraints across dimensions = concepts

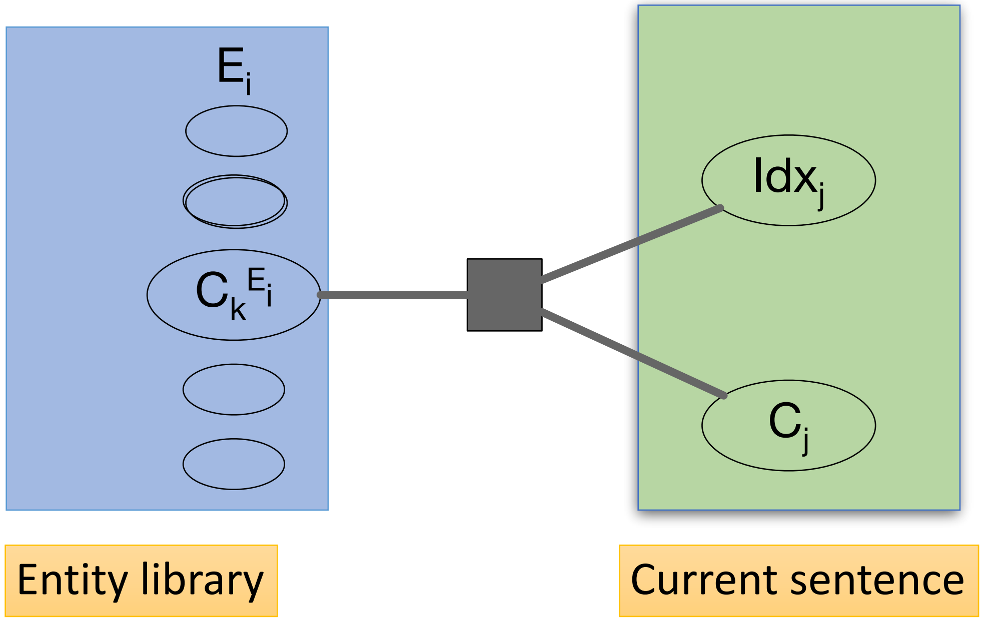


Constraints linking disambiguation and reference resolution: Concept values must match

- Utterance "the w_j": content word w_j, underlying concept random variable C_j , poss. values $c_1 \ldots c_m$
- Discourse referent with index Idx_j
- Stored entities $E_1, \ldots E_n$ E_i has Bernoulli variables $C^{E_i}_1 .. C^{E_i}_m$
- Factor: Index for w_j can be i if the underlying concept of w_j matches concepts of entity E_i

  1 iff:

  - $idx_j = i \wedge C_j = c_k \wedge C_k^{E_i} = 1$
  - $idx_j = i \wedge C_j \neq c_k \wedge C_k^{E_i} = 0$
  - $idx_j \neq i$

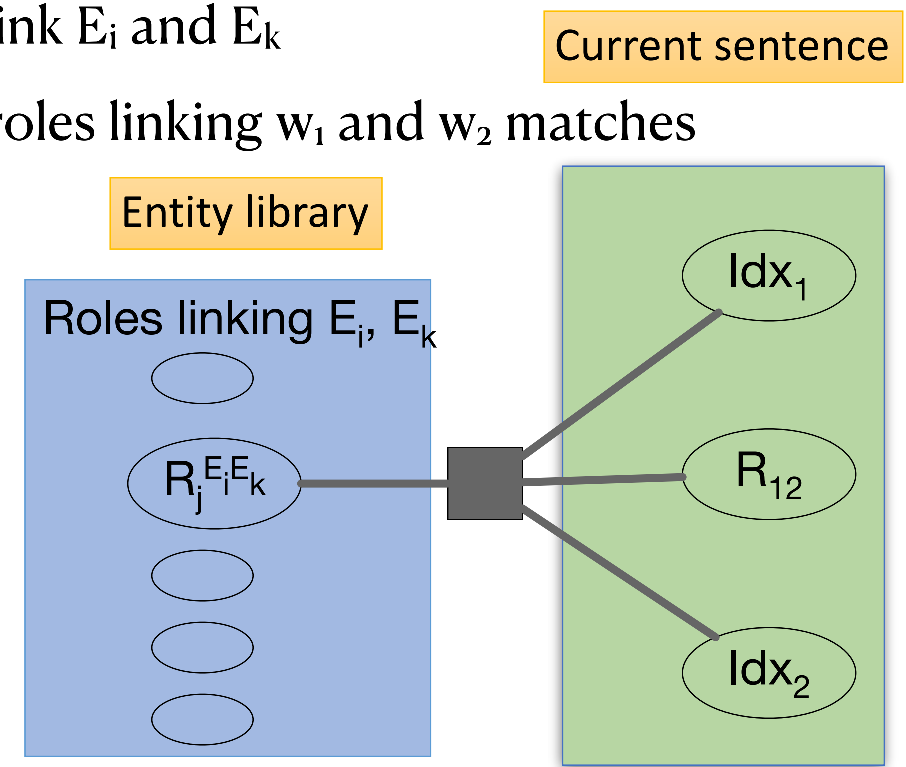


Constraints linking disambiguation and reference resolution: Roles must match

- Utterance contains w_1, w_2, both definite, w_2 is the argument of w_1. e.g., "the brown bear"
- Conceptual graph: Explicitly represent role linking C_1 and C_2, random variable R_{12}, poss. values $r_1, \ldots, r_m$
- Entity library: One Bernoulli variable for each role that could link E_i and E_k
- Factor: Index for w_1 can be i, and index for w_2 can be k, if the roles linking w_1 and w_2 matches the roles linking E_i and E_k

  1 iff:

  - $idx_1 = i \wedge idx_2 = k \wedge R_{12} = r_j \wedge R_j^{E_i, E_k} = 1$
  - $idx_1 = i \wedge idx_2 = k \wedge R_{12} \neq r_j \wedge R_j^{E_i, E_k} = 0$
  - $idx_1 \neq i$ or $idx_2 \neq k$



Scaling up probabilistic inference with the Visual Genome

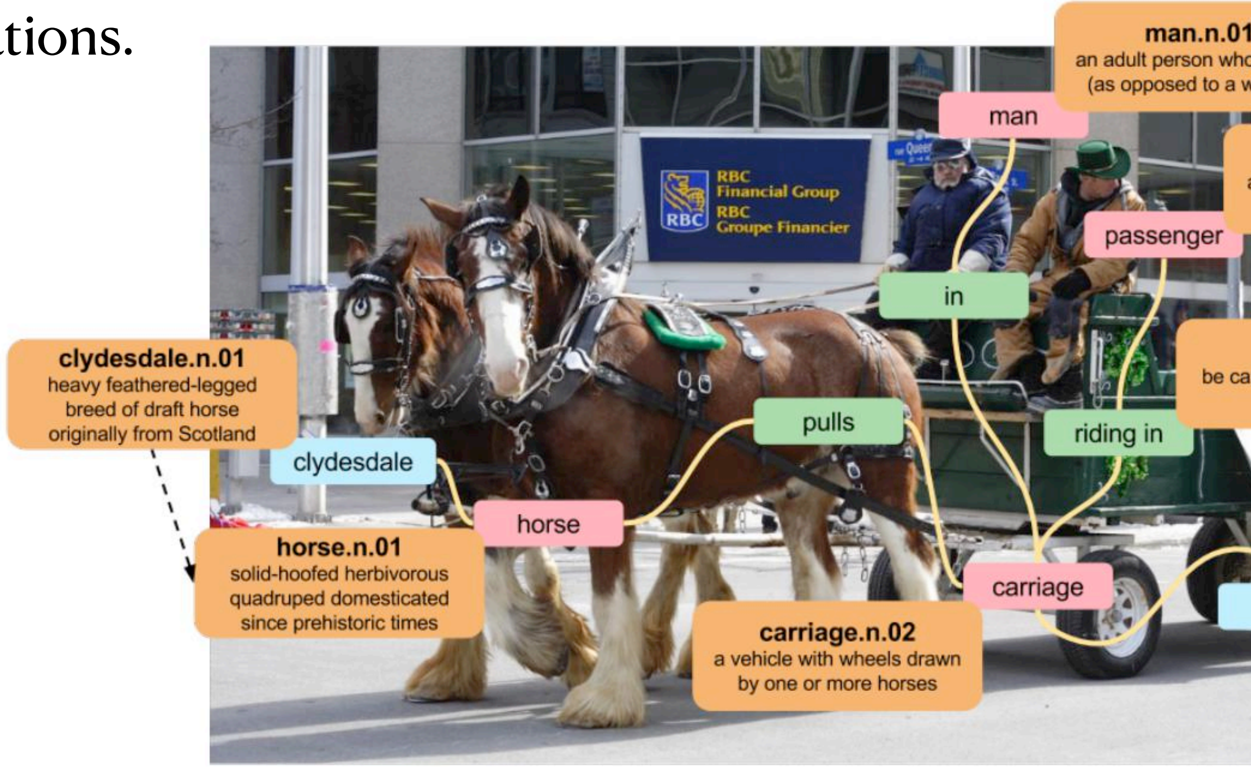
“Denotational embeddings” from the Visual Genome

Visual Genome: over 100K images, labeled with objects, attributes, relations (and WordNet senses)

Herbelot 2020, Herbelot & Copestake 2021, Merrill et al 2022:
With an “ideal corpus” of all true statements about each entity,
we can learn embeddings that characterize denotations.

> Embedding for “cat”: from all true statements
> involving entities in the extension of “cat”

Herbelot 2020: Learn such embeddings from
Visual Genome image labels,
Word2Vec model

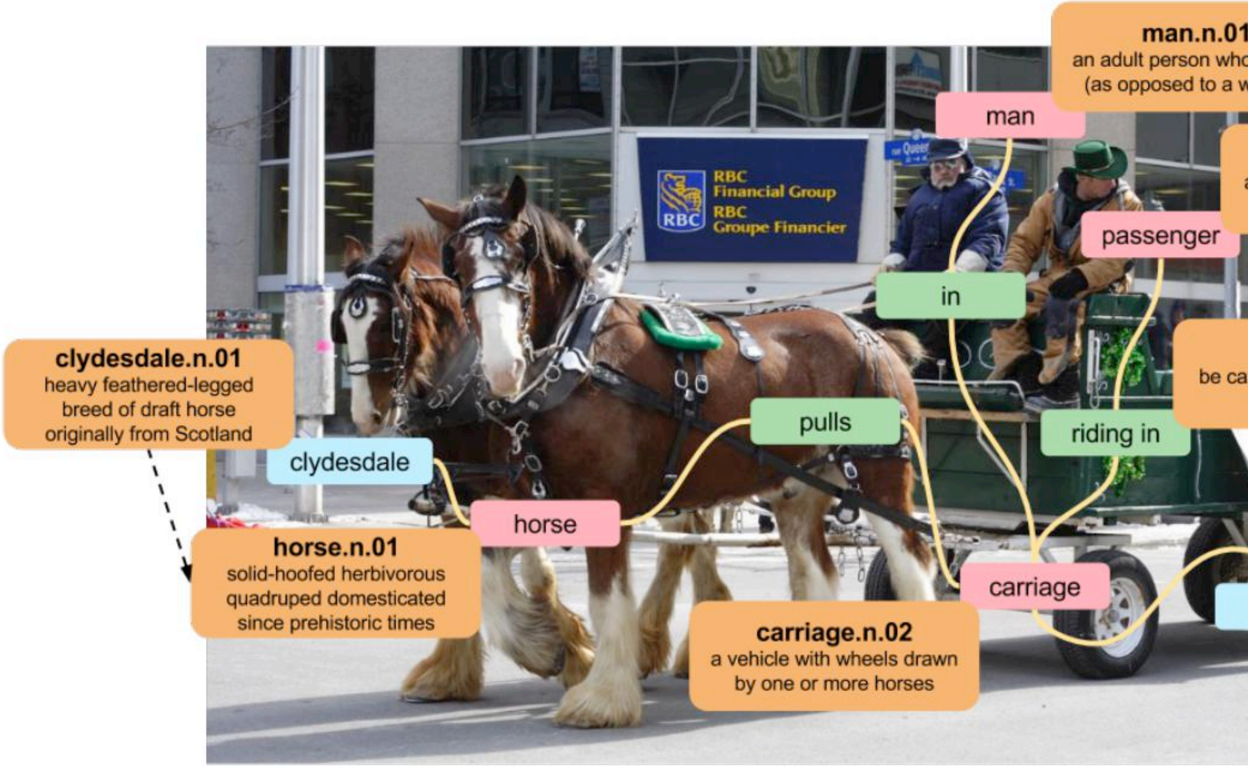


Experiments with the Visual Genome

Words concepts: embeddings as in Herbelot 2020

Selectional constraints of relations (green in the picture) and attributes (blue): centroid of vectors of its filler objects

Scenario = co-occurrence in an image,
Latent Dirichlet Allocation

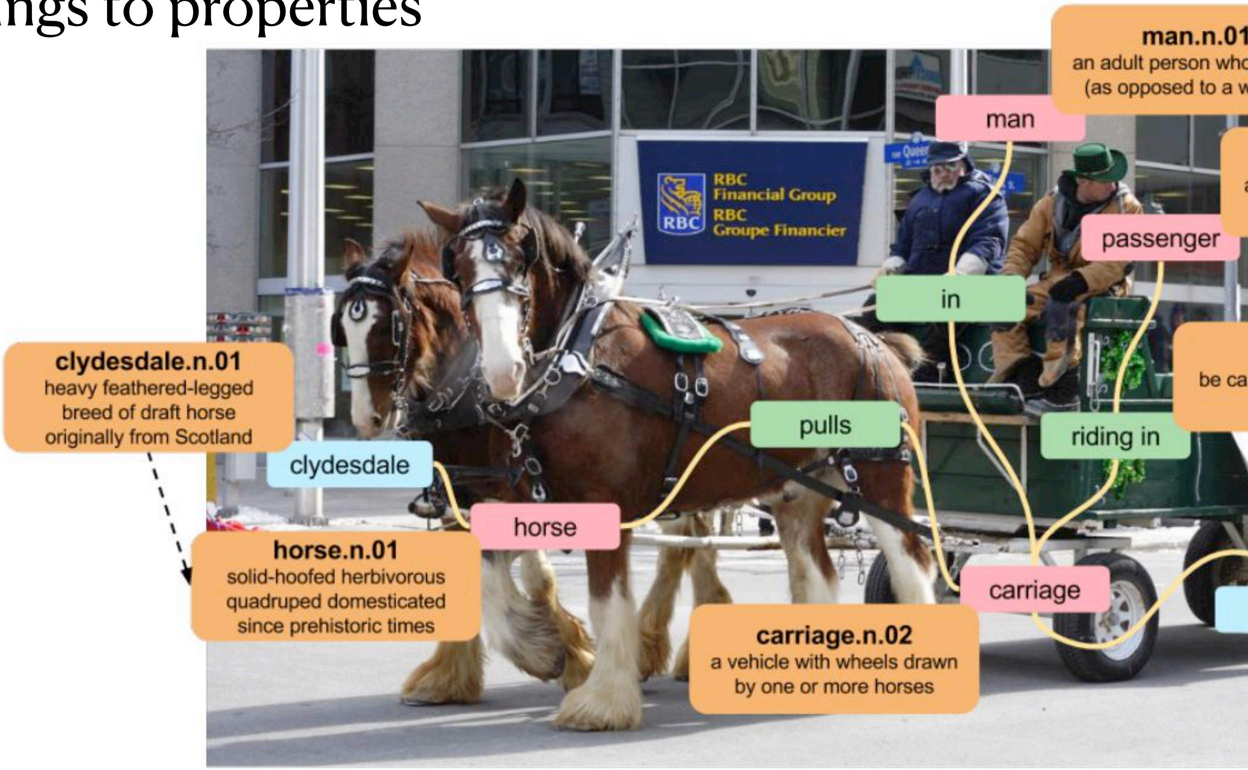


Experiments with the Visual Genome

What we can do with this data:

- Guess additional objects in images: scenario-based probabilistic “imagination”
- Guess attributes of objects: mapping embeddings to properties (Herbelot&Vecchi 2015, Rosenfeld&Erk 2022)
- Disambiguation:
  Basically no naturally occurring polysemy,
  so generate synthetically
- Reference resolution: “the horse”,
  “the clydesdale horse”, “the animal”

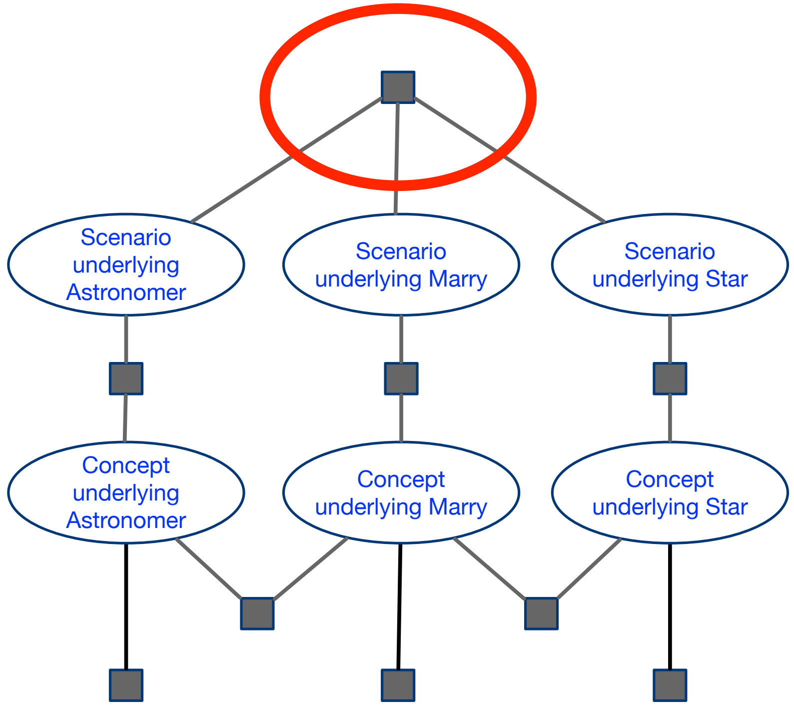
Work in progress, showing preliminary results



Scalable inference with factor graphs

- Categorical concepts, scenarios
  - Recent machine learning methods: large improvements for working with continuous values
  - Categorical values, not so much
  - So, use factor graphs,
    as there are packages for large-scale discrete models
- We use pgmax (Python):
  Efficient, scalable loopy belief propagation for discrete factor graphs

Scalable inference for Situation Description Systems



But: Factor graph implementations optimize for many nodes, few values.

We have: few nodes, many values, esp. factor linking scenario nodes.

Current simplification: Independence assumption, similar to variational inference.
Pre-compute distribution over scenario distributions, then sample scenarios from marginals

Future: Sampling messages in sum-product algorithm (Ihler/McAllester 2009)

Experiments with the Visual Genome

Scaling up:

- Lexicon of 3500 objects, 1900 attributes, 800 relations
- processing
  - about 4 seconds for one Visual Genome "sentence" with up to 25 objects, 1 ambiguous word
  - about 13 seconds with up to 1/3 of words in the sentence ambiguous

Factor graph inference:

- MAP assignment: Overall, what is the likeliest valuation? max-product algorithm
- Marginal probabilities: How likely, across all valuations, is this node to be valued "Sun" vs. "Well-known Person"? max-product algorithm
- MAP with evidence: "Most likely valuation where the concept for word 5 is Sun"
- Cannot compute weights for all valuations: too many

Experiment: Imagining objects

Guessing additional objects in an image:

- 90% of images for training: compute selectional preferences, scenario model
- Each test picture, mask up to 25 objects:
  - Compute MAP assignment of scenarios for the image
  - Scenario distribution for the picture: assume to match frequencies
  - Repeatedly: Sample scenario from scenario distribution, sample object from scenario
- Image: **OBJ:** truck(x0), road(x1), letters(x2), rim(x3), tire(x4), windshield(x5), truck(x6), side mirror(x7), ambulance(x8), license plate(x9), letters(x10), license plate(x11), writing(x12), street(x13), **ATTR:** white(x2), silver(x3), white and green(x6), green(x7), white(x8), black(x10), black(x12), concrete(x13), **REL:** driving down(x0,x1), with(x4,x3), OF(x5,x0)
- Prediction: car, street, road, bus, sidewalk, truck, wheel, tire, bike, building

	Model: 50 scenarios, selpref = centroid	Frequency baseline
Mean average precision	0.245	0.108
Average rank of highest correct	8.0	28.5

Experiment: Imagining attributes

Guessing properties for instances of a concept: mapping concept embeddings to properties (Herbelot&Vecchi 2015, Rosenfeld&Erk 2022)

- Partial Least Squares Regression: Learn interactions between input dimensions
- Trained on object embeddings to predict relative frequencies of attributes
- 500 most frequent attributes, 80% of object types for training

	Embeddings from attributes, relations	Embeddings from attributes, relations, co-occurring objects	Frequency baseline
Spearman's rho	0.245	0.222	0.228

Object: walkway

- Top predicted attributes: brown, wooden, concrete, red, gray, black, grey, green, paved, metal

- Strong frequency bias in the data and the model. Nor all applicable attributes annotated.
  Maybe predict association weight instead of relative frequency?

Experiment: Polysemy

Disambiguate words in context.

- Almost no naturally occurring polysemy in the data. So: 100 synthetically merged word pairs, sampled by frequency bins. Up to 50 test sentences per pair
- 90% of images for training: compute selectional preferences, scenario model

	50 scenarios. selpref = centroid (attr/rel embeddings)	50 scenarios, selpref = relative frequency	Frequency baseline
Accuracy	0.65	0.78	0.70

- Why worse performance with centroid for selectional preference? Maybe restrict to centroid of most frequent arguments
- Example image: **OBJ:** pipe(x1), urinal(x2), bathroom(x4), urinals(x5), cardboard/**flooring(x8)**, tile(x10), urinal(x11), pipe(x15), tile floor(x17), toilets(x18), toilets(x19), toilets(x20), wall(x21), urinal(x22), pipe(x28),
  ATTR: bathroom(x2), second(x2), white(x2), bathroom(x5), tile(x8), gray(x15),
  REL: for(x1, x2), for(x1,x4), ON(x18,x19), ON(x18,x21)

Next steps

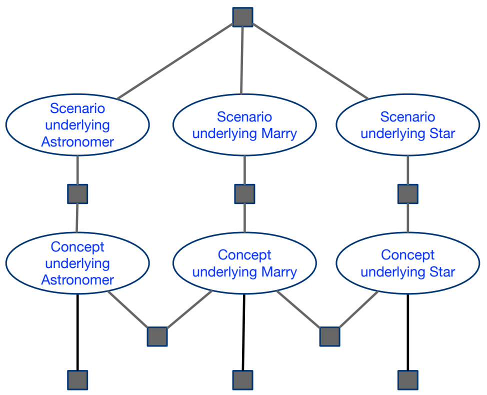
Next steps

- Reference and polysemy with the Visual Genome: “the brown bat” refers to image #2, and disambiguates “bat” to bat/animal
    - Again, no natural polysemy, so again make synthetically



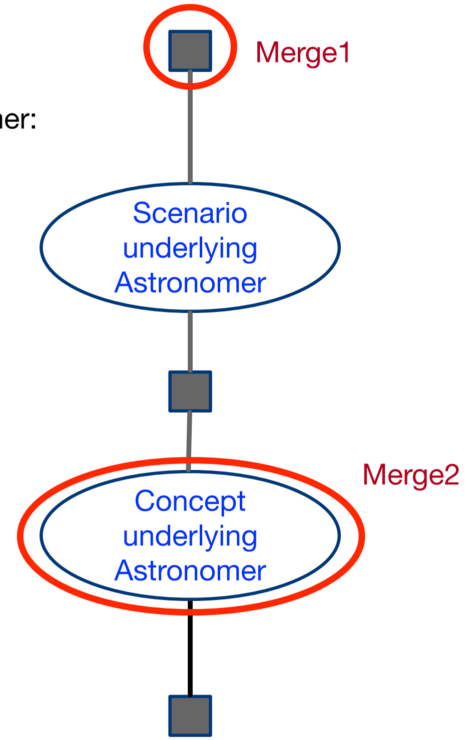

- Better synthetic polysemy for the Visual Genome: More ambiguous words per sentence, different degrees of similarity
- Text-based embeddings, to explore naturally occurring polysemy
    - Lexicalized concepts: clusters of contextualized embeddings

Next steps



- Semantics construction: graph combination on factor graphs. Use graph algebra, similar to Groschwitz et al 2015 for AMR
  - Cf. Cooper et al 2015: (partial) factor graphs as situation types?

astronomer:



marry: