Natural Language Processing

Paper Code: CS-821/6
UG 8th Sem

Dr. Samit Biswas, *Assistant Professor*

Department of Computer Sc. and Technology

Indian Institute of Engineering Science and Technology, Shibpur

Email: samit@cs.iiests.ac.in

Plan for Today

- Some Grammar Rules for English
- Features and Unification

Aspects of Phrase Structure of English

- Sentences with ***declarative*** Structure:
  - **a subject noun phrase followed by a verb phrase**
    - I prefer a morning flight.
    - I want a flight from Ontario to Chicago.
    - The flight should be eleven a.m. tomorrow.
    - The return flight should leave at around seven p.m.
- Sentences with ***imperative*** structure
  - often begin with a verb phrase and have no subject
    - Show the lowest fare.
    - Give me Sunday's flights arriving in Las Vegas from New York City.
    - List all flights between five and seven p.m.

S → VP

- Sentences with **yes-no question** structure
  - often (though not always) used to ask questions; they begin with an **auxiliary verb**, followed by a subject NP, followed by a VP.
    - Do any of these flights have stops?
    - Does American's flight eighteen twenty five serve dinner?
    - Can you give me the same information for United?

S → Aux NP VP

- various ***wh-structures***
  - ***wh-word*** *(who, whose, when, where, what, which, how, why)*
    - *What airlines fly from Burbank to Denver?*
    - *Which flights depart Burbank after noon and arrive in Denver by six p.m?*
    - *Whose flights serve breakfast?*

S → Wh-NP VP

Features

Constraints & Compactness

- Constraints in grammar
  - S → NP VP
  - But *They runs, *He run, *He disappeared the flight
  - Violate ***agreement*** (number), sub categorization
- Enforcing constraints
  - Add categories, rules

Why features?

- Need compact, general constraints
  - S → NP VP
    Only if NP and VP agree
- How can we describe agreement, subcat?
  - Decompose into elementary features that must be consistent
  - E.g. Agreement
    - Number, person, gender, etc

Feature Structure

Feature Structures

- Fundamentally, Attribute-Value pairs
  - Values may be symbols or feature structures
  - Feature path: list of features in structure to value
  - “Reentrant feature structures” : share same struct
- **Represented as**
  - Attribute-value matrix (**AVM**), or
  - Directed acyclic graph (**DAG**)

- Feature structures are sets of feature-value pairs where:
  - The features are atomic symbols and
  - The values are either atomic symbols or feature structures

$$\begin{bmatrix} FEATURE_1 & Value_1 \\ FEATURE_2 & Value_2 \\ FEATURE_3 & Value_3 \\ \vdots & \vdots \\ FEATURE_n & Value_n \end{bmatrix}$$

Example of Feature Structures, AVM:

- Number Feature

$$[Number \quad sg]$$

- Number-person features

$$\begin{bmatrix} Number & sg \\ Person & 3rd \end{bmatrix}$$

- Number-person-category features

$$\begin{bmatrix} CAT & NP \\ Number & sg \\ Person & 3rd \end{bmatrix}$$

Bundles of Features

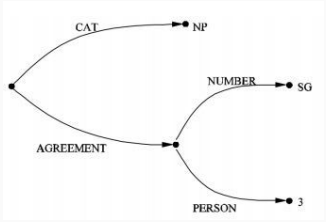
- Feature Values can be feature structures themselves.
- This is useful when certain features commonly co-occur, as number and person.

$$\begin{bmatrix} CAT & NP \\ AGREEMENT & \begin{bmatrix} NUMBER & sg \\ PERSON & 3rd \end{bmatrix} \end{bmatrix}$$

Feature Structure as DAG

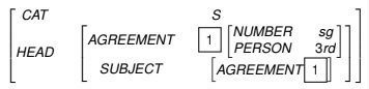
Feature Structure as DAG

$$\begin{bmatrix} CAT & NP \\ AGREEMENT & \begin{bmatrix} NUMBER & sg \\ PERSON & 3rd \end{bmatrix} \end{bmatrix}$$



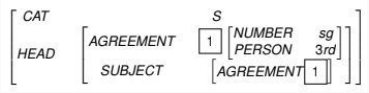
Reentrant Structure

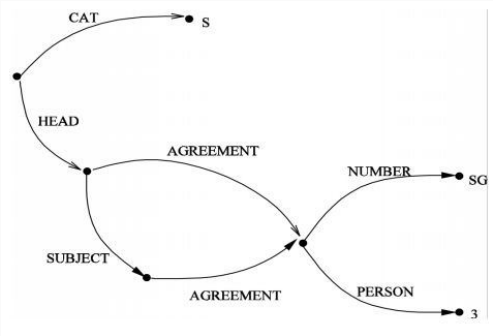
- We'll allow multiple features in a feature structure to share the same values. By this we mean that they share the same structure, not just that they have the same value.
- Numerical indices indicate the shared value.



Reentrant DAGs

Reentrant DAGs

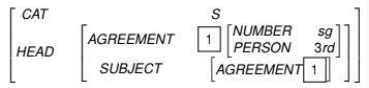




Reentrant Structure

Reentrant Structure

- It will also be useful to talk about paths through feature structures. As in the paths:
  - <HEAD AGREEMENT NUMBER >
  - < HEAD SUBJECT AGREEMENT NUMBER >



Unification

Unification

- So what do we want to do with these things…
  - check the compatibility of two structures
  - merge the information in two structures
- We can do both with an operation called Unification.
  - Merging two feature structures produces a new feature structure that is more specific (has more information) than, or is identical to, each of the input feature structures.

Unification Operation

- We say two feature structures can be unified if the component features that make them up are compatible.

$$[NUMBER \quad sg] \cup [NUMBER \quad sg] = [NUMBER \quad sg]$$

$$[NUMBER \quad sg] \cup [NUMBER \quad pl] = Fail$$

- Unification is a binary operation (represented as $\cup$)
- Unification returns the union of all feature/value pairs.

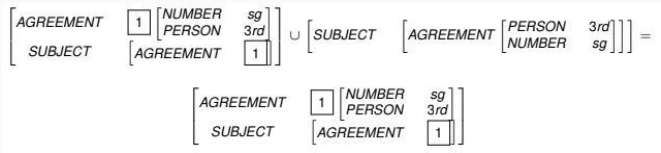
Unification Operation

$$[NUMBER \quad sg] \cup [NUMBER \quad []] = [NUMBER \quad sg]$$

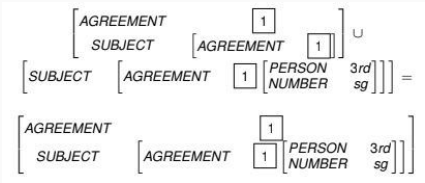
$$[NUMBER \quad sg] \cup [PERSON \quad 3rd] = \begin{bmatrix} NUMBER & sg \\ PERSON & 3rd \end{bmatrix}$$

Unification Operation
Examples of Unification for more complex reentrant structure

- Equality check complicated by the presence of a reentrant structure in the first argument.



- **Copying capabilities of unification**



Features, Unification and Grammar

How do we incorporate feature structures into our grammars?

- Assume that constituents are objects which have feature-structures associated with them.
- Associate sets of unification constraints with grammar rules.
- Constraints must be satisfied for rule to be satisfied.
- For a grammar rule: $\beta \rightarrow \beta_1 \ldots \beta_n$
  - $< \beta_i$ feature path $>$ = Atomic value
  - $< \beta_i$ feature path $> = < \beta_j$ feature path $>$

- To enforce subject/verb number agreement
  - S → NP VP
    < NP NUMBER >= < VP NUMBER >
    if the number of the NP is equal to the number of the VP

- Agreement
  - This flight serves breakfast.
    S → NP VP
    < NP AGREEMENT > = < VP AGREEMENT >
  - Does this flight serve breakfast?
    S → Aux NP VP
    < Aux AGREEMENT > = < NP AGREEMENT >

- Det/Nom agreement can be handled similarly
  - This flight, Those flights
    NP → Det Nominal
    < Det AGREEMENT > = < Nominal AGREEMENT >
    < NP AGREEMENT > = < Nominal AGREEMENT >
- And so on for other constituents and rules

Head Features

- Features of most grammatical categories are copied from head child to parent (e.g. from Verb to VP, Nominal to NP, N to Nominal, . . .)
- These normally written as head features, e.g.
  - VP → Verb NP
    < VP HEAD > = < Verb Head >
  - NP → Det Nominal
    < NP HEAD > = < Nominal Head >
  - < Det HEAD AGREEMENT> = < Nominal Head AGREEMENT>
  - Nominal → Noun
    < Nominal HEAD > = < Noun Head >
- Lexical entries that introduce this feature now reflect this HEAD notion.
  - Noun → flights
    < Noun HEAD AGREEMENT NUMBER> = pl
  - Verb → serves
    < Verb HEAD AGREEMENT NUMBER> = sg
    < Verb HEAD AGREEMENT PERSON> = 3rd

How can we parse with feature structures?

- Unification operator: takes 2 features structures and returns either a **merged feature structure** or **fail**
  - Input structures represented as DAGs
    - Features are labels on edges
    - Values are atomic symbols or DAGs
  - Unification algorithm goes through features in one input
  - DAG1 trying to find corresponding features in DAG2 - **if all match, success, else fail**

Summing Up

- Feature structures encoded rich information about components of grammar rules.
- Unification provides a mechanism for merging structures and for comparing them.
- Unification parsing:
  - Merge or fail.

References

References

- Daniel Jurafsky & James H. Martin, "*Speech and Language Processing*", Prentice Hall
- Feature Structures & Unification, http://www.nltk.org/howto/featstruct.html
- Diana Maynard, Kalina Bontcheva, Isabelle Augenstein, "*Natural Language Processing for the Semantic Web*", A Publication in the Morgan & Claypool Publishers series.
- Steven Bird, Ewan Klein and Edward Loper, "*Natural Language Processing with Python*", By O'Reilly.
- Christopher Manning and Hinrich Schtze, "*Foundations of Statistical Natural Language Processing*", The MIT Press.

Thank You

Contacts:

Dr. Samit Biswas
Department of CST,
IIEST, Shibpur

samit@cs.iiests.ac.in
https://www.iiests.ac.in/IIEST/Faculty/cs-samit