Computational Linguistics

Emily M. Bender

CSE 473

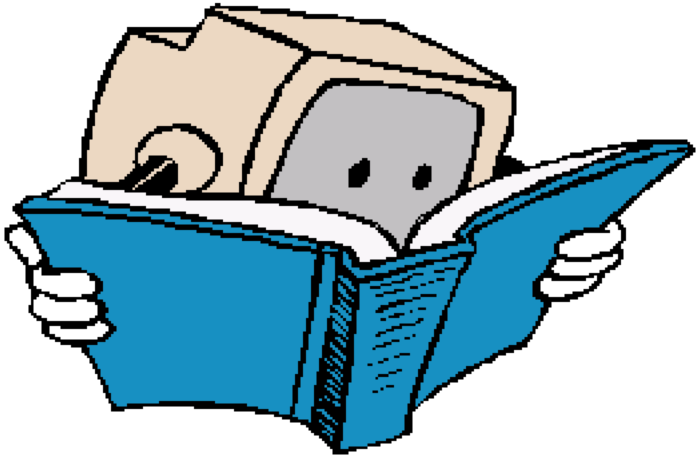
November 9, 2007

Overview

- What is computational linguistics
- A spectrum of approaches
- One project: The Grammar Matrix
- Resources/links

What is NLP?

- NLP: The processing of natural language text by computers
  - for practical applications
  - ... or linguistic research
- NLU: NLP with the goal of extracting meaning from the text for further machine processing



Human Language Understanding

- Relies on a wealth of intricate grammatical knowledge
- Is supported by an even greater wealth of world knowledge
- This means that information stored in natural language text requires a complex set of keys

Levels of linguistic structure

- Phonetics: Speech sounds, how we make them, how we perceive them
- Phonology: The grammatical structure of sounds and sound systems
- Morphology: How meaningful sub-word units combine to make words
- Syntax: How words combine to make sentences
- Semantics (lexical, propositional): What words mean and how those meanings combine to make sentence meanings
- Pragmatics: How sentence meanings are used to convey communicative intent
- ...

Pervasive ambiguity

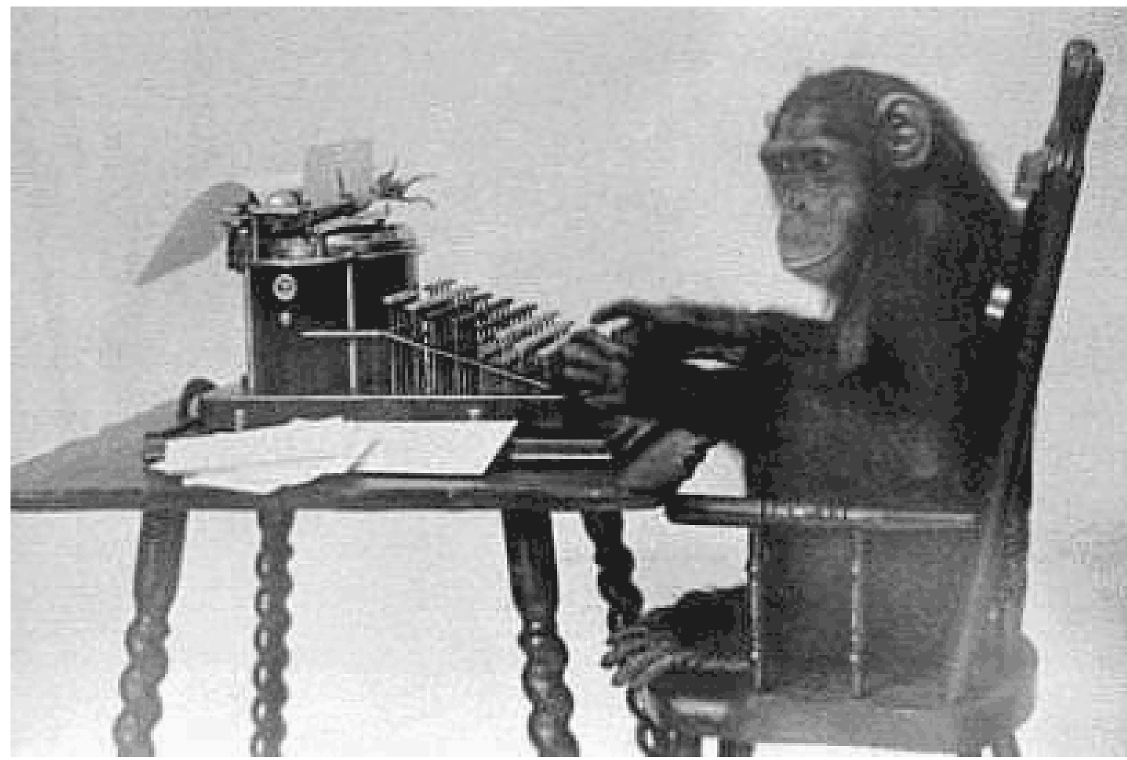
- Phonetic: *It's hard to wreck a nice beach.*
- Morphological: *This choice is undoable.*
- Syntactic: *Time flies like an arrow.*
- Semantic: *Every person read some book.*
- Pragmatic: *You should take those penguins to the zoo!*

And that's only the tip of the iceberg!

- Ambiguities are typically independent, leading to combinatorial explosions.
- *Have that report on my desk by Friday* (32-ways ambiguous)
- Humans are generally bad at detecting ambiguity, a consequence of being so good at *resolving* it.
- In NLP, stochastic models usually stand in for the common sense knowledge people use.

NLP: Spectrum of approaches

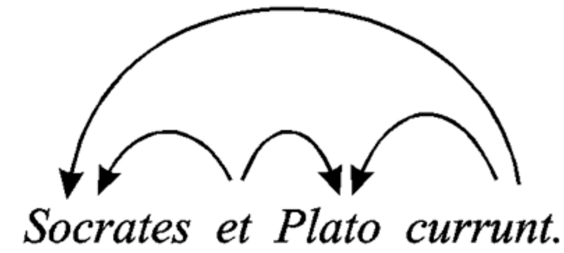
- Rule-based systems
- Stochastic models
    - Supervised v. unsupervised training
    - Incorporation of hand-made resources
    - Active learning
- Hybrid approaches



Evaluation in computational linguistics/NLP

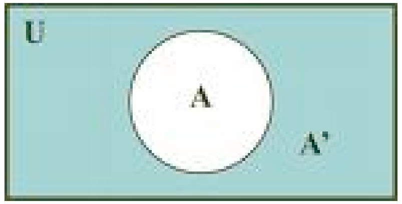
- Test performance of system against a gold standard
- But often the ‘right’ answer is not obvious:
  - Different approaches to linguistics suggest different answers
  - Multiple answers are right
- How to construct a gold standard for:
  - Speech recognition systems
  - Parsers
  - Machine translation
  - Summarization
  - Dialogue systems

Natural language syntax & semantics



- Constituent structure
- Mapping of linear string to predicate-argument structure (word order, case, agreement)
- Long distance dependencies
  - What did Kim think Pat said Chris saw?
- Idioms, collocations

Formal/'Generative' Grammars



- Characterize a set of strings (phrases and sentences)
- These strings should correspond to those that native speakers find acceptable
- Assign one or more syntactic structures to each string
- Assign one or more semantic structures to each string
- No complete generative grammar has ever been written for any language

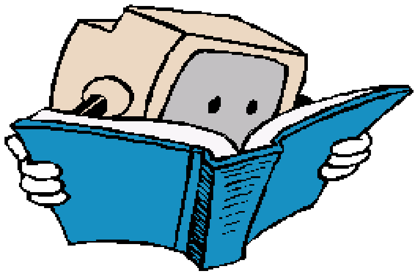
Precision Computational Grammars

- Knowledge engineering of formal grammars, for:
  - Parsing: assigning syntactic structure and semantic representation to strings
  - Generation: assigning surface strings to semantic representations

Why build precision grammars?

- Linguistic hypothesis testing
  - Test interacting analyses for consistency
  - Test grammar against test suites and naturally occurring text
  - Richer language documentation

Why build precision grammars?



- ‘Deep’ NLP/NLU
  - Automated customer service response
  - Machine translation (symbolic, hybrid)
  - Speech prostheses
  - Hybrid Q&A systems
  - Human-computer dialog/collaboration
  - Machine mediated human-human interaction
  - Better treebanks

Hurdles

- Efficient processing (Oepen et al 2002)
- Ambiguity resolution (Baldridge & Osborn 2003, Toutanova et al 2005, Riezler et al 2002)
- Domain portability (Baldwin et al 2005)
- Lexical acquisition (Baldwin & Bond 2003, Baldwin 2005)
- Extragrammatical/ungrammatical input (Baldwin et al 2005)
- Scaling to many languages

The Grammar Matrix: Overview

- Motivation
- HPSG
- Semantic representations
- Cross-linguistic core
- Libraries
- MMT: Massively Multilingual Translation/Matrix Machine Translation

Matrix: Motivation

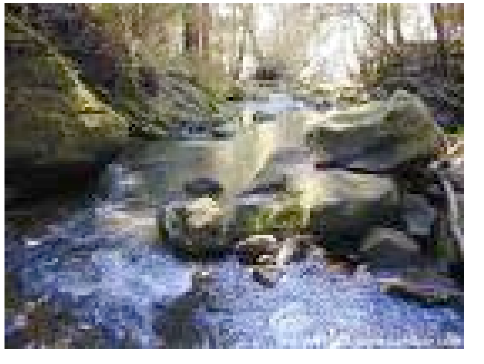
- English Resource Grammar:
  - 140,000 lines of code (25,000 exclusive of lexicon)
  - ~3000 types
  - 16+ person-years of effort
- Much of that is useful in other languages
- Reduces the cost of developing new grammars

Matrix: Motivation

- Hypothesis testing (monolingual and cross-linguistic)
  - Interdependencies between analyses
  - Adequacy of analyses for naturally occurring text

Matrix: Motivation

- Promote consistent semantic representations
  - Reuse downstream technology in NLU applications while changing languages
  - Transfer-based (symbolic or stochastic MT)



HPSG

- Head-Driven Phrase Structure Grammar (Pollard & Sag 1994)
- Typed feature-structures
- Declarative, order-independent, constraint-based formalism

An HPSG consists of

- A collection of feature-structure descriptions for phrase structure rules and lexical entries
- Organized into a type hierarchy, with supertypes encoding appropriate features and constraints inherited by subtypes
- All rules and entries contain both syntactic and semantic information

An HPSG is used

- By a parser to assign structures and semantic representations to strings
- By a generator to assign structures and strings to semantic representations
- Rules, entries, and structures are DAGs, with type name labeling the nodes
- Constraints on rules and entries are combined via unification

Example rule type

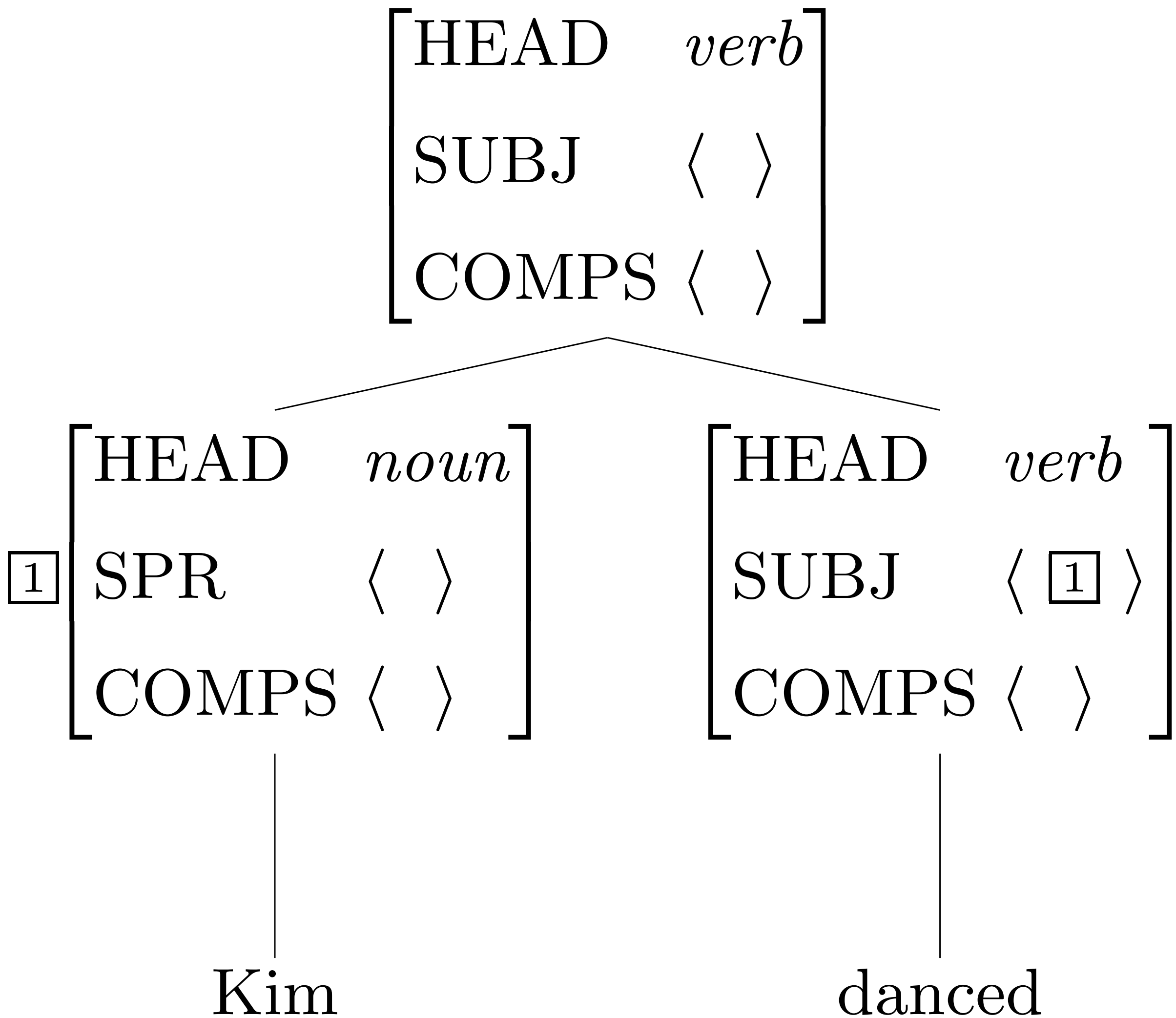
$$
\textit{head-subj-phrase}: \left[\begin{array}{ll}
\textit{binary-headed-phrase} \ \& & \\
\textit{head-compositional} & \\
\text{SUBJ} & \langle\ \rangle \\
\text{COMPS} & \boxed{1} \\
\text{HEAD-DTR} & \left[\begin{array}{ll} \text{SUBJ} & \langle\boxed{2}\rangle \\ \text{COMPS} & \boxed{1} \end{array}\right] \\
\text{NON-HEAD-DTR} & \boxed{2}
\end{array}\right]
$$

Example rule type

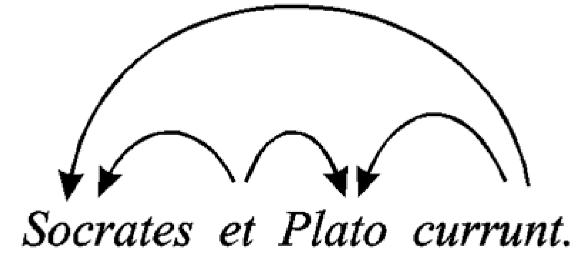
$$\textit{head-final}: \begin{bmatrix} \textit{binary-headed-phrase} \ \& \\ \text{HEAD-DTR} & \boxed{1} \\ \text{NON-HEAD-DTR} & \boxed{2} \\ \text{ARGS} & \langle\ \boxed{2}, \boxed{1}\ \rangle \end{bmatrix}$$

subj-head: *head-subj-phrase* & *head-final*

Example parse



Semantic Representations



- Not going for an interlingua
- Not representing connection to world knowledge
- Not representing lexical semantics (the meaning of life is life’)
- Making explicit the relationships among parts of a sentence
  - Kim gave a book to Sandy
  - give(e,x,y,z), name(x,‘Kim’), book(y), name(z,‘Sandy’), past(e)

Semantic Representations

- Sandy was given a book by Kim
- Kim continues to give books to Sandy
- This is the book that Kim gave Sandy
- Which book did Kim give Sandy?
- Which book do people often seem to forget that Pat knew Kim gave to Sandy?
- This book was difficult for Kim to give to Sandy.

Semantic representations

- Minimal Recursion Semantics (Copestake et al 2005)
  - Expressive adequacy
  - Computational tractability
  - Grammatical compatibility
  - Underspecifiability

Semantic representations

- MRS specifies well-formedness
- Matrix specifies representations
  - Nominal v. verbal predicates
  - Quantifiers
  - Illocutionary force
  - Coordination

Semantic representations

- Languages may still differ:
  - Lexical predicates
    - Japanese: kore, sore, are
  - Grammaticized tense/aspect, discourse status
  - Ways of saying
    - make a wish, center divider

Design criteria

- Strip all syntactic information
- Stay lexically close to the surface (for hybrid deep/shallow systems)
- Encode all distinctions marked in the surface from
- Leave underspecified all else that can be computed

Matrix: Cross-linguistic core

- Types defining feature geometry
- Types encoding compositional semantics
- General classes of phrase structure rules
- General classes of lexical items
- Configuration and parameter files for LKB (Copestake 2002) and PET (Callmeier 2000)

Matrix: Hypothesized universals

- Words and phrases combine to make larger phrases.
- The semantics of a phrase is determined by the meaning of its parts and how they're put together.
- Some rules for phrases add semantics, some don't.
- No rule can remove semantic information.
- Most phrases have an identifiable head daughter.
- Heads determine the type of arguments they require, and how they combine semantically with those arguments.
- Modifiers determine the type of heads they modify, and how they combine semantically with the head.

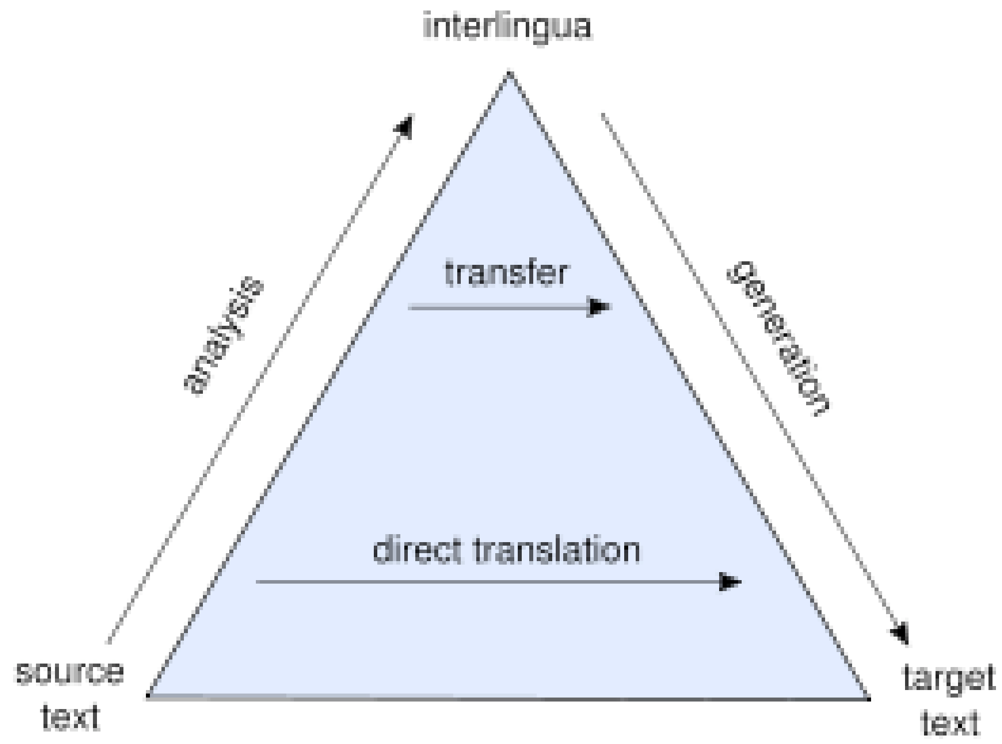
Libraries: Motivation

- Many patterns are not universal, yet recurring
  - Systems represented in every language: word order, negation, questions
  - Systems/patterns represented in some languages: noun incorporation, numeral classifiers, verb particle construction
- Promote reuse of code
- Promote consistency of analyses
- Crosslinguistic hypothesis testing:
  - Does the same analysis of SVO work in all SVO languages?

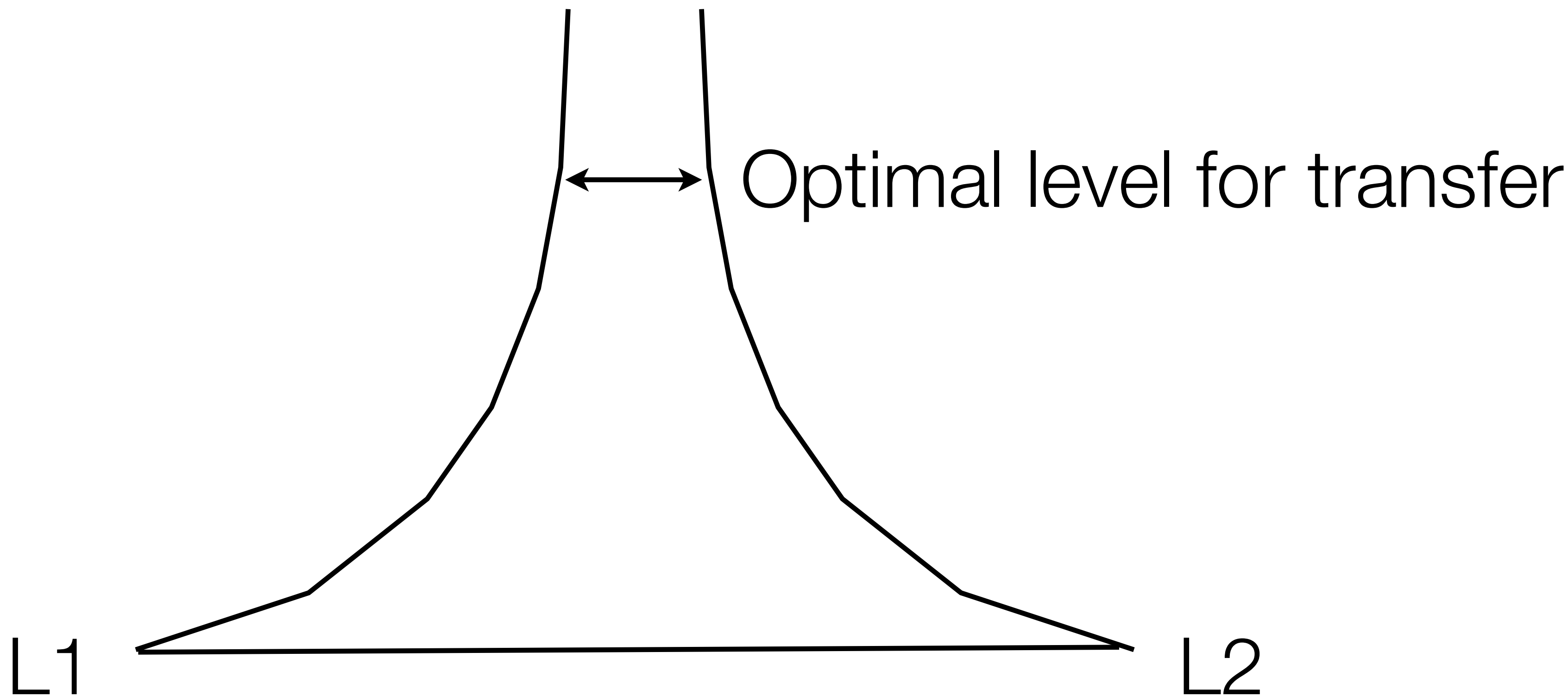
Application: MMT

- Most approaches to machine translation have a problem with scaling:
  - Statistical MT: Need for aligned corpora (‘bitexts’) for every language pair
  - Rule-based MT: Need for transfer grammars for every language pair

Machine Translation: Vauquois triangle



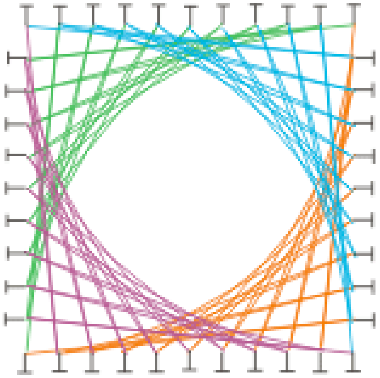
Machine Translation: Copestake Volcano



Example transfer rule

```
[ INPUT [ RELS < [ PRED “_adkmost_n_rel”,
                   LBL #label,
                   ARG0 #arg ] > ],
  OUTPUT [RELS < [ PRED “_access_n_1_rel”,
                   LBL #label,
                   ARG0 #arg ] > ] ]
```

Application: MMT



- Most approaches to machine translation have a problem with scaling:
    - Statistical MT: Need for aligned corpora (‘bitexts’) for every language pair
    - Rule-based MT: Need for transfer grammars for every language pair
- Can the normalization promoted by the Matrix facilitate moving MT to a panlingual scale?

MMT design goals

- Normalize semantic representations as much as possible
  - Within constraints of single step string-semantics mapping for each language
- Map into shared predicate space
- Handle remaining semantic differences with one transfer grammar per target language

Predicate-predicate mapping not sufficient in the general case

- Multiword expressions

  Ça ne me fait pas mal (fra)
  That not me make not bad
  'That doesn't hurt me'

- Different mappings of arguments

  Ça me plaît (fra)
  That me like
  'I like that.'

- Head switching

  Han fiske gjerne (nor)
  He fishes happily
  'He likes to fish.

But how far can we get with a naïve system?

- Existing MMT system has 10 languages each with tiny lexicons
- Connect to TransGraph (cite)
- How much coverage over open domain text?
- How useful as a toy translation system for cooperative users?
- How easy to add additional languages?

Overview

- What is computational linguistics
- A spectrum of approaches
- One project: The Grammar Matrix
- Resources/links

To learn more

- Courses: Ling 472, 570-573, 566, 567; EE 516, 517
- CLMA: Professional MA program in computational linguistics http://compling.washington.edu./
- Turing Center: http://turing.cs.washington.edu/
- Computational Linguistics lab: http://depts.washington.edu/uwcl/
- ACL Wiki: http://aclweb.org, http://aclweb.org/aclwiki