



# Robust Parsing <sub>1</sub>

---

- Introduction
- Chunking

# Robust Parsing <sub>2</sub>

---

- Conventional methods of parsing are insufficient for dealing with non restricted texts

- Adequate segmentation
- Disambiguation
- Coverage

Which are the units to parse

Which is the most likely parse

Parsing beyond the lexical coverage

- What to do

- Not parsing all
- Not parsing in depth

fragmental parsing

shallow parsing

# Robust Parsing <sub>3</sub>

---

- Problems when parsing non restricted corpus
  - Adaptation of a grammar to a corpus or sublanguage
  - Selection of the correct (!?) parse between the ones allowed by the grammar.
  - Production of good parses for entries outside the coverage of the grammar (Robustness)

# Robust Parsing <sup>4</sup>

---

- Partial parsers
  - phrasal parsers
    - chunkers, spotters
    - Church,1988
  - cooccurrence parsers
    - Church,Hanks,1989, Brent,1993
  - fragmental parsers
    - Fidditch, Hindle,1994, MITFP, Abney,1991
  - constraint-based parsers
    - Voutilainen,1995
- Probabilistic parsers
- Treebanks

# Robust Parsing <sup>5</sup>

- Chunking

- detection of phrases nominal, verbal, adjectival, adverbial basic (without recursion)
- FS techniques
- Performance of a cascade of transducers
- HMM
- ML

Abney, 1996  
Argamon et al, 1998  
Cardie, Pierce, 1998  
Church, 1988  
Ramshaw, Marcus, 1995  
Skut, Brants, 1998

- Bracketting

- Obtention of dependences between chunks and syntactic relations

Alembic  
• Vilain, 1999  
Pinocchio  
• Ciravegna, Lavelli, 1999  
TiMBL  
• Daelemans et al, 2000  
Supertagging  
• Bangalore, Joshi, 1999

# Robust Parsing <sup>6</sup>

---

- Definition of chunk
  - With linguistic basis: Abney
  - Only pragmatic:
    - Contiguous sequences of related tokens
      - Not confusing with terms
    - e.g. Base NP
- Approaches to chunking
  - Look for (include) information
  - Remove information
    - e.g. Chink

# Robust Parsing <sup>7</sup>

---

- Representing chunks
  - Labels
    - e.g. BIO tags
    - BEGIN, INSIDE, OUTSIDE
  - Trees
- Chunk parser
  - Looking for non overlapped chunks for reaching a maximal coverage

# Robust Parsing <sup>8</sup>

---

- Frequently regular expressions over sequences of POS tags
- agglomerative (chunk rules) vs divisive (chunk rules)
- Rules for fusion of adjacent chunks
- Rules for splitting a chunk in smaller components.
- Cascade approach for chunk detection



# Robust Parsing 9

---

- Related tasks
  - Bracketing
  - Term candidate extraction
  - Named Entity Recognition (NER)
  - Named Entity Classification (NEC)
  - NERC

# Robust Parsing <sup>10</sup>

---

- Example Church chunker
- Statistical tagger followed by chunker
- Between any pair of contiguous tokens, a chunk delimiter can be inserted
- chunks delimiters have to satisfy consistency conditions probabilities of inserting any tag are learned using supervised learning
- The result is the string with maximum probability assuming independence.

—————→ [ , ] , ] [ , -