Natural parsing: a psycholinguistically motivated computational language processing model

Gábor Prószéky – Balázs Indig

Pázmány Péter Catholic University, Faculty of Information Technology and Bionics
Budapest, Hungary
MTA-PPKE Hungarian Language Technology Research Group

{proszeky.gabor,indig.balazs}@itk.ppke.hu

Introduction: Our parser

- Input: natural language sentences
  - In raw or tokenised form
- Output: semantic representation
  - Identify facts from text that are not explicitly available
- Method: human parser motivated parallel system
  - Use any information that is available (statistical or rule-based)
- Prototype for Hungarian
  - Hand-written rules (norm) and statistics (simulate experience)

Psycholinguistic parsing: A computational model

- Bever (1970): “Parsing” is not inverted generation
- Kimbal (1973): Seven principles on parsing
  - Existence of a “supreme decision-maker”
- Frazier and Fodor (1978): delayed decisions
  - “Garden path” sentences
- Prószéky (2000): Parser can override the lexicon
  - Linguistic decisions during parsing have higher priority
- Pléh & Lukács (2014): Human parsing has patterns
  - Some of them are identified and **we** follow them

Psycholinguistic parsing: A computational model

- Instead of sentences we process utterances
  - Can span multiple coherent sentences
- Representation is not a traditional “parse Tree”
  - But a DAG with colored edges and labels on them

- Distant from traditional computational models
  - Which are plagued by generative grammar
  - And only care about the result of the parser

Architecture: Traditional vs. Parallel

Raw Text

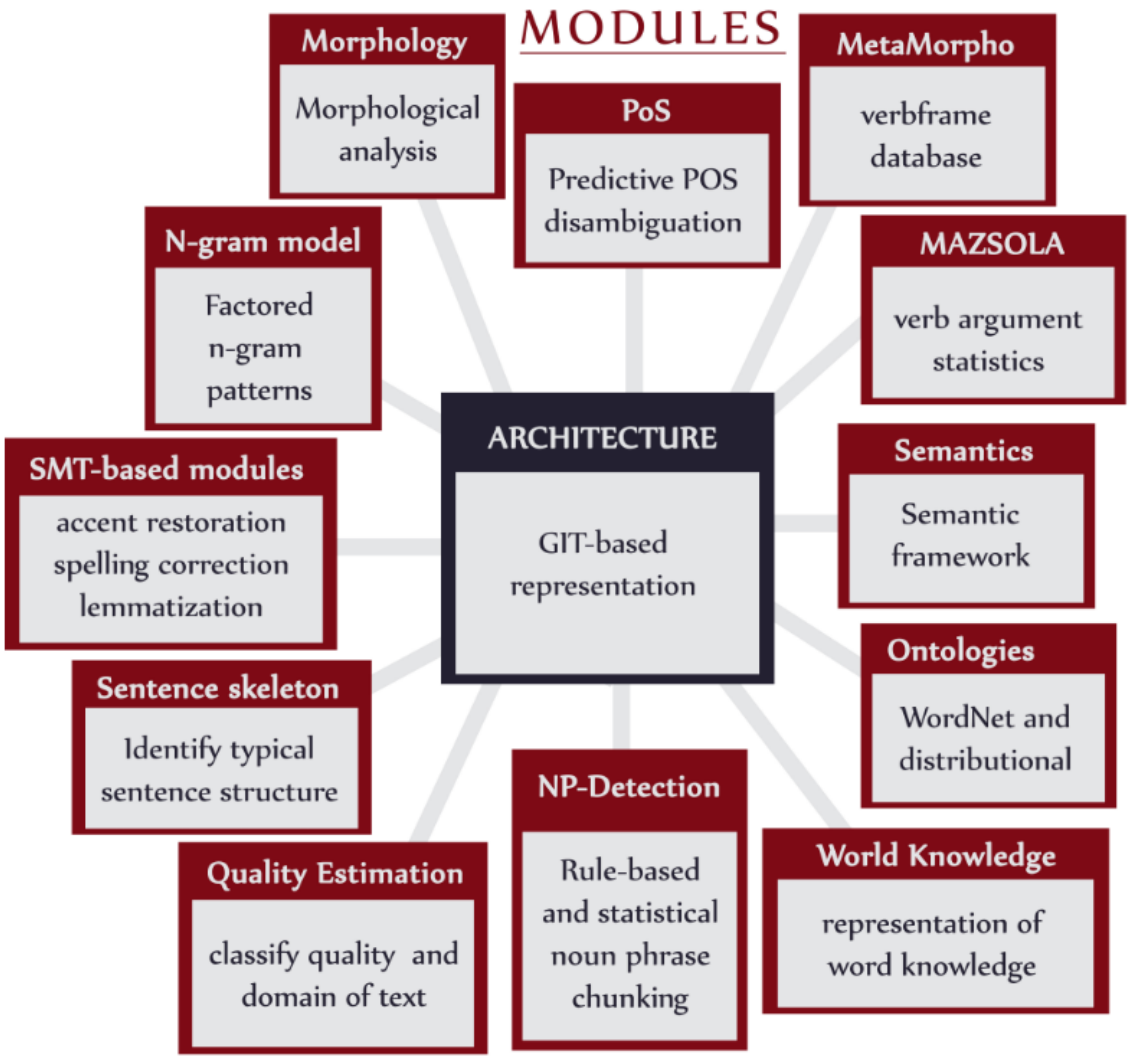
Tokenization

POS tagging

NP-Chunking

NER

Parsing & Semantics



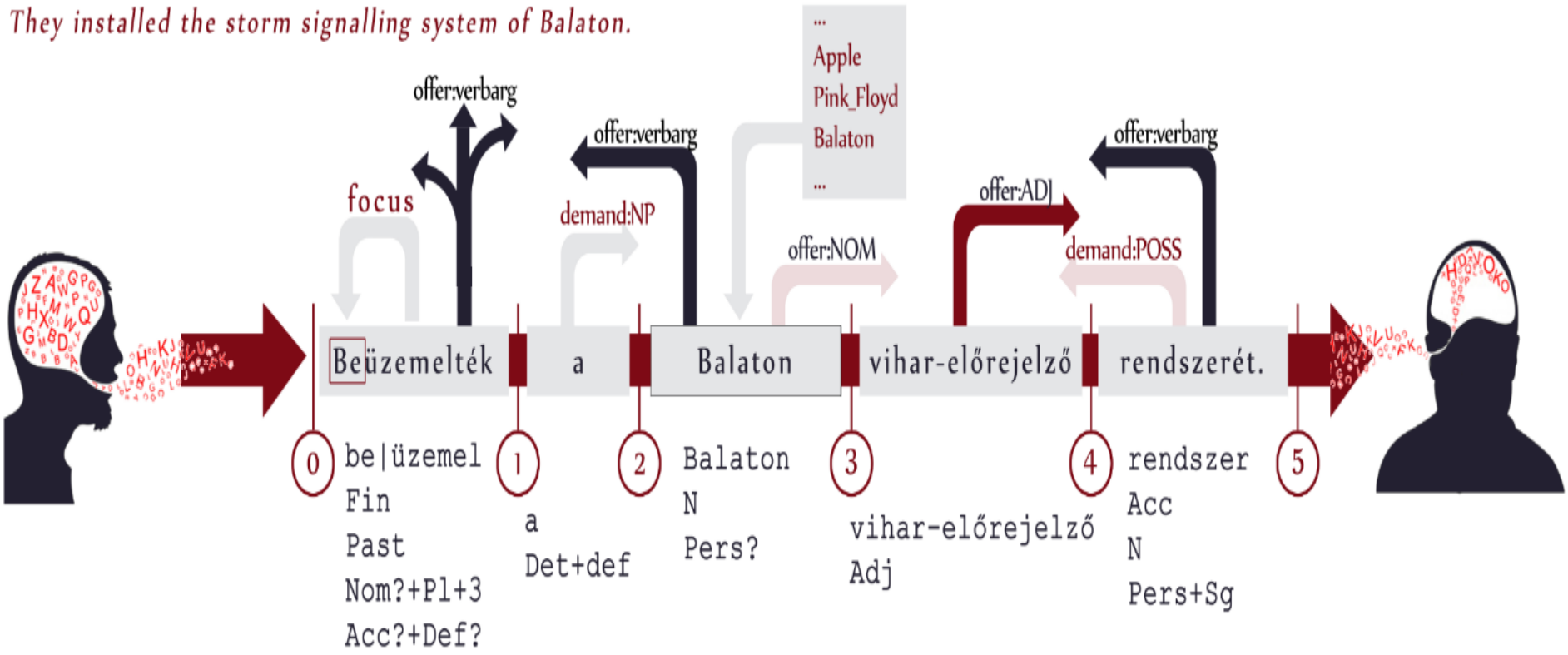
Supply and Demand style processing

- Every word may have demands
  - Verbs demand their arguments to form a frame
  - Nouns demand their attributes to form a Noun Phrase
- Every word may have features to supply
  - Nouns have grammatical case
  - Possession offers itself to a matching possessor
  - Determinant offers itself to the next Noun
- Similarly to Word Expert Parser (Small 1983)

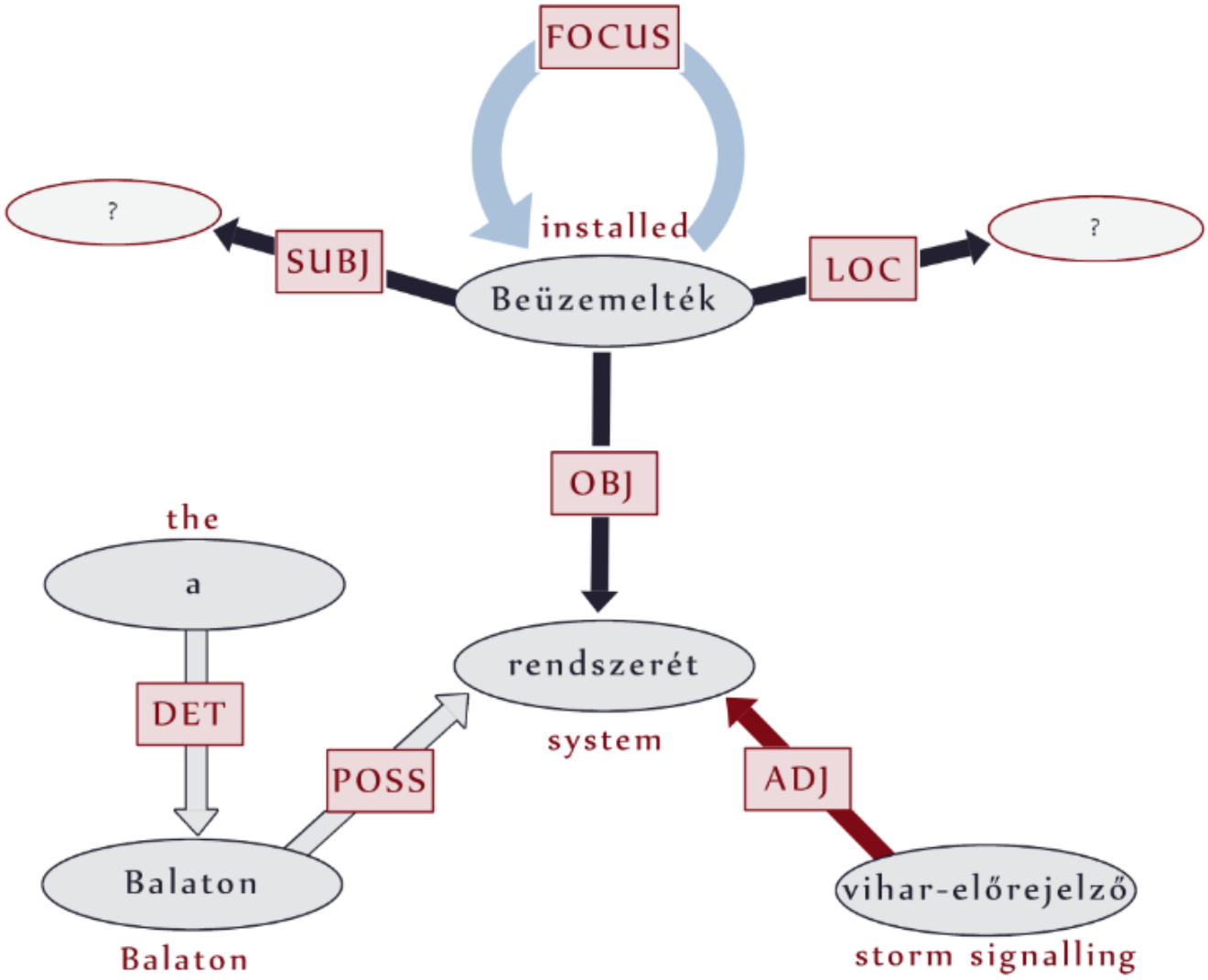
Properties of the parser

- Does not have part-of-speech disambiguation
  - Only an n-gram model of the previous words
  - Still avoids combinatoric explosion as most paths can not continue after one or two steps
- Mixes syntactic, semantic (and other) structures
  - Use everything to aid parsing just like a human parser
- Makes exactly the same mistakes as humans do
  - Backtrack only when it's really needed (Grice 1975)

Example


They installed the storm signalling system of Balaton.
offer:verbarg
focus
offer:verbarg
demand:NP
...
Apple
Pink_Floyd
Balaton
...
offer:NOM
offer:ADJ
offer:verbarg
demand:POSS
Beüzemelték
a
Balaton
vihar-előrejelző
rendszerét.
0
be|üzemel
Fin
Past
Nom?+Pl+3
Acc?+Def?
1
a
Det+def
2
Balaton
N
Pers?
3
vihar-előrejelző
Adj
4
rendszer
Acc
N
Pers+Sg
5

Example Output



Modules

- Automatically build a corpus from the web
- Text normalisation (unaccented text, spell check)
  - A module is designed to estimate text quality
- Building a sentence skeleton bank
  - Noun phrases are stripped from the sentence
- Factored n-gram patterns used for “caching”
- Verb frame database with semantic information
  - With the help of rule-based and statistic resources

See you at the poster stand!