INF5820

Natural Language Processing - NLP

H2009

Jan Tore Lønning

jtl@ifi.uio.no

Dependensparsing

INF5830

Lecture 12

Nov 2, 2009

Today

- Dependency grammar
- Dependency parsing
  - w/MaltParser

Dependency grammar

- Syntax as dependency structure between words
- Asymmetric relationship between head and dependent
- Different from phrase structure: only terminals
- Lucien Tesnière (1893-1954)
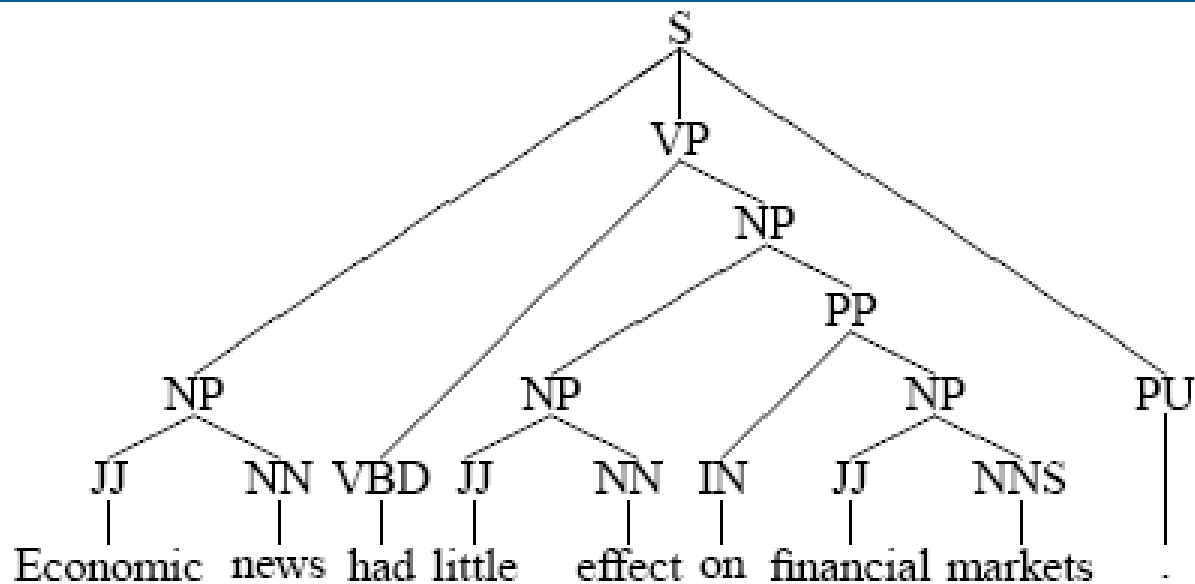 - *Éléments de syntaxe structurale* (1959)

Figure 1: Constituent structure for English sentence from the Penn Treebank
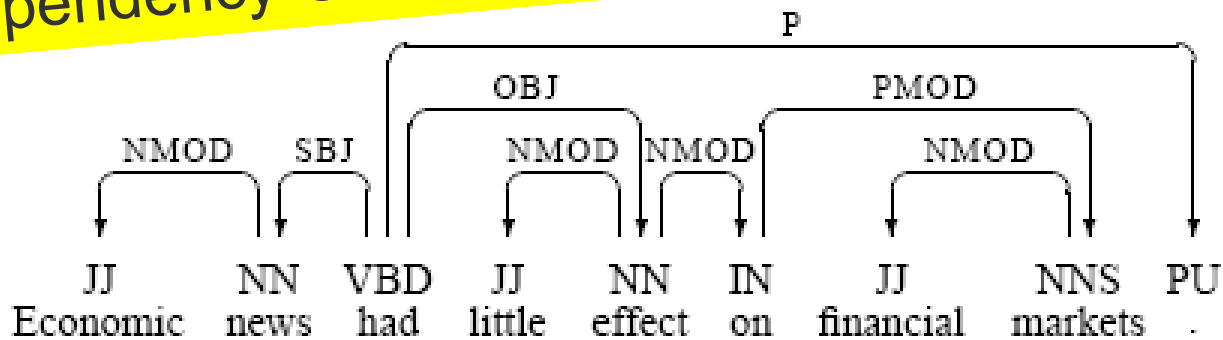
Nivre: Dependency Grammar and Dependency Parsing

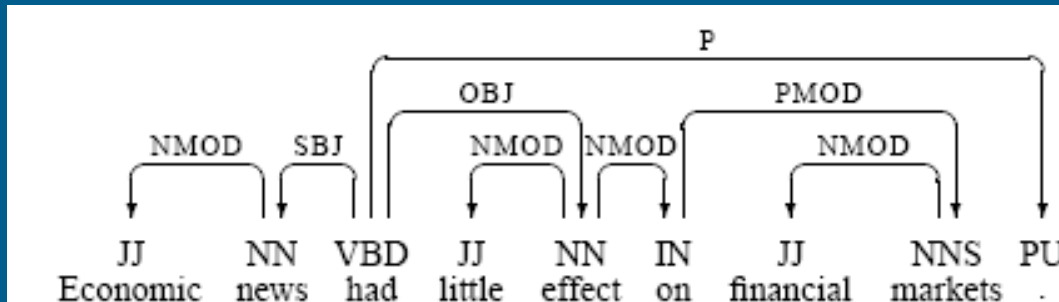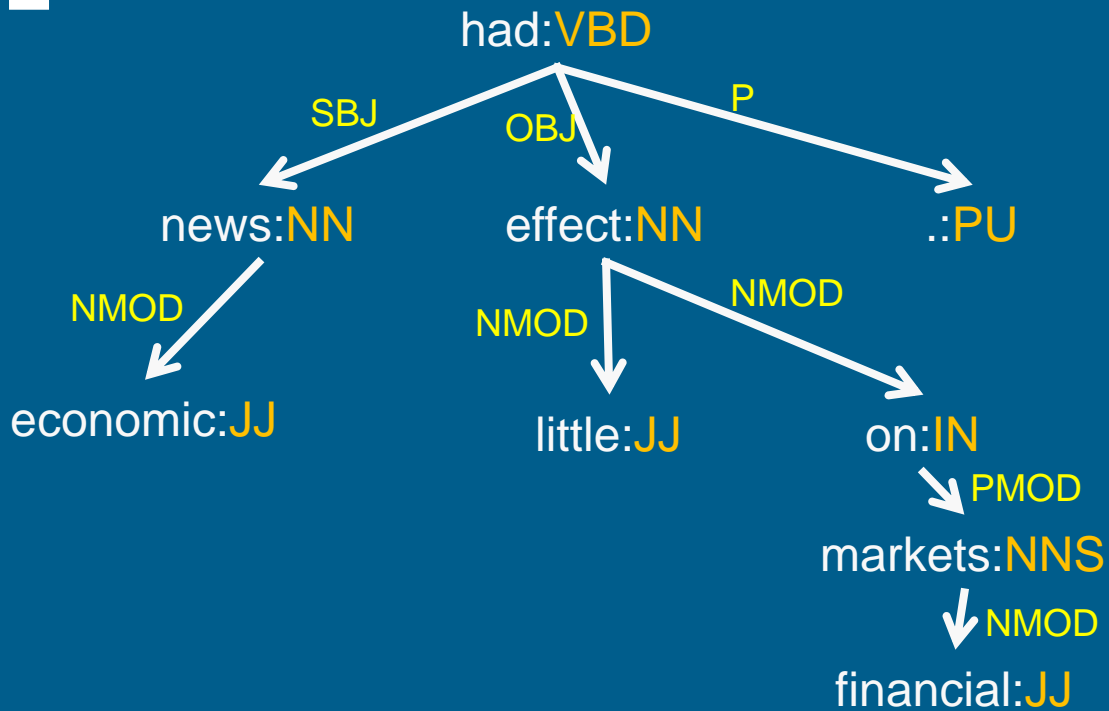Figure 2: Dependency structure for English sentence from the Penn Treebank

Alternativ grafikk

had:VBD

SBJ — news:NN
OBJ — effect:NN
P — .:PU

news:NN
NMOD — economic:JJ

effect:NN
NMOD — little:JJ
NMOD — on:IN

on:IN
PMOD — markets:NNS

markets:NNS
NMOD — financial:JJ

Sammenheng
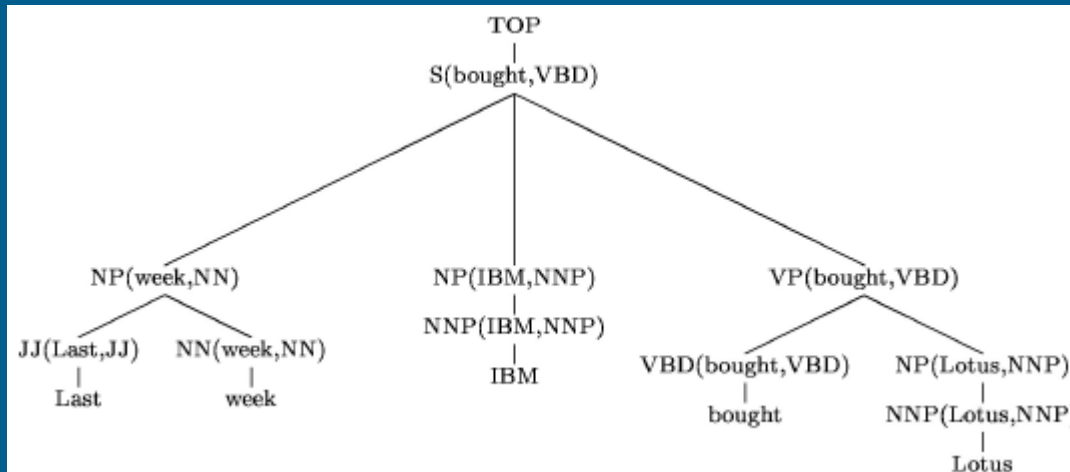
- Fra et hodemerket PS-tre:
- Lag depedensstruktur ved å
  - slå sammen hodet med sin mor
  - Bestem dependenstype
- Fra depedens til frasestruktur ikke rett frem

Formell definisjon

Definition 1

Given a set R of dependency types (arc labels), a *dependency graph* for a sentence $x = (w_1, \ldots, w_n)$ is a labeled directed graph $G = (V, E, L)$, where:

1. $V = \mathbf{Z}_{n+1}$
2. $E \subseteq V \times V$
3. $L : E \to R$

Definition 2

A dependency graph G is *well-formed* if and only if:

1. The node 0 is a root (ROOT).
2. G is connected (CONNECTEDNESS).[4]

Styrker

- Færre ting å ta stilling til
- Velegnet for språk med friere ordstilling
  - Men ikke trivielt å parse fri ordstilling
- Har den vesentlige informasjonen for videre semantisk prosessering:
  - "hvem gjorde hva med hvem"
- Velegnet for parsing:
  - Liknende informasjon som i leksikalsk PCFG tilgjengelig: ord og POS-tagg for hodet

Projektiv?
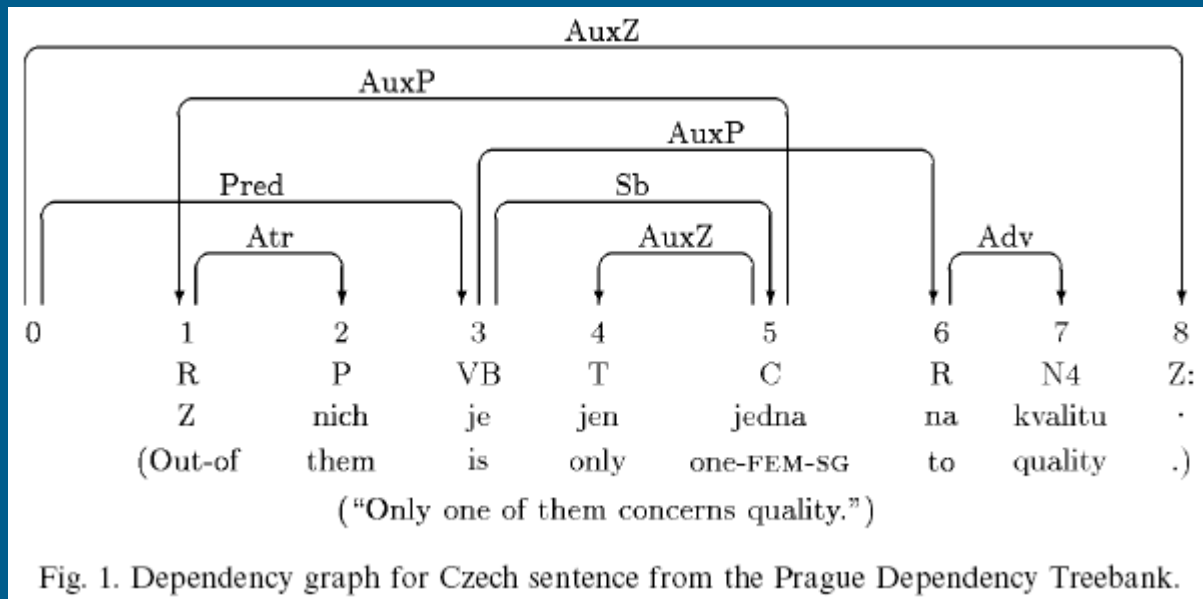
Fig. 1. Dependency graph for Czech sentence from the Prague Dependency Treebank.

- **Mulige krav**
  3. Single-Head: Ikke i→ j og i → k
  4. Acyklisk: ikke i → j og j →* i
  5. Projektiv: Hvis i → j og
     i < k < j eller j < k < i, så i →* k

MaltParser

1. Deterministisk SR-parser for depedenssturktur

2. ”History-based” trekkmodell for å predikere neste trekk (a la Collins)

3. Diskriminativ læring av orakel for å styre parservalgene

Economic news had little effect on financial markets .

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets.SHIFT

NMOD

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets. SHIFT
|| news had little effect on financial markets.   LEFT-ARCH

NMOD

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets. SHIFT
|| news had little effect on financial markets.   LEFT-ARCH
news || had little effect on financial markets.   SHIFT

NMOD SBJ

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets. SHIFT
|| news had little effect on financial markets.   LEFT-ARCH
news || had little effect on financial markets.   SHIFT
|| had little effect on financial markets.   LEFT-ARCH

NMOD   SBJ

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets. SHIFT
|| news had little effect on financial markets. LEFT-ARCH
news || had little effect on financial markets. SHIFT
|| had little effect on financial markets. LEFT-ARCH
had little || effect on financial markets. SHIFT + SHIFT

NMOD   SBJ   NMOD

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets. SHIFT
|| news had little effect on financial markets.   LEFT-ARCH
news || had little effect on financial markets.   SHIFT
|| had little effect on financial markets.   LEFT-ARCH
had little || effect on financial markets.   SHIFT + SHIFT
had || effect on financial markets.   LEFT-ARCH

Economic news had little effect on financial markets.

|| Economic news had little effect on financial markets.

Economic || news had little effect on financial markets.   SHIFT

|| news had little effect on financial markets.   LEFT-ARCH

news || had little effect on financial markets.   SHIFT

|| had little effect on financial markets.   LEFT-ARCH

had little || effect on financial markets.   SHIFT + SHIFT

had || effect on financial markets.   LEFT-ARCH

had effect || on financial markets.   RIGHT-ARCH

NMOD   SBJ   OBJ   NMOD   NMOD

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets.   SHIFT
|| news had little effect on financial markets.   LEFT-ARCH
news || had little effect on financial markets.   SHIFT
|| had little effect on financial markets.   LEFT-ARCH
had little || effect on financial markets.   SHIFT + SHIFT
had || effect on financial markets.   LEFT-ARCH
had effect || on financial markets.   RIGHT-ARCH
had effect on || financial markets.   RIGHT-ARCH

NMOD   SBJ   OBJ   NMOD   NMOD

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.
Economic || news had little effect on financial markets.   SHIFT
|| news had little effect on financial markets.   LEFT-ARCH
news || had little effect on financial markets.   SHIFT
|| had little effect on financial markets.   LEFT-ARCH
had little || effect on financial markets.   SHIFT + SHIFT
had || effect on financial markets.   LEFT-ARCH
had effect || on financial markets.   RIGHT-ARCH
had effect on || financial markets.   RIGHT-ARCH
had effect on financial || markets.   SHIFT

Economic news had little effect on financial markets.

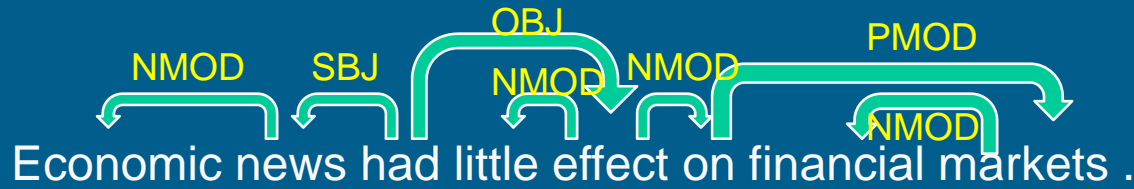|		
|| Economic news had little effect on financial markets.	
Economic || news had little effect on financial markets.	SHIFT
|| news had little effect on financial markets.	LEFT-ARCH
news || had little effect on financial markets.	SHIFT
|| had little effect on financial markets.	LEFT-ARCH
had little || effect on financial markets.	SHIFT + SHIFT
had || effect on financial markets.	LEFT-ARCH
had effect || on financial markets.	RIGHT-ARCH
had effect on || financial markets.	RIGHT-ARCH
had effect on financial || markets.	SHIFT
had effect on || markets.	LEFT-ARCH

|| Economic news had little effect on financial markets.

Economic || news had little effect on financial markets. SHIFT

|| news had little effect on financial markets. LEFT-ARCH

news || had little effect on financial markets. SHIFT

|| had little effect on financial markets. LEFT-ARCH

had little || effect on financial markets. SHIFT + SHIFT

had || effect on financial markets. LEFT-ARCH

had effect || on financial markets. RIGHT-ARCH

had effect on || financial markets. RIGHT-ARCH

had effect on financial || markets. SHIFT

had effect on || markets. LEFT-ARCH

had effect on markets||. RIGHT-ARCH

Economic news had little effect on financial markets.

|| Economic news had little effect on financial markets.		
Economic || news had little effect on financial markets.	SHIFT	
|| news had little effect on financial markets.	LEFT-ARCH	
news || had little effect on financial markets.	SHIFT	
|| had little effect on financial markets.	LEFT-ARCH	
had little || effect on financial markets.	SHIFT + SHIFT	
had || effect on financial markets.	LEFT-ARCH	
had effect || on financial markets.	RIGHT-ARCH	
had effect on || financial markets.	RIGHT-ARCH	
had effect on financial || markets.	SHIFT	
had effect on || markets.	LEFT-ARCH	
had effect on markets||.	RIGHT-ARCH	
had || .	3 x REDUCE	

Economic news had little effect on financial markets .

|| Economic news had little effect on financial markets.	
Economic || news had little effect on financial markets.	SHIFT
|| news had little effect on financial markets.	LEFT-ARCH
news || had little effect on financial markets.	SHIFT
|| had little effect on financial markets.	LEFT-ARCH
had little || effect on financial markets.	SHIFT + SHIFT
had || effect on financial markets.	LEFT-ARCH
had effect || on financial markets.	RIGHT-ARCH
had effect on || financial markets.	RIGHT-ARCH
had effect on financial || markets.	SHIFT
had effect on || markets.	LEFT-ARCH
had effect on markets||.	RIGHT-ARCH
had || .	3 x REDUCE
had . ||	RIGHT-ARCH

Economic news had little effect on financial markets.

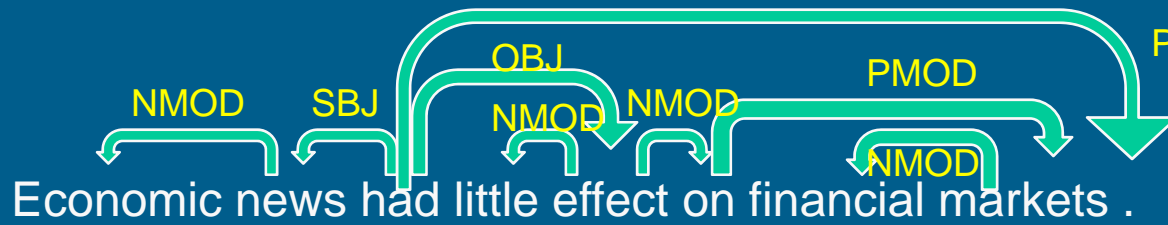| State	Action
|| Economic news had little effect on financial markets.	
Economic || news had little effect on financial markets.	SHIFT
|| news had little effect on financial markets.	LEFT-ARCH
news || had little effect on financial markets.	SHIFT
|| had little effect on financial markets.	LEFT-ARCH
had little || effect on financial markets.	SHIFT + SHIFT
had || effect on financial markets.	LEFT-ARCH
had effect || on financial markets.	RIGHT-ARCH
had effect on || financial markets.	RIGHT-ARCH
had effect on financial || markets.	SHIFT
had effect on || markets.	LEFT-ARCH
had effect on markets||.	RIGHT-ARCH
had || .	3 x REDUCE
had . ||	RIGHT-ARCH
||	2 x REDUCE

0

NMOD SBJ OBJ NMOD NMOD PMOD P

NMOD

Economic news had little effect on financial markets .
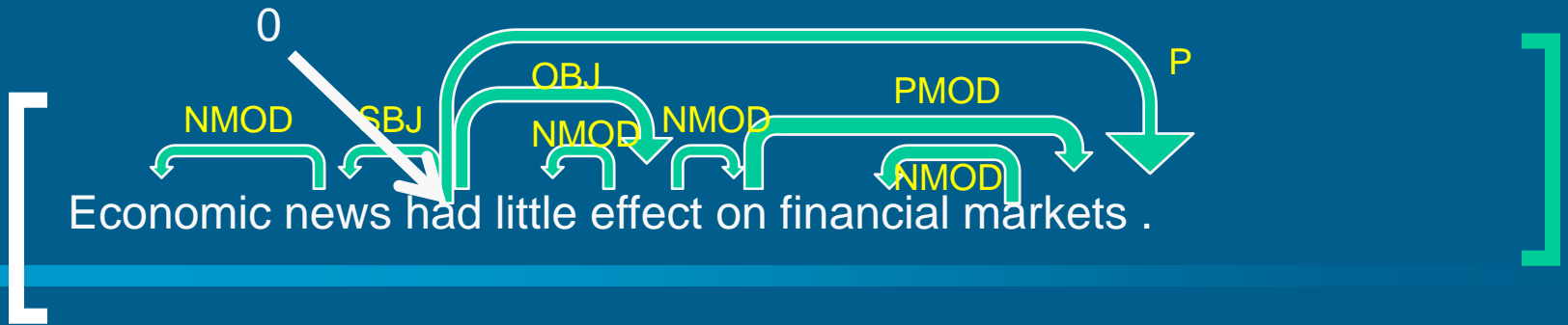
Detalj:
- starter med kant fra 0 til alle andre noder
- fjerner kanten fra 0 til n når vi legger inn ny kant til n (SINGLE-HEAD)
- står igjen med kant fra 0 til had

Deterministic

- So far: non-deterministic algorithm
- Deterministic:
  - An oracle decides the next move when stack is non-empty
  - The oracle is learned from a treebank

Features

σ_j is j-th token on stack

τ_j is j-th token in input

- h(j), head of j in graph
- l(j), leftmost child of j
- r(j), rightmost child of j

- p(j) is tag of j
- d(j) is dependency type of j
- w(j) is word of j

Standard modell:

- p(.):
  - σ_0, σ_1
  - τ_0, τ_1, τ_2, τ_3
- w(.):
  - σ_0, h(σ_0)
  - τ_0, τ_1
- d(.):
  - l(σ_0), σ_0,
  - r(σ_0), l(τ_0)

Economic news had little effect on financial markets .

had effect || on financial markets.

$p(\sigma_0) = p(\text{effect}) = \text{NN}$

$p(\sigma_1) = p(\text{had}) = \text{VBD}$

$p(\tau_0) = p(\text{on}) = \text{IN}$

$p(\tau_1) = p(\text{financial}) = \text{JJ}$

$p(\tau_2) = p(\text{markets}) = \text{NNS}$

$p(\tau_3) = p(.) = \text{PU}$

$w(\sigma_0) = \text{effect}$

$w(h((\sigma_0)) = \text{had}$

$w(\tau_0) = \text{on}$

$w(\tau_1) = \text{financial}$

$d(l(\sigma_0)) = \text{NMOD}$

$d(\sigma_0) = \text{OBJ}$

$d(r(\sigma_0)) = \text{nil}$

$d(l(\tau_0)) = \text{nil}$

Læring

- Trekkene læres fra en trebank med eksempler
- Forenlig med ulike læringsteknikker:
  - Memory-based learning, som TiMBL
  - Support-Vector Machines
  - Osv.

Evaluering

Table 1. *Data sets. AS = Annotation scheme (C = Constituency, D = Dependency, G = Grammatical functions); Pro = Projective; #D = Number of dependency types; #P = Number of PoS tags; TA =Tagging accuracy; #W = Number of words; #S = Number of sentences; SL = Mean sentence length; EM = Evaluation method (T = Held-out test set, CV_k = k-fold cross-validation)*

Language	AS	Pro	#D	#P	TA	#W	#S	SL	EM
Bulgarian	C	no	14	51	93.5	72k	5.1k	14.1	CV_8
Chinese	CG	yes	12	35	100.0	509k	18.8k	27.1	T
Czech	D	no	26	28	94.1	1507k	87.9k	17.2	T
Danish	D	no	54	33	96.3	100k	5.5k	18.2	T
Dutch	CD	no	23	165	95.7	186k	13.7k	13.6	T
English	CG	yes	12	48	96.1	1174k	49.2k	23.8	T
German	CG	no	31	55	100.0	382k	22.1k	17.3	CV_{10}
Italian	D	no	17	89	93.1	42k	1.5k	27.7	CV_{10}
Swedish	CG	yes	17	46	95.6	98k	6.3k	15.5	T
Turkish	D	no	24	484	100.0	48k	5.6k	8.6	CV_{10}

Results

Table 3. *Overview of results. Model = Best feature model ($-$ = omitted, $+$ = added, $\rightarrow$ = replaced by); Settings = TiMBL settings; AS_U = Unlabeled attachment score; AS_L = Labeled attachment score*

Language	Model	Settings	AS_U	AS_L
Bulgarian	$\forall a\,[w(a) \rightarrow s_6(w(a)))]$	Standard	81.3	73.6
Chinese	Standard	$k = 6,\, l = 8$	81.1	79.2
Czech	Standard	Standard	80.1	72.8
Danish	$[w(h(\sigma_0)) \rightarrow s_6(w(h(\sigma_0))); -w(\tau_1)]$	Standard	85.6	79.5
Dutch	Standard	$k = 10$	84.7	79.2
English	Standard	$k = 7,\, l = 5$	88.1	86.3
German	$[-w(h(\sigma_0)); -w(\tau_1); +p(\sigma_2)]$	$k = 13$, IL	88.1	83.4
Italian	Standard	Standard	82.9	75.7
Swedish	Standard	Standard	86.3	82.0
Turkish	$[-p(\sigma_1); -p(\tau_2); -p(\tau_3); -w(h(\sigma_0)); -w(\tau_1)]$	Standard	81.6	69.0

Time complexity

- $O(n^3)$:
  - CKY-recognition
  - CKY w/parse forest
  - CKY for PCFG
  - n-best CKY with PCFG
- Collins-parser:
  - $O(n^5)$ (an adaption is $O(n^4)$)
- MaltParser
  - $O(n)$ with standard model
  - $O(n^2)$ with unrestricted look-ahead

Generative model

- E.g. PCFG
- Generates tree with probability:
  - $P(t) = P(t, s)$ if s is the yield of t
  - $P(t, s) = 0$ for other s
- Hence:
  $$P(t|s) = \frac{P(s|t)P(t)}{P(s)} = \frac{P(t)}{P(s)}$$
  - and
  $$argmax_s P(t|s) = argmax_s P(t)$$
-

  Also language model:
  $$P(s) = \sum_t P(t, s)$$

Discriminative model

- Does not construct the tree
- Only compares P(t1 |s) to P(t2 | s) for candidate trees
- MaltParser
  - does not compare full trees
  - but moves:
    - local decisions