Lexical Acquisition

- Classification of verbs in semantic classes
- Learning of the difference between argument and adjuncts

Automatic lexical learning (1/2)

Many application require very rich lexical information. For example, in machine translation, a good translation is, among other things, determined by

1- The class of the verb:

The rider ***raced*** the horse > Le jockey a fait courir le cheval
'manner of motion' *Le jockey a couru le cheval

The cook ***burnt*** the cake > Le cuisinier a brûlé le gateau
'change of state' *Le cuisinier a fait brûler le gateau

Automatic lexical learning (2/2)

The difference between arguments and adjuncts

a) Gianni ha messo i documenti **nella cartella** > Jean filed the documents

b) Gianni ha trovato i documenti nella cartella > Jean found the documents in the file

In a) the PP is part of the meaning of the verb and in English it can undergo incorporation, while in b) it is extra information.

This knowledge must be stored in the lexicon.

Practical Motivation

- Incorrect attachment of prepositional phrases often constitutes the main source of errors in current parsing systems.

- Correct attachment of PPs is necessary to construct a parse tree which will support the proper interpretation of constituents in the sentence.

- It has been assumed that the ambiguity that needs to be resolved is the **attachment site** of the PP

V-attachment I [saw [the man] [with the telescope]]

N-attachment I [saw [the man [with the telescope]]]

Practical Motivation

•But we also need to know the **function** of the PP in the sentence.

•Minimally, we want to know if the PP is an argument (of the verb or noun) or an adjunct.

Jane baked a cake for her daughter (benefactive argument)

Jane baked a cake for 5 dollars (adjunct)

•Distinguishing arguments from adjuncts is key to identifying the semantic kernel of a sentence.

Practical Motivation

Needed for automatic acquisition of	subcategorisation frames
	argument structure
Several NLP tasks and applications	parsing
	machine translation
	information extraction

First difficulty: Defining argumenthood

I put the book **on the table** (argument)

I read the book **in the room** (adjunct)

Function in the sentence

Argument fills a role in the relation described by the head

Modifier predicates a separate property from the head

Interpretation

Argument interpretation depends exclusively on its governor

Modifier interpretation remains constant with different heads

Second difficulty: contradictory diagnostics

The diagnostics to determine if a PP is an argument or an adjunct are

often contradictory

give rise to relative judgements

Native speaker have a hard time giving a global judgement, but have no trouble judging each diagnostic independently.

Machine learning algorithms are good at finding global solutions and combining partial information.

The method

- detemine the linguistic factors that give rise to the distinction between arguments and adjuncts
- determine a quantification of the diagnostics
- collect data in a corpus
- use collected data as features in a vector to train a binary classifier

Head dependence

Arguments depend on a lexical head

Modifiers do not depend on a specific lexical head, hence they co-occur with many heads

a man/dog/woman/scarecrow with gray hair (ajout)

a member/*dog/*woman/*scarecrow from Parliament (argument)

$$Entropy_{PP}(X) = -\sum_{h \in X} p(h) \log_2 p(h)$$

Countable indicator in corpus: |{v1,…, vn}| **(Merlo and Leybold 2001)**

Range does not take distribution of verbs into account

Expectation Entropy arguments << Entropy adjuncts

Optionality

Arguments tend to be obligatory as they contribute to the semantics of the verb. Adjuncts do not.

John put the book in the room (argument)

*John put the book

John saw/read the book in the room (ajout)

John saw/read the book

Countable indicator: *P(PP|v)*

Expectation *P(PP Arg|v) >> P(PP Mod |v)*

Iterativity and Order

Iterativity: Arguments fill a specific role in the sentence, hence they are not repeated, adjuncts do not

*Chris rented the gazebo to yuppies, to libertarians

Order: In English, arguments are always closer to the governor

Kim met Sandy in Baltimore in the hotel lobby in a corner

The second PP in a sequence is an adjunct P(adj|PP1) ~ P(PP2)

John gave the book to Mary at six

John arrived at six

Copular Paraphrase

Arguments cannot be replaced by a relative sentence
(Only for PPs attached to nouns)

a man from Paris	a man who is from Paris
the destruction of the city	*the destruction that was of the city

Approximation

if PP follows copular verb then it is an adjunct

P(ADJ|PP) =~ P(copula < PP)

Deverbal Nominalisation

PPs following a deverbal noun are likely to be arguments

the destruction of the city

Identify deverbal noun, based on predefined suffixes

ANT	inhabitant, contestant
EE	appointee, employee
ER,OR	singer, writer

The diagnostics and their quantification

(elaboration of Merlo and Leybold 2001)

Head dependence: adjunct PPs co-occur with a larger set of verbs $P(ADJ|PP) \approx Entropy_{PP}(X)$

Optionality: arguments tend to be obligatory $P(ADJ|PP) \approx P(PP|h)$

Iterativity/Order: arguments do not iterate, are closest to head $P(ADJ|PP_1) \approx P(PP_2)$

Copular paraphrase: arguments cannot be replaced by copular relative clause $P(ADJ|PP) \approx P(copula < PP)$

Deverbal nominalisation: PPs following deverbal noun likely to be arguments $P(ADJ|PP) \approx P(deverbnoun < PP)$

The Counts

Corpus: two corpora from the Penn Treebank
- single PP corpus
- two PPs corpus

Counts: estimated on all constructions of interest
(including unambiguous, but no overlapping PPs)

Example of tree in the PTB:

```
(S (NP-SBJ (NNS Workers))
  (VP (VP (VBD dumped)
        (NP (NP (JJ large)
              (NNS sacks))
           (PP (IN of)                    (attachment au nom)
              (NP (DT the)
                 (VBN imported)
                 (NN material))))
        (PP-DIR (IN into)                 (attachment au verbe)
              (NP (DT a)
                 (JJ huge)
                 (NN bin))))
     (, ,)
```


The diagnostics

Head dependence	Entropy of distribution of verbs or nouns co-occurring with a given PP. Back off to classes
Optionality	Conditional probability of PP given verb/noun. Back-off to classes
Iterativity/Order	Count of how many times PP found in second position over all PPs in 2nd position. Back-off to classes
Copular paraphrase	Count how many times PP found following a copula, even if not in relative clause. Back off to classes
Deverbal nominalisation	How many times following a deverbal noun. Deverbal noun based on morphology

Computation of the diagnostics: head dependence

- Problem: often sequences of same PP are rare
- eg: *on Monday* (frequent), *for Catherine* (rare)

- Solution:

Take semantic classes into account (types) instead of words (tokens):

Write for Catherine (person)
Run for John (person) — *write/run/speak for person*
Speak for president (person)

Class Features

Heads with similar semantics and similar syntax can be grouped into classes (Levin 1993, Pinker 1989)

Argument structure is a property of the class

We expect class information to be highly correlated to whether a PP is an argument or an adjunct of the predicate

Class features: Annotation with WordNet

•Nouns: Acts or actions, Animals, Man-made objects, Attributes of people and objects, Body parts, Cognitive processes and contents, …

•Verbs: verbs of grooming, dressing and bodily care, verbs of size, temperature change, intensifying, …

- Top Wordnet nodes are semantic classes
- Most frequent sense

Back- off algorithm

If I have not found the sequence of preferred priority, I look for the next preferred sequence
If I have not found the next preferred sequence, I look for the even next preferred sequence
….
I take the majority class

If vectors are not naturally in a total order, then we can define a back-off lattice

1- verb, preposition, noun

2- verb, preposition, noun class

3- verb class, preposition, noun class

The Experimental Data

Materials: All ambiguous sentences

Total examples 13906

'Rich parents **buy** too many **presents to** their **children**'

V, CLV, N1, CLN1, P, N2, CLN2, <DIAGNOSTICS>, Target

buy, *40*, *present*, *14*, *to*, *child*, *18*, *0.026*, *…*, *1.84*, *2.28*, *0*, *argument*

Method

C5.0 Decision Tree Induction

Training/testing

Annotation of the target attribute

•The PTB does not distinguish between arguments and adjuncts

We assign the target label based on the semantic functional labels of the PPs

Arguments: all untagged PPs (probably quite noisy)
or PPs tagged with CLR, PUT,DTV,BNF, or LGS
Adjuncts: all PPs tagged with DIR, LOC, MNR, PRP, TMP

Validation: PUT,DTV,PRD (args); DIR,LOC,TMP,PRP (adjs) ok
MNR,CLR,BNF,no tag: more doubtful

Native speaker: 7 tests, 8 sentences per tag
Results: MNR adj ok; CLR 5/8 arg; BNF arg ok, no tag 6/8 arg

Experimental hypotheses

• The distinction between arguments and adjuncts can be performed on the basis of information collected from an annotated corpus

•The learning of the distinction between argument and adjunct can be improved by lexical classes and linguistic diagnostic features

over a simple baseline

over a model using lexical features

Experiments

Attached to verb:

Features used	Accuracy (%)
1. Preposition (BASELINE)	67.9
2. Preposition and verb classes	73.8
3. Preposition and Classes	75.0
4. Vcl,P,Hdep1,Hdep2,Opt1,Opt2,Opt3,Iter	**79.8**

Attached to noun (without *of*)

Features used	Accuracy (%)
1. Preposition (BASELINE)	93.3
2. Preposition and classes	93.3
3. Preposition and Hdep1	**94.1**
4. Prep and Iter	93.3

Discussion

• Distinction between arguments and adjuncts can be learned with high accuracy

• For noun attached PPs, baseline performance is so high already, that improvement is less than 1%.

•For verb attached PPs improvement over the baseline is greater than 10%, reducing the errror rate by 1/3.

•The most informative features are lexical classes, head dependence, optionality and iterativity.

General Conclusions

• The obtained results show that it is possible to show complex linguistic tasks automatically, using mainly corpus-based statistics.

• Experiments that include the notion of class give better results than those where classes are not used, confirming

+ the usefulness of the notion of class for learning

+ the correlation between the class of a verb and its argument structure

Future work will lie in integrating a finer-grained classification of argument and adjuncts and finer-grained classes.

THANK YOU