Lexical Acquisition in NLP

Attachment Ambiguity

and

Selectional Preferences

Attachment Ambiguity

Problem:

(1) The children ate the cake *with a spoon*.

Does the PP attach to the VP or to the NP?

→ syntactic ambiguity

→ easy to answer for humans, difficult for computers

Attachment Ambiguity

Solution:

Lexical preferences can be used for disambiguation

Calculation of likelihood ratio (λ):

(2)

$$\lambda(v, n, p) = \log \frac{P(p|v)}{P(p|n)}$$

P(p|v) = the probability of seeing a PP with p after v
P(p|n) = the probability of seeing a PP with p after n

→ Attachment to **verb** for **λ > 0**, to **noun** for **λ < 0**

Attachment Ambiguity

Problem:

There is a preference (at least in the English language) for attaching phrases low in the parse tree
→ Late Closure Strategy

(3) Chrysler confirmed that it would end its troubled venture *with Maserati*.

P(p|v) ≈ 0.118 (NYTC)
P(p|n) ≈ 0.107

→ Attachment to the verb is predicted!

Attachment Ambiguity

Problem:

There is a preference (at least in the English language) for attaching phrases low in the parse tree
→ Late Closure Strategy

(3) Chrysler confirmed that it would end its troubled venture *with Maserati*.

P(p|v) ≈ 0.118 (NYTC)
P(p|n) ≈ 0.107

→ Attachment to the verb is predicted! WRONG

Attachment Ambiguity

Solution:

Probabilistic Models (e.g. Hindle and Rooth (1993))

Question: How likely is it for a P to attach to V or N?

VA*ρ*=1/0: There is/isn't a PP headed by P following and attaching to V

NA*ρ*=1/0: There is/isn't a PP headed by P following and attaching to N

Both can equal 1:

(4) He put the book *[on WWII] [on the table]*.

Attachment Ambiguity

Calculating the probability of PP attachment to N and V:

(5)

$$\begin{aligned} P(\mathrm{VA}_p, \mathrm{NA}_p | v, n) &= P(\mathrm{VA}_p | v, n) P(\mathrm{NA}_p | v, \mathrm{n}) \\ &= P(\mathrm{VA}_p | v) P(\mathrm{NA}_p | n) \end{aligned}$$

We assume **conditional independence** of the two attachments (see (4))

Attachment Ambiguity

Calculating the probability of PP attachment to N ($\text{NA}\rho=1$):

(6)

$$\begin{aligned} \mathrm{P}(\mathrm{Attach}(\mathrm{p}) = n|v,n) &= P(\mathrm{VA}_p = 0 \text{ v } \mathrm{VA}, = 1/\mathrm{v}) \text{ x } P(\mathrm{NA}_p = 1|n) \\ &= 1.0 \text{ x } P(\mathrm{NA}_p = 1|n) \\ &= P(\mathrm{NA}_p = 1|n) \end{aligned}$$

→ PPs modifying V in the same sentence are of no importance for determining the status of the PP immediately following N

Attachment Ambiguity

Calculating the probability of PP attachment to V (VA*ρ*=1):

(7)

$$\begin{aligned}\mathrm{P}(\mathrm{Attach}(\mathrm{p}) = \nu|\nu, \mathrm{n}) &= P(\mathrm{VA}_p = 1, \mathrm{NA}_p = 0|\nu, n) \\ &= P(\mathrm{VA}_p = 1|\nu)P(\mathrm{NA}_p = 0|n)\end{aligned}$$

→ NA*ρ*=0 must hold in order to avoid crossing lines in the phrase structure!

Attachment Ambiguity

Likelihood ratios, again, but more complex:

$$\begin{aligned}\lambda(v,n,p) &= \log_2 \frac{P(\text{Attach}(p) = v \mid v, n)}{P(\text{Attach}(p) = n \mid v, n)} \\ &= \log_2 \frac{P(\text{VA}_p = 1 \mid v) P(\text{NA}_p = 0 \mid v)}{P(\text{NA}_p = 1 \mid n)}\end{aligned} \quad (8)$$

→ again, attachment to **verb** for **λ > 0**, to **noun** for **λ < 0**
→ probability of error decreases with greater values

Attachment Ambiguity

Question: How does one estimate the probabilities P(VA*ρ*=1|V) and P(NA*ρ*=1|V)?

Answer: Maximum Likelihood estimates:

(9)

$$P(\mathrm{VA}_p \; = \; 1(\mathrm{v}) \; = \; \frac{C(v,p)}{C(v)}$$

$$P(\mathrm{NA}_p \; = \; 1|n) \; = \; \frac{C(n,p)}{\mathrm{C(n)}}$$

C(v), C(n) = the number of occurrences of V and N in the corpus

C(v,p), C(n,p) = the number of attachments

Attachment Ambiguity

Problem: Unlabeled Corpus (=no attachments specified)

Solution (Hindle and Rooth (1993)):

1. Build an initial model from **unambiguous** cases (e.g. “The road *to London* is long and winding.”)

2. Apply initial model to **ambiguous** cases and apply them to the appropriate counts if λ exceeds a threshold

3. Divide remaining cases evenly between the counts

Attachment Ambiguity

Example:

(10) Moscow sent more than 1000.000 soldiers *into Afghanistan* ...

(11)

$$P(\mathrm{VA}_{into} = 1 | send) = \frac{C(send, \mathrm{into})}{C(send)} = \frac{86}{1742.5} \approx 0.049$$

$$P(\mathrm{NA}_{into} = 1 | soldiers) = \frac{C(soldiers, into)}{C(soldiers)} = \frac{1}{1478} \approx 0.0007$$

(12)

$$P(\mathrm{NA}_{into} = 0 | soldiers) = 1 - P(\mathrm{NA}_{into} = 1 | soldiers) \approx 0.9993$$

(13)

$$A\ (send, soldiers, into) \approx \log, \frac{0.049 \text{ x } 0.9993}{0.0007} \approx 6.13$$

→ Attachment to the verb is $2^{6.13}$ ≈ 70 times more likely.

Attachment Ambiguity

Limitations of the presented model:

- only considers the identity of P, V and N. However:

(13) I saw a man *with a telescope*.
I saw a man *with a broken leg*.

- considers only the most basic case of PP attachment (PP occurs immediately after the modified element)

Attachment ambiguities also present in noun compounds, adverbial phrases and participipial phrases.

Selectional Preferences

(14) Susan had never eaten a fresh *durian* before.

durian is missing from our dictionary, but we can infer that it is a type of food from the selectional restrictions of *eat*

How can selectional preferences be defined?

Resnik (1993,1996):
→ selectional preference strength
→ selectional association

Selectional Preferences

Selectional preference strength: how strongly does the verb constrain its direct object?

(15) $$S(v) = D(P(C|v)||P(C)) = \sum_{c} P(c|v) \log \frac{P(c|v)}{P(c)}$$

P(C) = the overall probability distribution of noun classes

P(C|v) = the probability distribution of noun classes in the direct object position of v

Selectional Preferences

Selectional association: Contribution of a verb class to the overall preference strength of a verb:

(16)
$$A(v,c) = \frac{P(c|v)\log\frac{P(c|v)}{P(c)}}{S(v)}$$

Association strength of a **noun** (as opposed to a noun class) = the highest association strength of any of its classes

Selectional Preferences

What is the probability that a direct object in a certain noun class occurs given a certain verb?

(19) $$P(v,c) = \frac{1}{N} \sum_{n \in \mathrm{words}(c)} \frac{1}{|\mathrm{classes}(n)|} C(v,n)$$

N = total number of verb-object pairs in the corpus

words(c) = set of all nouns in class c

classes(n) = number of noun classes containing the noun as a member

C(v,n) = number of verb-object pairs with v as the verb and n as the head of the object NP

Selectional Preferences

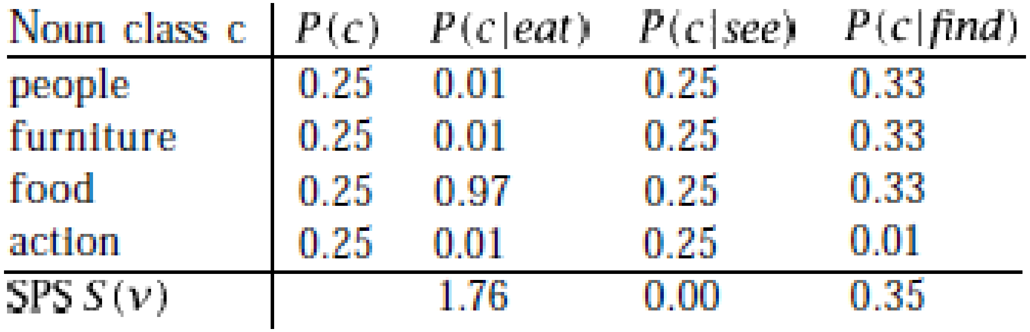
Example (Selectional Association):

(18) Susan interrupted the chair.

A (interrupt, people) >> A (interrupt, furniture)
Hence: A(interrupt, chair) = A(interrupt, people)

Example (Selectional Preference Strength):

(19)

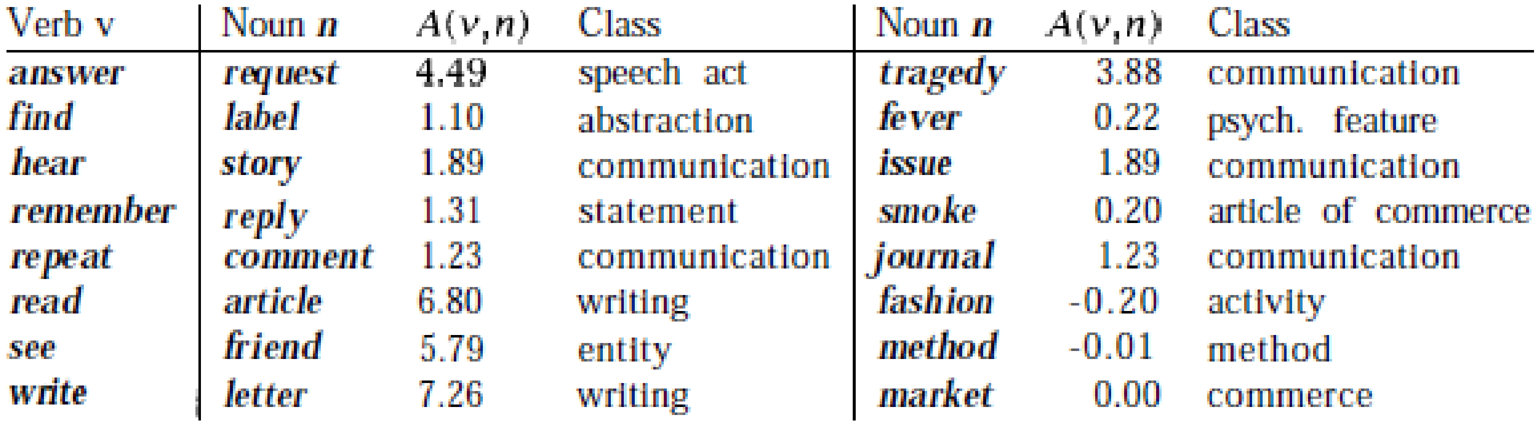


Noun class c	$P(c)$	$P(c \mid eat)$	$P(c \mid see)$	$P(c \mid find)$
people	0.25	0.01	0.25	0.33
furniture	0.25	0.01	0.25	0.33
food	0.25	0.97	0.25	0.33
action	0.25	0.01	0.25	0.01
SPS $S(v)$		1.76	0.00	0.35

Selectional Preferences

Example (Brown corpus):

(20)



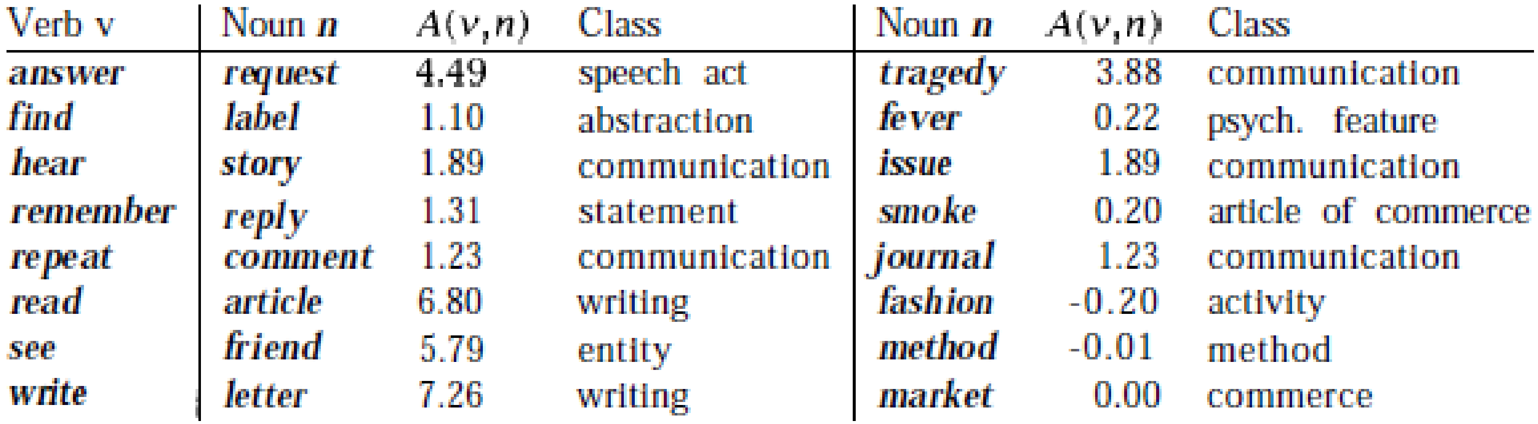
Verb v	Noun *n*	$A(v,n)$	Class	Noun *n*	$A(v,n)$	Class
answer	***request***	4.49	speech act	***tragedy***	3.88	communication
find	***label***	1.10	abstraction	***fever***	0.22	psych. feature
hear	***story***	1.89	communication	***issue***	1.89	communication
remember	***reply***	1.31	statement	***smoke***	0.20	article of commerce
repeat	***comment***	1.23	communication	***journal***	1.23	communication
read	***article***	6.80	writing	***fashion***	-0.20	activity
see	***friend***	5.79	entity	***method***	-0.01	method
write	***letter***	7.26	writing	***market***	0.00	commerce

Judgment errors mostly due to the model choosing the highest association strength among the possible classes of a noun (e.g. hear: story/issue)

Selectional Preferences

Example (Brown corpus):

(20)



Verb v	Noun n	$A(v,n)$	Class	Noun n	$A(v,n)$	Class
answer	*request*	4.49	speech act	*tragedy*	3.88	communication
find	*label*	1.10	abstraction	*fever*	0.22	psych. feature
hear	*story*	1.89	communication	*issue*	1.89	communication
remember	*reply*	1.31	statement	*smoke*	0.20	article of commerce
repeat	*comment*	1.23	communication	*journal*	1.23	communication
read	*article*	6.80	writing	*fashion*	-0.20	activity
see	*friend*	5.79	entity	*method*	-0.01	method
write	*letter*	7.26	writing	*market*	0.00	commerce

Judgment errors mostly due to the model choosing the highest association strength among the possible classes of a noun (e.g. hear: story/issue)

Implicit Object Alternation: The more constraints a verb puts on its object, the more likely it is to permit an implicit object (e.g. eat vs. find)