

Resolving Shell Nouns

EMNLP 2014

Doha, Qatar



by

Varada Kolhatkar and Graeme Hirst

University of Toronto

Shell nouns

(Vendler 1968; Halliday and Hasan 1976; Ivanic 1991; Asher 1994; Francis 1994; Schmid 2000, inter alia)

Shell noun resolution: Identifying shell content of a shell noun phrase in the given context

shell content

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is dividing communities across the country.

shell noun phrase

Background

Examples of shell nouns

Schmid provides a list of 670 shell nouns

fact	issue	problem	principle	decision
thing	concept	reason	notion	phenomenon
idea	rumour	legend	message	possibility
belief	plan	truth	theory	thought
order	trend	argument	proposal	certainty

Why do we care?

- Ubiquity of shell nouns
 - *fact, idea, problem*: among 100 most frequently occurring nouns in the BNC (Schmid 2000)
- Functions in discourse
 - **Characterize** and **label** information in the context
- Potential applications
 - Discourse understanding, text summarization, non-factoid question answering, ESL learning (Francis 1988; Flowerdew 2003; Hinkel 2004)

Different types of usages

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

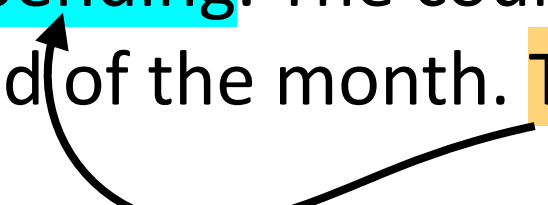
A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

Different types of usages

anaphoric (ASN)

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...



The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

Different types of usages

anaphoric (ASN)

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

cataphoric (CSN)

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

Different types of usages

anaphoric (ASN)

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

cataphoric (CSN)

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

indefinite shell content

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

Different types of usages

anaphoric (ASN)

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

cataphoric (CSN)

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

indefinite shell content

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

non-shell noun usage

Different types of usages

anaphoric (ASN)

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

cataphoric (CSN)

✓ The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

indefinite shell content

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.

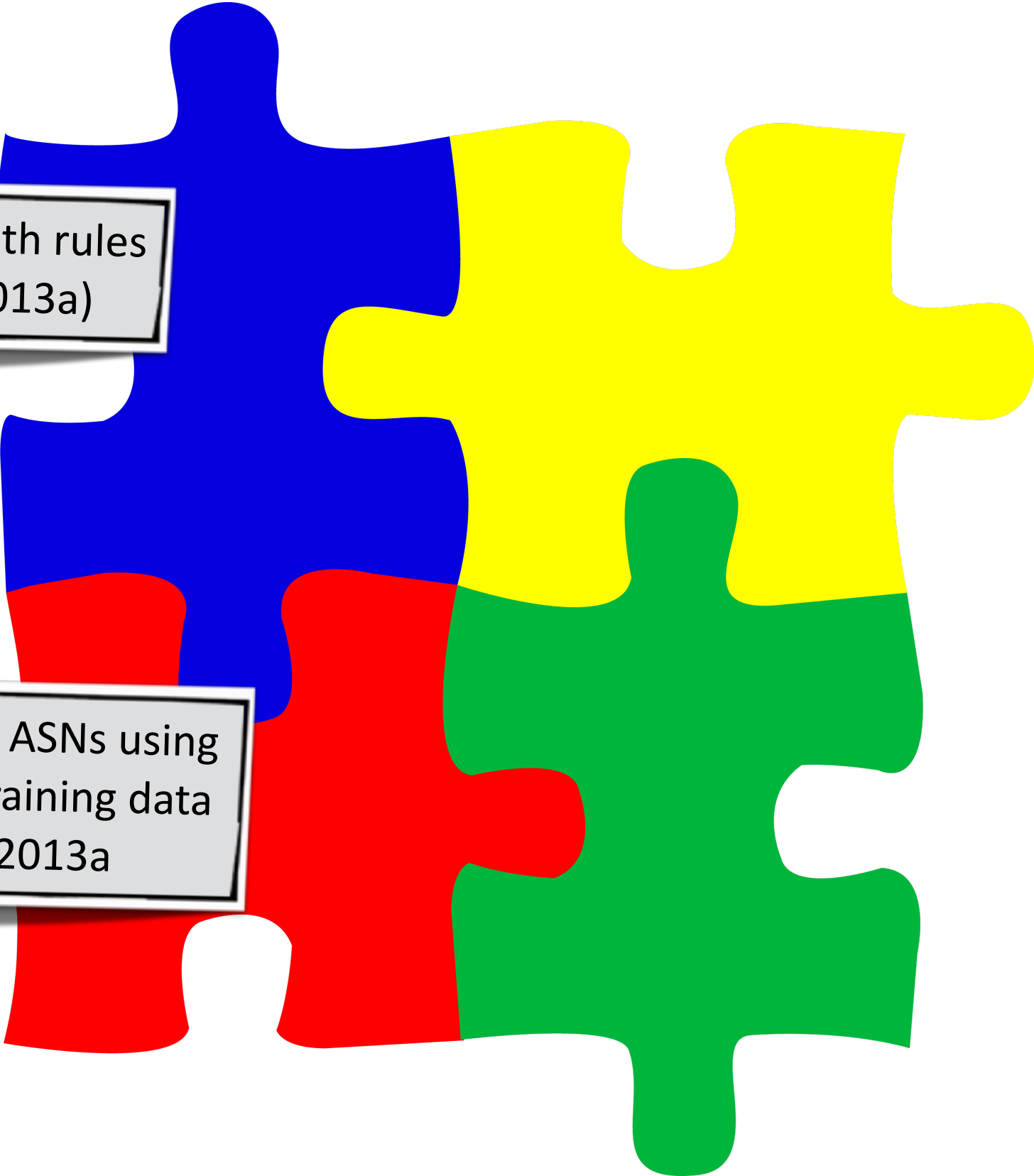
non-shell noun usage

State of the art in CL

- Fair amount of attention in Linguistics
(Vendler 1968; Halliday and Hasan 1976; Ivanic 1991;
Asher 1994; Francis 1994; Schmid 2000, inter alia)

... but largely ignored in CL

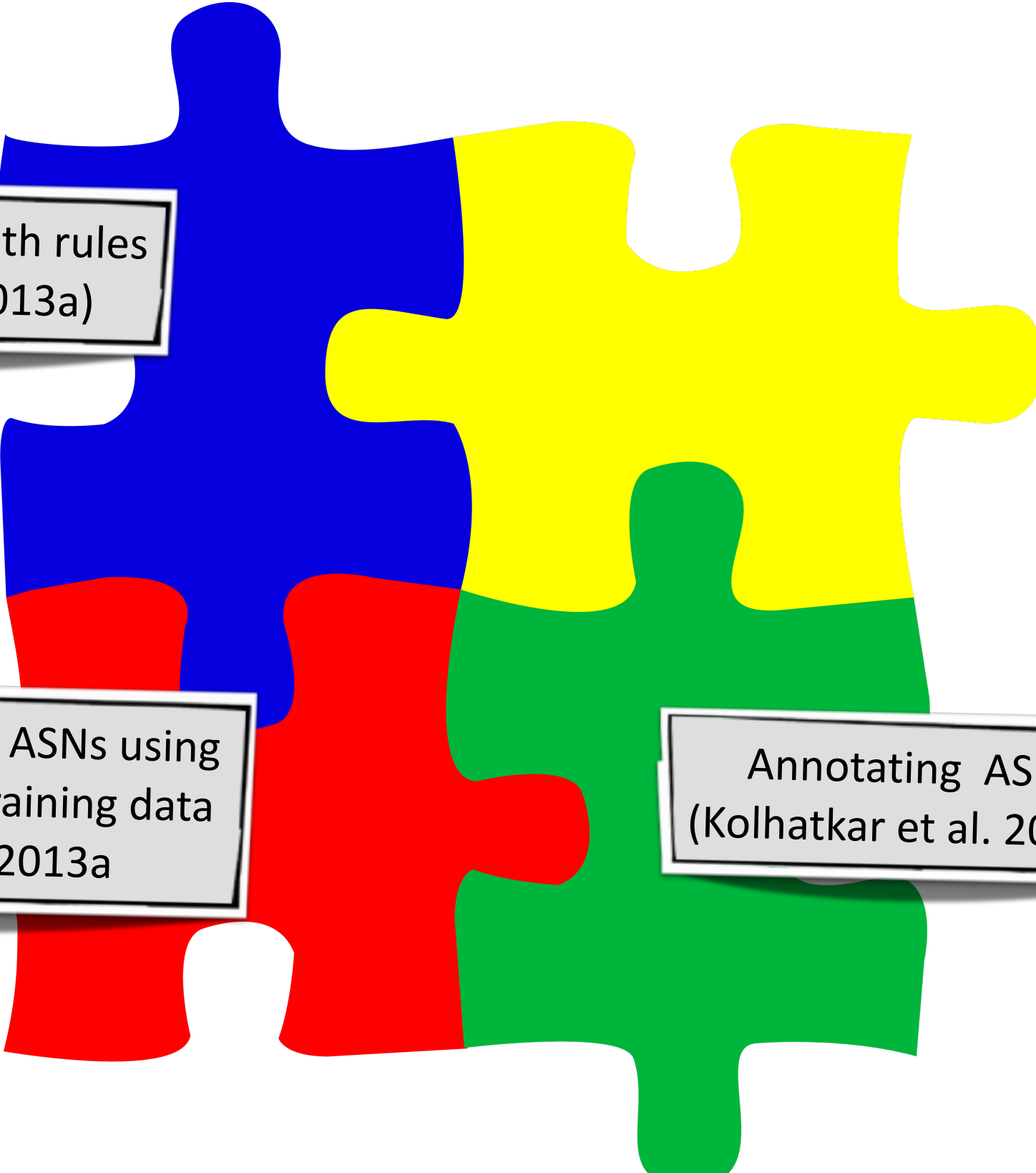
End-to-end shell noun resolution



Resolving six CSNs with rules
(Kolhatkar et al. 2013a)

Resolving the same six ASNs using
CSN shell content as training data
(Kolhatkar et al. 2013a)

End-to-end shell noun resolution

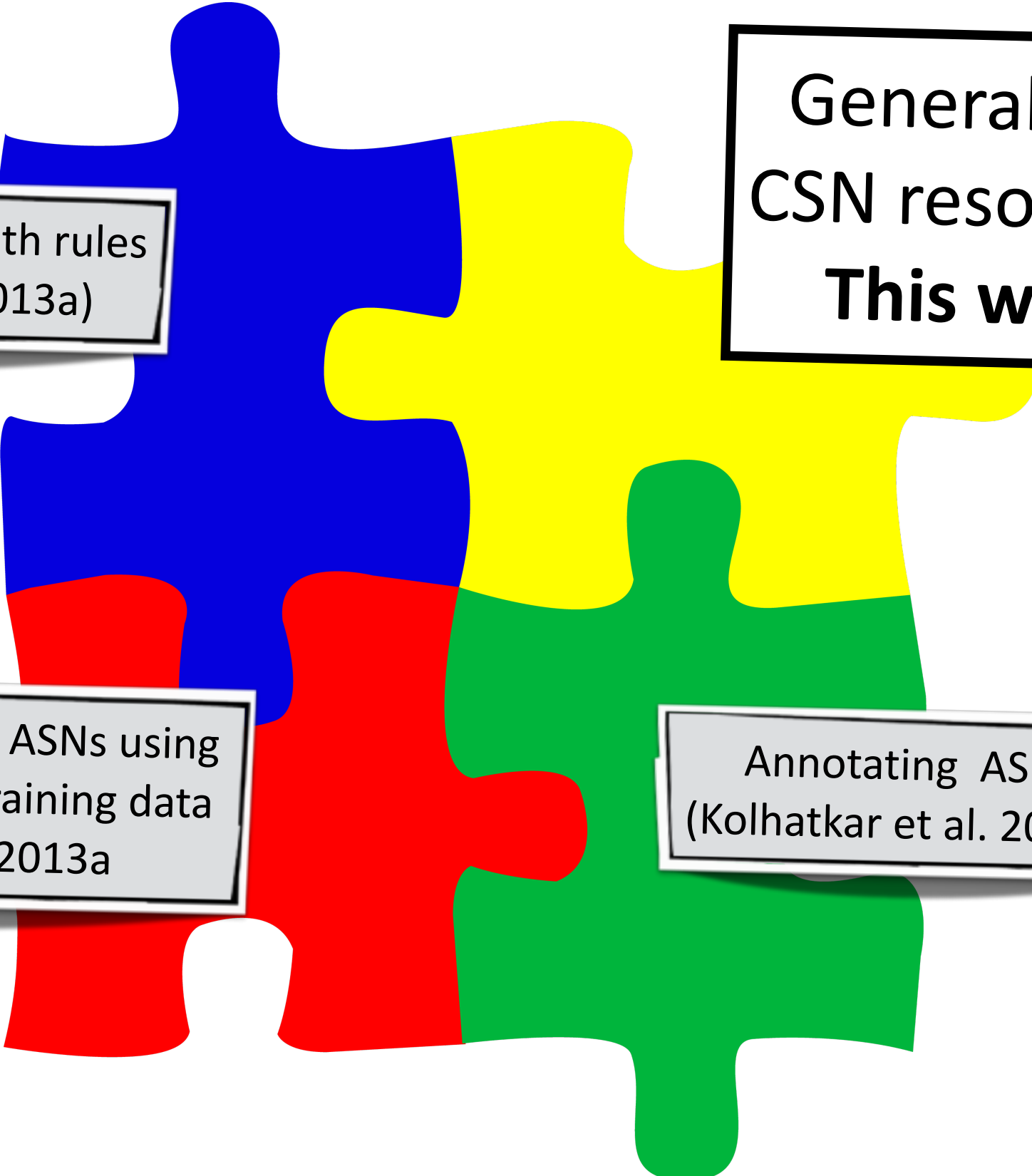


Resolving six CSNs with rules
(Kolhatkar et al. 2013a)

Resolving the same six ASNs using
CSN shell content as training data
(Kolhatkar et al. 2013a)

Annotating ASNs
(Kolhatkar et al. 2013b)

End-to-end shell noun resolution



Resolving six CSNs with rules
(Kolhatkar et al. 2013a)

Generalizing
CSN resolution
This work

Resolving the same six ASNs using
CSN shell content as training data
(Kolhatkar et al. 2013a)

Annotating ASNs
(Kolhatkar et al. 2013b)

CSN patterns

(Schmid, 2000)

Pattern Example

N-be-
clause

N-be-to Our plan is **to** hire and retain the best managers we can.

N-be-that The major reason is **that** doctors are uncomfortable with uncertainty.

N-be-wh Of course, the central issue is **whether** animal testing is cruel.

N-to The decision **to** disconnect the ventilator came after doctors found no brain activity.

N-that These challenges do not undermine the fact **that** museums are on a high.

N-wh If there ever is any doubt **whether** a plant is a poppy or not, break off a stem and squeeze it.

N-of The concept **of** having an outsider as Prime Minister is outdated.

N-clause

Challenges

Idiosyncrasies

CSN

shell content

The fact that a major label hadn't been at liberty to exploit and repackaging the material on CD meant that prices on the vintage LP market were soaring.

CSN

effect

One reason that 60 percent of New York City public-school children read below grade level is that many elementary schools don't have libraries.

cause = shell content

Limitations of current CL tools

Parsing errors with respect to
relative and complement clauses

relative clause

The fact that she pointed out is that nobody
on the panel was an expert.

complement clause

The fact that nobody on the panel was an
expert is the source of severe criticism.

CSNs: a semantic phenomenon

CSN

effect

One reason that 60 percent of New York City public-school children read below grade level is that many elementary schools don't have libraries.

cause = shell content

- Identify that *reason* expects two arguments: cause and effect
- Identify that the shell content is given in the cause argument
- Identify the syntactic constituent representing cause

Shell noun families

(Schmid, 2000)

- Where can we find this kind of semantic knowledge?

Answer:
Schmid's semantic families

- Schmid groups together different usages of 670 shell nouns into 79 semantic families

Shell noun families

(Schmid, 2000)

Idea

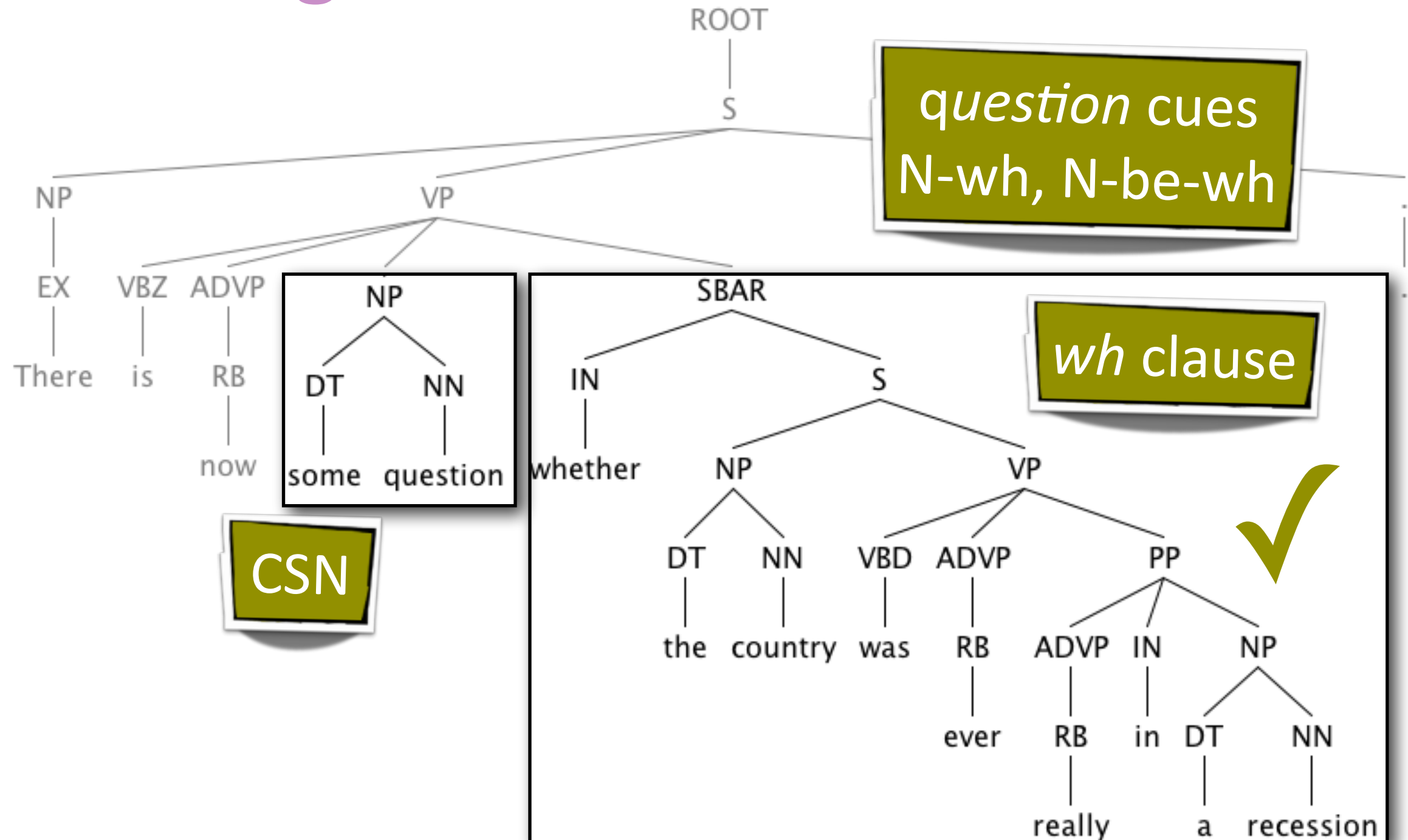
Shared semantic features [MENTAL], [CONCEPTUAL]

frame Mental; focus on propositional content of IDEA

Nouns *Point, idea, position, issue, theory, notion, thought, principle, rule, subject, image, myth, law, theme, concept, secret, scenario, wisdom, hypothesis, thesis,...*

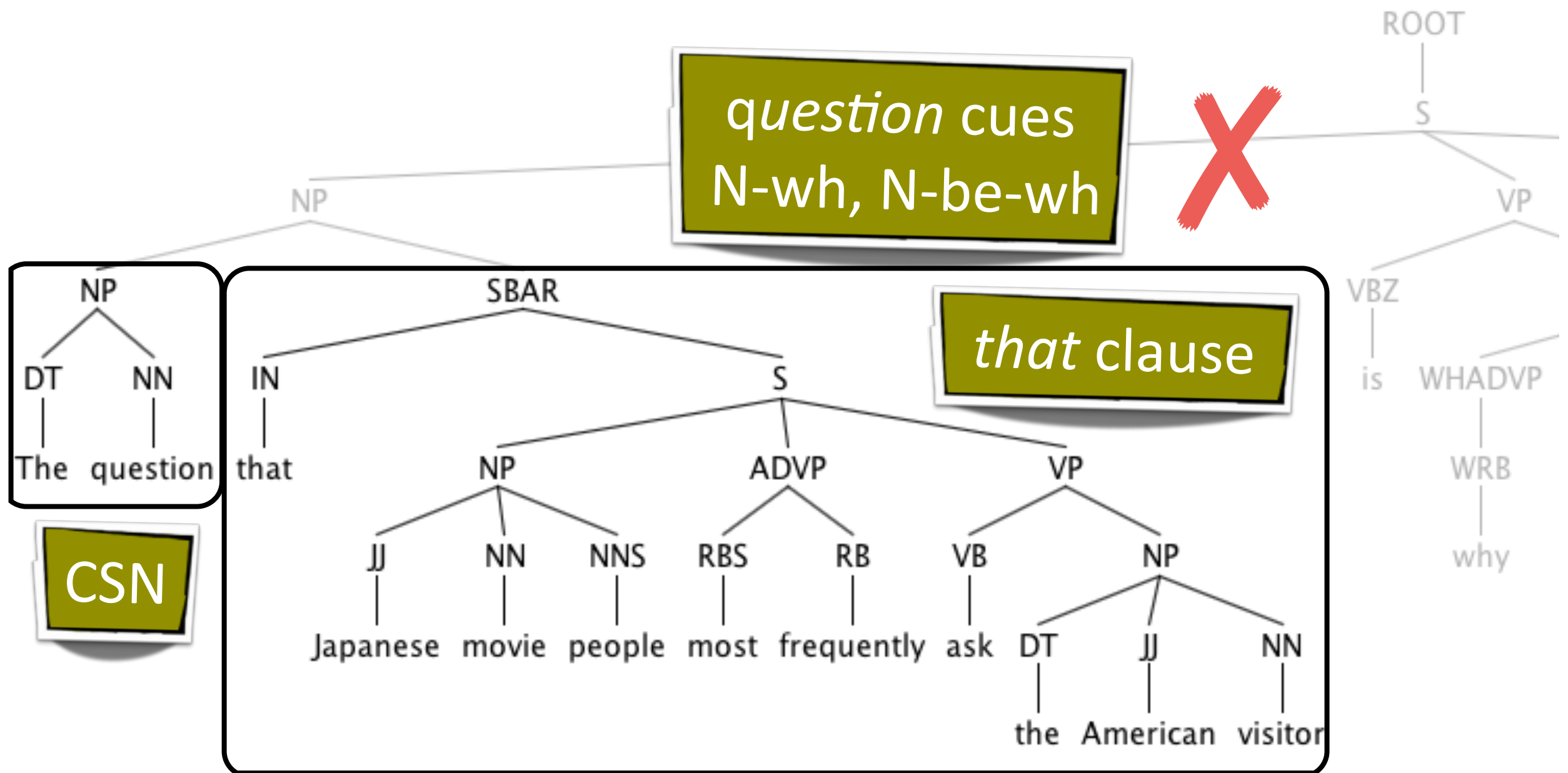
Pattern *N-be-that, N-that*

Using families for CSN resolution



There is now some **question** whether the country was ever really in a recession.

Using families for CSN resolution



The **question** that Japanese movie people most frequently ask the American visitor is why ...

Evaluation

Evaluation data

Idea family

Semantic features: [mental], [conceptual]

Frame: member of IDEA

Nouns: *idea* **idea, issue, concept** *theory, ...*

Patterns: *N-be-that/of, N-that/of*

Trouble family

Semantic features: [eventive], [attitudinal], [manner], [deontic]

Frame: general eventive

Nouns: *problem, trouble, difficulty,*

Patterns: *N-be-to*

Thing family

Semantic features: [factual]

Frame: general factive

Nouns: *fact, phenomenon* *business*

Patterns: *N-that, N-be-that*

Plan family

Semantic features: [mental], [volitional], [manner]

Frame:

Nouns: *decision, plan, policy*

Patterns: *N-be-to/that, N-to/that*

Problem family

Semantic features: [factual], [attitudinal], [impeding]

Frame: general factual

Nouns: *problem, trouble, difficulty, point, thing, snag,*

Patterns: *N-be-that/of*

Reason family

Semantic features: [factual], [causal]

Frame: causal attentional focus on CAUSE

Nouns: *reason* *thing*

Patterns: *N-be-that/why, N-that/why*

The CSN Corpus

- Base corpus: The New York Times corpus
(Sandhaus, 2008)
- ~100 instances per 12 selected shell nouns
idea, issue, concept, decision, plan, policy,
problem, trouble, difficulty, fact,
phenomenon, reason

Annotation methodology



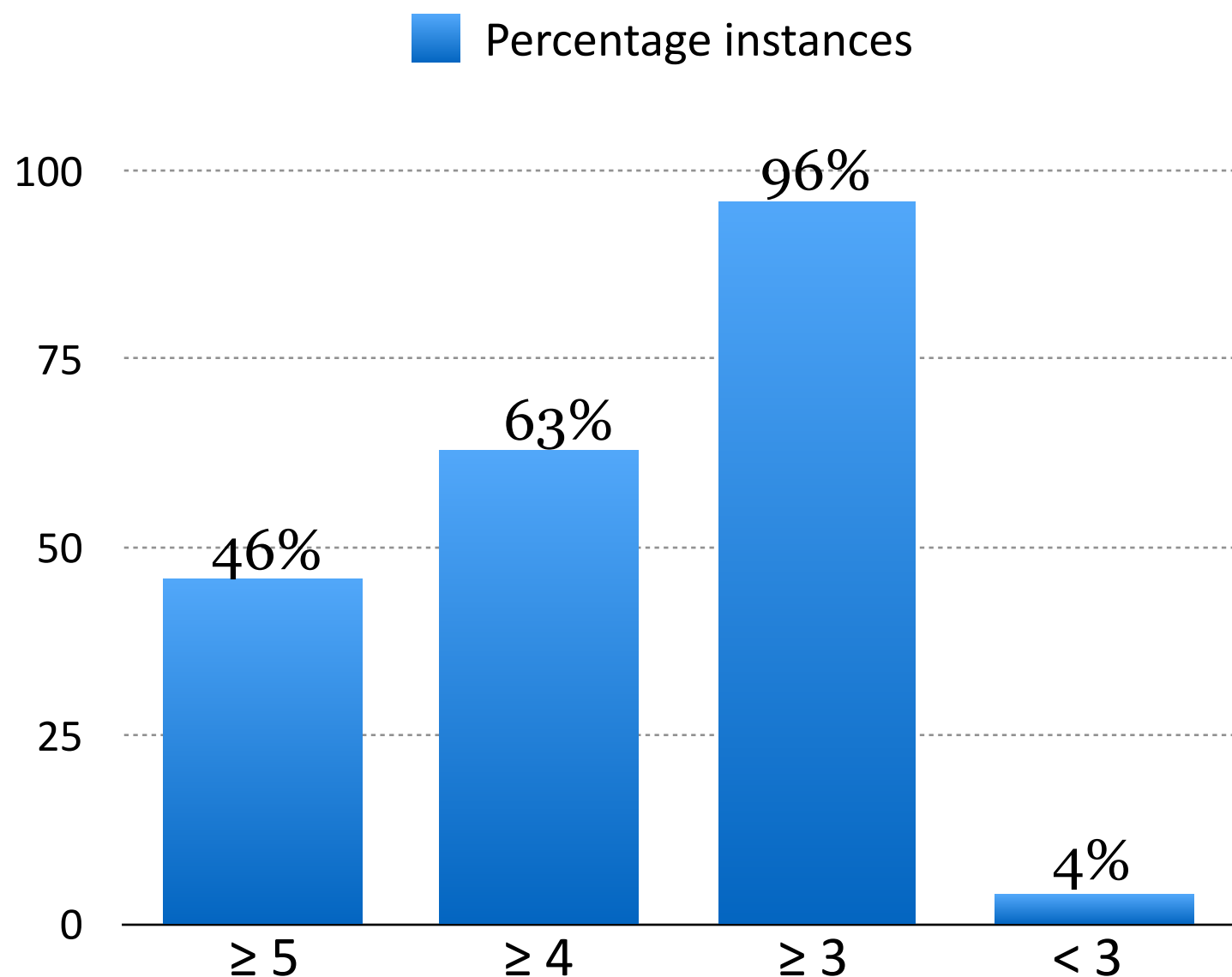
One reason that 60 percent of New York City public-school children read below grade level is that many elementary schools don't have libraries.

Select one of the following options that is the shell content of the highlighted shell noun phrase

- ✗ (a) that 60 percent of New York City public-school children read below grade level
- ✓ (b) that many elementary schools don't have libraries
- ✗ (c) libraries
- ✗ (d) None of the above

Annotation results

At least 3 out of 5 annotators agreed on a single answer for 96% of the instances



minimum 5
annotators
per instance

Baseline

Extract the complement clause depending upon the pattern the example follows



The fact that nobody on the panel was an expert is the source of severe criticism.

Women had good reason to expect a steady and inevitable progression toward gender parity in politics.

Baseline

Extract the complement clause depending upon the pattern the example follows



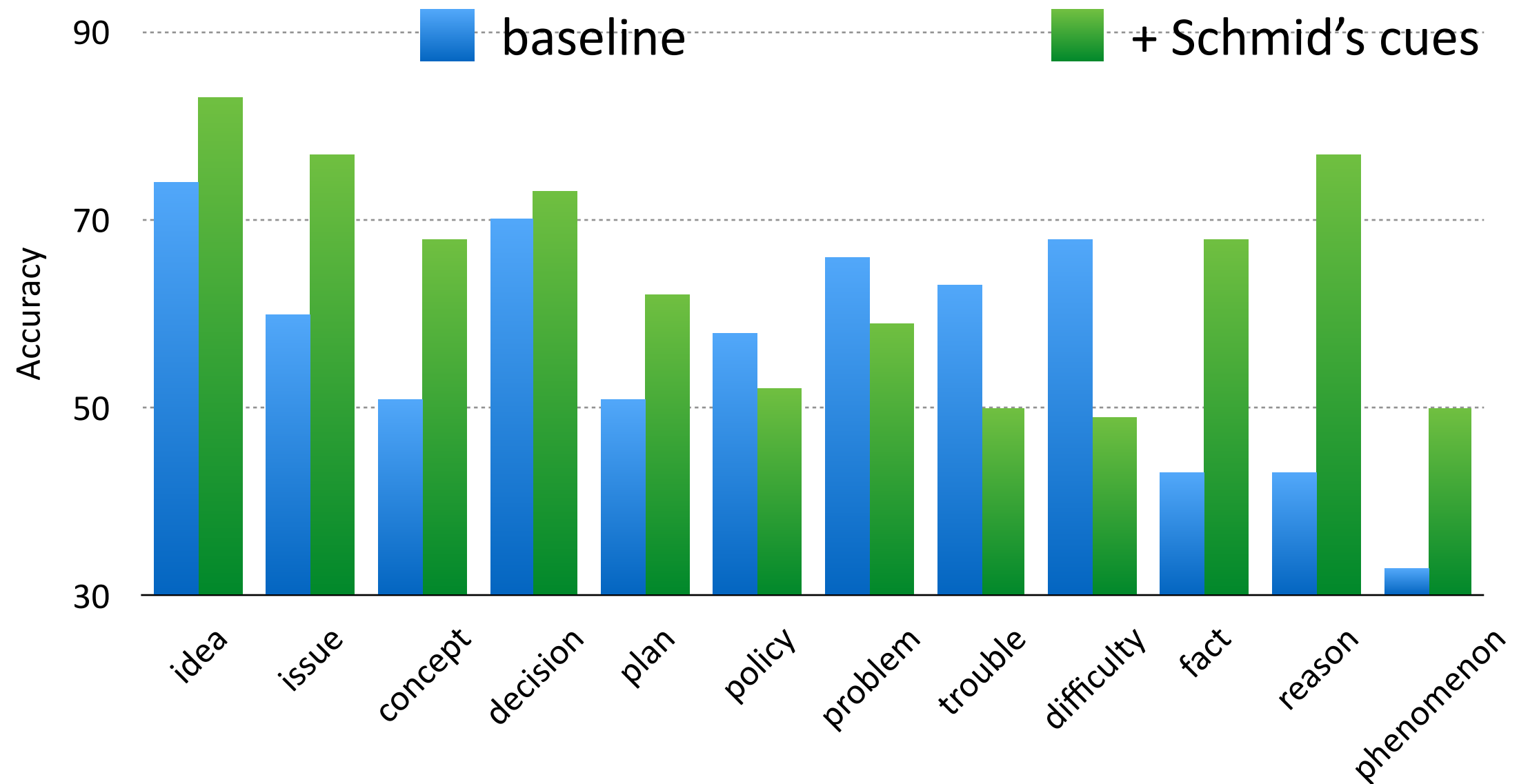
The fact that nobody on the panel was an expert is the source of severe criticism.



Women had good reason to expect a steady and inevitable progression toward gender parity in politics.

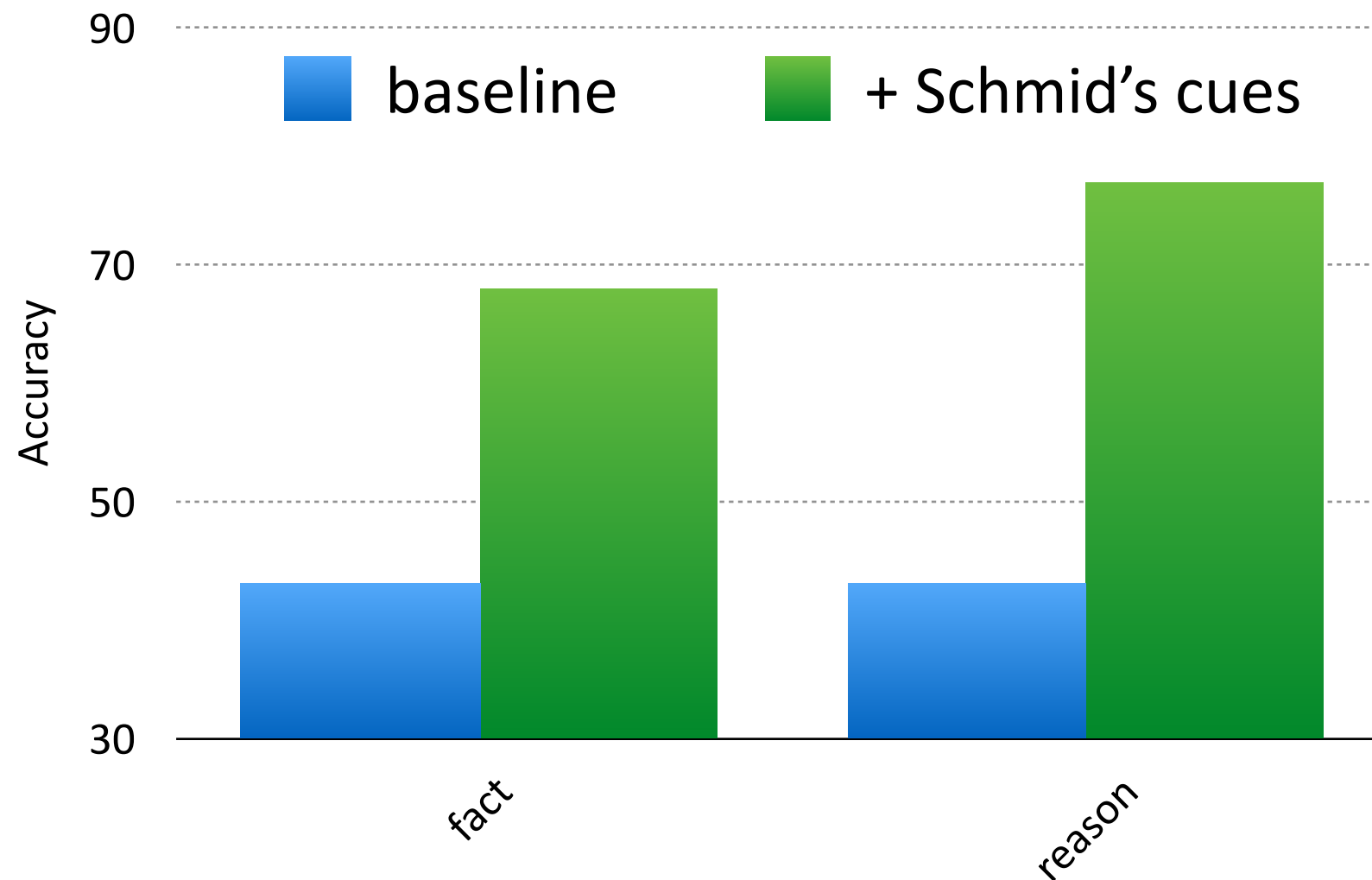
CSN resolution results

Semantic knowledge from Schmid's families is helpful in resolving CSNs



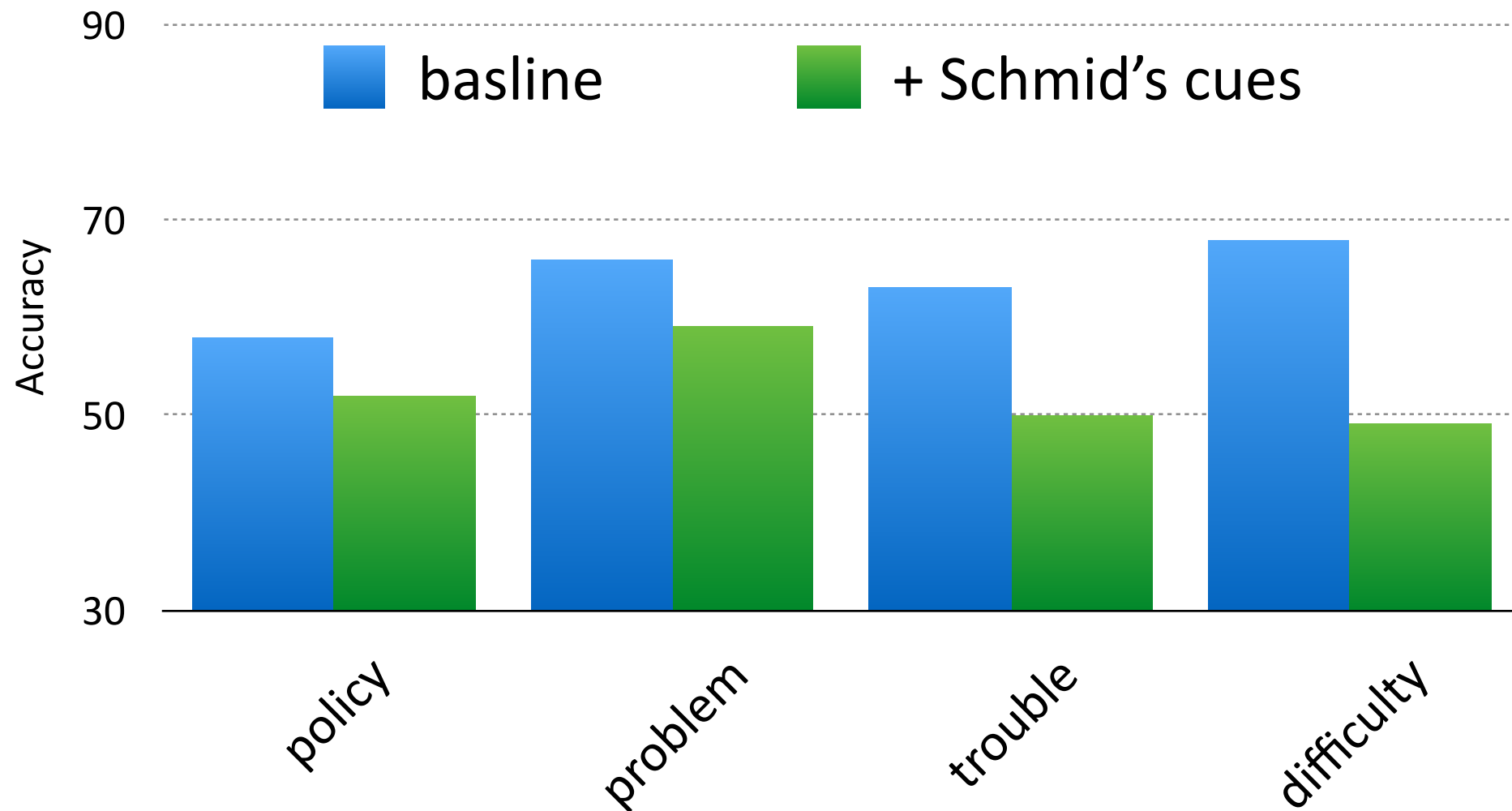
CSN resolution results

Schmid's framework particularly helps in resolving nouns with strict expectations



CSN resolution results

Schmid's cues were deleterious for more flexible nouns



Conclusion

- We proposed a general method to resolve shell nouns following cataphoric constructions
- Our results suggest that
 - Semantic knowledge from Schmid's families is helpful in resolving CSNs
 - Schmid's cues are defeasible
 - Possible to improve on Schmid's cues using crowdsourcing annotation and by exploiting lexico-syntactic patterns from a variety of corpora

Future directions

- Clustering shell nouns with similar semantic expectations similar to verb clustering
(Merlo and Stevenson, 2000; Schulte im Walde and Brew, 2002, inter alia)
- A robust machine learning approach that incorporates context and deeper semantics of the sentence, along with Schmid's cues

