https://presemo.helsinki.fi/nlp2020

LECTURE 14: PRAGMATICS AND FUTURE CHALLENGES IN NLP

Mark Granroth-Wilding

WHY NLP IS HARDER THAN IT SEEMS

- Real-life NLP presents more challenges than we've seen
- Methods here provide good starting point
- Today: examples of harder challenges
- Includes **unsolved problems**, ongoing research
- Sample of a few: there are plenty more!

Main take-away:

NLP is far from solved yet!

DIFFICULT PROBLEMS IN SEMANTICS

Saw a few in previous lecture:

- Time / events

  I arrived in NY *I will arrive in NY* *I arrive at 10*

- Aspect
  - Finished vs. ongoing
  - Point in time vs. interval
  - Causing change of state vs. not

    I need the cheapest fare *I drive a Mazda*

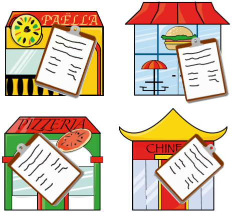
    He booked me a reservation *She found her gate*

QUANTIFIER SCOPING

Another difficult problem

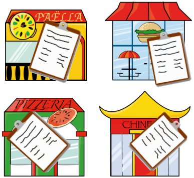
All restaurants have a menu

- Contains two quantifiers
- Leads to two readings: what?



QUANTIFIER SCOPING

All restaurants have a menu



- Can be distinguished by FOL representation

$$\forall x. Restaurant(x) \Rightarrow \exists y.(Menu(y) \wedge Has(x, y))$$

$$\exists y. Menu(y) \wedge \forall x.(Restaurant(x) \Rightarrow Has(x, y))$$

SCOPE AMBIGUITY

$$\forall x. Restaurant(x) \Rightarrow \exists y.(Menu(y) \wedge Has(x, y))$$

$$\exists y. Menu(y) \wedge \forall x.(Restaurant(x) \Rightarrow Has(x, y))$$

- No **syntactic** distinction
- Same *composition* must yield ambiguous readings
- Build into composition process:
  - generate multiple readings, OR
  - generate one reading that represents ambiguity

DIFFICULT SCOPE AMBIGUITIES

Combinations of quantifiers get *very* complicated very quickly!

- >2 quantifiers:

  Some representative of every company saw at least one sample

- Interaction with pronouns:

  The man who builds each clock also repairs it

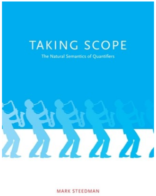
- '*Counting*' existentials:

  Every referee read at least three papers

QUANTIFIERS

- Complicated, language-specific rules
- Needs special formal mechanisms
- <u>But</u> humans do it with little difficulty
- Loads of theoretical work, still not solved!

Taking Scope: The Natural Semantics of Quantifiers.
Mark Steedman



PRAGMATICS

Real human language processing not like most NLP tasks

News article

Following a three-week trial at Bristol Crown Court, jurors told Judge Martin Picton they could not reach a verdict

Parse in isolation
Extract information

Phone call

A: Some of them disagree
A: I r ome
the o
B: Ex
B: bu
majority was
B: So I didn't know if that was just something for drama or that's truly the way it is
B: I always thought it had to be unanimous
A: I think it does have to be unanimous
B: So I I
B: But uh rather interesting

Individual utterances useless

Interpret structure of dialogue, speech acts, context, ...

PRAGMATICS

- **Pragmatics** concerns *meaning* in broader context
- Includes questions of e.g.
    - conversational **context**
    - speaker's **intent**
    - **metaphorical** meaning
    - **background knowledge**
- Abstract analysis
- Depends heavily on other NLP tasks we've seen
- <u>Many</u> unsolved problems
- Some tasks tackled in NLP: late in pipelines

DOABLE PRAGMATICS TASKS

- Certain pragmatics problems tackled in NLP
- Particularly in **discourse processing**
- Language use not isolated, unrelated sentences

Every restaurant has a menu

- **Structured**, **coherent** sequences

We sampled the local cuisine. Every restaurant has a menu and each one featured different specialities.

- Relatively doable subtasks:
  1. **Coreference resolution**
  2. **Discource structure**

COREFERENCE RESOLUTION

This month, a judge in California cleared thousands of criminal records with one stroke of his pen. He did it thanks to a new algorithm. The programmers behind it say: "we're just getting started solving America's urgent problems".

- Interpret sentences in **discourse**
- Must identify who's talked about
- Multiple references to same person, people, thing, ...
- Identifying **repeated** references to same entity:
  coreference resolution

COREFERENCE RESOLUTION

This month, a judge in California cleared thousands of criminal records with one stroke of his pen. He did it thanks to a new algorithm. The programmers behind it say: "we're just getting started solving America's urgent problems".

- These are **referring expressions**
- Refer to *judge in California* – **referent** (person)
- **Anaphora**: referring back to previously mentioned referent
- Often uses **pronouns**: *he*, *she*, . . .
- Some references license others:

  a judge in California is **antecedent** for *he*

COREFERENCE EXAMPLE

It belongs to Dennis Coffey, a 78-year-old who still rocks his wah-wah licks once a week. Dennis doesn't appear in Motown: The Musical. But he played on at least 100 million-selling records, and there was a year when he was on three of the Top 10 every month.

Pronouns most common form of reference

COREFERENCE EXAMPLE

Manchester City forward Raheem Sterling has been voted the 2019 Footballer of the Year by the Football Writers' Association. The England international topped the poll of the 400-strong FWA membership. He was more than 100 votes ahead of Virgil Van Dijk.

Not just pronouns: arbitrary phrases

NOT JUST ENTITIES

References not just to entities/individuals

According to Doug, Sue just bought a 1961 Ford Falcon.

1. *But that was a lie.*
2. *But that was false.*
3. *That struck me as a funny way to describe the situation.*
4. *That causes a financial problem for Sue.*

DIFFICULT CASES

What makes these hard?

The man who gave his paycheck to his wife was wiser than the man who gave it to his mistress.

The boy entered the room. The door closed automatically.

Georges Cipriani left a prison in Ensisheim on parole on Wednesday. He departed the prison in a police vehicle bound for an open prison near Strasbourg.

COREF SYSTEM EXAMPLE

Hugging Face

- **Classifier**-based approach
- Neural network
- Uses **word embeddings** instead of manually encoded features
- Designed to work with **informal dialogue**

COURSE FEEDBACK

- Almost finished course
- More to come still after break, but...
- This course is quite new
- We need your **feedback** for future versions!
    - Course structure
    - Content, lectures, slides, materials
    - Assignments...
- During (long) break: fill in course feedback form:
    - Follow link in email
    - Or log into **WebOodi**

DOABLE PRAGMATICS TASKS

- Certain pragmatics problems tackled in NLP
- Language use not isolated, unrelated sentences

Every restaurant has a menu

- **Structured**, **coherent** sequences

We sampled the local cuisine. Every restaurant has a menu and each one featured different specialities.

- Relatively doable subtasks:
  1. ~~Coreference resolution~~
  2. Discource structure

LARGER-SCALE STRUCTURE

- Most documents have high-level structure
    - Sections in *scientific papers*
    - *News articles*: summary at start, background later, ...
    - *Stories*: (mostly) temporal order
    - Different conventions
- Topics discussed by document
    - Tried to model with **topic modelling**
- Explicit sections, paragraphs, ...
- Relationships between sentences
- Seen in **NLG** for structuring output

DISCOURSE COHERENCE

- Longer discourse: **coherence** between sentences/utterances

  John hid Bill's car keys. He was drunk.

  * *John hid Bill's car keys. He likes spinach.*

- Different types of **relations**

  I went to the shop. I had run out of milk.

  I went to the shop. I was out when you called.

- Sometimes signalled explicitly

  I went to the shop, **because** *I had run out of milk.*

  I went to the shop, **so** *I was out when you called.*

DISCOURSE COHERENCE

I went to the shop. I had run out of milk.
I went to the shop. I was out when you called.

- Coherence relations often implicit
- Inference is important

Max fell. John pushed him.

Max fell **because** *John pushed him.*

Max fell **and then** *John pushed him.*

HIDDEN DISCOURSE STRUCTURE

Samuel Bronston vs. US trial[1]

a. Q: Do you have any bank accounts in Swiss banks, Mr. Bronston?
b. A: No, sir.
c. Q: Have you ever?
d. A: The company had an account there for about six months, in Zurich.

- Bronston used to have personal account in Swiss bank
- Answer interpreted as '*no*': indirect
- Did he lie?
- Literally truthful!
- Convicted for perjury, later overturned

[1] Solan & Tiersma (2005). *Speaking of Crime*

RHETORICAL RELATIONS

- Formalize coherence relations between phrases
- **Rhetorical relations**
- Set of types of relations

Max fell. John pushed him.

Max fell **because** *John pushed him.* — *Explanation*

Max fell **and then** *John pushed him.* — *Narration*

Cue phrases

Discourse/rhetorical relations

INTERPRETATION OF COHERENCE

- Assume discourse coherence
- Inference of relations affects interpretation

Kim's bike got a puncture. She phoned a mechanic.

Coherence assumption $\rightarrow$ *bike* $=$ *motorbike*

John likes Bill. He gave him an expensive Christmas present.

- **Explanation** $\rightarrow$ *he* $=$ *Bill*
- **Justification** $\rightarrow$ *he* $=$ *John*

SOME MORE RELATIONS

- **Result**

  Max tripped. He hit his head.

- **Parallel**

  Max tripped. John stumbled.

- **Elaboration**

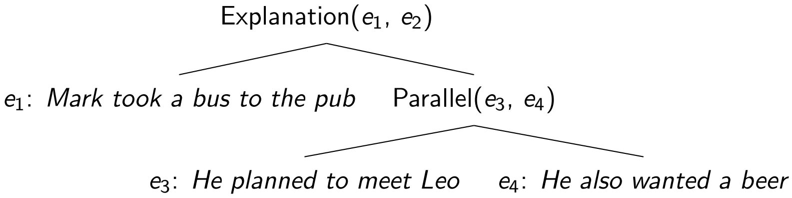
  Max fell. He tripped and took a terrible tumble.

- **Occasion**: state change, final state

  Max fell down. He lay on the floor.

NON-LINEAR STRUCTURE

- Discourse structure is non-linear



- Nodes: **discourse segments**
- Locally coherent clauses/sentences

CLUES TO INTERPRETATION

- **Cue phrases**: *because*, *so that*
- **Punctuation** and **text structure** ← Prosody in speech

  Max fell (John pushed him) and Kim laughed.
  Max fell, John pushed him and Kim laughed.

- Real-word **knowledge**

  Max fell. John pushed him as he lay on the ground.

- **Tense** and **aspect**

  Max fell. John had pushed him.
  Max was falling. John pushed him.

These can help, but interpretation remains a **hard problem**

MORE PRAGMATICS

News article

Following a three-week trial at Bristol Crown Court, jurors told Judge Martin Picton they could not reach a verdict.

Phone call

A: Some of them disagree
A: I mean some of them said one way and some the other
B: Exactly
B: but they took you know whatever the majority was
B: So I didn't know if that was just something for drama or that's truly the way it is
B: I always thought it had to be unanimous
A: I think it does have to be unanimous
B: So I I
B: But uh rather interesting

Isolated utterances useless

Discourse structure, pronoun resolution, coref resolution, speech acts, context, ...

What else do we need to do?

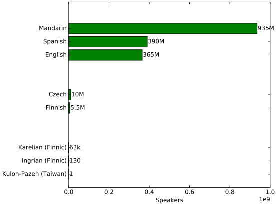
OTHER HARD TASKS

- Seen some **difficult problems** and (partial) solutions
  - Formal semantics: quantifier scope, negation
  - Coreference
  - Discourse structure
- Now a few more difficult NLP challenges
  - Open problems
  - Topics of current research
  - Some (partial) solutions...

MULTIPLE LANGUAGES

- This course: mostly English examples
- Until recently, most NLP in English only
- Can we do same things on other languages?
  - Depends on task
  - Depends on data
  - Particularly hard for **low-resource** languages

- Or, single method/model for *all* languages?
  - **Multilingual NLP**
  - Much more work in recent years



MULTIPLE LANGUAGES

- To some extent, same methods for other languages
- Linguistic differences easy to underestimate
- E.g. Chinese: what is a character, word, token?
- Still need data

High-resource	**Low-resource**
Few languages	Very many languages
Much linguistic study	Typically less studied
Large corpora	Small or no corpora
Annotated resources	Limited or no annotations
Commercial interest	Less commercial incentive

POSSIBLE SOLUTIONS

- **Annotate** data, **invest** in small languages
- **Unsupervised**, or minimally supervised, methods
  - Some useful methods: e.g. topic modelling
  - Realistic replacement for supervised in some areas
  - Often, *much* poorer quality
  - Hard to use output: what does it mean?
- **Language transfer**
  - Re-use what we know about *high-resource* language
  - Map language-specific parts, re-use others
  - E.g. POS tagger with word translations
  - Can work for **closely related** languages
- **Multilingual** methods:
  - Multilingual word embeddings...

BROAD-DOMAIN NLP

Syntactic parsing is *really* good on news text

- But I want to parse forum posts, conversational speech, Estonian, ...
- Answer: annotate a treebank for these
- Penn Treebank took years, huge effort and lots of money
- How to produce parsers that work on all these and more?
- Major area of research
- But there are some promising directions. . .

BROAD-DOMAIN PARSING

Some approaches:

- Unknown words: **backoff** to POS tags
  - Can make good guesses from context
  - Is POS tagger domain independent?
- **Domain transfer**
  - Build grammar/parser for one domain
  - Extend to others with *unlabelled data*
  - Can method discover new rules/constructions?
- **Transfer/expansion** on particularly domain-specific parts
  - E.g. (automatically) expand **lexicon**
  - Keep same rules

BROAD-DOMAIN PARSING

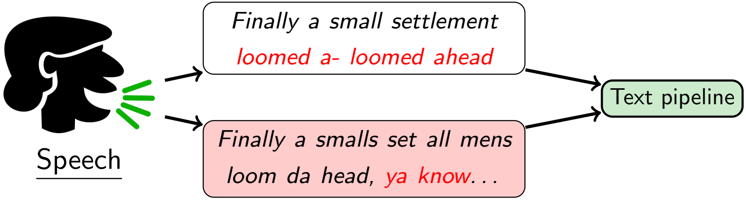
Some approaches:

- Train on **multiple domains**
  - Maybe learn **domain-general** characteristics
  - Better performance on new domain
  - Or **dynamically select** domain(s) when parsing
- Separate **domain-general** and **domain-specific** knowledge while learning
  - Reuse general parts
  - Adapt/ignore specific parts

Many solutions apply to same *domain-specificity* problem on other tasks

SPEECH

- Speech processing: **speech recognition** at start of pipeline
- Introduces further challenges:
    - More ambiguity
    - More noise
    - Disfluency
    - Wider variety of language (informal, dialectal)



CHALLENGES OF NLP ON SPEECH

- **Ambiguity**:
  - Many ways to interpret speech input
  - Choosing one-best may restrict later analysis
  - E.g. can't parse
  - Interaction between SR and later components
- **Language type** (register):
  - People don't speak like a newspaper
  - NLP tools must be able to handle **informal language**
  - Same issue with many text sources
- **Disfluency**
  - Real speech is not fluent: taster here

DISFLUENCY

- Fillers: *ya know*, *well*, *totaaa*
- Other non-lexical sounds: *hmm*, *um*, grunts, coughs
- **Repairs**:

the current plan is we take – okay let's start with the bananas

Alice saw the dog um the duck

All cause problems for standard tools. Proper handling essential to interpretation

SPEECH REPAIRS

- One type of disfluency
- Can be categorized:
  1. **Fresh start**
     the current plan is we take – okay let's start with the bananas
     → *let's start with the bananas*
  2. **Modification**
     A flight to um Berlin I mean Munich → *A flight to Munich*
  3. **Abridged repair**
     we need to um get the bananas → *we need to get the bananas*
- In some cases, identification permits cleaning

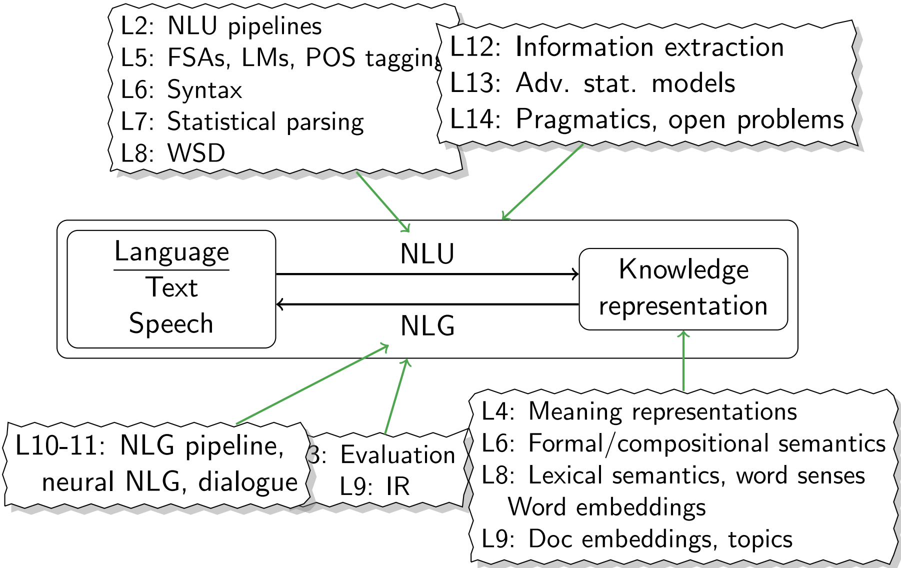
SPEECH REPAIRS

- Annotation efforts: *Switchboard corpus*
  - Part of PTB
  - Transcribed phone calls
  - Manually annotated with structures and types of disfluencies
- Train models to identify and clean up: difficult task
- Highly **language dependent** and only annotated for English
- **Transcribed speech** still much easier than SR output
- Some disfluencies can't be easily fixed:

  And the yeah the tour stays – you stay in the water but you do – we cruise out underneath the bridge

- *Repaired text* and *edit terms* **not meaningless**:
  lose meaning by cleaning up

NLP: WHAT HAVE WE SEEN?

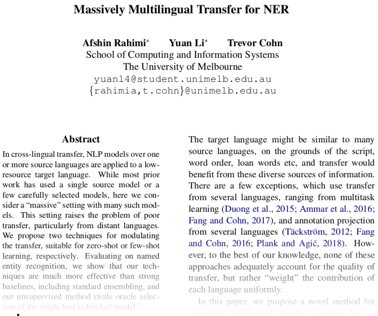


EXERCISE: NLP PAPERS

- Take a paper: ACL 2019 ← top NLP conference
- Randomly chosen from conference proceedings
- *Don't* read it all!
  - Abstract
  - Conclusion
  - Introduction?
  - Skim other sections?
- Look for things we've seen on course:
  - tasks, methods, models, metrics, . . .
- Look for ones we've not covered

EXERCISE: NLP PAPERS

My attempt


Massively Multilingual Transfer for NER

Afshin Rahimi* Yuan Li* Trevor Cohn
School of Computing and Information Systems
The University of Melbourne
yuanl4@student.unimelb.edu.au
{rahimia,t.cohn}@unimelb.edu.au

Abstract

In cross-lingual transfer, NLP models over one or more source languages are applied to a low-resource target language. While most prior work has used a single source model or a few carefully selected models, here we consider a "massive" setting with many such models. This setting raises the problem of poor transfer, particularly from distant languages. We propose two techniques for modulating the transfer, suitable for zero-shot or few-shot learning, respectively. Evaluating on named entity recognition, we show that our techniques are much more effective than strong baselines, including standard ensembling, and our unsupervised method rivals oracle selection of the single best individual model.[1]

The target language might be similar to many source languages, on the grounds of the script, word order, loan words etc, and transfer would benefit from these diverse sources of information. There are a few exceptions, which use transfer from several languages, ranging from multitask learning (Duong et al., 2015; Ammar et al., 2016; Fang and Cohn, 2017), and annotation projection from several languages (Täckström, 2012; Fang and Cohn, 2016; Plank and Agić, 2018). However, to the best of our knowledge, none of these approaches adequately account for the quality of transfer, but rather "weight" the contribution of each language uniformly.

In this paper, we propose a novel method for

<u>Things we've seen</u>

- NER
- Multilingual NLP
- Language transfer
- Baselines
- Corpora
- Error analysis
- Generative probabilistic model
- Confusion matrix
- Tokens
- Word embeddings
- F-score
- ...

<u>Things we've not seen</u>

- Zero-shot learning
- Truth inference
- Probabilistic graphical models
- ELBO
- Variational inference
- ...

EXERCISE: NLP PAPERS

- Read: abstract, conclusion, introduction?, methods?
  $\sim$5-10 mins
- Look for things we've seen on course:
  - tasks, methods, models, metrics, ...
- Look for ones we've not covered
- Post/upvote on Presemo
- 10 mins

https://presemo.helsinki.fi/nlp2020

NLP TODAY

- Hopefully this course has convinced you that:
  - NLP is **not a hopeless task**
  - existing work has a lot of **useful applications**
- Hopefully today has convinced you that:
  - NLP is a **very hard task**
  - there are *tonnes* of **unsolved problems**
- We can already do a lot with it
  - Recent years: much more useful for real-word applications

Now is a good time to be an NLP expert!

And now you are one...

FINAL PROJECT

- No exam!
- Task: extend one of assignments
  - Small extension – improvement, application, ...
  - Suggestions provided with assignments
  - Use techniques from course: link
- Submit code and short report (2-3 pages)
  - What did you do? Why?
  - Did it work? How do you know?
  - What would you do next?
- Due 1 week from today: **6.3**

FINAL PROJECT

From instructions. . .

Link

The main criterion is that you **display an understanding** of:

- the **theoretical content relevant** to the particular task, as given in lectures;
- how this can be **applied** in practice; and
- what its **limitations** might be.

We will **not** be grading the **success of your system**.

READING MATERIAL

- Formal semantics for NLP
  - *J&M3 ch 14*
  - *Representation and Inference for Natural Language*

    (Blackburn & Bos, 2005)
- Scope, negation, . . . :
  Taking Scope: The Natural Semantics of Quantifiers

  (Steedman)
- Coreference resolution: *J&M3, chap 22*
- Discourse coherence: *J&M3, chap 23*
- Speech: *J&M**2**, Part II* *Not yet in J&M3*