Our special friend, language

Seminar week #1: Non-information-density approaches to online processing, Winter 2015/2016

Asad Sayeed

Uni-Saarland

Today's agenda

- Introduction to the "concept"/purpose of this seminar.
- Getting to know each other.
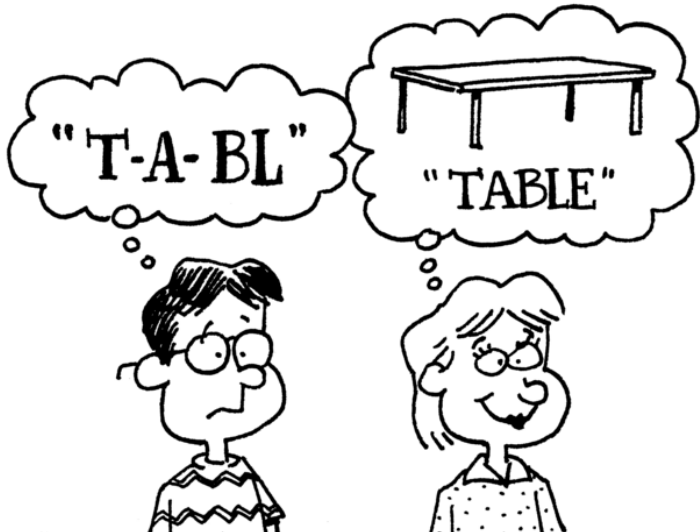- Preliminary organizational details.

So...

I have some questions for the audience. (Uh oh)

What is language?

No, seriously, what is language?

Actually, I don't have a serious answer.

But Wikipedia does!

Actually, I don't have a serious answer.

But Wikipedia does!

Wikipedia on "Language"

Language is the ability to acquire and use complex systems of communication, particularly the human ability to do so, and a language is any specific example of such a system.

Actually, I don't have a serious answer.

But Wikipedia does!

Wikipedia on "Language"

Language is the ability to acquire and use complex systems of communication, particularly the human ability to do so, and a language is any specific example of such a system.

Really?

The real question I want to ask is. . .

. . . what makes language "linguistic"?

Seriously, there are other ways to communicate complex thoughts.

But language has particular characteristics that go beyond communication.

Hey, more Wikipedia, why not?

- "Mental faculty, organ, or instinct."
- "Formal symbolic system."
- "Tool for communication."

So where in these does the "linguistic" part of language exist?

Theoretical linguistics focused on the "mental organ" part.

Generative linguistics: Chomsky and beyond.

- Divide language into (roughly speaking):
  - "Competence": the abstract knowledge of linguistic structure and the processes required to assemble it.
  - "Performance": the articulatory and perceptual vicissitudes involved in "producing" and "consuming" language.

But these are implicitly held to be specific to the human organism.

Theoretical linguistics focused on the "mental organ" part.

Generative linguistics: Chomsky and beyond.

- Divide language into (roughly speaking):
  - "Competence": the abstract knowledge of linguistic structure and the processes required to assemble it.
  - "Performance": the articulatory and perceptual vicissitudes involved in "producing" and "consuming" language.

But these are implicitly held to be specific to the human organism. BUT WAIT:

- Computation/logic/formalism: Constrains the representation of. . . competence?

This requires an implicit methodology of research.

Performance isn't the point: it's the grimy mirror through which we see the mental organ of language.

- We have no "direct" access to competence, so there is a "necessary evil" of making use of what we have access to.
- Formalism: allows us to write down generalizations.

You should not be upset by your dirty mirror. If it wasn't for the mirror, you would not be aware of the dust.

- Sheng-Yen

www.facebook.com/ShortZenPoetry

So what about online processing?

What sort of relationship can we have between a "competence" theory and the observations of linguistic behaviour?

- We need a "linking theory".
- For a competence theory, the real question:
  - Words arrive in a sequence.
  - But the rules that define possible utterances (via formal representations) are not necessarily sequential.
  - Can we make a minimal inference between these facts?

Yes: exploit the derivation.

1960s and 1970s: the Derivational Theory of Complexity (DTC).

- Reliant on now-outdated notions of "deep structure" (DS) and "surface structure." (SS)
- Difficulty in processing $\approx$ derivational distance between DS and SS.

Yes: exploit the derivation.

1960s and 1970s: the Derivational Theory of Complexity (DTC).

- Reliant on now-outdated notions of "deep structure" (DS) and "surface structure." (SS)
- Difficulty in processing $\approx$ derivational distance between DS and SS.
- Alleged to have failed.
  - Some derivations at the time didn't correlate well to processing measures.
  - (We'll touch on this in the next session.)

(Just to give a flavour of the DTC.)

A thing that probably worked: passive constructions.
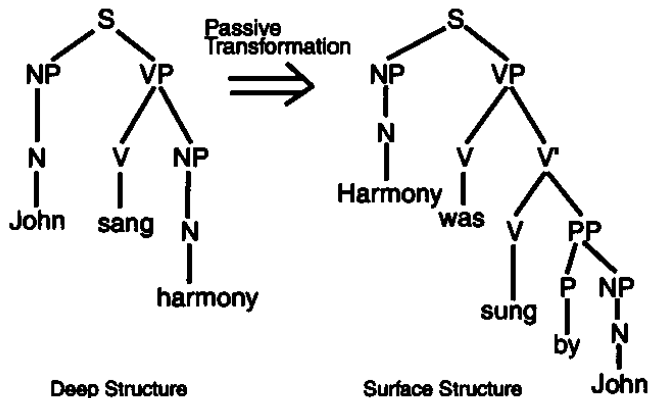
But performance is where communication happens!

People *use* language for a *purpose* beyond their abstract knowledge of language.

- Are these purposes part of the "grimy mirror"?
- Do they belong to some other area of cognition?
- How far must we extract away from them, before we can dig down to "the linguistic", narrowly construed?

But performance is where communication happens!

People *use* language for a *purpose* beyond their abstract knowledge of language.

- Are these purposes part of the "grimy mirror"?
- Do they belong to some other area of cognition?
- How far must we extract away from them, before we can dig down to "the linguistic", narrowly construed?

The world-related content of an utterance: a "black box" on which the "language organ" imposes structure."

But performance is where communication happens!

People *use* language for a *purpose* beyond their abstract knowledge of language.

- Are these purposes part of the "grimy mirror"?
- Do they belong to some other area of cognition?
- How far must we extract away from them, before we can dig down to "the linguistic", narrowly construed?

The world-related content of an utterance: a "black box" on which the "language organ" imposes structure."
Is that so?

The implication/presupposition: language structure is not (fully? mostly?) conditioned on use!

But not everyone is happy with this.

What would the world look like...

- ...if "folk" beliefs about language-as-communication were true?
- ...if "world-relevant" content of the language were what language is *for*?
- ...if the supposed structural details of "competence" were actually artifacts of function-driven behaviour?

Who would be happy if that were true?

Linguistics also contains a variety of functional approaches:

- Examples include the Prague School, Systemic Functional Grammar, etc.
- Often driven by a demand to analyze language in sociological or literary terms.
- Varying degrees of formalism – not really concerned with The Steps To Generate An Utterance.

. . . but ALL focused on the idea of language as a communicative tool, and any putative "mental organ" as itself a servant of that tool.

But if structure is not **the overriding** theme. . .

. . . for traditional functionalist approaches, then how do we account for observed structure.

- Functionalist approaches account for the presence of "linguistic functions".
- Focus on how structures are *used*, not how they are *learnable* or even *possible*.
- (Obligatory mention of "poverty of the stimulus", etc.)

So we're *still* left with the question of where structure comes from.

...EXCEPT ...

People have come up with theories that DO cover this.

The advent of statistical NLP changed things.

- Can now come up with large-scale statistical models of linguistic behaviour.
- Can now associate linguistic structures with these statistics.
- These statistics can be used to represent "information" in the Shannon sense.

People have come up with theories that DO cover this.
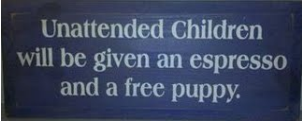
The advent of statistical NLP changed things.

- Can now come up with large-scale statistical models of linguistic behaviour.
- Can now associate linguistic structures with these statistics.
- These statistics can be used to represent "information" in the Shannon sense.

And so a new approach was born.

So how do we use statistics...

...to come up with a theory of function?
A popular answer these days: **surprisal**: a measure of (un)predictability.

Surprisal (Hale 2001)
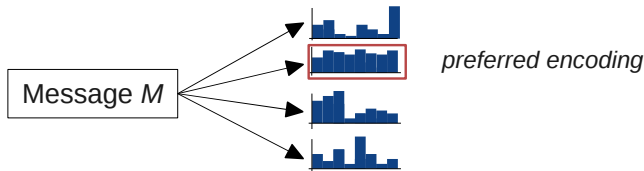
$$-\log P(w_i|w_{1\ldots i-1})$$

---

w words, can generalize to other linguistic events.

Can surprisal explain linguistic structure?

According to some, it can!
Uniform information density (UID) hypothesis:

- Speakers tend to maintain a constant rate of information transfer.

preferred encoding

Message *M*

We can measure information transfer as surprisal.

Then we can condition the fine-grained structure of language on attempting to maintain a "flat" surprisal!
But does it work?

We can measure information transfer as surprisal.

Then we can condition the fine-grained structure of language on attempting to maintain a "flat" surprisal!
But does it work?
Some "low-level" results:

- n-gram surprisal affects lexical/syntactic choices:
    - UID account can explain use of reduced forms in English.
      Levy and Jaeger (2007), Frank and Jaeger (2008) (e.g. "cannot" vs. "can't")

We can measure information transfer as surprisal.

Then we can condition the fine-grained structure of language on attempting to maintain a "flat" surprisal!
But does it work?
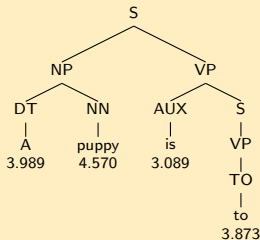Some "low-level" results:

- n-gram surprisal affects lexical/syntactic choices:
  - UID account can explain use of reduced forms in English.
    Levy and Jaeger (2007), Frank and Jaeger (2008) (e.g. "cannot" vs. "can't")
- n-gram surprisal vs. orthographic length:
  - surprisal from Google trigrams predicts word length.
    Piantadosi et al. (2011)

But it works at even higher levels.

Syntactic surprisal example, from Roark parser (Roark 2001, Roark et al. 2009):

Top-ranked partial parse.

Sentence: *A puppy is to ‖ a dog what a kitten is to a cat.*

```
                              S
                  ┌───────────┴───────────┐
                 NP                        VP
            ┌─────┴─────┐            ┌──────┴──────┐
           DT          NN          AUX             S
            │           │           │              │
            A         puppy         is             VP
          3.989       4.570        3.089            │
                                                    TO
                                                    │
                                                    to
                                                  3.873
```
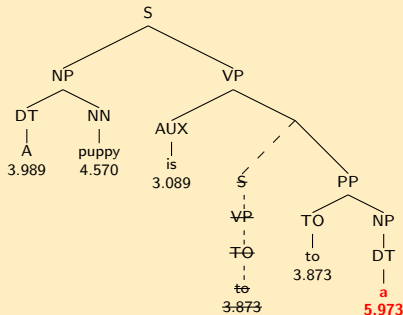
But it works at even higher levels.

Syntactic surprisal example, from Roark parser (Roark 2001, Roark et al. 2009):

Top-ranked partial parse.

Sentence: *A puppy is to a ‖ dog what a kitten is to a cat.*

But it works at even higher levels.

- Demberg, Sayeed, Gorinski, Engonopoulos [2012]: Roark syntactic surprisal correlated with word pronunciation duration!
- Speakers are conditioning their fine-grained linguistic structure on a rational attempt at managing communication.

But it works at even higher levels.

- Demberg, Sayeed, Gorinski, Engonopoulos [2012]: Roark syntactic surprisal correlated with word pronunciation duration!
- Speakers are conditioning their fine-grained linguistic structure on a rational attempt at managing communication.

How far can we go with this?

- Encroaches on DTC territory – with information density!
- We even recently managed semantics (crudely: Sayeed, Fischer, and Demberg 2015).

It's a matter of ambition.

I divide UID approaches into two categories (my names for them):

- **Strong UID hypothesis**
  - Information density governs most if not all linguistic structure.
  - Radical focus on communicative efficiency: language is structured around ensuring optimal communication, everything is performance.
- **Weak UID hypothesis** ($\Leftarrow$ my preference, obviously)
  - Information density governs part of linguistic performance.
  - Communicative efficiency doesn't directly touch on structural competence.

It's a matter of ambition.

I divide UID approaches into two categories (my names for them):

- **Strong UID hypothesis**
  - Information density governs most if not all linguistic structure.
  - Radical focus on communicative efficiency: language is structured around ensuring optimal communication, everything is performance.
- **Weak UID hypothesis** ($\Leftarrow$ my preference, obviously)
  - Information density governs part of linguistic performance.
  - Communicative efficiency doesn't directly touch on structural competence.

(Obviously these are strawmen/extremes.)

We immediately have a research question.

How much of linguistic behaviour is "left over" for explanation in structural terms?

We immediately have a research question.

How much of linguistic behaviour is "left over" for explanation in structural terms?

Because there's lots of reasons to think UID is not the last word!

- Statistical approaches that are **not** yet fully connected to information theory – e.g. distributional semantics.
- Structural approaches – still not dead (!), often related to memory-limit approaches.

And I would like to remind everyone, myself included, about what's there.

And I would like to remind everyone, myself included, about what's there.
—particularly given SFB1102, which is about information density.

Goals of the seminar

- To gain a broad understanding of the aspects of sentence processing not focused on communicative function — particularly not direct information-transfer concerns.
- To explore recent work in theory, experiment, computation focused on linguistic structure and mental organization.
- To remind the local community of on-going activity in non-ID research.
- To give students further practice in reading, presenting, and understanding a diverse range of research papers.

Ideally: a discussion group atmosphere.

Who should take this seminar?

I want to avoid this seminar being too specialized.

- Sentence processing can be dealt with a very technical way, want to minimize the requirement to know a lot of formalism.
- Unfortunately, technical details are unavoidable, I'm hoping this is all we'll need:
    - Basic introductory understanding of syntax and semantics.
    - A little exposure to psycholinguistics/psych experimental methodology.
- Presenters: should assess their audience well.

In this seminar, I want to focus on BREADTH, not DEPTH.

And the details. . .

- Meeting: Wednesday 12-2 (ct) – can we change this to later on Wednesday?
- First meeting: Today.
- Location: C7.2 2.11.
- And there may be the occasional doge.
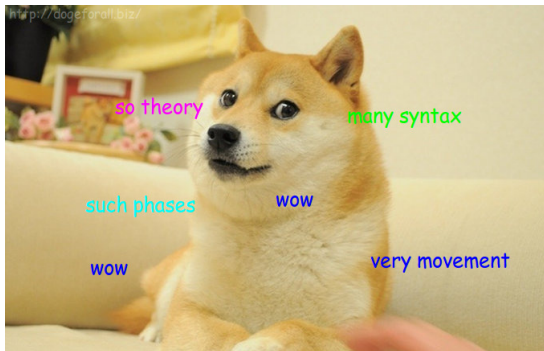
And yet more details...

- Please read the web page:
  http://www.coli.uni-saarland.de/courses/nondensity15/
- **Must** sign up for the mailing list (on the web page).
- The schedule will unfold "dynamically" – I will list readings on the mailing list as well as on the web site.
- I have listed some suggested papers for presentation on the web site with links, but the list is not exhaustive!

"Getting to know you" pause.

Yes, we have requirements.

I'm not 100% sure how all of this stuff works, but I know:

- For any credit at all, all of you HAVE to do a presentation.
    - And I expect most of you to attend the presentations, most of the time.
- Some of you will want/need additional credit: write a term paper, requirements to be discussed based on how challenging your topic is.
- Yes, I do oral exams for those who need the credit for that.

How will presentations be evaluated?

I want to formalize this a little bit:

- Demonstrated understanding of the topic: 40%.
- Ability to impart the material to other students: 30%.
- Quality of presentation materials: 15%.
- Discussion leadership/participation: 15%.

The "quality" part is a bit dangerous: see later.

What can you present/write a term paper about?

Possible (non-exhaustive) types of topics:

- A specific **paper** in the literature, including but not limited to what I put up on the web page. (**NB**: some of the longer papers can be presented by two people.
- A survey of opinions on a particular phenomenon, theory, controversy, etc.
- A survey of a particular (claimed) **principle**.
- Your own original idea.
- ...

Length of presentation: depends on participation.

How should you present it?

Possible presentation style:

- Good old slideshow – safest, I suppose.
- Detailed handout (traditional for syntax papers in particular).
- Even a well-presented whiteboard discussion is OK.

You will also *lead discussion* of the idea/paper/etc.

For evaluation: super-duper use of technology is nice but won't necessarily get you a better grade.

The (rough/tentative) schedule

- Week 1 (today): this intro stuff.
- Week 2: I will talk about recent controversies in the context of a paper by Philips and Lewis (2012).
- Remaining weeks: student presentations.

The Wednesday right after New Year's may have to be rescheduled.

Some things to consider

We are only trying to skim the highest levels: breadth more than depth!

- I don't expect you to understand deeply everything you read for the course.
- Theory changes over time and the "intro material" can become outdated, other than the basic axioms of logic.
- I want you to instead learn to "appreciate" the material at an abstract level and be able to pursue the "useful" threads.

USEFUL?!?!!

THIS IS MAH JOB