Semantic representation and world knowledge

Asad Sayeed

Saarland University

Overview of today's talk

1. a bird's-eye view of research programme
2. on-going work in distributional semantics
3. experimental work in scope ambiguity resolution

Part 1: overall agenda and background context

What are my main research interests?

My research programme has been focused on integrating the following areas in language science:

What are my main research interests?

My research programme has been focused on integrating the following areas in language science:

- **Linguistic theory and formalisms**: the foundations of grammar and the syntax/semantic interface.

What are my main research interests?

My research programme has been focused on integrating the following areas in language science:

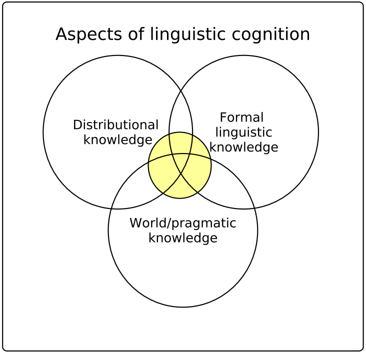
- **Linguistic theory and formalisms**: the foundations of grammar and the syntax/semantic interface.
- **Natural language processing**: representing and automating linguistic processes and applying them to problems of interaction.

What are my main research interests?

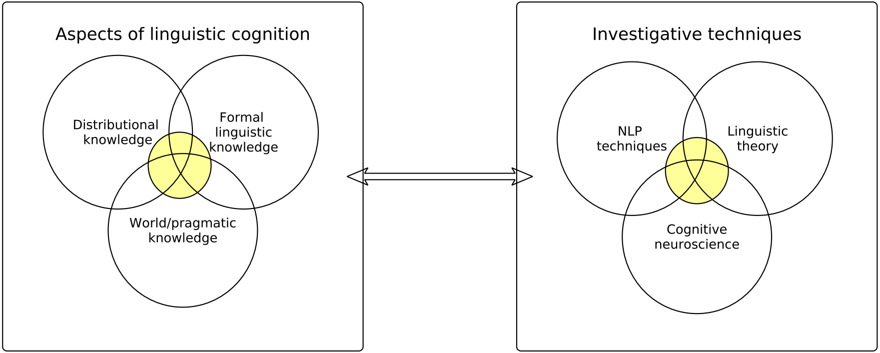
My research programme has been focused on integrating the following areas in language science:

- **Linguistic theory and formalisms**: the foundations of grammar and the syntax/semantic interface.
- **Natural language processing**: representing and automating linguistic processes and applying them to problems of interaction.
- **Psycholinguistics and cognitive modeling**: Doing the above in a way that reveals scientific insights about linguistic cognition.

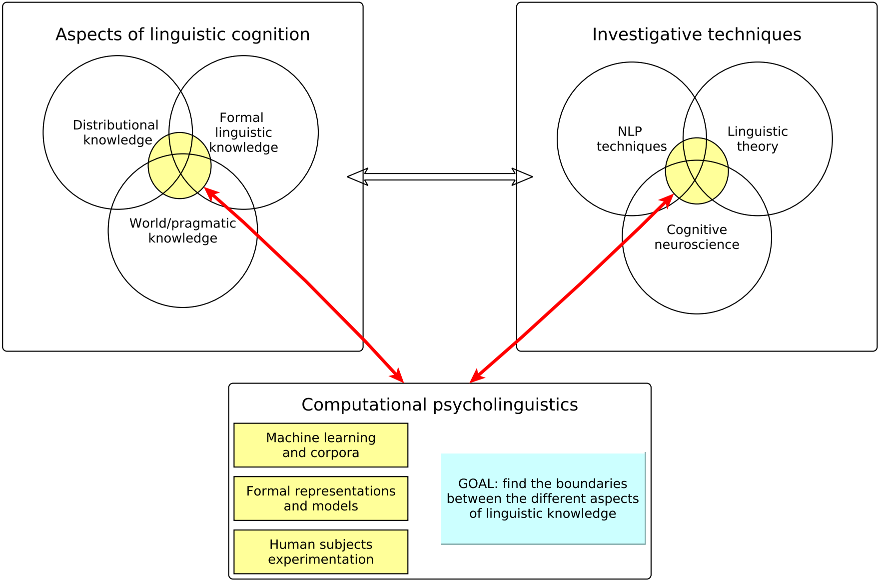
Overall research vision



Overall research vision



Overall research vision



Larger societal context: adaptiveness

Dialog systems are an increasing part of daily life.

Larger societal context: adaptiveness

Dialog systems are an increasing part of daily life. Consider Siri, Amazon Alexa, etc. Explicitly intended to handle general language.

Larger societal context: adaptiveness

- Incrementally responsive recognition systems: respond faster to user requests. [e.g. Schlangen et al., 2010]

Larger societal context: adaptiveness

- Incrementally responsive recognition systems: respond faster to user requests. [e.g. Schlangen et al., 2010]
  $\Rightarrow$ Should already be preparing queries/responses during user input.

Larger societal context: adaptiveness

- Incrementally responsive recognition systems: respond faster to user requests. [e.g. Schlangen et al., 2010]
  ⇒ Should already be preparing queries/responses during user input.
- Consider the following partial query:

Talk to your in-car GPS.

The route blocked. . .

Routes don't normally block!

Larger societal context: adaptiveness

- Incrementally responsive recognition systems: respond faster to user requests. [e.g. Schlangen et al., 2010]
  ⇒ Should already be preparing queries/responses during user input.
- Consider the following partial query:

Talk to your in-car GPS.

The route blocked this

Routes don't normally block!

Larger societal context: adaptiveness

- Incrementally responsive recognition systems: respond faster to user requests. [e.g. Schlangen et al., 2010]
  $\Rightarrow$ Should already be preparing queries/responses during user input.
- Consider the following partial query:

Talk to your in-car GPS.

The route blocked this morning by the accident, I want to avoid.

Routes don't normally block!

- But "route" was not the subject of the sentence.

Larger societal context: adaptiveness

Ambiguity matters to adaptiveness.

Larger societal context: adaptiveness

Ambiguity matters to adaptiveness.

Topicalization

The route blocked this morning by the accident, I want to avoid.

Recognize that "route" is not a plausible agent of "block"
$\Rightarrow$ recognize that "block" is not a main verb.

Larger societal context: adaptiveness

Ambiguity matters to adaptiveness.

Topicalization

The route blocked this morning by the accident, I want to avoid.

Recognize that "route" is not a plausible agent of "block"
$\Rightarrow$ recognize that "block" is not a main verb.

Syntax and semantics: mutual disambiguation.

Larger societal context: usability

Scope ambiguity with personal assistant.

Send every restaurant a reservation request.

Is there one reservation sent to all the restaurants (i.e. for single mass event)

or

does each restaurant receive a separate reservation request (as alternates)?

Larger societal context: usability

Scope ambiguity with personal assistant.

Send every restaurant a reservation request.

Is there one reservation sent to all the restaurants (i.e. for single mass event)

or

does each restaurant receive a separate reservation request (as alternates)? Normally the latter – our "common sense" tells us.

Larger societal context: usability

Scope ambiguity with personal assistant.

Send every restaurant a reservation request.

Is there one reservation sent to all the restaurants (i.e. for single mass event)

or

does each restaurant receive a separate reservation request (as alternates)? Normally the latter – our "common sense" tells us.

⇒ **The system should start preparing the right request as early as possible.**

Application context feeds back into cognition

Coercion

The author was starting the book in the house.
(Reading? Writing?)

Application context feeds back into cognition

Coercion

The author was starting the book in the house.
(Reading? Writing?)

Attachment ambiguity vs. semantic mismatch

Bob cut the cake with a flower.
(The flower was made of cake icing?)

Application context feeds back into cognition

Coercion

The author was starting the book in the house.
(Reading? Writing?)

Attachment ambiguity vs. semantic mismatch

Bob cut the cake with a flower.
(The flower was made of cake icing?)

Syntax/semantics mismatches interact with reading time [McElree et al. 2001; Zarcone et al. 2014]

Psycholinguistic effects of scope ambiguity

"Higher-order" semantics: scope ambiguity.

Two quantifiers

- "Every child climbed a tree."
- Two interpretations:
  - "For every child, there is a tree that that child climbed." (linear scope: $\forall > \exists$)
  - "There is a tree such that for every child, that child climbed the tree." (inverse scope: $\exists > \forall$)

Psycholinguistic effects of scope ambiguity

"Higher-order" semantics: scope ambiguity.

Two quantifiers

- "Every child climbed a tree."
- Two interpretations:
  - "For every child, there is a tree that that child climbed." (linear scope: $\forall > \exists$)
  - "There is a tree such that for every child, that child climbed the tree." (inverse scope: $\exists > \forall$)

People **very** strongly prefer the linear reading. [Dwivedi 2013]

Psycholinguistic effects of scope ambiguity

Inverse scope possible.

Psycholinguistic effects of scope ambiguity

Inverse scope possible.

From Dwivedi [2013]

- "Every jeweller appraised a diamond."
- Both readings available, but inverse ($\exists > \forall$) more often preferred than with "child/tree".

Psycholinguistic effects of scope ambiguity

Inverse scope possible.

From Dwivedi [2013]

- "Every jeweller appraised a diamond."
- Both readings available, but inverse ($\exists > \forall$) more often preferred than with "child/tree".

Lexical-pragmatic bias $\Rightarrow$ What is the interaction between distributional knowledge and world-knowledge?

Contributions and research questions

Contributions discussed in this talk:

Contributions and research questions

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).

Contributions and research questions

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.

Contributions and research questions

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.

Leading to the following research questions:

Contributions and research questions

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.

Leading to the following research questions:

- How far can distributional approaches take us in representing selectional preference, given world knowledge?

Contributions and research questions

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.

Leading to the following research questions:

- How far can distributional approaches take us in representing selectional preference, given world knowledge?
- In what way does world knowledge interact with distributional knowledge and "higher-order" formal structure?

Part 2: distributional semantics and thematic fit

Modeling incremental linguistic reasoning. . .

. . . needs an incremental semantic interpretation.

Modeling incremental linguistic reasoning...

...needs an incremental semantic interpretation.

- Sayeed and Demberg [2013]: Neo-Davidsonian incremental semantics.
- A formal way of saying: commit to only what we know about roles and role-participants.

Modeling incremental linguistic reasoning. . .

. . . needs an incremental semantic interpretation.

- Sayeed and Demberg [2013]: Neo-Davidsonian incremental semantics.
- A formal way of saying: commit to only what we know about roles and role-participants.

"The route blocked. . . "

- $\exists x \text{route}(x) \& \exists e \text{block}(e) \& \text{ROLE?}(e, x)$
- "There exists a route and there exists an event of blocking such that the route is the ROLE? of blocking."

Syntax insufficient by itself

Probabilistic parsing: highest ranked parse will always make “route” the subject/agent.

Syntax insufficient by itself

Probabilistic parsing: highest ranked parse will always make "route" the subject/agent.

- $\exists x \text{route}(x) \& \exists e \text{block}(e) \& \text{agent}(e, x)$
- "There exists a route and there exists an event of blocking such that the route is the agent of blocking."

Syntax insufficient by itself

Probabilistic parsing: highest ranked parse will always make “route” the subject/agent.

- $\exists x \mathrm{route}(x) \& \exists e \mathrm{block}(e) \& \mathrm{agent}(e, x)$
- “There exists a route and there exists an event of blocking such that the route is the agent of blocking.”

Need to detect that lower-probability structure may exist. . .

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

- By a catalogue of entity/event features.

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

- By a catalogue of entity/event features.
  ⇒ not robust, incompatible with statistical approaches.

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

- By a catalogue of entity/event features.
  $\Rightarrow$ not robust, incompatible with statistical approaches.
- **Distributional hypothesis**: the meaning of a word is its co-occurrence with other words.

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

- By a catalogue of entity/event features.
  $\Rightarrow$ not robust, incompatible with statistical approaches.
- **Distributional hypothesis**: the meaning of a word is its co-occurrence with other words.
  - Quantifiable, potentially robust to data.

Needed: concept of plausibility

How do we measure plausibility of an entity in a role in an event?

- By a catalogue of entity/event features.
  $\Rightarrow$ not robust, incompatible with statistical approaches.
- **Distributional hypothesis**: the meaning of a word is its co-occurrence with other words.
  - Quantifiable, potentially robust to data.
  - **Great! So how do we leverage role distributions?**

Evaluating role-filling

The thematic fit problem

Given a verb/event-type v, an entity x, how well does v fit x in role r?

- Slightly different from *selectional preference*.
- Need to get ratings. Possible questions:
  - "How common is it for a cake to bake something?"
  - "An oven is something you can use for baking."

  Rate from 1-7.

Existing evaluation data

Sample 1-7 ratings from Padó [2007]:

Verb	**Noun**	**Semantic role**	**Score**
advise	doctor	subj	6.8
advise	doctor	obj	4.0
confuse	baby	subj	3.7
confuse	baby	obj	6.0
eat	lunch	subj	1.1
eat	lunch	obj	6.9
kill	lion	subj	2.7
kill	lion	obj	4.9

We evaluate thematic fit models with Spearman's ρ (pos/neg percentage).
Widely used, e.g. Erk et al. (2010).

Model construction

A direct corpus-based approach: distributional memory (DM) [Baroni and Lenci, 2010]:

Model construction

A direct corpus-based approach: distributional memory (DM) [Baroni and Lenci, 2010]:

- Construct third-order tensor (three-way table) representing semantic dependencies.

Model construction

A direct corpus-based approach: distributional memory (DM) [Baroni and Lenci, 2010]:

- Construct third-order tensor (three-way table) representing semantic dependencies.
- Statistically reweighted (local mutual information) counts of occurrences of $< w_1, l, w_2 >$, where w_1, w_2 words, l the dependency between them.

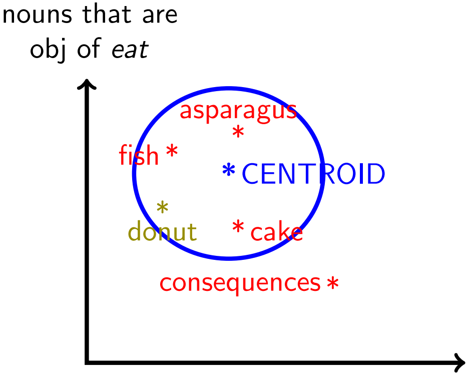
Model construction

A direct corpus-based approach: distributional memory (DM) [Baroni and Lenci, 2010]:

- Construct third-order tensor (three-way table) representing semantic dependencies.
- Statistically reweighted (local mutual information) counts of occurrences of $< w_1, l, w_2 >$, where w_1, w_2 words, l the dependency between them.
- What we can get from this: sparse word vectors whose dimensions represent semantic connections to other words.
- Corpus: UKWAC, BNC (about 2bn words).

Thematic fit measurement

Query: how good is "donut" as an object of "eat"?

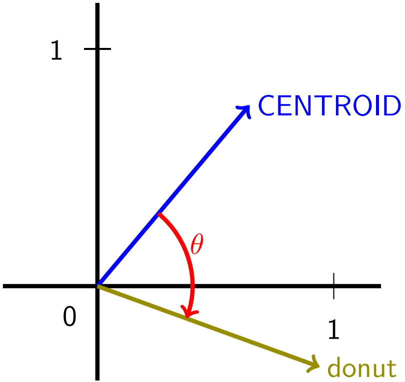


(Special thanks to A. Zarcone.)

Find a centroid based on 20 nouns that are typical "eat"-objects.

Thematic fit measurement

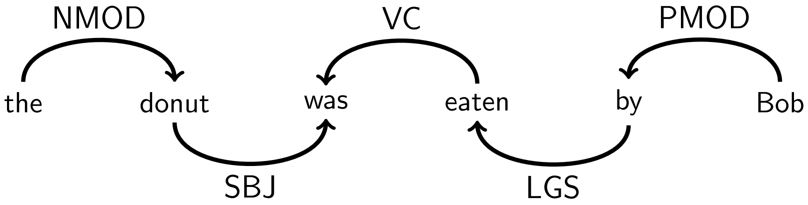
Query: how good is “donut” as an object of “eat”?



Then take the cosine of “donut” with the centroid.

Feature source

Baroni and Lenci's TypeDM model: "semantic" features hand-crafted from syntactic dependencies.



Feature source

My challenge: can we get a better-fitting model by going directly to semantically-labelled features?

the donut | was | eaten | by Homer
PATIENT | | VERB | AGENT

Feature source

My challenge: can we get a better-fitting model by going directly to semantically-labelled features?

the donut was eaten by Homer

PATIENT VERB AGENT

Does it help?

Sample of experiments

- . . . the answer is **yes** [Sayeed and Demberg, 2014; Sayeed, Demberg, and Shkadzko, 2015]:
    - Best model on agent/patient data: combine syntax-derived [Baroni and Lenci, 2010] features with SRL-based features.
        - Our result: $\rho = 59$
        - Syntax-only baseline: $\rho = 53$.

Sample of experiments

- . . . the answer is **yes** [Sayeed and Demberg, 2014; Sayeed, Demberg, and Shkadzko, 2015]:
  - Best model on agent/patient data: combine syntax-derived [Baroni and Lenci, 2010] features with SRL-based features.
    - Our result: $\rho = 59$
    - Syntax-only baseline: $\rho = 53$.
  - Best model on instrument role data [Ferretti et. al, 2001]:
    - Our result: $\rho = 45$
    - Syntax-only baseline: $\rho = 36$

Sample of experiments

- ... the answer is **yes** [Sayeed and Demberg, 2014; Sayeed, Demberg, and Shkadzko, 2015]:
  - Best model on agent/patient data: combine syntax-derived [Baroni and Lenci, 2010] features with SRL-based features.
    - Our result: $\rho = 59$
    - Syntax-only baseline: $\rho = 53$.
  - Best model on instrument role data [Ferretti et. al, 2001]:
    - Our result: $\rho = 45$
    - Syntax-only baseline: $\rho = 36$

Other best unsupervised results in Greenberg, Sayeed, and Demberg. [2015], Sayeed, Demberg, and Shkadzko [2015] *inter alia*.

Larger sentence contexts

Can rate verbs with single arguments with strong correlations with human judgements (e.g. "doctor advises").

Larger sentence contexts

Can rate verbs with single arguments with strong correlations with human judgements (e.g. "doctor advises").

- Can we do more than two? (i.e. real sentences)

Larger sentence contexts

Can rate verbs with single arguments with strong correlations with human judgements (e.g. "doctor advises").

- Can we do more than two? (i.e. real sentences)
- Vector spaces: multiplication/intersection operations necessary for composition.

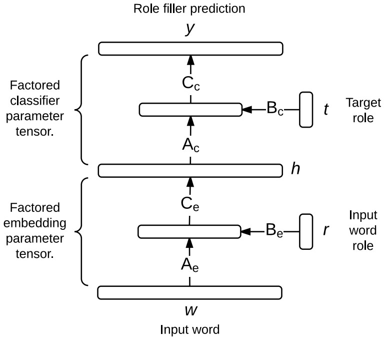
Larger sentence contexts

Can rate verbs with single arguments with strong correlations with human judgements (e.g. “doctor advises”).

- Can we do more than two? (i.e. real sentences)
- Vector spaces: multiplication/intersection operations necessary for composition.
- Multiple arguments: too many zeros in representation! (due to multiplication)

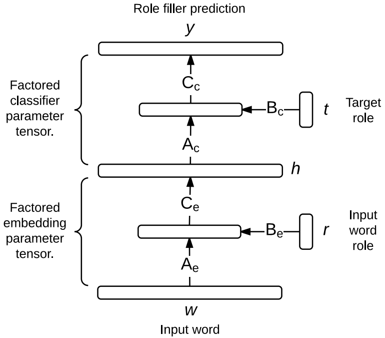
Larger sentence contexts

- Deep learning for compositionality across multiple roles [Tilk, Demberg, Sayeed, Klakow, Thater, 2016].
    - Simulate compositionality, prediction with neural network.

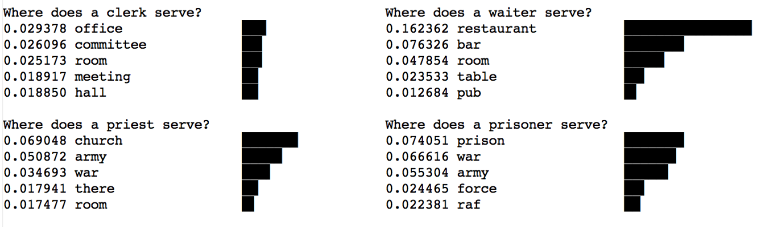


Larger sentence contexts

- Deep learning for compositionality across multiple roles [Tilk, Demberg, Sayeed, Klakow, Thater, 2016].
  - Simulate compositionality, prediction with neural network.



Top locations for serving given a subject:



Neural network model

- Factored tensor trained via feed-forward NN and RNN.

Neural network model

- Factored tensor trained via feed-forward NN and RNN.
- Feed-forward model on single verb-role pairs:
  - Comparable results to DMs ($\rho = 45$ on instruments, **best** $\rho = 44$ on locations).

Neural network model

- Factored tensor trained via feed-forward NN and RNN.
- Feed-forward model on single verb-role pairs:
  - Comparable results to DMs ($\rho = 45$ on instruments, **best** $\rho = 44$ on locations).
  - Fewer trade-offs (e.g. DM models that do well on agent/patient do worse on instruments. . .)

Neural network model

- Factored tensor trained via feed-forward NN and RNN.
- Feed-forward model on single verb-role pairs:
  - Comparable results to DMs ($\rho = 45$ on instruments, **best** $\rho = 44$ on locations).
  - Fewer trade-offs (e.g. DM models that do well on agent/patient do worse on instruments. . .)
- Current application project: nominal predicate role-filler prediction:
  - e.g. "**Bob** wanted to get rid of his car. From the proceeds of the **sale**. . . "

Limits of distributional knowledge

- Observed distributions don't necessarily match world knowledge.

Limits of distributional knowledge

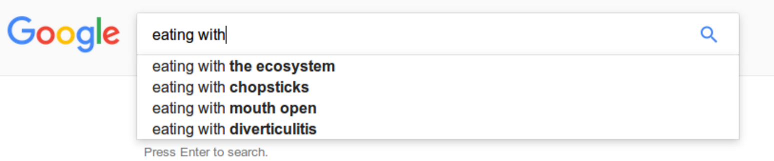
- Observed distributions don't necessarily match world knowledge.
- Consider eating with chopsticks vs eating with forks.

Limits of distributional knowledge

- Observed distributions don't necessarily match world knowledge.
- Consider eating with chopsticks vs eating with forks.
  - English text mentions eating with chopsticks more often!

Limits of distributional knowledge

- Observed distributions don't necessarily match world knowledge.
- Consider eating with chopsticks vs eating with forks.
    - English text mentions eating with chopsticks more often!



Limits of distributional knowledge

We started with plausibility.

Limits of distributional knowledge

We started with plausibility.

(1) Bob cut a cake

a. with a knife. (typical/frequent, possible)
b. with a hammer. (distributionally similar to *knife*, impossible)
c. with floss. (atypical/dissimilar/infrequent, possible)
d. with a towel. (atypical/dissimilar/infrequent, impossible)

Limits of distributional knowledge

We started with plausibility.

(1) Bob cut a cake

a. with a knife. (typical/frequent, possible)
b. with a hammer. (distributionally similar to *knife*, impossible)
c. with floss. (atypical/dissimilar/infrequent, possible)
d. with a towel. (atypical/dissimilar/infrequent, impossible)

Does "distributional knowledge" really reflect relative plausibility?

Limits of distributional knowledge

We started with plausibility.

(1) Bob cut a cake

a. with a knife. (typical/frequent, possible)
b. with a hammer. (distributionally similar to *knife*, impossible)
c. with floss. (atypical/dissimilar/infrequent, possible)
d. with a towel. (atypical/dissimilar/infrequent, impossible)

Does "distributional knowledge" really reflect relative plausibility? People don't talk about cutting a cake with floss very often!

Limits of distributional knowledge

(But people do cut cakes with floss.)

Types of knowledge

My definitions:

Types of knowledge

My definitions:

Implicit world-knowledge

Latent knowledge about the world that can be induced from indirect information sources (e.g. distributional characteristics of language).

Types of knowledge

My definitions:

Implicit world-knowledge

Latent knowledge about the world that can be induced from indirect information sources (e.g. distributional characteristics of language).

Explicit world-knowledge

Knowledge about the world that is coded explicitly, deduced formally, innate, learned by being told, etc.

Types of knowledge

My definitions:

Implicit world-knowledge

Latent knowledge about the world that can be induced from indirect information sources (e.g. distributional characteristics of language).

Explicit world-knowledge

Knowledge about the world that is coded explicitly, deduced formally, innate, learned by being told, etc.

Future research question: Distributional semantics acquires implicit world-knowledge, but can we acquire the knife-floss distinction this way?

Objective: more targeted data

Existing data sources don't explicitly code for this.

Proposal: collect ratings for an instrument data set.

Objective: more targeted data

Existing data sources don't explicitly code for this.

Proposal: collect ratings for an instrument data set.

(2) Rate from 1-7:

a. How common is it to use a knife to cut something? (probability-biased)

b. Can a knife be used to cut? (possibility-biased)

c. A knife is something that you use to cut. ("equipoise")

Objective: more targeted data

Existing data sources don't explicitly code for this.
Proposal: collect ratings for an instrument data set.

(2) Rate from 1-7:

a. How common is it to use a knife to cut something? (probability-biased)
b. Can a knife be used to cut? (possibility-biased)
c. A knife is something that you use to cut. ("equipoise")

- Previous data sets never explicitly distinguished between probable events and possible events!

Objective: more targeted data

Existing data sources don't explicitly code for this.
Proposal: collect ratings for an instrument data set.

(2) Rate from 1-7:

a. How common is it to use a knife to cut something? (probability-biased)
b. Can a knife be used to cut? (possibility-biased)
c. A knife is something that you use to cut. ("equipoise")

- Previous data sets never explicitly distinguished between probable events and possible events!
- Then use these ratings for correlation studies **both** across computational models and psycholinguistic measures (EEG, reading time, etc).

But can we go beyond verb-argument relations? Yes.

Part 3: scope ambiguity resolution and world knowledge

Beyond verb-argument relations

- Recap, types of world knowledge and their sources:

Beyond verb-argument relations

- Recap, types of world knowledge and their sources:
  - Thematic fit: typicality of verb-argument set.
  - Implicit vs. explicit information.

Beyond verb-argument relations

- Recap, types of world knowledge and their sources:
  - Thematic fit: typicality of verb-argument set.
  - Implicit vs. explicit information.
- But how does knowledge interact with higher-order semantic structure?

Beyond verb-argument relations

- Recap, types of world knowledge and their sources:
  - Thematic fit: typicality of verb-argument set.
  - Implicit vs. explicit information.
- But how does knowledge interact with higher-order semantic structure?
- One place to look: scope ambiguity resolution.

Scope ambiguity in processing

Readings for: Every child climbed a tree.

- For each child, that child found a tree and climbed it.
  Linear scope
- There is a tree such that all the children climbed that tree.
  Inverse scope

Scope ambiguity in processing

Readings for: Every child climbed a tree.

- For each child, that child found a tree and climbed it.
  Linear scope
- There is a tree such that all the children climbed that tree.
  Inverse scope

English speakers prefer the linear reading much more strongly than the inverse readings **even when strongly prompted otherwise**. (Dwivedi, 2013)

Scope ambiguity in processing

Readings for: Every jeweller examined a diamond.

- For each jeweller, that jeweller had a diamond to examine.
  Linear scope
- There is a diamond such that all the jewellers examined that diamond.
  Inverse scope

Scope ambiguity in processing

Readings for: Every jeweller examined a diamond.

- For each jeweller, that jeweller had a diamond to examine.
  Linear scope
- There is a diamond such that all the jewellers examined that diamond.
  Inverse scope

English speakers prefer the linear reading much more strongly than the inverse reading **but not as strongly as with the "children-tree" example**. (Dwivedi, 2013)

Scope ambiguity in processing

- English-speakers intuition: many children $\Rightarrow$ many trees. Less so for jewellers and diamonds.

Scope ambiguity in processing

- English-speakers intuition: many children $\Rightarrow$ many trees. Less so for jewellers and diamonds.
- Role of world-knowledge – concepts bound up with words constrain likelihood of inverse readings.

Scope ambiguity in processing

- English-speakers intuition: many children $\Rightarrow$ many trees. Less so for jewellers and diamonds.
- Role of world-knowledge – concepts bound up with words constrain likelihood of inverse readings.

But is that the only constraint?

Scope ambiguity in processing

Consider the following sentence and continuations:

(3) A caregiver comforted a child every night.

a. The caregiver wanted the child to get some rest.
b. The caregivers wanted the child to get some rest.
c. The caregiver wanted the children to get some rest.
d. The caregivers wanted the children to get some rest.

There are four plausible readings of the first sentence, based on the scope of "every night".

Scope ambiguity in processing

Consider the following sentence and continuations:

(3) A caregiver comforted a child every night.

a. The caregiver wanted the child to get some rest.
b. The caregivers wanted the child to get some rest.
c. The caregiver wanted the children to get some rest.
d. The caregivers wanted the children to get some rest.

There are four plausible readings of the first sentence, based on the scope of "every night".

With three quantifiers: can investigate whether there is a preferred specified order in incremental context.

Scope ambiguity in processing

(4) A caregiver comforted a child every night.

- The linear order is $\exists = \exists > \forall$, ie, child and caregiver singular.

Scope ambiguity in processing

(4) A caregiver comforted a child every night.

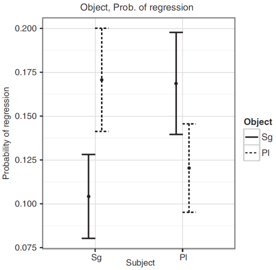
- The linear order is $\exists = \exists > \forall$, ie, child and caregiver singular.
- English speakers can be easily prompted to all other orders.

Scope ambiguity in processing

(4) A caregiver comforted a child every night.

- The linear order is $\exists = \exists > \forall$, ie, child and caregiver singular.
- English speakers can be easily prompted to all other orders.
- Dotlačil and Brasoveanu (2015):

The caregiver wanted the child to get some rest.
The caregivers wanted the child to get some rest.
The caregiver wanted the children to get some rest.
The caregivers wanted the children to get some rest.

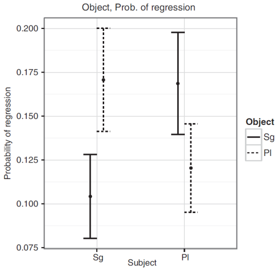


Scope ambiguity in processing

(4) A caregiver comforted a child every night.

- The linear order is $\exists = \exists > \forall$, ie, child and caregiver singular.
- English speakers can be easily prompted to all other orders.
- Dotlačil and Brasoveanu (2015):

The caregiver wanted the child to get some rest.
The caregivers wanted the child to get some rest.
The caregiver wanted the children to get some rest.
The caregivers wanted the children to get some rest.



Evidence for algorithmic processing (as opposed to purely pragmatic considerations).

Focus on role of world knowledge

Experimental work: eye-tracking to detect where scope specification really occurs.

Focus on role of world knowledge

Experimental work: eye-tracking to detect where scope specification really occurs.

- Exploit non-English linguistic phenomena; for example, German verb-second.

(5) a. Jeder Spion hat diesen/einen/diese Auftrag/Aufträge erhalten. Der/die Auftrag/Aufträge war(en) gefährlich und riskant.
Every spy-NOM has this/a/these order(s)-ACC received. The order(s) was/were dangerous and risky.
'Every spy received this/a/these order(s). The order(s) was/were dangerous and risky.'

b. Diesen/Einen/Diese Auftrag/Aufträge hat jeder Spion erhalten. Der/die Auftrag/Aufträge war(en) gefährlich und riskant.
This/A/These order(s)-ACC has every spy-NOM received. The order(s) was/were dangerous and risky.
'Every spy received this/a/these order(s). The order(s) was/were dangerous and risky.'

Focus on role of world knowledge

Experimental work: eye-tracking to detect where scope specification really occurs.

- Exploit non-English linguistic phenomena; for example, German verb-second.

(5) a.

Jeder	Spion	hat	diesen/einen/diese	Auftrag/Aufträge	erhalten.	Der/die	Auftrag/Aufträge	
Every	spy-NOM	has	this/a/these	order(s)-ACC	received.	The	order(s)	
war(en)	gefährlich	und	riskant.					
was/were	dangerous	and	risky.					

‘Every spy received this/a/these order(s). The order(s) was/were dangerous and risky.’

b.

Diesen/Einen/Diese	Auftrag/Aufträge	hat	jeder	Spion	erhalten.	Der/die	Auftrag/Aufträge	
This/A/These	order(s)-ACC	has	every	spy-NOM	received.	The	order(s)	
war(en)	gefährlich	und	riskant.					
was/were	dangerous	and	risky.					

‘Every spy received this/a/these order(s). The order(s) was/were dangerous and risky.’

Manipulate word order in Dwivedi-style experiment to test whether world knowledge truly dominates linear/inverse distinction.

First step: judgement study

24 stimuli in either German word order (unscrambled vs. scrambled):

(6) a.

Jeder	Spion	hat	diesen/einen/diese	Auftrag/Aufträge	erhalten.	*Der/die*	*Auftrag/Aufträge*
Every	spy-NOM	has	this/a/these	order(s)-ACC	received.	*The*	*order(s)*

war(en)	gefährlich	und	riskant.
was/were	dangerous	and	risky.

'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

b.

Diesen/Einen/Diese	Auftrag/Aufträge	hat	jeder	Spion	erhalten.	*Der/die*	*Auftrag/Aufträge*
This/A/These	order(s)-ACC	has	every	spy-NOM	received.	*The*	*order(s)*

war(en)	gefährlich	und	riskant.
was/were	dangerous	and	risky.

'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

First step: judgement study

24 stimuli in either German word order (unscrambled vs. scrambled):

(6) a. Jeder Spion hat diesen/einen/diese Auftrag/Aufträge erhalten. *Der/die Auftrag/Aufträge*
Every spy-NOM has this/a/these order(s)-ACC received. *The order(s)*
war(en) gefährlich und riskant.
was/were dangerous and risky.
'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

b. Diesen/Einen/Diese Auftrag/Aufträge hat jeder Spion erhalten. *Der/die Auftrag/Aufträge*
This/A/These order(s)-ACC has every spy-NOM received. *The order(s)*
war(en) gefährlich und riskant.
was/were dangerous and risky.
'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

- Subjects use online tool (N=68; 24 fillers) and fill in subject of second sentence (italicized).

First step: judgement study

24 stimuli in either German word order (unscrambled vs. scrambled):

(6) a.

Jeder	Spion	hat	diesen/einen/diese	Auftrag/Aufträge	erhalten.	*Der/die*	*Auftrag/Aufträge*
Every	spy-NOM	has	this/a/these	order(s)-ACC	received.	*The*	*order(s)*

war(en)	gefährlich	und	riskant.
was/were	dangerous	and	risky.

'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

b.

Diesen/Einen/Diese	Auftrag/Aufträge	hat	jeder	Spion	erhalten.	*Der/die*	*Auftrag/Aufträge*
This/A/These	order(s)-ACC	has	every	spy-NOM	received.	*The*	*order(s)*

war(en)	gefährlich	und	riskant.
was/were	dangerous	and	risky.

'Every spy received this/a/these order(s). *The order(s) was/were* dangerous and risky.'

- Subjects use online tool (N=68; 24 fillers) and fill in subject of second sentence (italicized).
- Native speaker assists in judging plurality of response.

Experimental goals

Hypothesis

For indefinite article: plural bias of existentially quantified noun should be higher in unscrambled than in scrambled sentences.

Experimental goals

Hypothesis

For indefinite article: plural bias of existentially quantified noun should be higher in unscrambled than in scrambled sentences.

- If effect observed, surface order competes with world knowledge bias of plurality.

Experimental goals

Hypothesis

For indefinite article: plural bias of existentially quantified noun should be higher in unscrambled than in scrambled sentences.

- If effect observed, surface order competes with world knowledge bias of plurality.
- If plural bias still present, even if effect holds, world-knowledge bias must overcome word order.
  - Evidence for interaction between world knowledge and reanalysis process.

Evidence for hypothesis

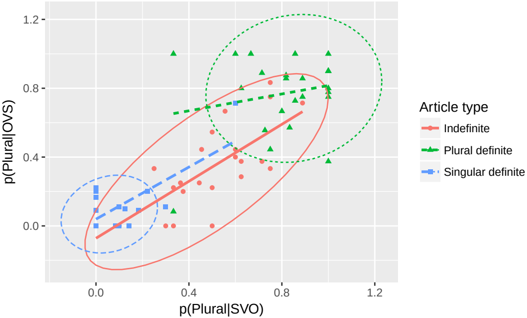
Result from logistic mixed-effects modeling of judgements I collected:

- Significant main effect of surface order for indefinite condition ($b = 0.93$, $p = 0.001$, $z = 3.26$) such that SVO order results more often in a plural response.

Evidence for hypothesis

Result from logistic mixed-effects modeling of judgements I collected:

- Significant main effect of surface order for indefinite condition ($b = 0.93$, $p = 0.001$, $z = 3.26$) such that SVO order results more often in a plural response.
- Plurality correlation between OVS vs. SVO order strong for indefinite articles.



Evidence for hypothesis

- German scrambling provides evidence for world knowledge forcing reanalysis of surface syntax.

Evidence for hypothesis

- German scrambling provides evidence for world knowledge forcing reanalysis of surface syntax.
- Respondents prefer singular interpretation when existentially quantified object fronted.

Evidence for hypothesis

- German scrambling provides evidence for world knowledge forcing reanalysis of surface syntax.
- Respondents prefer singular interpretation when existentially quantified object fronted.
- Competition between surface order and world-knowledge may be activated by presence of indefinite.

Evidence for hypothesis

- German scrambling provides evidence for world knowledge forcing reanalysis of surface syntax.
- Respondents prefer singular interpretation when existentially quantified object fronted.
- Competition between surface order and world-knowledge may be activated by presence of indefinite.

Next steps: Self-paced reading, eye-tracking to investigate time course of competition effect.

Part 4: concluding remarks

Contributions: reprise

Contributions discussed in this talk:

Contributions: reprise

Contributions discussed in this talk:

- Separating “properly” semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).

Contributions: reprise

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
  - DM tensors built from SRL directly seem to encode different/useful part of the space.
  - NN models allow compositionality without sacrificing performance.

Contributions: reprise

Contributions discussed in this talk:

- Separating “properly” semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
  - DM tensors built from SRL directly seem to encode different/useful part of the space.
  - NN models allow compositionality without sacrificing performance.
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.

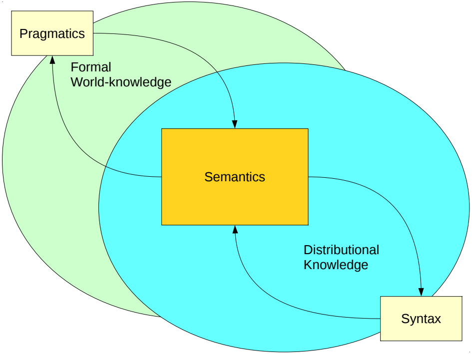
Contributions: reprise

Contributions discussed in this talk:

- Separating "properly" semantic knowledge from syntactic distributional approaches to selectional preference eval (thematic fit).
  - DM tensors built from SRL directly seem to encode different/useful part of the space.
  - NN models allow compositionality without sacrificing performance.
- Identifying the influence of syntax over pragmatic factors in scope disambiguation.
  - Judgement study points towards mechanism for syntax/semantics/pragmatics interaction.

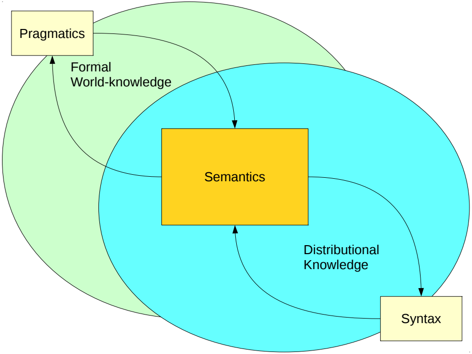
Semantics in context

Bird's-eye view of challenge:



Semantics in context

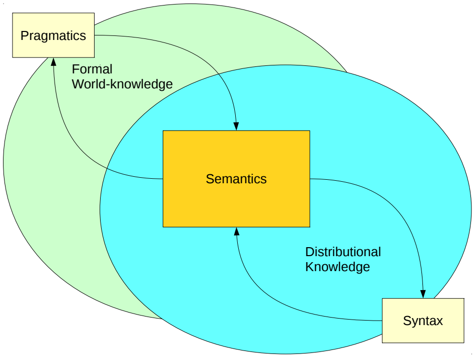
Bird's-eye view of challenge:



- Semantics sits between structural/syntactic and pragmatic levels.

Semantics in context

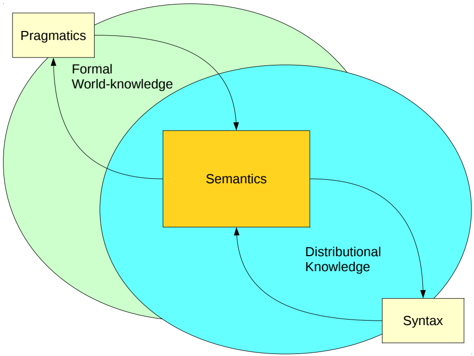
Bird's-eye view of challenge:



- Semantics sits between structural/syntactic and pragmatic levels.
- Computational techniques reveal what semantics we can extract directly from syntactic observation.

Semantics in context

Bird's-eye view of challenge:

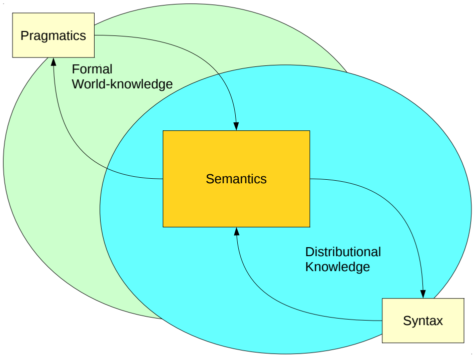


- Semantics sits between structural/syntactic and pragmatic levels.
- Computational techniques reveal what semantics we can extract directly from syntactic observation.

- Psycholinguistic techniques reveal what world knowledge needs to be made formally explicit, interaction with formal structure.

Semantics in context

Bird's-eye view of challenge:



- Semantics sits between structural/syntactic and pragmatic levels.
- Computational techniques reveal what semantics we can extract directly from syntactic observation.
- Psycholinguistic techniques reveal what world knowledge needs to be made formally explicit, interaction with formal structure.

Aim: integrate these lines of research into a single scientific program.

Larger societal context: adaptiveness

⇒ level of representative richness required for effective interaction?

Let's find the line between formal world-knowledge and distributional linguistic knowledge.

Let's find the line between formal world-knowledge and distributional linguistic knowledge.

THANKS!