Prediction in Language Comprehension

Vera Demberg

Cluster of Excellence “Multimodal Computing and Interaction”
Universität des Saarlandes

– Oberseminar ”Linguistic Modeling and its Interfaces” –

Tübingen, April 19th, 2013

Linguistic Processing Difficulty

Measuring and modelling linguistic processing difficulty

- How does the brain process language?
- Readability assessment
- Natural language generation

What is prediction?

Example

(1) a. *Peter ironed his new shirt.*
    b. *Peter bought a new shirt.*

- People anticipate upcoming linguistic content / structure.
- Why would they do that?
  - historically, people have argued against prediction
  - helps with noisy speech signal
  - cognitive plausibility
- Is it really prediction, or just facilitated integration?
- How do prediction and predictability relate to processing difficulty?
  - facilitation at high predictability
  - difficulty when prediction turns out to be incorrect

What is prediction?

Example

(1) a. *Peter ironed his new shirt.*
    b. *Peter bought a new shirt.*

- People anticipate upcoming linguistic content / structure.
- Why would they do that?
  - historically, people have argued against prediction
  - helps with noisy speech signal
  - cognitive plausibility
- Is it really prediction, or just facilitated integration?
- How do prediction and predictability relate to processing difficulty?
  - facilitation at high predictability
  - difficulty when prediction turns out to be incorrect

What is prediction?

Example

(1) a. *Peter ironed his new shirt.*
    b. *Peter bought a new shirt.*

- People anticipate upcoming linguistic content / structure.
- Why would they do that?
  - historically, people have argued against prediction
  - helps with noisy speech signal
  - cognitive plausibility
- Is it really prediction, or just facilitated integration?
- How do prediction and predictability relate to processing difficulty?
  - facilitation at high predictability
  - difficulty when prediction turns out to be incorrect

What is prediction?

Example

(1) a. *Peter ironed his new shirt.*
    b. *Peter bought a new shirt.*

- People anticipate upcoming linguistic content / structure.
- Why would they do that?
  - historically, people have argued against prediction
  - helps with noisy speech signal
  - cognitive plausibility
- Is it really prediction, or just facilitated integration?
- How do prediction and predictability relate to processing difficulty?
  - facilitation at high predictability
  - difficulty when prediction turns out to be incorrect

Table of Contents

Table of Contents

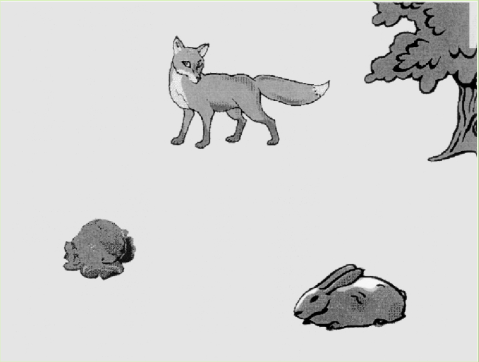
Empirical Evidence for Prediction

Visual world experiment: **anticipatory eye-movements** show that people predict subsequent input [Kamide et al. 2003]

Experimental Findings: Incrementality and Prediction

"Der Hase frisst gleich den Kohl."
The Hare-nom will eat soon the cabbage-acc.

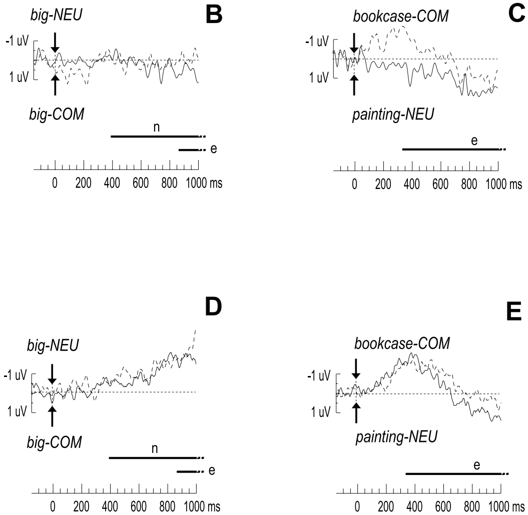
"Den Hasen frisst gleich der Fuchs."
The Hare-acc will eat soon the fox-nom.



Evidence for Prediction: EEG [van Berkum et al., 2005]

The burglar had no trouble whatsoever to locate the secret family safe.
Of course, it was situated behind a...

[no predictive discourse context]
Of course, it was situated behind a ...



Empirical Evidence for Incrementality and Prediction

Either...or processing: faster reading at or-NP [Staub & Clifton, 2006]

Experimental Finding: Prediction

- **processing facilitation** through prediction
- The presence of “either” leads to shorter fixation times on “or” and the second conjunct.

Peter read either a book ***or an essay*** *in the school magazine.*	
Peter read a book ***or an essay*** *in the school magazine.*	

- More general treatment of expectation-raising constructions at discourse level [Cristea & Webber, 1997]

Table of Contents

Surprisal [Hale 2001, 2003; Levy, 2008]

Surprisal as a measure for capturing predictability effects

$$s_{w_n} = -\log P(w_n | context)$$

Key idea: Processing difficulty at w_i $\propto$ amount of Surprisal at perceiving w_i

Surprisal and Processing Difficulty

Experimental support

- predictability effects
- facilitating ambiguity effects (Traxler, 1998)
  "The daughter of the colonel who shot herself had been very depressed."
  "The daughter of the colonel who shot himself had been very depressed."
  "The son of the colonel who shot himself had been very depressed."
- anti-locality effects (Konieczny, 2000)
  "Die Einsicht, dass der Freund dem Kunden das Auto aus Plastik verkaufte ..."
  "Die Einsicht, dass der Freund des Kunden das Auto aus Plastik verkaufte ..."

Support from reading times in naturalistic texts

- on Dundee Corpus (Demberg and Keller, 2008; Roark et al., 2009; Frank, 2009; Fossum and Levy, 2012; Demberg et al., 2014)
- on Potsdam Sentence Corpus (Boston et al., 2008)

Uniform Information Density

Surprisal has also been related to language production:

Uniform Information Density Hypothesis

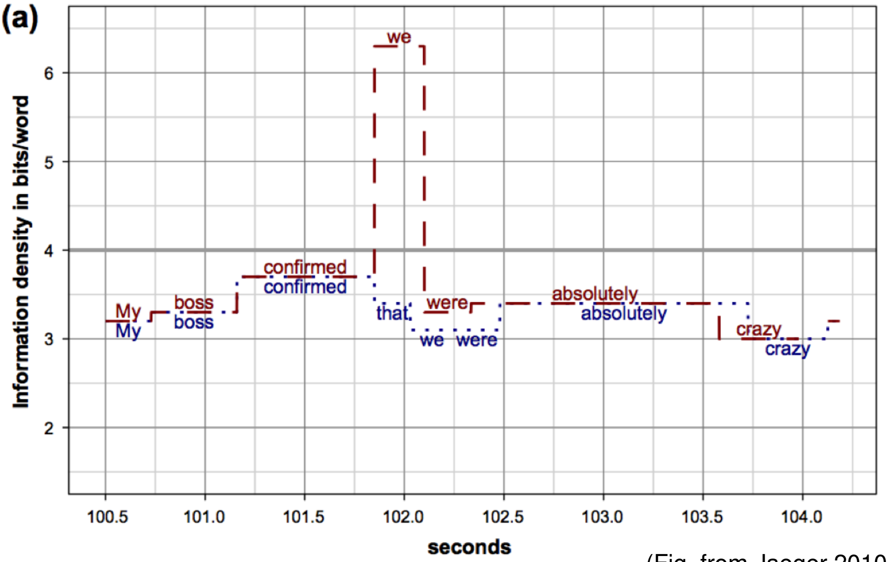
humans tend to spread information evenly across a text, and can use linguistic devices (word length, optional markers, alternative lexicalizations etc.) to achieve this.

(Frank & Jaeger 2008, Jaeger 2010)

Example:

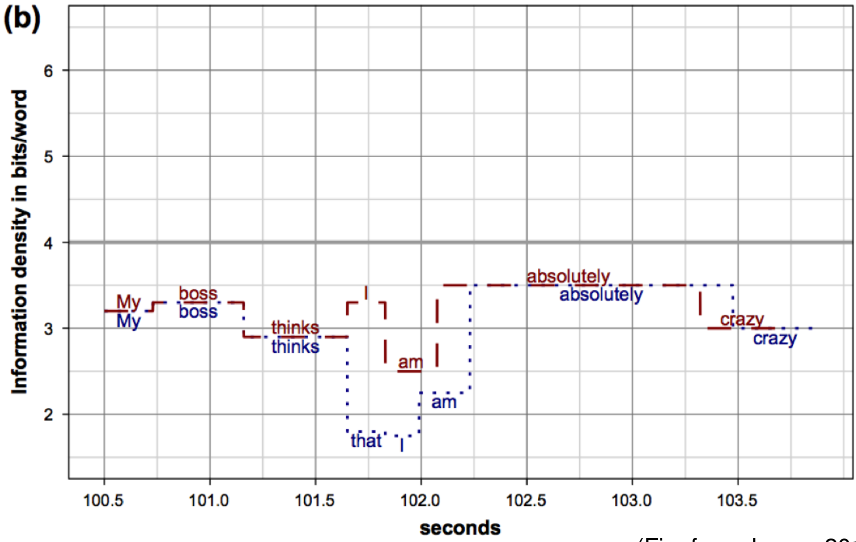
(1) a. My boss confirmed (that) we were absolutely crazy.
    b. My boss thinks (that) we were absolutely crazy.

Uniform Information Density



(Fig. from Jaeger 2010)

Uniform Information Density



(Fig. from Jaeger 2010)

But surprisal alone doesn't explain everything

Surprisal can't account for

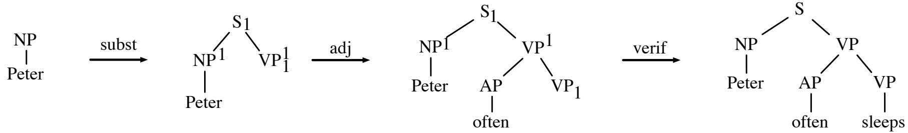
- locality effects
  (Gibson, 1998; e.g., English subject vs. object relative clauses)
- digging-in effects
  "As the author wrote the book describing Babylon grew."
  (Ferreira and Henderson, 1991; Tabor and Hutchins, 2004)
- center-embedding

Effects	Surprisal	DLT
Either-or Prediction	+	–
English Relative Clause	–	+
German Relative Clause	+	–
Facilitating Ambiguity	+	–
Storage Cost Effects	–	+
Center Embedding	NA	+

Combining Surprisal and Memory-based account

Several suggestions to **combine surprisal with a locality/memory** based account (Demberg and Keller, 2008; Levy, 2008; Patil et al., 2009; Staub 2010).

- **Unified model**: Prediction Theory (Demberg & Keller, 2009; Demberg et al., 2014)
- **Idea:** Predictions have to be verified, cost modulated by memory decay



Summary So Far

- Evidence for prediction in language comprehension
- Surprisal can explain predictability effects
- Prediction Theory
  - explicitly represents predictions
  - contains an operation for verifying predictions
  - (currently being extended to semantics)
- $\rightarrow$ if we model predictions explicitly, this can improve explanative power of model.

Table of Contents

Predicting Discourse Relations

- Do any of these ideas hold outside syntax?
- ... for linguistic structures above the sentence?
- Cristea and Webber (1997) observe that certain discourse connectors "raise expectations" (e.g. *on the one hand...on the other hand*)

Discourse Connectors

Discourse Connectors and Processing Difficulty

- discourse connectors can facilitate language processing (e.g., Murray, 1997)
- (if used correctly)
- some types of connectors (e.g., contrast, concessive) have larger effect than others (e.g., elaboration, causal).

Discourse Connectors

Discourse Connectors and Processing Difficulty

- discourse connectors can facilitate language processing (e.g., Murray, 1997)
- (if used correctly)
- some types of connectors (e.g., contrast, concessive) have larger effect than others (e.g., elaboration, causal).

Discourse Connectors

Discourse Connectors and Processing Difficulty

- discourse connectors can facilitate language processing (e.g., Murray, 1997)
- (if used correctly)
- some types of connectors (e.g., contrast, concessive) have larger effect than others (e.g., elaboration, causal).

Discourse Connectives and Incremental Processing

Are discourse connectors processed **incrementally**?

Can people make **predictions** based on discourse connectors?

- Connective Integration Model (Millis & Just, 1994): When connective encountered, preceding part buffered, integration at the end
- Incremental processing (Traxler, Bybee, & Pickering, 1997)
- Evidence for incremental processing of causals, but without connector (Kuperberg, Paczynski, & Ditman, 2011)

Experiment on time course of integration of causal and concessive connectors (therefore / however)

Discourse Connectives and Incremental Processing

Are discourse connectors processed **incrementally**?

Can people make **predictions** based on discourse connectors?

- Connective Integration Model (Millis & Just, 1994): When connective encountered, preceding part buffered, integration at the end
- Incremental processing (Traxler, Bybee, & Pickering, 1997)
- Evidence for incremental processing of causals, but without connector (Kuperberg, Paczynski, & Ditman, 2011)

Experiment on time course of integration of causal and concessive connectors (therefore / however)

Experiment on Discourse Connectors

[Köhne and Demberg, 2013]

Steffen denkt über einen kleinen Snack nach. Er hat gerade Lust, etwas Süsses zu essen.

Daher holt er sich aus der Küche die appetitliche Waffel.

Dennoch holt er sich aus der Küche die appetitliche Bretzel.

Experiment on Discourse Connectors

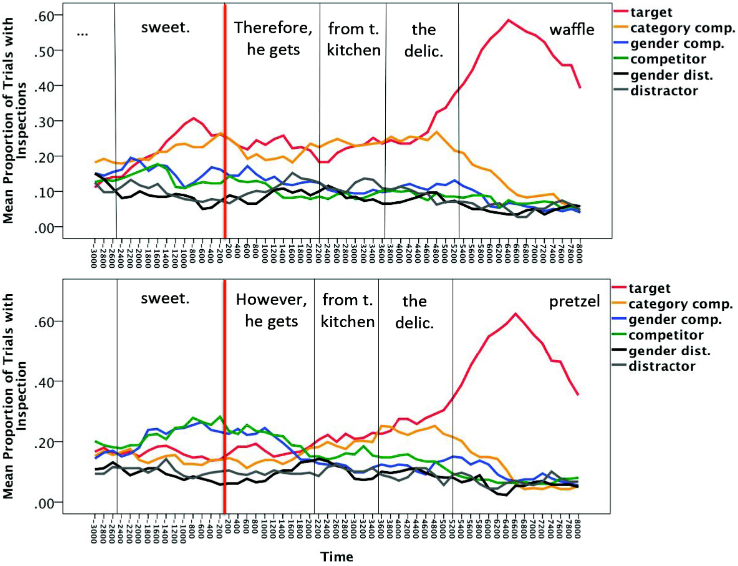
[Köhne and Demberg, 2013]

Steffen denkt über einen kleinen Snack nach. Er hat gerade Lust, etwas Süsses zu essen.

Daher holt er sich aus der Küche die appetitliche Waffel.

Dennoch holt er sich aus der Küche die appetitliche Bretzel.

Experiment on Discourse Connectors



Results of Visual World Experiment

Results:

- Discourse connector is integrated incrementally.
- Evidence for prediction based on discourse cue (at least in a strongly predictive context).
- Concessive connector gives rise to search for alternatives (similar to negation; Kaup (2006)).
- Concessives processed more slowly than causals.

For more details, see my talk here in June: DETEC 2013

Results of Visual World Experiment

Results:

- Discourse connector is integrated incrementally.
- Evidence for prediction based on discourse cue (at least in a strongly predictive context).
- Concessive connector gives rise to search for alternatives (similar to negation; Kaup (2006)).
- Concessives processed more slowly than causals.

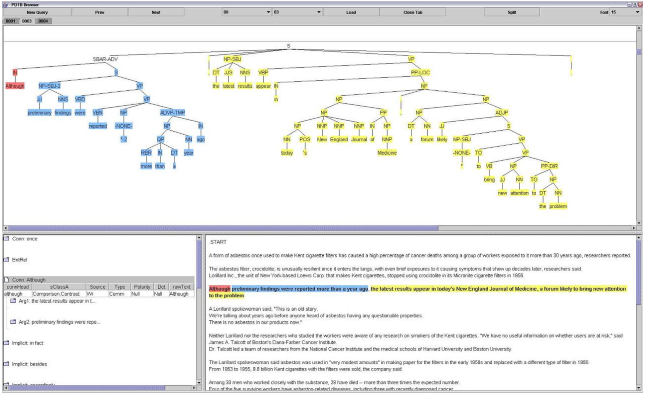
For more details, see my talk here in June: DETEC 2013

Table of Contents

Discourse Relations

Do people anticipate discourse relations?

We approached this question via a corpus study of the Penn Discourse Treebank (Prasad et al., 2008)



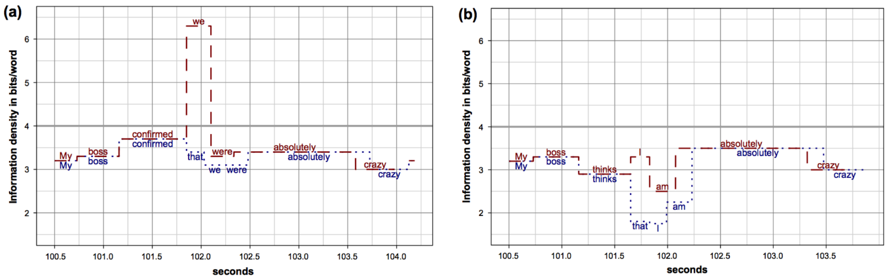
Reminder

Uniform Information Density Hypothesis

humans tend to spread information evenly across a text, and can use linguistic devices to achieve this. (Frank & Jaeger 2008, Jaeger 2010)

Information Density

$$surprisal(event) = -\log P(event|context)$$ (Surprisal, Hale 2001)



Implicit vs. explicitly marked discourse relations

Translating the UID observations about the optionality of “that” etc. to the context of discourse connectors:

Discourse relations can be:

- **Explicitly marked**
  “Sarah got a sunburn **because** she forgot to put on sun screen.”
- **Implicit**
  “Sarah got a sunburn. She forgot to put on sun screen.”

Distribution in Penn Discourse Treebank (Prasad et al., 2008)

Relations in WSJ	Frequency
Explicit	18459
Implicit	16224

Implicit vs. explicitly marked discourse relations

Translating the UID observations about the optionality of “that” etc. to the context of discourse connectors:

Discourse relations can be:

- **Explicitly marked**
  “Sarah got a sunburn **because** she forgot to put on sun screen.”
- **Implicit**
  “Sarah got a sunburn. She forgot to put on sun screen.”

Distribution in Penn Discourse Treebank (Prasad et al., 2008)

Relations in WSJ	Frequency
Explicit	18459
Implicit	16224

UID hypothesis and implicit discourse relations

Can we use the **Uniform Information Density Hypothesis** to explain when **discourse connectors are explicit vs. implicit**?

We need to know what is an expected (little-surprising) event.

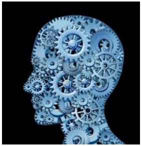
Corpus study testing: Expected discourse events should be less likely to be marked explicitly with a discourse cue than unexpected discourse events.

(Asr & Demberg, 2012)

UID hypothesis and implicit discourse relations

Can we use the **Uniform Information Density Hypothesis** to explain when **discourse connectors are explicit vs. implicit**?

We need to know what is an expected (little-surprising) event.



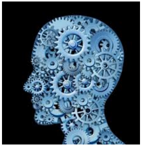
Corpus study testing: Expected discourse events should be less likely to be marked explicitly with a discourse cue than unexpected discourse events.

(Asr & Demberg, 2012)

UID hypothesis and implicit discourse relations

Can we use the **Uniform Information Density Hypothesis** to explain when **discourse connectors are explicit vs. implicit**?

We need to know what is an expected (little-surprising) event.



Corpus study testing: Expected discourse events should be less likely to be marked explicitly with a discourse cue than unexpected discourse events.

(Asr & Demberg, 2012)

Literature on Discourse Relations

Continuity Hypothesis: readers expect a sentence to be causally congruent and continuous with respect to its preceding context.

– (Segal et al.,1991; Murray 1997)

Supporting Evidence

People have tendency to **identify continuous relations** between adjacent sentences (Segal et al., 1991)

More **reading facilitation for signals of discontinuity** (continuity is already expected) (Murray, 1994)

More **salient effect of inappropriate discontinuous** discourse markers (Murray, 1997)

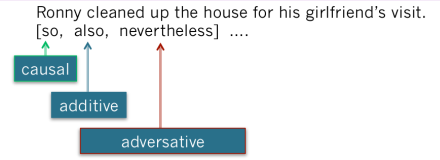
Literature on Discourse Relations

Causality-by-default Hypothesis: readers start out assuming a causal relation between two consecutive sentences.

– (Sanders, 2005)

Supporting Evidence

Tendency to choose **causal** sentence completion (Murray 1997)



Ronny cleaned up the house for his girlfriend's visit.
...

Semantic processing difficulty (larger N400) when reading causally unrelated sentences. (Kuperberg et al., 2011)

Continuity and Causality (Example)

Causal

Gary's daughter was sick (cause) so he took her to the hospital (consequence)

Gary took his daughter to the hospital (consequence) because she was sick (cause)

Continuity and Causality (Example)

Causal

Forward

$\underbrace{\text{Gary's daughter was sick}}_{\text{cause}}$ so $\underbrace{\text{he took her to the hospital}}_{\text{consequence}}$

Backward

$\underbrace{\text{Gary took his daughter to the hospital}}_{\text{consequence}}$ because $\underbrace{\text{she was sick}}_{\text{cause}}$

Continuity and Causality (Example)

Causal

Forward

Gary's daughter was sick so he took her to the hospital

cause consequence

Backward

Gary took his daughter to the hospital because she was sick

consequence cause

Concessive (=Negative Causal)

Forward

Gary's daughter was sick but he sent her to the kindergarten

cause neg-consequence

Backward

Gary sent his daughter to the kindergarten although she was sick

neg-consequence cause

Our Hypotheses

Taking together

- Uniform Information Density Hypothesis
- Causality-by-default Hypothesis
- Continuity Hypothesis

Going to test this on the **Penn Discourse Treebank** (Prasad et al., 2008)

Our Hypotheses

Taking together

- Uniform Information Density Hypothesis
- Causality-by-default Hypothesis
- Continuity Hypothesis

causal concessive

Going to test this on the **Penn Discourse Treebank** (Prasad et al., 2008)

Our Hypotheses

Taking together

- Uniform Information Density Hypothesis
- Causality-by-default Hypothesis
- Continuity Hypothesis

causal	concessive
continuous	discontinuous

Going to test this on the **Penn Discourse Treebank** (Prasad et al., 2008)

Our Hypotheses

Taking together

- Uniform Information Density Hypothesis
- Causality-by-default Hypothesis
- Continuity Hypothesis

Implicit ⟷ Explicit

Implicit	Explicit
causal	concessive
continuous	discontinuous
forward temporal	backward temporal

Going to test this on the **Penn Discourse Treebank** (Prasad et al., 2008)

Our Hypotheses

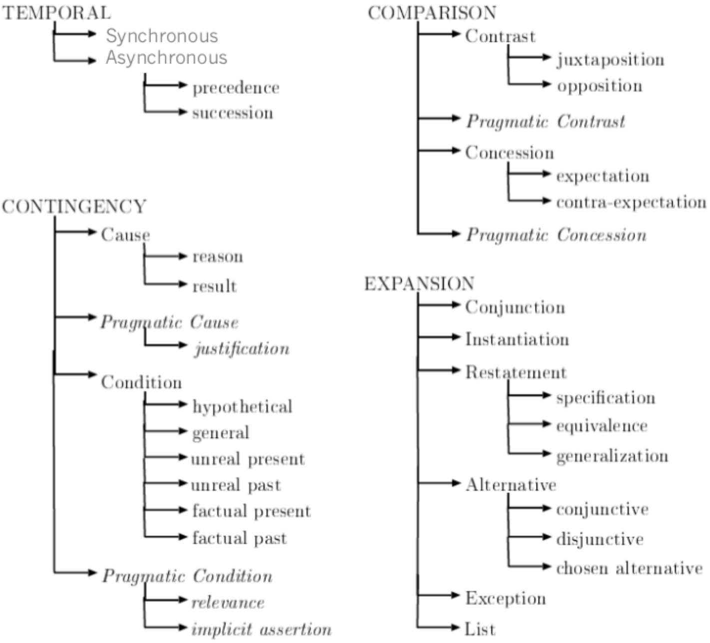
Taking together

- Uniform Information Density Hypothesis
- Causality-by-default Hypothesis
- Continuity Hypothesis

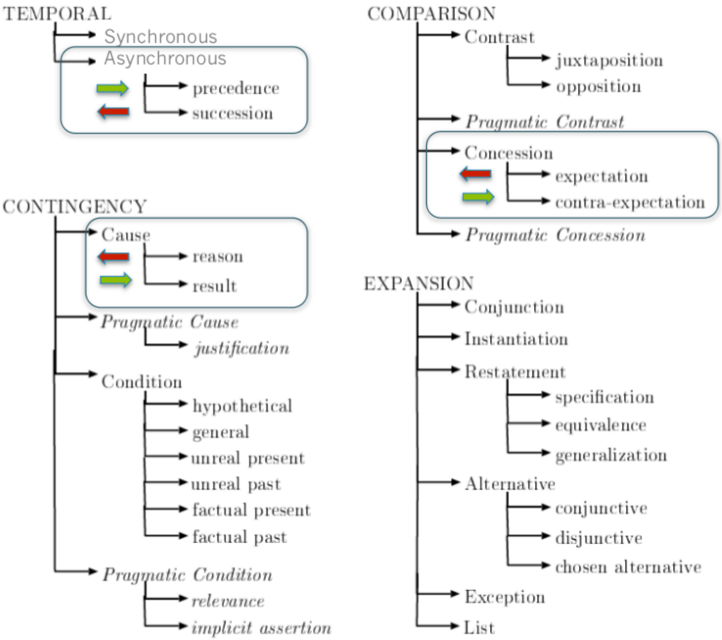
causal	concessive
continuous	discontinuous
forward temporal	backward temporal

Going to test this on the **Penn Discourse Treebank** (Prasad et al., 2008)

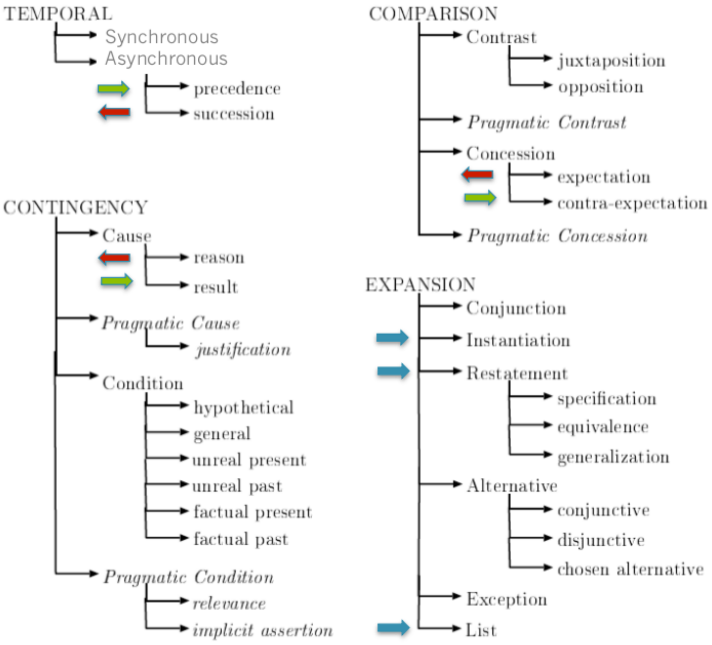
PDTB relation sense hierarchy



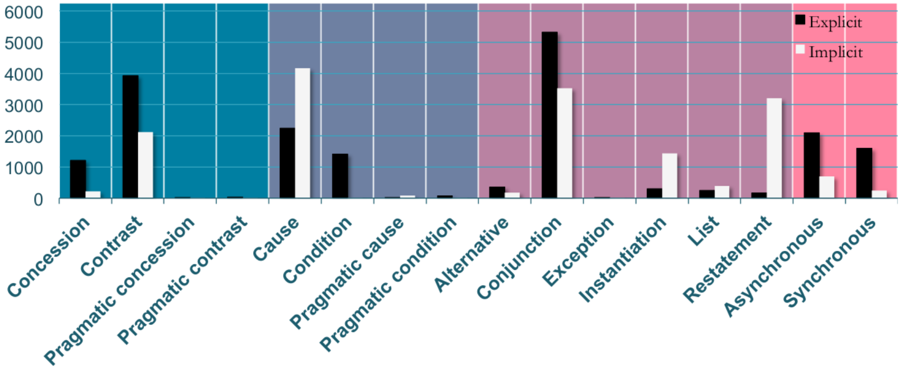
PDTB relation sense hierarchy



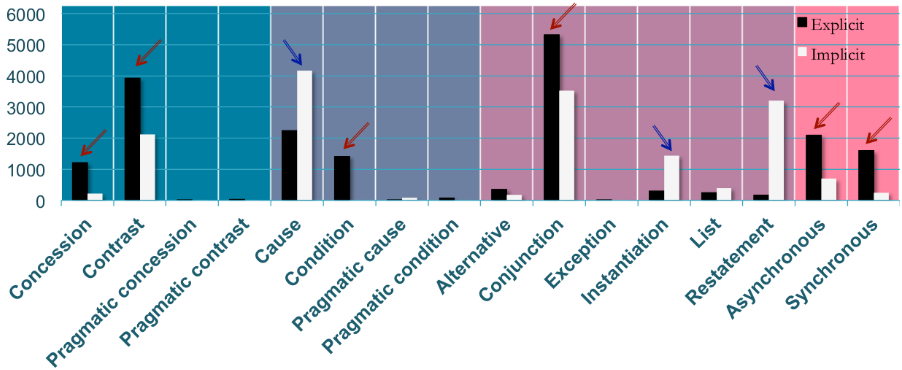
PDTB relation sense hierarchy



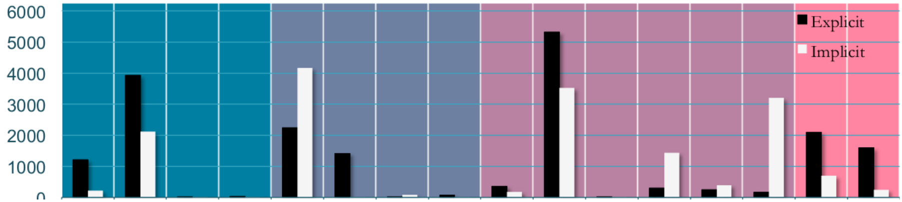
Overall frequencies



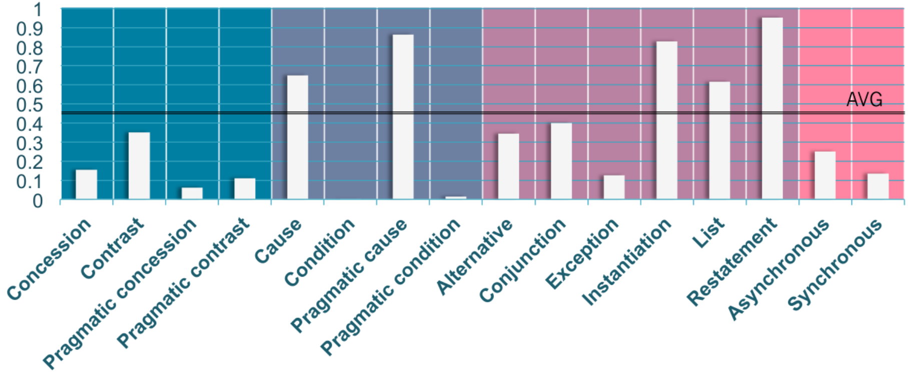
Overall frequencies



Overall frequencies



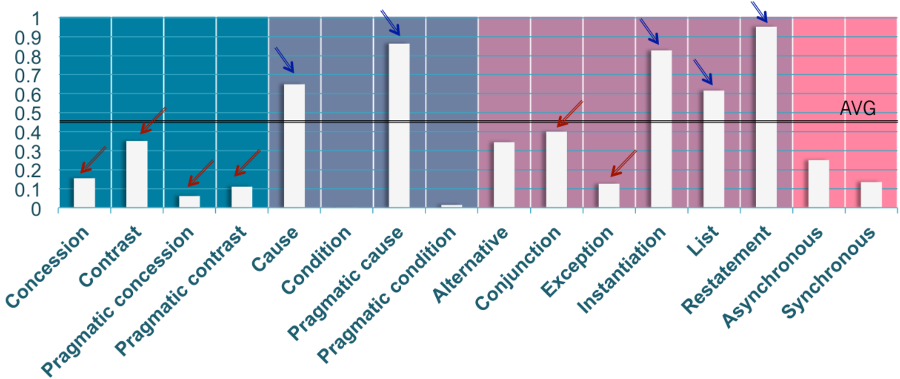
Implicitness ratio



Implicit ⟷ Explicit

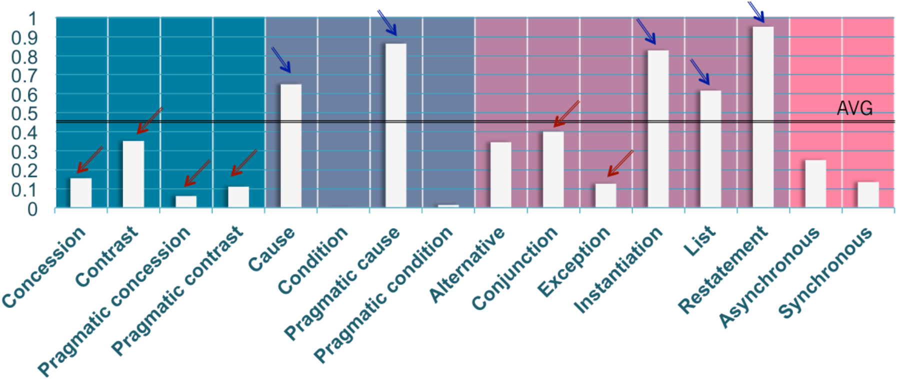
Implicit	Explicit
causal	concessive
continuous	discontinuous
forward temporal	backward temporal

Implicitness ratio



causal	yes	concessive
continuous		discontinuous
forward temporal		backward temporal

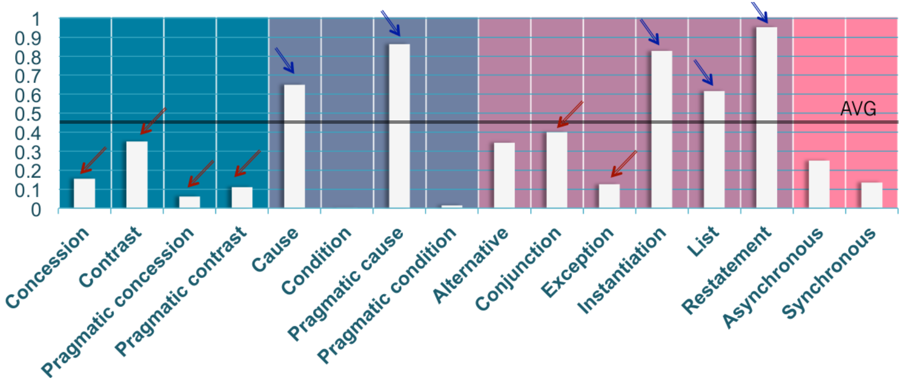
Implicitness ratio



Implicit ⟷ Explicit

causal	yes	concessive
continuous	yes	discontinuous
forward temporal		backward temporal

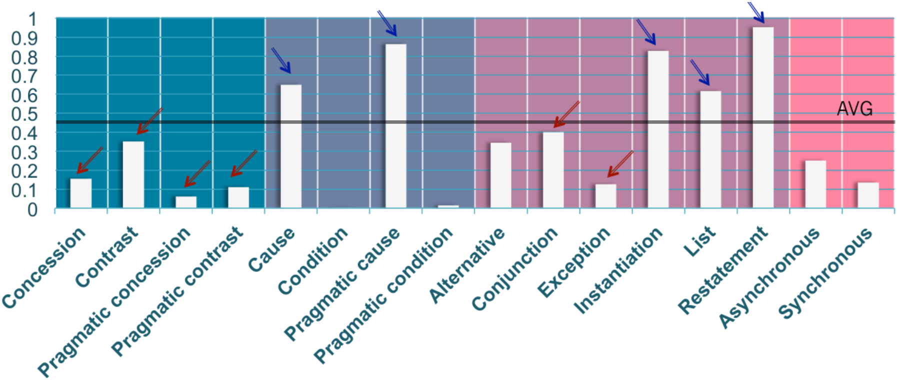
Implicitness ratio



Implicit ⟷ Explicit

Implicit		Explicit
causal	yes	concessive
continuous	yes	discontinuous
forward temporal	?	backward temporal

Implicitness ratio

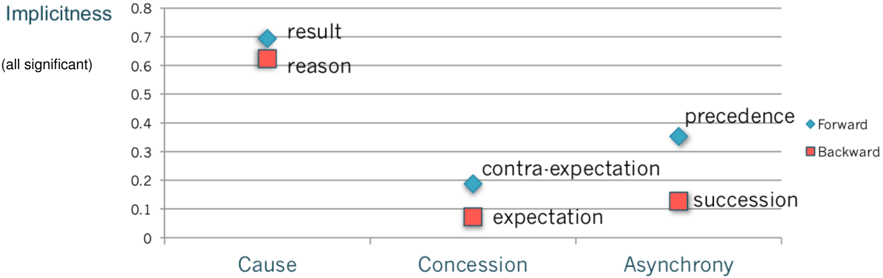


Implicit ⟷ Explicit

causal	yes	concessive
continuous	yes	discontinuous
forward temporal	?	backward temporal

Forward vs. backward temporality

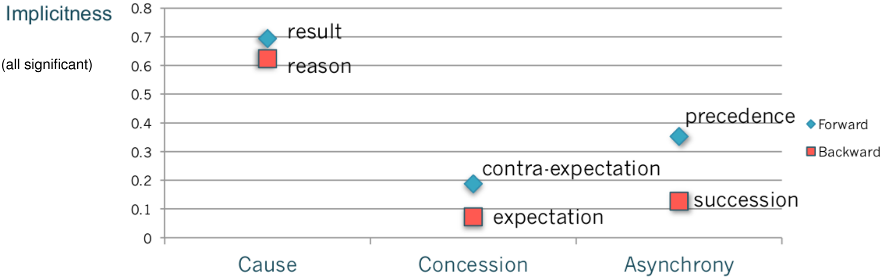
(all significant)



Implicit ⟷ Explicit

causal	yes	concessive
continuous	yes	discontinuous
forward temporal	yes	backward temporal

Forward vs. backward temporality



So far...

- **Hypothesis:** Predictable relations need not be expressed explicitly (UID)
- **Finding:** Relations that are more expected due to cognitive biases (causality, continuity) are more often implicit.
- **But:** no *local* context taken into account
- **Next:** let's take a look at local context.

Local Factors

Implicit causality (IC) verbs trigger a discourse expectation for a reason
Kehler et al. (2008); Rohde & Horton (2010)

Example

Dawn **amazed** Malcolm...

She was playing the piano with her eyes closed.$_{reason}$
He applauded her talents.$_{other}$

Are causal relations more likely to be implicit if the ARG1 contains an IC verb?
Test on PDTB inconclusive.

Local Factors

Implicit causality (IC) verbs trigger a discourse expectation for a reason
Kehler et al. (2008); Rohde & Horton (2010)

Example

Dawn **amazed** Malcolm...

She was playing the piano with her eyes closed.$_{\textit{reason}}$
He applauded her talents.$_{\textit{other}}$

Are causal relations more likely to be implicit if the ARG1 contains an IC verb?
Test on PDTB inconclusive.

Local Factors

Webber (2013) shows that `Chosen Alternative` relations are usually licensed by negation, modals, downward-entailing verbs.

Example

If the flex is worn, **do not use insulating tape to repair it.** Instead, *you should replace it ....*

If these are strong local cues, we expect that an explicit cue is not necessary.

Feature	Implicit tokens	Explicit tokens
Negation marker	116 (67.8%)	47 (39.8%)
Downward-entailment	24 (14.0%)	18 (15.3%)
Event Modal	9 (5.3%)	13 (11.0%)
Other	22 (12.9%)	40 (33.9%)
Total	171	118

(table taken from Webber, 2013)

Local Factors

Webber (2013) shows that `Chosen Alternative` relations are usually licensed by negation, modals, downward-entailing verbs.

Example

If the flex is worn, **do not use insulating tape to repair it.** Instead, *you should replace it ....*

If these are strong local cues, we expect that an explicit cue is not necessary.

Feature	Implicit tokens	Explicit tokens
Negation marker	116 (67.8%)	47 (39.8%)
Downward-entailment	24 (14.0%)	18 (15.3%)
Event Modal	9 (5.3%)	13 (11.0%)
Other	22 (12.9%)	40 (33.9%)
Total	171	118

(table taken from Webber, 2013)

Is negation a good cue?

Are these features good cues only for `Chosen Alternative` or also for other discourse relations?

- also: significantly more **reasons**, especially **implicit** ones.
- significantly **fewer temporals**
- significantly more conditional unreal (can only be explicit)

Interesting to look at subtype of specifications:

more	**less**
implicit generalization	implicit instantiation
explicit specification	implicit specifications

Is negation a good cue?

Are these features good cues only for `Chosen Alternative` or also for other discourse relations?

- also: significantly more **reasons**, especially **implicit** ones.
- significantly **fewer temporals**
- significantly more conditional unreal (can only be explicit)

Interesting to look at subtype of specifications:

more	**less**
implicit generalization	implicit instantiation
explicit specification	implicit specifications

Is negation a good cue?

Are these features good cues only for `Chosen Alternative` or also for other discourse relations?

- also: significantly more **reasons**, especially **implicit** ones.
- significantly **fewer temporals**
- significantly more conditional unreal (can only be explicit)

Interesting to look at subtype of specifications:

more	**less**
implicit generalization	implicit instantiation
explicit specification	implicit specifications

Take-home points from this exercise

Results

- People can generate predictions of upcoming content given a discourse connector.
  $\rightarrow$ our syntactic models may be applied to processing above the sentence
- Concessives (= negative causals) are more difficult to process than causal connectors).
  $\rightarrow$ better estimates of processing difficulty
- Uniform information density can account for use of optional discourse connectives
  $\rightarrow$ useful for language generation
- First indications that local cues might help humans in anticipating discourse relations.
  $\rightarrow$ useful for automatic text comprehension / relation labelling

Take-home points from this exercise

Results

- People can generate predictions of upcoming content given a discourse connector.
  $\rightarrow$ our syntactic models may be applied to processing above the sentence
- Concessives (= negative causals) are more difficult to process than causal connectors).
  $\rightarrow$ better estimates of processing difficulty
- Uniform information density can account for use of optional discourse connectives
  $\rightarrow$ useful for language generation
- First indications that local cues might help humans in anticipating discourse relations.
  $\rightarrow$ useful for automatic text comprehension / relation labelling

Take-home points from this exercise

Results

- People can generate predictions of upcoming content given a discourse connector.
  $\rightarrow$ our syntactic models may be applied to processing above the sentence
- Concessives (= negative causals) are more difficult to process than causal connectors).
  $\rightarrow$ better estimates of processing difficulty
- Uniform information density can account for use of optional discourse connectives
  $\rightarrow$ useful for language generation
- First indications that local cues might help humans in anticipating discourse relations.
  $\rightarrow$ useful for automatic text comprehension / relation labelling

Take-home points from this exercise

Results

- People can generate predictions of upcoming content given a discourse connector.
  $\rightarrow$ our syntactic models may be applied to processing above the sentence
- Concessives (= negative causals) are more difficult to process than causal connectors).
  $\rightarrow$ better estimates of processing difficulty
- Uniform information density can account for use of optional discourse connectives
  $\rightarrow$ useful for language generation
- First indications that local cues might help humans in anticipating discourse relations.
  $\rightarrow$ useful for automatic text comprehension / relation labelling

Table of Contents

Taking it to the wild

Does this matter for processing difficulty effects in real life?

Taking it to the wild

Does this matter for processing difficulty effects in real life?

Can we detect a measurable ...

... effect on cognitive load in dual task scenarios?

... performance drop in the driving task?

... performance drop in language task?

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving
  - language
- Task performance
  - driving
  - language
- Cognitive load

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving
  - language
- Cognitive load

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving $\longrightarrow$ events in the driving task (e.g., steering performance)
  - language $\longrightarrow$ comprehension tests (reaction time, answer accuracy)
- Cognitive load

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving $\longrightarrow$ events in the driving task (e.g., steering performance)
  - language $\longrightarrow$ comprehension tests (reaction time, answer accuracy)
- Cognitive load $\longrightarrow$ **difficult in realistic dual-task setting!**

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving $\longrightarrow$ events in the driving task (e.g., steering performance)
  - language $\longrightarrow$ comprehension tests (reaction time, answer accuracy)
- Cognitive load $\longrightarrow$ pupillometry, skin conductance

Experimental Design

Simultaneous driving and language experiment, manipulating

- the difficulty of the driving task
- the complexity of the language

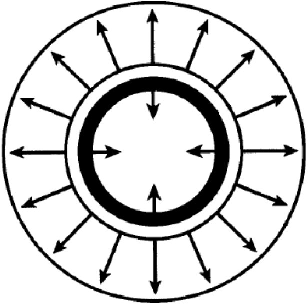
Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving $\longrightarrow$ events in the driving task (e.g., **steering performance**)
  - language $\longrightarrow$ comprehension tests (reaction time, answer accuracy)
- Cognitive load $\longrightarrow$ **pupillometry**, skin conductance

Table of Contents

Background on Pupillometry

- pupil dilation = activation / inhibition of two muscles (Dilator Pupillae & Sphincter Pupillae)
- response time: 200-300msec; peak after about 1200ms



from: Beatty & Lucero-Wagoner 2000

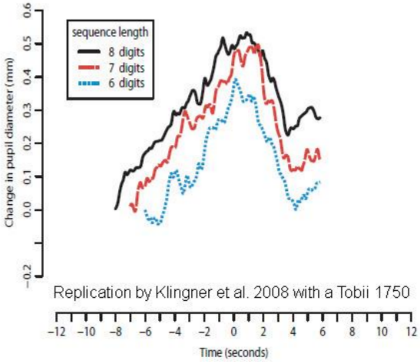
Pupil size has been argued to reflect

- arithmetic problems
  (Hess & Polt 1964)
- digit recall, memory tasks
  (Kahnemann & Beatty 1966)
- attention (Beatty, 1982)
- inference
- language
  - syntactic complexity
    (Just & Carpenter 1993)
  - translation
    (Hyönä, Tomola & Alaja, 1995)
  - grammaticality violations
    (Gutierrez & Shapiro 2010)
  - context integration
    (Engelhardt et al. 2010)

Pupil size has been argued to reflect

- arithmetic problems (Hess & Polt 1964)
- digit recall, memory tasks (Kahnemann & Beatty 1966)
- attention (Beatty, 1982)
- inference
- language
  - syntactic complexity (Just & Carpenter 1993)
  - translation (Hyönä, Tomola & Alaja, 1995)
  - grammaticality violations (Gutierrez & Shapiro 2010)
  - context integration (Engelhardt et al. 2010)

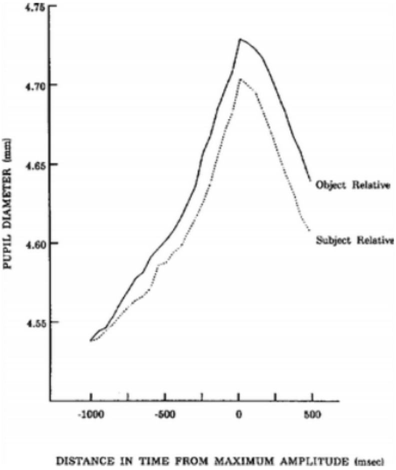
memorization: dilation
recall: constriction



Replication by Klingner et al. 2008 with a Tobii 1750

Pupil size has been argued to reflect

- arithmetic problems
  (Hess & Polt 1964)
- digit recall, memory tasks
  (Kahnemann & Beatty 1966)
- attention (Beatty, 1982)
- inference
- language
  - syntactic complexity
    (Just & Carpenter 1993)
  - translation
    (Hyönä, Tomola & Alaja, 1995)
  - grammaticality violations
    (Gutierrez & Shapiro 2010)
  - context integration
    (Engelhardt et al. 2010)



Difficulties

Difficulties when working with pupil size

- need constant lighting of room
- must control for luminance of stimuli
- must normalize wrt. pupil size

Also problematic for driving task

Difficulties

Difficulties when working with pupil size

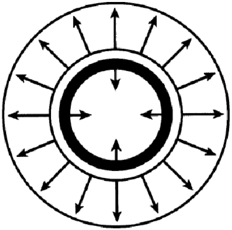
- need constant lighting of room
- must control for luminance of stimuli
- must normalize wrt. pupil size

Also problematic for driving task

Index of Cognitive Activity

Pupillometry – *Index of Cognitive Activity* (ICA; Marshall, 2002)

- Frequency of rapid changes in pupil size (up to 20%)
- Factors out changes due to ambient light
- Different from traditionally used overall dilation
- Not previously used for language



from: Beatty & Lucero-Wagoner 2000

Drei Experimente: ICA & Sprache (Demberg et al., 2013)

Can we measure linguistic processing difficulty using the ICA?

Three self-paced reading experiments with pupil size measurement

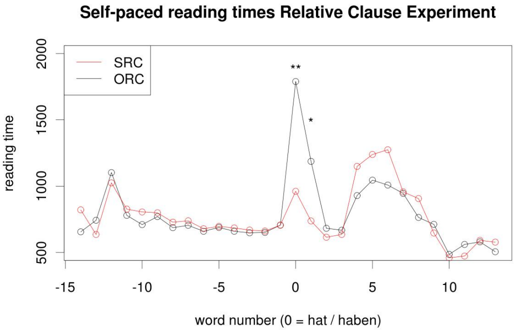
- German subject vs. object relative clause
- Semantic Processing
- Grammar Processing (Gender match / mismatch)

Drei Experimente: ICA & Sprache (Demberg et al., 2013)

Can we measure linguistic processing difficulty using the ICA?

Three self-paced reading experiments with pupil size measurement

- German subject vs. object relative clause



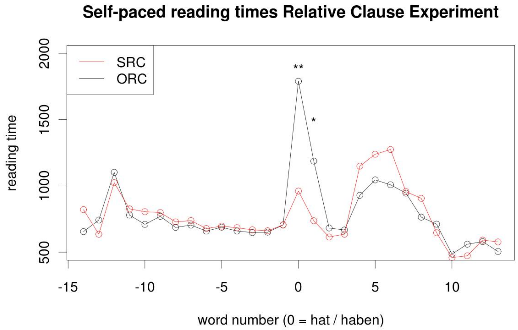
- Semantic Processing
- Grammar Processing (Gender match / mismatch)

Drei Experimente: ICA & Sprache (Demberg et al., 2013)

Can we measure linguistic processing difficulty using the ICA?

Three self-paced reading experiments with pupil size measurement

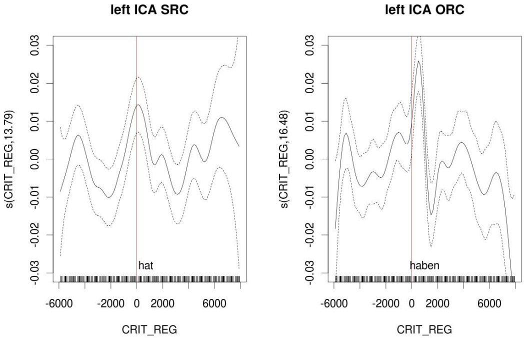
- German subject vs. object relative clause



- Semantic Processing
- Grammar Processing (Gender match / mismatch)

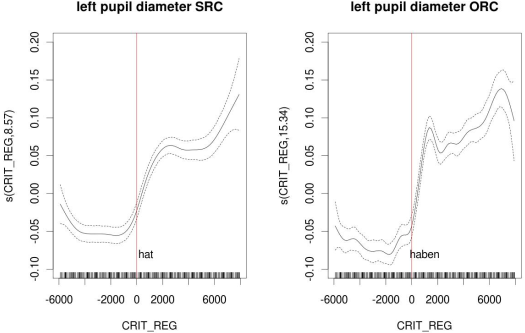
Results: ICA Left Eye

*"Die Lehrerin, die einige Eltern wegen einer solchen Kleinigkeit angerufen **hat/haben**, rief neulich eine Elternversammlung ein."*



	Estimate	Std. Error	t value
(Intercept)	0.831987	0.008271	100.60
Subject RC	-0.013866	0.006414	-2.16

Results: Pupil Size Left Eye



	Estimate	Std. Error	t value
(Intercept)	4.407e-03	2.584e-02	0.171
Subject RC	-8.353e-03	1.770e-02	-0.472
Zeit	7.397e-05	8.398e-06	8.809
Subject RC:Zeit	-3.136e-05	1.181e-05	-2.654

Table of Contents

Experimental Design

Simultaneous driving and language experiment, manipulating

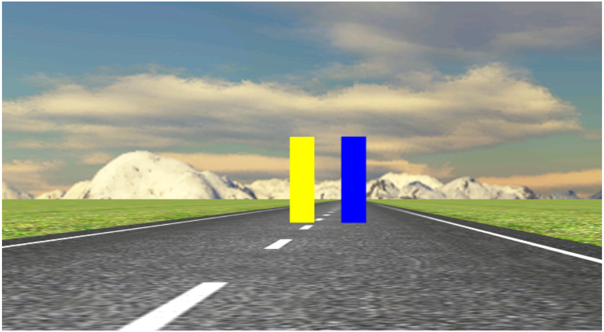
- the difficulty of the driving task
- the complexity of the language

Measure:

- Task difficulty
  - driving $\longrightarrow$ difficulty of driving course
  - language $\longrightarrow$ psycholinguistic measures (e.g., surprisal)
- Task performance
  - driving $\longrightarrow$ events in the driving task (e.g., **steering performance**)
  - language $\longrightarrow$ comprehension tests (reaction time, answer accuracy)
- Cognitive load $\longrightarrow$ pupillometry, skin conductance

(Engonopoulos, Sayeed and Demberg, 2013; Demberg, 2013)

Driving task



- Desktop-based simulator provided by DFKI (Mahr et al., 2012)
- Yellow bar moves at random intervals
- Difficulty manipulation
- Participants control steering object (blue)

Linguistic task

Linguistic stimuli:

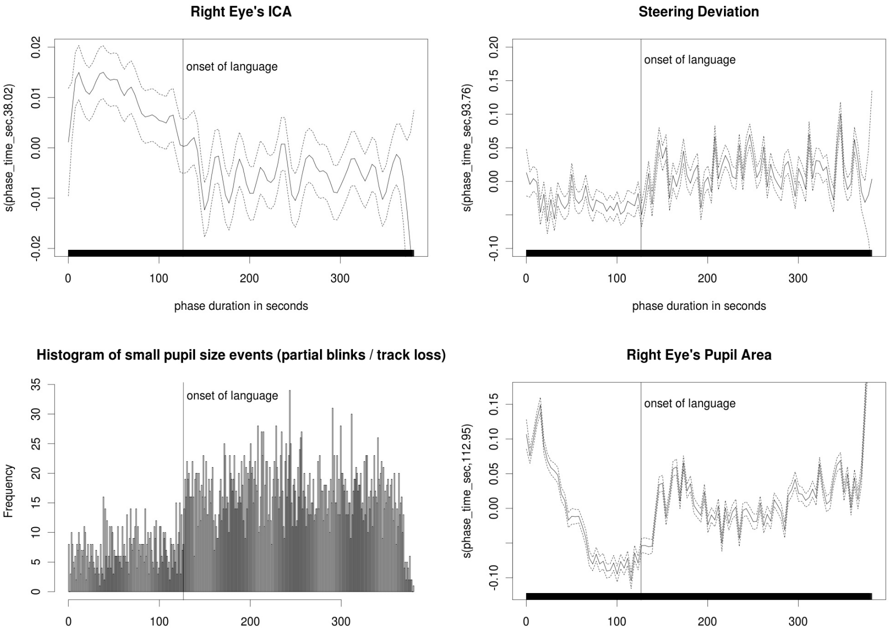
- 40 pairs of German sentences
- loosely based on Bader & Meng (1999)
- local subject-object relative clause ambiguity

Example item

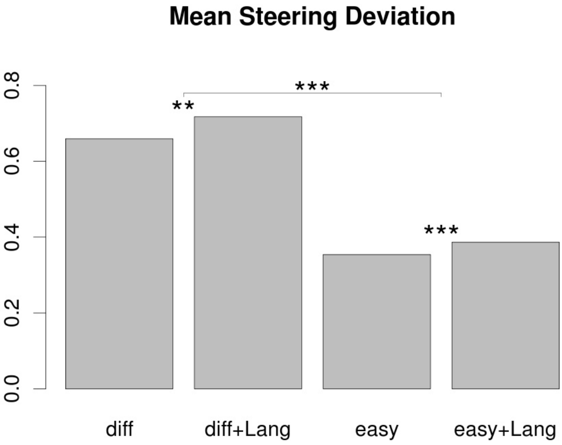
Die Lehrerin, die einige Eltern wegen einer solchen Kleinigkeit angerufen **haben**/**hat**, rief neulich eine Elternversammlung ein.

- Synthesized using MARY TTS
- Critical region forced to be equal by manipulating pause duration.
- Comprehension questions asked (yes/no)

Table of Contents


Right Eye's ICA
onset of language
s(phase_time_sec,38.02)
phase duration in seconds
Steering Deviation
onset of language
s(phase_time_sec,93.76)
phase duration in seconds
Histogram of small pupil size events (partial blinks / track loss)
onset of language
Frequency
Right Eye's Pupil Area
onset of language
s(phase_time_sec,112.95)

How is steering affected by experimental manipulations?



Effect of experimental manipulation

- We find an effect of RC type on ICA measure.
- ... bot not on other measures (pupil size, skin conductance)

	left ICA		right ICA	
	Estimate	t-value	Estimate	t-value
(Intercept)	0.7504	35.71 ***	0.736	37.82 ***
subject RC	**-0.0354**	**-2.12** *		
phase time	-1.16×10^{-7}	-2.59 *		
time wrt. onset	-2.78×10^{-5}	-6.38 ***	-1.84×10^{-5}	-4.36 ***
steering veloc	0.0257	5.37 ***	0.0226	4.88 ***
steering accel	0.0108	2.00 *		
SRC:phase time	1.34×10^{-7}	2.12 *		

Table: Mixed effects regression analysis with left and right ICA as response variable, 100–1800msec after critical region onset. (Critical region duration: 0-600msec)

Steering Accuracy during ambiguous region

Steering performance significantly **worse during ambiguous region**.

	Estimate	t-value	
(Intercept)	3.562e-01	17.07	***
phase time	8.459e-08	3.44	***
target velocity	3.832e-01	205.08	***
critical region	**1.396e-02**	**2.88**	**
easy driving	-2.248e-01	-64.91	***
target acceleration	-2.680e-02	-5.90	***

Table: Mixed effects regression analysis with steering deviation as response variable, for region of 2s before the onset till 2s after end of the critical region.

Discussion

Processing difficulty in dual task setting

- manipulate difficulty of driving task, effect on language processing
- methodological challenges
- fine-grained measures
- robust measures
- first indications that ICA measure might be useful

Conclusions

- Seen evidence for prediction at syntax / semantics level
- Modelling explicit prediction can also help to account for locality effects
- Prediction also occurs at discourse level:
  - based on discourse connectors
  - uniform information density provides an account for implicitness of discourse cues
- Processing difficulty effects in realistic tasks.
- ICA pupillometric measure might be just what we need.

End of Presentation

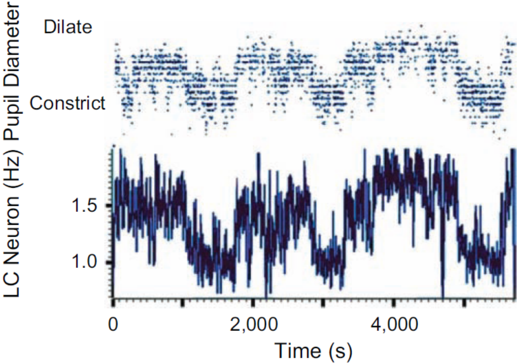
Thank you for your attention!

Thanks to my collaborators and students in Saarbrücken:
Fatemeh Torabi Asr, Judith Köhne, Asad Sayeed,
Nikolaos Engonopoulos, Evi Kiagia

Relation between pupil dilation and cognitive load

Laeng, Sirois, Gredebäck (2012):

- LC neuron activated by stress, engaged during memory retrieval
- LC sends innervations to brain areas involved in selective attention processing
- thought to promote adequate levels of activation for cognitive performance



(Aston-Jones & Cohen, 2005)

Coherence and online reading in a less constraining context

- Discourse connectors facilitate reading and comprehension (when used correctly), while incoherent discourse connectors make reading slower (e.g. Millis & Just 1994, Murray 1997)
- Prior work compared major category violations
- Our goal: investigate further the possible facilitatory effect (and its time course) of minimally different discourse cues.

Cue	but	although
COMPARISON(contrast)	2419	157
COMPARISON(pragmatic contrast)	28	0
COMPARISON(concession)	494	153
COMPARISON(pragmatic concession)	0	0
COMPARISON	260	0
EXPANSION(conjunction)	77	0
OTHER	1	0

- *but* and *although* different type of concession (expectation vs. contra-expectation), but couldn't find a pair with same exact meaning but different distribution.