# (Mis-)understanding

Seminar: Hot & Odd Topics in Semantics

*David Meier*

16.07.2018

# Content

# Motivation

- when agents (speakers/listeners, writers/readers) engage in discourse, they typically have a *goal* (ordering a drink, asking for information...)
- an *utterance* is an attempt to reach this goal
- a *misunderstanding* happens when one of the agents fails to convey his goal to his partner, or if the partner on the other hand fails to interpret an utterance of the speaker correctly (given the original intention of the speaker)

# Motivation

- A misunderstanding may happen on an *auditory* or an *intentional* level:
- *auditory:* the listener could not hear what has been said
- *intentional:* the listener did not interpret the meaning of an utterance correctly

# Meaning

Where does the meaning of a text lie?

What is the *locus* of meaning?

# History

Traditional answers:

- **Writer/Speaker**: A text means whatever it's author intended it to mean
- **Reader/Listener**: Meaning is defined by the recipient's response
- **Text**: Meaning lies in the text itself (*conduit view*)

# Meaning

Non-verbal / non-textual actions can affect meaning and infer knowledge:

**Event**: Nadia says "I'm going to take the garbage bin out."
**Circumstance**: The garbage collectors come Wednesday's.
**Conclusion**: Nadia assumes it's Wednesday.

# History

Rising importance and availability of large collections of text
$\rightarrow$ In text understanding, the "conduit view" is the modern one

*"Text is to be processed, not understood."*

# Detection of misunderstanding

How to detect misunderstanding?

$\rightarrow$ Occurence might be hypothesized by a listener if he cannot interpret the text given, or other party responds in an unexpected fashion

# Example

**A**: When is the dinner tonight?
**B**: Is it at 7:30?
**A**: I'm asking you.
**B**: Oh, I don't know.

# Detection of misunderstanding

Occurence might be hypothesized by a listener if he cannot interpret the text given, or other party responds in an unexpected fashion.

→ listener/speaker perspective (world knowledge, knowledge about the previous discourse...) is important
→ can misunderstanding be detected on a purely textual level?

# Detection of misunderstanding

How do we detect and prevent misunderstandings in text?

**Automatic Identification of Nocuous Ambiguity**: pre-evaluate risk of misunderstanding (vs. recognizing it retrospectively) using the concept of **Nocuity**.

# Inhalt

# Overview

**Ambiguity:** text that does not have single meaning, but allows for multiple interpretations

**Nocuity:** Ambiguity is **nocuous** if actual *readers* have varying interpretations of its meaning. Otherwise, it's **innocuous**.

# Examples

*"Hans gave Peter a present. He was very happy about it."*

# Examples

*"Hans gave Peter a present. **He** was very happy about it."*

(Referential) Ambiguity: *He* can refer to both *Hans* and *Peter*

***innocuous:*** A human reader can easily identify the correct reference

# Examples

*"Hans gave Peter a present. He thought it would be a surprise."*

# Examples

*"Hans gave Peter a present. **He** thought it would be a surprise."*

(Referential) Ambiguity: *He* can refer to both *Hans* and *Peter*

***innocuous:*** ditto

# Examples

*"Hans gave Peter a present. He put it on the table."*

# Examples

*"Hans gave Peter a present.* ***He*** *put it on the table."*

(Referential) Ambiguity: *He* can refer to both *Hans* and *Peter*

***nocuous:*** purely subjective, the context doesn't give enough information

# Introduction

Previous work assumes there is a "correct" or "intended" Disambiguation

*But:* that is not always the case. Often, there is no consensus about the meaning of a text.

→ possible cause for misunderstanding?

**Assumption:** the more nocuous an ambiguity is, the more likely it is to lead to misunderstanding

# Identifying nocuous ambiguities

We distinguish between *syntactic* and *interpretative* ambiguities

***Syntactic ambiguity*** is caused by the structure of the text (examples)

***Interpretative ambiguity*** is caused by the reader/listener (e.g. homonyms)

# Human judgement

*Nocuous* if multiple readers give various interpretations
$\rightarrow$ let human judges read and evaluate text passages

***Certainty:*** percentage of people who assign corresponding meaning to text

***Ambiguity Threshold $\tau$:*** models tolerance of Nocuity (cookbook vs. safety manual)

# Human judgement

***Coordination ambiguity***

In a coordinative environment (*and*, *or*...), do modifiers apply to just one or both/multiple conjuncts?

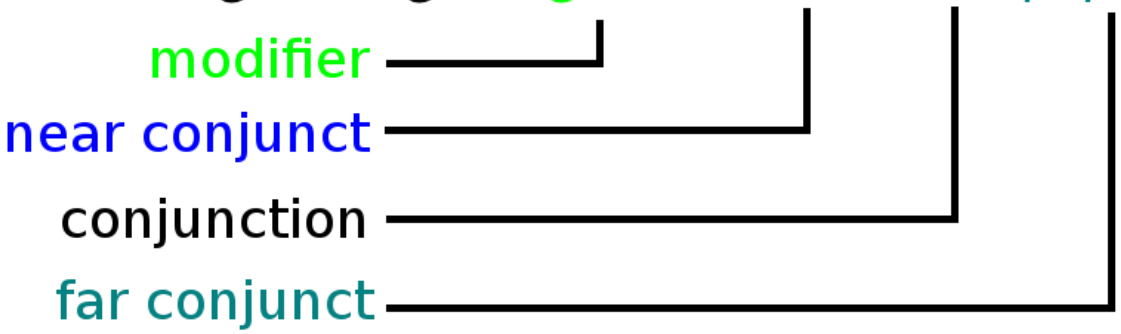
*"He spent all night reading through linguistic books and papers."*

# Human judgement

***Coordination ambiguity***

In a coordinative environment (*and*, *or*...), do modifiers apply to just one or both/multiple conjuncts?

*"He spent all night reading through linguistic books* ***and*** *papers."*



# Human judgement

*"He spent all night reading through linguistic books **and** papers."*

Possible modifier attachment:

*low attachment* (LA):
→ **and** ( linguistic ( books ) , papers )

*high attachment* (HA):
→ linguistic ( **and** ( books , papers ) )

# Human Judgement

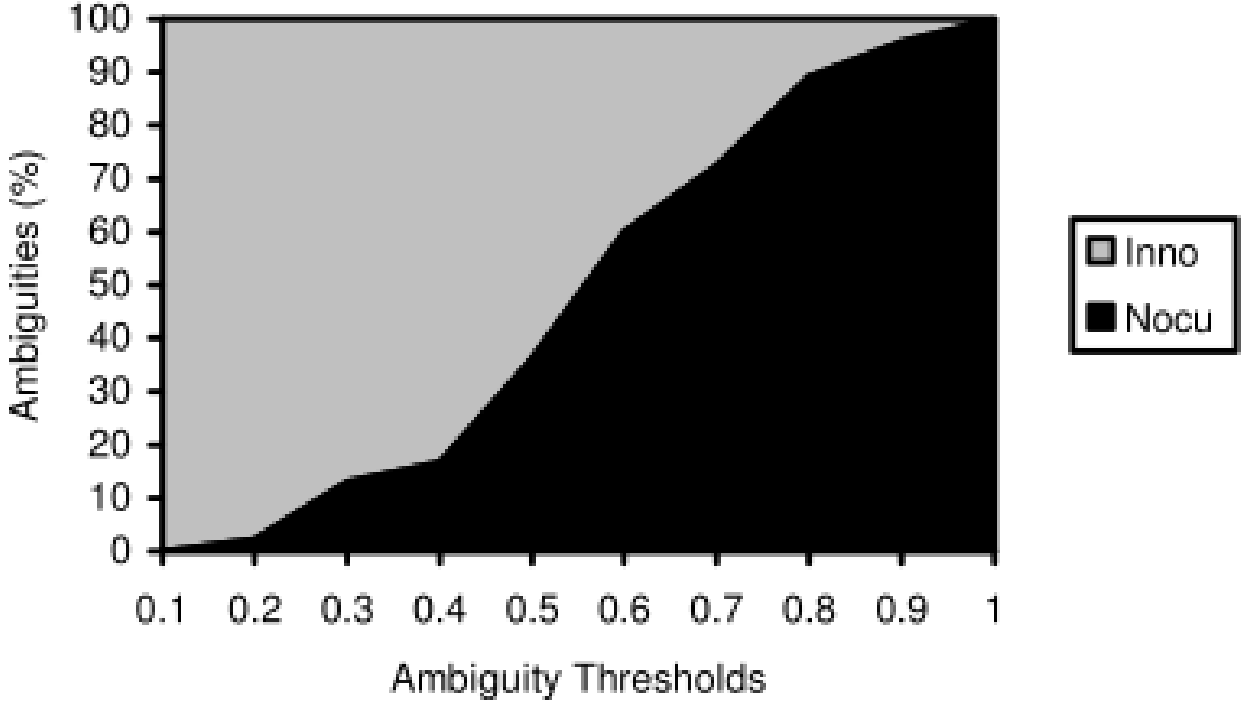
- 138 instances of coordination ambiguities
- 17 computing professionals (academic staff and students)
- High Attachment, Low Attachment or Ambiguous
- Ambiguity Threshold $\tau = 0.7$
- If the certainty of neither HA and LA is greater than $\tau$, it is considered to be **nocuous**

# Human Judge Results

| | Judgments | | |
|---|---|---|---|
| | HA | LA | A |
| (a) *security* and *privacy* requirements | 12 | 1 | 4 |
| (b) electrical *characteristics* and *interface* | 4 | 4 | 9 |

Table 1. Judgment count for the sample instances (HA=high attachment; LA=low attachment; and A=Ambiguous)

# Ambiguity Threshold



# Heuristics

**Heuristic:** simplified/"relaxed" problem that is far easier to compute, but still gives a reliable answer that points us in the correct direction

We use 5 heuristics that can help us:

# Heuristics

- **Coordination matching**
  If both conjuncts are often found coordinated in the dataset we use, this indicates HA
- **Distribution similarity**
  If both conjuncts are often found in the same context, this indicates HA
- **Collocation frequency**
  If the modifier is more often found with the near conjunct than the far conjunct, this indicates LA
- **Morphology**
  If the conjuncts have the same morphological marker (e.g. plural or tense suffix), this indicates HA
- **Semantic similarity**
  If the conjuncts are semantically similar (i.e. are structurally closely connected in WordNet), this indicates HA

# Anaphora ambiguity

- 200 anaphora instances, mostly 3rd person pronouns (examples)
- 30 professional judges
- Similar procedure, but different shape of the output, as now we cannot reduce it to a binary model

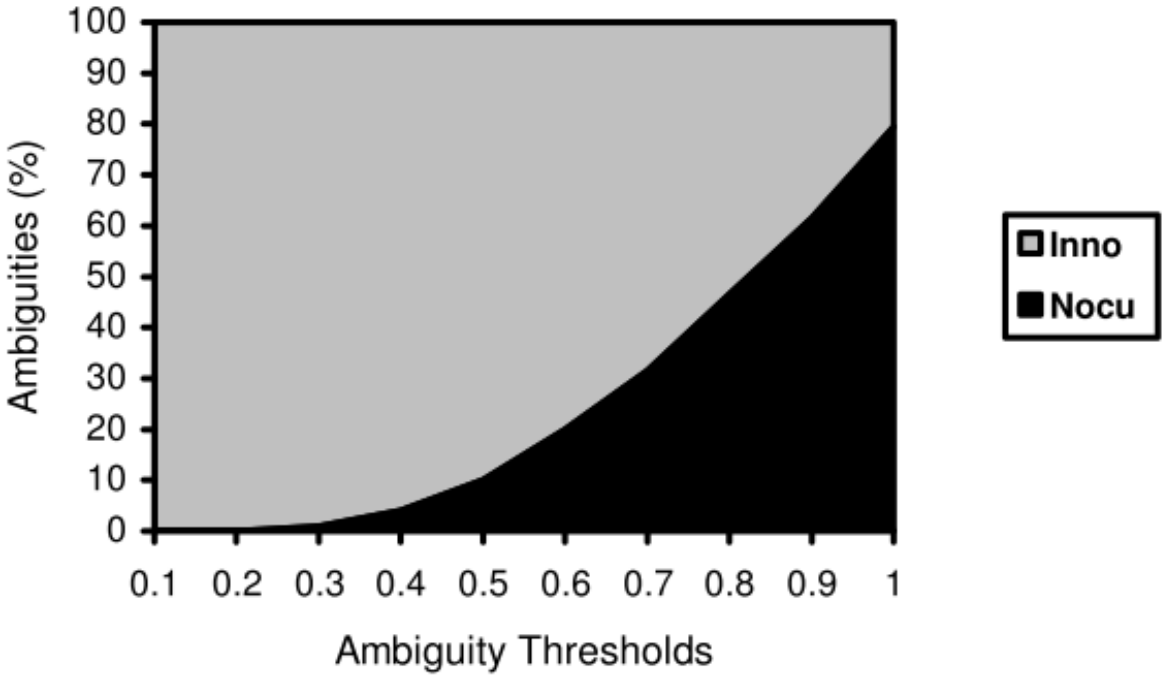
# Anaphora ambiguity

Instead of a binary decision HA vs. LA, judges now decide to which antecedent the ambiguous anaphora should resolve to

# Human Judge Results

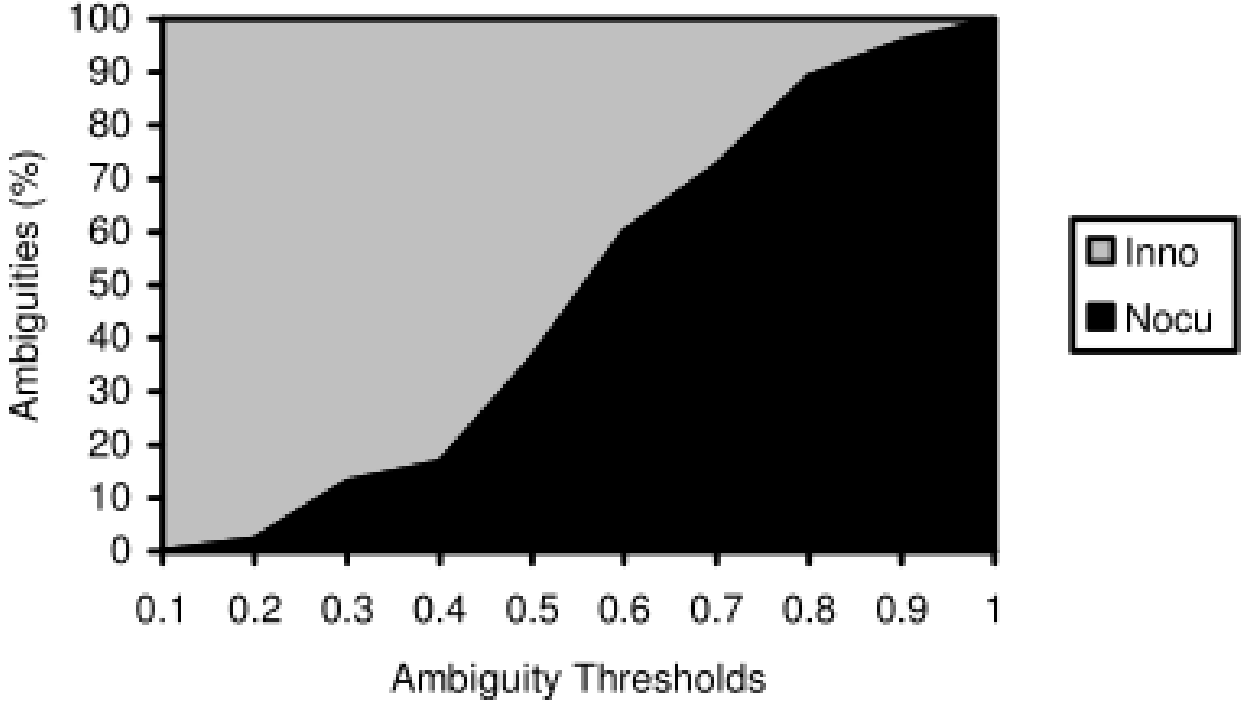
| 1. Supervisors may only modify tasks **they** supervise to the agents they supervise. | | |
|---|---|---|
| | Response Percent | Response Count |
| (a) supervisors | 92.3% | 12 |
| (b) tasks | 7.7% | 1 |

$\rightarrow$ innocuous given $\tau <= 0.9$




→ anaphoric ambiguities less likely to lead to misunderstandings

# Ambiguity Threshold



# Anaphora Heuristics

- Using a variety of heuristics that favor certain antecedents
- The classifier assigns weights to each candidate antecedent
- These weights are used to decide whether or not the ambiguity is defined as nocuous

# Anaphoric Ambiguity

| 2. Testing performed to demonstrate to the acquirer that a CSCI system meets **its** specified requirements. | | |
|---|---|---|
| | Response Percent | Class Label |
| (a) Testing | 0% | N |
| (b) the acquirer | 16.7% | N |
| (c) a CSCI system | 83.3% | Y |

| 1. The LPS operational scenarios represent sequences of activities performed by operations personnel as **they** relate to the LPS software. | | |
|---|---|---|
| | Response | Label |
| (a) the LPS operational scenarios | 33.3% | Q |
| (b) sequences of activities | 66.7% | Q |
| (c) activities | 0% | N |
| (d) operations personnel | 0% | N |

# Weak Thresholds

Antecedent preference reflects a spectrum rather than binary values, so they also add a *weak positive threshold* $W_Y$ and a *weak negative threshold* $W_N$ set at 0.5 and 0.4. $\rightarrow$ act as indication for a false negative for the Q label

# Procedure

- one **Y**, no **Q**, at least one **N**;
- no **Y**, one **Q**, at least one (non-weak) **N**;
- one (non-weak) **Y**, any **Q** or **N**;

→ ***innocuous***

any other:
→ ***nocuous***

# Coordination Results

**Baselines:**

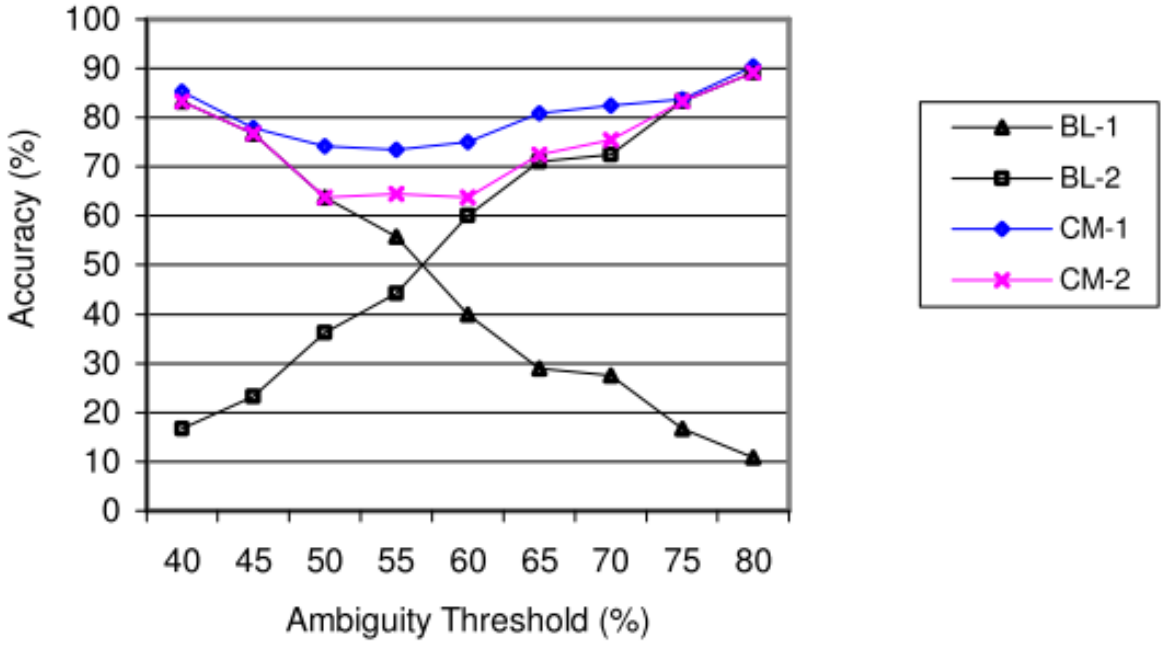
**BL-1**: assumes all ambiguity to be innocuous

**BL-2**: assumes all ambiguity to be nocuous

**CM-1**: classifier results

**CM-2**: results of a prior project

# Coordination Results



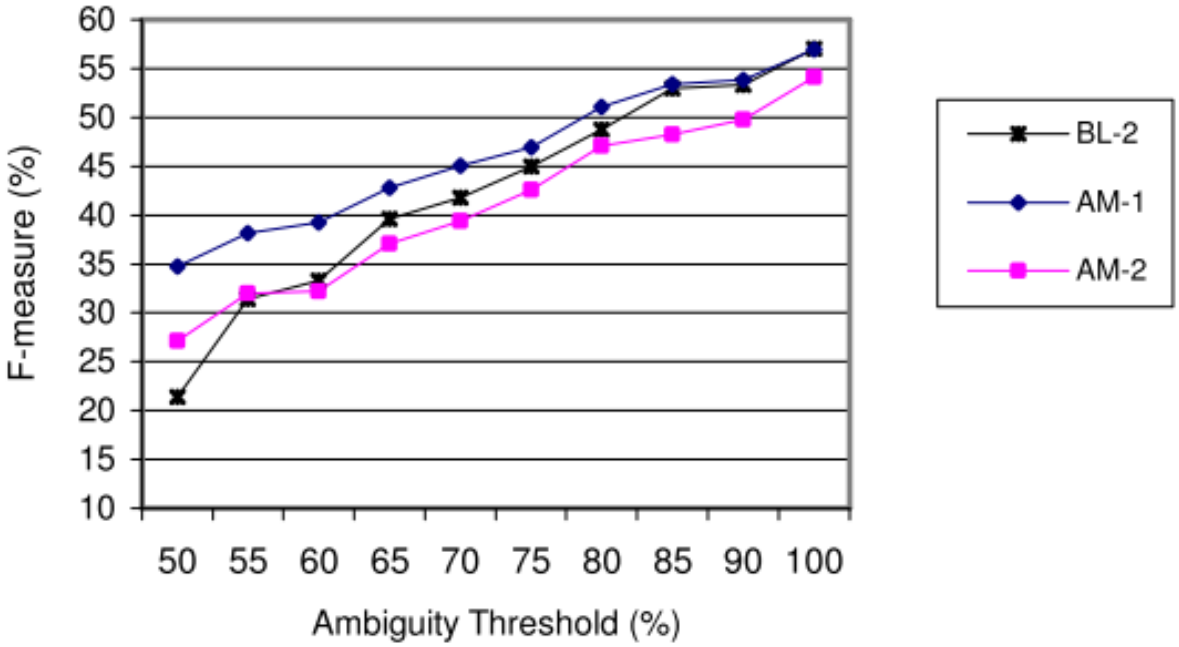

$\rightarrow$ accuracy $>= 75\%$ on all thresholds $\rightarrow$ at high thresholds, no improvement over the baselines

# Anaphora Results

**AM-1**: with weak thresholds

**AM-2**: without weak thresholds

# Anaphora Results




→ model with weak thresholds produces best F-measure → but lower accuracy then AM-2 due to more false positives (innocuous falsely classified as nocuous)

# Result Interpretation

- some ambiguities are more likely to lead to misunderstanding between human judges (coordination vs anaphora)
- these ambiguities are more *nocuous*
- using heuristics, it is possible to automatically determine nocuity with relatively decent results
- this approach can be used and expanded to systems that know when it is likely for a misunderstanding to occur, which may help to repair these in the future

# Critical Remarks

- judges were tasked to judge sentences *without context*
- language competence affects perceptions of ambiguity
- only a small subset of possible ambiguities was looked at
- different heuristics for different forms of ambiguity
- role of the ambiguity thresholds and their values remain unclear
- knowledge about the reader was not at all used (humans merely as a "classification machine")

# Conclusion

- Nocuity plays a clear role in misunderstanding.
- **BUT:** ambiguity/nocuity is only one of many factors that can lead to misunderstanding, and even then we only discussed two forms of ambiguity

→ more research needs to be done

# Sources

- **Identifying Nocuous Ambiguities in Natural Language Requirements** *(Chantree et al.)*
- **The Future of Text-Meaning in Computational Linguistics** *(Graeme Hirst, 2008)*
- **A Methodology for Automatic Identification of Nocuous Ambiguity** *(Yang et al., 2010)*