Computational Semantics and Pragmatics

Autumn 2011

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

Overview

Things we have seen up to now:

- textual entailment – and tools for automated reasoning
- distributional semantic models
  - ∗ main practical aspects involved in building them
  - ∗ theoretical implications
- word sense disambiguation
  - ∗ supervised and unsupervised machine learning methods for classification tasks

Coming days: towards linguistic interaction.
Today:

- intro to Gricean pragmatics and conversational implicature
- computational explorations of implicature: generation of referring expressions

Gricean Pragmatics

When we use language, we very often mean more than what we literally say:

(1) A: Are you going to Paul's party?
B: I have to work.
$\rightsquigarrow$ *I am not going.*

- B *implies* that she's not going to the party *without saying it.*
- This enrichment of the literal meaning is not a logical implication or entailment of B's utterance – it depends on features of the conversational context $\rightarrow$ *conversational implicature*
- Grice proposes that conversational implicatures can be systematically accounted for by a set of general rationality principles for the efficient and effective use of language in conversation.

The CP and the Maxims

The Cooperative Principle: Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

- Maxim of Quality: be truthful
  - ∗ Do not say what you believe to be false.
  - ∗ Do not say that for which you lack adequate evidence.
- Maxim of Quantity:
  - ∗ Make your contribution as informative as is required (for the current purposes of the exchange).
  - ∗ Do not make your contribution more informative than is required.
- Maxim of Relation: be relevant
- Maxim of Manner: be perspicuous.
  - ∗ Avoid obscurity of expression / Avoid ambiguity.
  - ∗ Be brief / Be orderly.

Grice's point is not that we adhere to these maxims on a superficial level, rather that we interpret utterances assuming that the principles are being followed at some deeper level, often contrary to appearances.

Conversational Implicature

Theoretical definition:

Speaker A conversationally implicates q with utterance U iff:

(i) A believes that it is common knowledge between A and the hearer B that A is following the CP;

(ii) In order to maintain that (i) is the case given U, B must assume that A believes q;

(iii) A believes that (ii) is common knowledge between A and B.

Calculability: the hearer B must be able to calculate implicatures based on the literal content of what was said, its context, and the specific assumptions encoded by the CP and the maxims.

B must know the following facts:

1. the conventional content of what was said (U)
2. the CP and the maxims
3. the context of U (e.g. its relevance)
4. some background information (e.g. that U is blatantly false)
5. that (1-4) are mutual knowledge between A and B.

Types of Implicature

Grice distinguishes three main types of conversational implicature:

- The speaker is directly observing the maxims
- The speaker violates a maxim that clashes with another one
- The speaker is openly flouting a maxim to exploit it

Example 1

The speaker is directly observing the maxims:

Kyle to Ellen: "I have \$9"
Implicature: *Kyle does not have \$10.*

a. Contextual premise: Both Kyle and Ellen need \$10 for their movie tickets.
b. Contextual premise: It is mutual, public information that Kyle has complete knowledge of how much money he has on him
c. Assume Kyle is cooperative at least insofar as he is obeying Quantity and Quality.
d. Then he will assert what is maximally relevant, informative, and true.
e. By (a), "I have \$10" is more informative and relevant in this context than "I have \$9."
f. Therefore, Kyle must lack sufficient evidence to assert "I have \$10".
g. By contextual premise (b), he must lack evidence "I have \$10" because it is false.

Examples from Chris Potts (with some adaptations)

http://compprag.christopherpotts.net/implicature.html

Example 2

The speaker violates a maxim that clashes with another principle:

A: In which city does Kim live?
B: She lives somewhere in Spain.
Implicature: *B does not know which city Kim lives in.*

a. Contextual premise: B is forthcoming about Kim's personal life.
b. Assume B is cooperative.
c. Assume, towards a contradiction, that B does know which city Kim lives in (the negation of the implicature).
d. Supplying the city's name would do better on Relevance and Quantity than supplying just the country name.
e. The contextual assumption is that B will supply such information.
f. This contradicts the cooperativity assumption (b).
g. We can therefore conclude that the implicature is true.

Examples from Chris Potts (with some adaptations)

http://compprag.christopherpotts.net/implicature.html

Example 3

The speaker is openly flouting a maxim to exploit it:

A newspaper review of a newly opened play says that, in the third act, “Soap opera star Rose Singer produced a series of sounds corresponding closely to the score of an aria from Rigoletto.”
Implicature: *the reviewer believes that Rachel Singer’s performance was not good.*

a. Assume the reviewer is cooperative.
b. There is a shorter form, *sang*, competing with *produced a series of sounds corresponding closely to the score of*.
c. The use of the unusual form violates Manner and indicates that the performance must have been unusual somehow.
d. If it was unusually good, the reviewer would have said so directly, by cooperativity.
e. If it was unusually bad, the reviewer would have said so directly, unless politeness was preventing him from doing so.
f. Therefore, we conclude that the performance was unusually bad.

Examples from Chris Potts (with some adaptations)

http://compprag.christopherpotts.net/implicature.html

Cancelability of Implicature

One of the defining properties of implicatures is that they are defeasible inferences – they are cancelable: a potential implicature can be denied or suspended, directly or via background contextual assumptions or later clarifications.

(2) A: Was the movie good?
B: It was outstanding!
$\leadsto$ *Yes*

(3) A: Are you going to Paul's party?
B: I have to work.
$\leadsto$ *No*

- In (3), B's utterance conveys "No" though "No" is not explicitly encoded. This is an implicature that follows from conversational principles and it's cancelable (B could continue by saying *"but I'll try to make."*)
- In (2), B's utterance conveys "Yes" though "Yes" is not explicitly encoded. However this is not cancelable (*"It was outstanding but it was bad"* is contradictory) and therefore the inference is an entailment.

$\Rightarrow$ *Next week we'll look into indirect question-answer pairs*

Computational Exploration of CI?

Grice's proposals were brief and only suggestive of how work on the underlying ideas may proceed.

Work has indeed proceeded in several directions:

- Formal pragmatics: neo-gricean approaches, relevance theory, ...
- Experimental pragmatics: what do speakers/hearers actually do?
- Computational pragmatics: can we account computationally for some phenomena that seem related to conversational implicature?

As mentioned, today we'll look at the generation of referring expressions (GRE), going over a classic paper in the field:

Dale & Reiter (1995) Computational interpretation of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

Natural Language Generation

NLG is a subfield of Natural Language Processing. At an abstract level, one can think of it as the reverse of the process of Natural Language Understanding (NLU):

- NLU: Mapping human language into non-linguistic representations.
- NLG: Mapping non-linguistic representations of information into human language.

Some of the problems NLG is concerned with are the following:

- What constitutes "appropriate" language in a given communicative situation? How can the relevant pragmatic, semantic, syntactic, and psycholinguistic constraints be formalised?
- What is the best way for a system to communicate information to a human? What kind of behaviour does a person expect from a computer and how can it be implemented?

These issues have no obvious counterpart in NLU research.

Generation of Referring Expression

GRE is concerned with the production of linguistic expressions that enable the hearer to identify one or more entities in a given context.

GRE is an issue for NLG because the same entity may be referred to in many different ways.

There are two main aspects that need to be taken into account:

- Content determination: deciding what information should be communicated – which properties of an entity should be used in describing it?
- Lexicalisation or lexical realisation: how should properties be realised and with what kind of structures?

Content determination vs. Lexicalisation

Possible properties for content determination	Possible lexicalisations
`function=lamp, shape=teacup`	*the teacup lamp*
	the lamp that's a teacup
`shape=teacup, position=upsidedown`	*the upside down teacup*
	the teacup that's upside down

Dale & Reiter 1995

D&R focus on referring expressions that:

- are realised as definite NPs - but they focus on content determination
- refer to physical objects
- their communicative goal is solely to identify a target object

the black dog	the woman with the glasses	the upside-down cup

D&R focus on three criteria an algorithm for GRE should satisfy:

1. it should produce expressions that satisfy the communicative goal: that allow the hearer to identify the intended object
2. it should produce expressions that do not lead the hearer to derive false implicatures
3. it should be computationally efficient

Satisfying the Communicative Goal

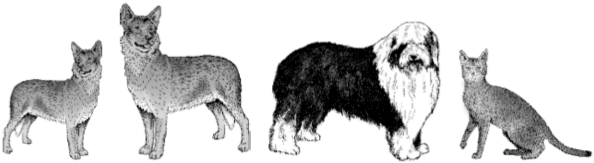
A referring definite description satisfies its communicative goal if it is a distinguishing description.

- Let D be the set of entities that are in the focus of attention of speaker and hearer (the context set).
- Each entity in D is characterised by means of a set of properties or attribute-value pairs such as $\langle \texttt{colour}, \texttt{red} \rangle$ or `colour=red`.
- If a property p does *not* apply to an entity $d \in D$, we say that p rules out d.
- Let $r \in D$ be the target referent, and C the contrast set: the set of all elements in D except r.

A set L of properties is a distinguishing description if the following conditions hold:

C1. Every property in L applies to r.

C2. For every $c \in C$, there is at least one property in L that rules out c.



$D = \{d_1, d_2, d_3, d_4\}$	Knowledge base representing the scene:
$r = d_1$	d_1 : `type=dog, size=small, color=brown`
$C = \{d_2, d_3, d_4\}$	d_2 : `type=dog, size=large, color=brown`
	d_3 : `type=dog, size=large, color=black+white`
	d_4 : `type=cat, size=small, color=brown`

Some examples of possible descriptions in this scenario:

content determination	possible realisation	distinguishing
$L = \{$`type=dog,size=small`$\}$	*'the small dog'*	✓
$L = \{$`type=dog,colour=brown`$\}$	*'the brown dog'*	×
$L = \{$`type=dog,size=small,colour=brown`$\}$	*'the small brown dog'*	✓

Are all distinguishing descriptions equally felicitous or appropriate?

Appropriateness Conditions

REs that are *distinguishing* in the sense defined above satisfy a condition of sufficiency – they provide enough information to pick out the intended referent

But one can think of other appropriateness conditions:

- Efficiency: the RE does not provide more information that is needed to satisfy the communicative goal ($\rightsquigarrow$ is not redundant).
- Sensitivity: the RE does not make use of properties of the target referent that the addressee may not be able to determine ($\rightsquigarrow$ takes into account the common ground).
- Lexical preference: the RE makes use of preferred (basic-level) lexical classes whenever possible.

Sufficiency and efficiency are concerned with the informational aspect of referring expressions ($\rightsquigarrow$ Quality, Quantity, Brevity?)

Sensitivity is related to the concept of *audience design*.

Maxims for GRE

In the context of GRE, the Gricean maxims can be reinterpreted as:

- Quality: an RE must be an accurate description of the target referent.
- Quantity: an RE should contain

  Q1: enough information to enable the hearer to identify the target

  Q2: no more information than required.
- Relevance: an RE should not mention attributes that have no discriminatory power
  - *sensitivity?*
  - *lexical preference?*
- Manner (Brevity): an RE should be short whenever possible

By definition, a distinguishing description satisfies Quality & Q1 - sufficiency

Q2 and Brevity are related to efficiency (there is some duplication...).

Sensitivity and lexical preference might be related to Relevance, but may be considered orthogonal constraints.

Computational Interpretations of the Maxims

D&R95 present three algorithms for GRE that differ essentially in their interpretation of Q2 / Brevity:

1. Full Brevity
2. Greedy Heuristic
3. Local Brevity
4. The Incremental Algorithm

- Full Brevity interprets efficiency (i.e. brevity, Q2,...) literally.
- Greedy Heuristic and Local Brevity are computationally tractable approximations to Full Brevity.
- The Incremental Algorithm attempts to mimic human behaviour, without direct use of brevity.

Computational Efficiency

How computationally costly are these GRE algorithms?

Parameters to measure computational complexity ($\approx$ *the time or steps it may take the algorithm to produce a solution*)

- n : the number of elements in the domain
- n_d: the number of distractor elements given a target
- n_a: the number of properties known to be true of the target referent
- n_l: the number of properties used in the final description

Full Brevity: Generating Minimal Descriptions

According to the FB interpretation of Q2, an RE is optimal if it is *minimal* – the shortest possible description that is distinguishing.

- The algorithm discussed does an exhaustive search:
  - for all properties of the target referent (n_a), it first tries to generate a distinguishing description using only one property; if this fails, it considers all possible combinations of two properties, and so on.
  - The run-time grows exponentially ($\approx n_a^{n_l}$)

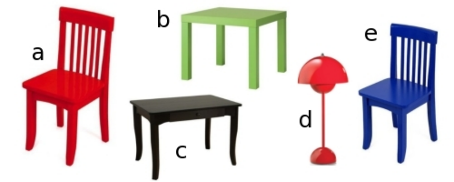
Two problems with this strict interpretation:

- computationally very costly (NP hard) and hence not feasible
- psychologically unrealistic since humans do not always produce minimal descriptions.

What do people do?

- Humans often include "unnecessary" modifiers in REs. For instance, in the example below, where `d` is the target, the property `colour=red` seems redundant. However:
  - ∗ in itself it has discriminatory power (it rules out some elements in the contrast set, those that are not red)
  - ∗ including it may help the hearer in their search

'the red lamp'



- Eye-tracking experiments show that humans start producing REs before scanning the scene completely: they produce REs incrementally without backtracking

The Incremental Algorithm

- Dale & Reiter (1995) present the incremental algorithm, which has become a sort of standard in the field.
- The algorithm relies on a list of preferred attributes, e.g. $\langle \texttt{colour}, \texttt{size}, \texttt{material} \rangle$
- The assumption is that for each domain we can identify a set of attributes that are conventionally useful to produce REs, because of previous usage, perceptual salience, etc.
- The algorithm iterates through this domain-dependent list of preferred attributes
  - it adds a property to the description if it rules out any distractors not yet ruled out
  - it terminates when a distinguishing description is found.
- It may produce "redundant" descriptions, but since the extra attributes are used by speakers (and hence are expected) they should not lead to false implicatures

The Incremental Algorithm - simplified

Let:

- r be the target referent;
- A be the set of properties $a{=}v$ that characterise r;
- C be the set of distractors (the contrast set);
- $\texttt{RulesOut}(a{=}v)$ be the subset of C ruled out by property $a{=}v \in A$;
- P be an ordered list of task-dependent *preferred attributes*; and
- L be the set of properties to be realised in our description.

$\boxed{\texttt{MakeReferringExpression}(r, C, P)}$

$L \leftarrow \{\}$
for each member a_i of list P **do**
  if $\texttt{RulesOut}(a_i{=}v) \neq \emptyset$ (for some $a_i{=}v \in A$)
  then $L \leftarrow L \cup \{a_i{=}v\}$
    $C \leftarrow C - \texttt{RulesOut}(a_i{=}v)$
  endif
  if $C = \{\}$ **then**
    if $\{\texttt{type=}v\} \in L$ (for some value v such that $\texttt{type}{=}v \in A$)
    then return L
    else return $L \cup \{\texttt{type=}v\}$
    endif
  endif
return failure

The Incremental Algorithm - in words

In the previous slide, you have a simplified version of the Incremental Algorithm in pseudo-code. Here are the steps in words:

- We start with an empty description (an empty L)
- We then go through the attributes in the list of preferred attributes P, starting with the first attribute in the list.
  - ∗ We select the property of the target referent that has to do with the attribute we are dealing with. If it rules out some elements in the contrast set, then
    - ▶ we add that property to L, and
    - ▶ substract from the contrast set the elements that have been ruled out
  - ∗ If the contrast set is empty, then we are done. But we still want to make sure the attribute `type` is in there because we need a head noun for the description. So:
    - ▶ if a property with attribute `type` is in L, we are indeed done;
    - ▶ if not, we add it to L and are also done.

An Example

To see the impact of the order of attributes in the list of preferred attributes P, consider our earlier example scenario:

Knowledge base representing the scene:

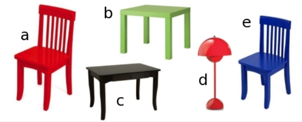
d_1 : `type=dog, size=small, color=brown`
d_2 : `type=dog, size=large, color=brown`
d_3 : `type=dog, size=large, color=b+w`
d_4 : `type=cat, size=small, color=brown`

$r = d_1$
$A = \{\texttt{type=dog,size=small,color=brown}\}$
$C = \{d_2, d_3, d_4\}$

- If $P = \langle \texttt{colour}, \texttt{size}, \texttt{type} \rangle$ then $L = \{\texttt{colour=brown, size=small, type=dog}\}$
- If $P = \langle \texttt{size}, \texttt{type}, \texttt{colour} \rangle$ then $L = \{\texttt{size=small, type=dog}\}$
- If $P = \langle \texttt{type}, \texttt{size}, \texttt{colour} \rangle$ then $L = \{\texttt{size=small, type=dog}\}$
- If $P = \langle \texttt{type}, \texttt{colour}, \texttt{size}, \rangle$ then $L = \{\texttt{colour=brown, size=small, type=dog}\}$

A Possible Caveat

In our earlier example, where
$r = d$ and $C = \{a, b, c, e\}$:



If $P = \langle \texttt{type}, \texttt{colour} \rangle$ then $L = \{\texttt{type=lamp}\}$
If $P = \langle \texttt{colour}, \texttt{type} \rangle$ then $L = \{\texttt{colour=red, type=lamp}\}$

- Recall that we are dealing with content determination, not with lexicalisation / realisation!
- In principle, content determination should be language independent
- But the same attributes do not have the same discriminatory power during processing in languages with different word order

  $L = \{\texttt{colour} = \texttt{red}, \texttt{type} = \texttt{lamp}\} \rightsquigarrow$ *the red lamp / la làmpara vermella*

  In Catalan, the post-nominal adjectve does not seem to reduce the search effort for the hearer as much as in English, so presumably it would be added less often. . .

What's Next

- HW#3 (short exercise) will be available later today
  Due next week before class: Thursday 17 Nov, 13:00h.
- Required reading for next week:

de Marneffe, Manning & Potts (2010) "Was it good? It was provocative." Learning the meaning of scalar adjectives. In *Proceedings of ACL 2010*, pp. 167-176.

- Check information about final project on the website.