Computational Semantics and Pragmatics

Autumn 2013

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam

Yesterday

- NLG and GRE are about making choices to satisfy a communicative goal concerning *what* to say and *how* to say it.
  - content determination
  - linguistic realisation
- Gricean pragmatics: the maxims of conversation are formulated as directives for the speaker ⇝ relevant for NLG
  - cooperative speakers adhere to the maxims (in non trivial ways)
  - the maxims and general adherence to them are common knowledge
  - this leads to special inferences called *implicatures*
- Dale & Reiter (1995) investigate the impact that the maxims have on content determination for GRE: computational interpretations of the maxims
  - complexity of algorithms implementing different interpretations
  - descriptive adequacy (what people actually do) of different interpretations ⇝ *NLG as a modelling human production*

Terminology and Task Definition

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).
- Let $r \in D$ be the target referent, and $C \subset D$ the contrast set: the set of all elements in D except r.
- 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 it has discriminatory power and that it rules out d.

At the content determination stage, a description can be modelled as a set L of properties. L is a distinguishing description iff:

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.

The Maxims in the Context of GRE

- 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 ($\approx$ Q2)
  - are not available to the hearer
- Manner (Brevity): an RE should be short whenever possible ($\approx$ Q2)

Sit by *the brown wooden table*.

Assuming that (1) the communicative goal is exclusively to single out the referent and (2) all the maxims are followed, several implicatures are licensed:

$\leadsto$ *there are other objects that are not brown / wooden* (Relevance)
$\leadsto$ *there is at least one other table that is not brown and wooden* (Q2)

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 Q2 / brevity (efficiency) 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 ($\approx n_a^{n_l}$)

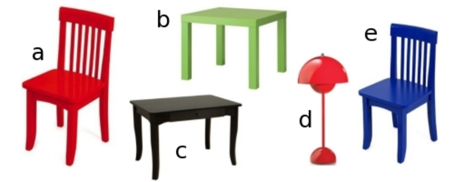
Two problems with this strict interpretation:

- 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.

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.

The IA and the Maxims

The IA is computationally efficient and can produce non-minimal descriptions.

- the latter point is in accordance to human behaviour
- what does this tell us about the Maxims? why do some "overspecified" descriptions not lead to false implicatures?

Quantity: a referring description should contain

- Q1: enough information to enable the hearer to identify the target
- Q2: no more information than required.

Extensions of D&R95's Approach

This approach to GRE relies on a number of simplifying assumptions, which more recent research has tried to lift:

- the target referent is one single entity - no generation of plural descriptions (reference to sets)
- the context is represented as a very simple knowledge base consisting of atoms
- properties are fixed, not context-dependent or vague (e.g. *small*)
- all objects in the domain are assumed to be equally salient

Krahmer & van Deemter (2012) Computational Generation of Referring Expressions: A Survey. *Computational Linguistics*, 38(1):173–218

Sets & More Sophisticated KR

- The easiest extension to refer to set S would be to find those properties that are true of all elements in S: *the chairs*
- What if there aren't any shared (and distinguishing) properties? a better solution is to consider the union of those subsets of S that do share properties: *the blue chairs and the table*
- When referring to sets, coherence of perspective may be important: *the man and the teacher* vs. *the cook and the teacher* or *the man and the woman*
- If we can use set union we may as well use other operations such as complementation (*the chairs that are not by the table*) $\rightsquigarrow$ we may use Boolean operations for sets and for singletons too.
- Rather than using only atomic propositions possibly with Boolean operations, we may use modern knowledge representation frameworks like description logic (*chair a is in house b* could be inferred)

Gatt & van Deemter (2007) Lexical choice & conceptual perspective in the generation of plural referring expressions
van Deemter (2002) Generating Referring Expressions: Boolean Extensions of the Incremental Algorithm, *CL*
Areces et al. (2008) Referring expressions as formulas of description logic.

Context Dependency & Vagueness

- Early models don't make justice to concepts like *young* or *tall*, which are gradable, context-dependent (relative), and vague.
- One possibility is to include in the KB the relevant scale (e.g. `height`) with numerical values.
  - ∗ A possible distinguishing description at content determination stage: `{type=man,height=180cm}`
    - ▸ it could be relised as *The man who is 180cm tall*
    - ▸ but when can it be realised as *The tall / taller / tallest man* ?
  - ∗ Context dependence gets more complicated with several gradable properties: *the small heavy box in the expensive room*
- Are relative properties dispreferred? Do they involve an extra cost? Are they ever used 'redundantly'?

van Deemter (2006) Generating referring expressions that involve gradable properties
Horacek (2005) Generating referential descriptions under conditions of uncertainty

Corpus-based Methods

GRE research has tranditionally been rather formal and mathematical. But corpus-based methods are also used:

- Evaluation
  - ∗ in terms of overlap with human-produced descriptions (Dice, MASI)
  - ∗ human judgements of appropriateness or naturalness
  - ∗ see section 5 of Krahmer & van Deemter (2012) on Evaluation
- Estimation of parameters: preference orders
  - ∗ deriving the preference order from the frequency of attributes in a corpus of referring expression such as TUNA

Gatt et al. (2007) Evaluating Algorithms for the Generation of Referring Expressions
Koolen et al. (2012) Learning Preferences for REG: Effects of Domain, Language and Algorithm
van Deemter et al. (2012) Generation of Referring Expressions: Assessing the Incremental Algorithm

Nickerson et al. (2006) Referring-Expression Generation Using a Transformation-Based Learning Approach

Resources

- Semantically and pragmatically transparent corpora of referring expressions (from Krahmer & van Deemter 2012):

Table 2
Overview of dedicated Referring Expression corpora (alphabetical), with for each corpus a representative reference, an indication of the domain, and the number of participants and collected distinguishing descriptions.

Corpus Name	Reference	Domain	Participants	Descriptions
Bishop	Gorniak & Roy (2004)	Colored cones in 3D scene	9	447
Drawer	Viethen & Dale (2006)	Drawers in filing cabinet	20	140
GRE3D3	Viethen & Dale (2008)	Spheres, Cubes in 3D scene	63	630
iMap	Guhe & Bard (2008)	Various objects on a map	64	9,567
TUNA	van Deemter et al. (in press)	Furniture, People	60	2,280

(TUNA examples: `http://staff.science.uva.nl/~raquel/teaching/rid/tunaexamples/`)

- NLG / REG shared tasks and challenges: with datasets, evaluation metrics, etc.

  http://www.nltg.brighton.ac.uk/research/genchal09/

  http://www.give-challenge.org

Gatt & Belz (2010) Introducing shared task evaluation to NLG: The TUNA shared task evaluation challenges. In *Empirical Methods in Natural Language Generation*.

Next Week

- Discussion of an approach to colour reference

Bert Baumgaertner, Raquel Fernández, and Matthew Stone (2012) Towards a Flexible Semantics: Colour Terms in Collaborative Reference Tasks. In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM), Montreal, Canada.

- An important limitation of the NLG approach we have seen is that it ignores the interactive character of referring
  - ∗ referring in interactive settings (dialogue)
  - ∗ will email you details of relevant readings