Constraints on Definite Descriptions with Vague Adjectives: Towards a Computational Model

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

LoMoReVI Conference – Čejkovice Sept 09

Definite descriptions with gradable adjectives

Definite descriptions with gradable adjectives

the big triangle
the cheap laptop
the short basketball player

- Referential expressions used to pick up an entity in a relevant domain

Definite descriptions with gradable adjectives

the big triangle
the cheap laptop
the short basketball player

- Referential expressions used to pick up an entity in a relevant domain

"Suppose you enter a vet's surgery in the company of two dogs: a big one on a leash, and a tiny one in your arms. The vet asks 'Who's the patient?', and you answer 'the big dog'. This answer will allow the vet to pick out the patient just as reliably as if you had said 'the one on the leash'; the fact that *'big'* is a vague term is irrelevant. You omit the exact size of the dog, just like some of its other properties (e.g., the leash), because they do not improve the description. This shows how *vague* properties can contribute to the *precise* task of identifying a referent." [van Deemter, 2006]

Definite descriptions with gradable adjectives

the big triangle
the cheap laptop
the short basketball player

- Referential expressions used to pick up an entity in a relevant domain

"Suppose you enter a vet's surgery in the company of two dogs: a big one on a leash, and a tiny one in your arms. The vet asks 'Who's the patient?', and you answer 'the big dog'. This answer will allow the vet to pick out the patient just as reliably as if you had said 'the one on the leash'; the fact that *'big'* is a vague term is irrelevant. You omit the exact size of the dog, just like some of its other properties (e.g., the leash), because they do not improve the description. This shows how *vague* properties can contribute to the *precise* task of identifying a referent." [van Deemter, 2006]

- Vague expressions are effciently used in communication

Kees van Deemter (2006) Generating Referring Expressions that involve Gradable Properties. *Computational Linguistics* (32) 2: 195–222.

Overview

- Motivation: natural language generation
- Focus: what contexts licence the positive form?
- Proposal based on relative similarity
- Problems
- Conclusions

Generation of Referring Expressions

Generation of Referring Expressions

- In Computational Linguistics, the production of referring expressions is studied in the subfield of Natural Language Generation (NLG)
- The goal is to generate a distinguishing description of a target object that singles it out from the set of distractors
- This is known as the content determination problem and requires finding a set of properties that jointly identify the intended referent

Generation of Referring Expressions

- In Computational Linguistics, the production of referring expressions is studied in the subfield of Natural Language Generation (NLG)
- The goal is to generate a distinguishing description of a target object that singles it out from the set of distractors
- This is known as the content determination problem and requires finding a set of properties that jointly identify the intended referent

```
Object1:  ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2:  ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3:  ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩
```

Generation of Referring Expressions

- In Computational Linguistics, the production of referring expressions is studied in the subfield of Natural Language Generation (NLG)
- The goal is to generate a distinguishing description of a target object that singles it out from the set of distractors
- This is known as the content determination problem and requires finding a set of properties that jointly identify the intended referent

```
Object1:  ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2:  ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3:  ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩
```

- Usually a description is considered optimal if it is *minimal* (Grice's Maxim of Quantity) – eq. to minimal set cover, i.e. NP-hard

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

Generation of Referring Expressions

Object1: ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2: ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3: ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩

Generation of Referring Expressions

```
Object1:  ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2:  ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3:  ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩
```

- NLG systems may have vague information as input (as above)

Generation of Referring Expressions

```
Object1:  ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2:  ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3:  ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩
```

- NLG systems may have vague information as input (as above)
- Or they may have more precise input and produce vague output

Input: ⟨day, tuesday⟩, ⟨rain, 45mm⟩ Output: *Heavy rain fell on Tuesday*

Generation of Referring Expressions

Object1: ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2: ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3: ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩

- NLG systems may have vague information as input (as above)
- Or they may have more precise input and produce vague output

Input: ⟨day, tuesday⟩, ⟨rain, 45mm⟩ Output: *Heavy rain fell on Tuesday*

- We need a model that specifies how precise information is mapped onto descriptions containing vague adjectives (in their positive, comparative or superlative forms) in a given context

Generation of Referring Expressions

Object1: ⟨type, dog⟩, ⟨size, small⟩, ⟨colour, black⟩
Object2: ⟨type, dog⟩, ⟨size, large⟩, ⟨colour, white⟩
Object3: ⟨type, cat⟩, ⟨size, small⟩, ⟨colour, black⟩

- NLG systems may have vague information as input (as above)
- Or they may have more precise input and produce vague output

Input: ⟨day, tuesday⟩, ⟨rain, 45mm⟩ Output: *Heavy rain fell on Tuesday*

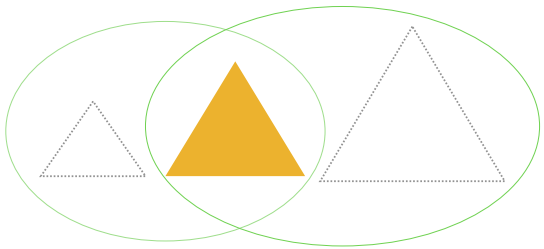
- We need a model that specifies how precise information is mapped onto descriptions containing vague adjectives (in their positive, comparative or superlative forms) in a given context

- **What kind of contexts license the production of a definite description involving a vague adjective in the positive form?**

Context dependence

Context dependence

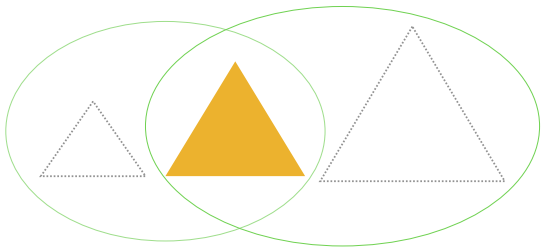
- Well known fact: The positive form can be used to refer to an object that differs from others only in the extent to which it possesses the property expressed by the adjective
  — does not hold for absolute adjectives, e.g. *"the full glass"*



the big triangle / the small triangle

Context dependence

- Well known fact: The positive form can be used to refer to an object that differs from others only in the extent to which it possesses the property expressed by the adjective
  — does not hold for absolute adjectives, e.g. *"the full glass"*



the big triangle / the small triangle

- Lots of psycholinguistics research confirming this

Ebeling & Gelman (1994) Children's Use of Context in Interpreting "Big" and "Little." *Child Development*, 65: 1178–1192.
Syrett, Lidz & Kennedy (to appear) Meaning and Context in Children's Understanding of Gradable Adjectives, *Journal of Semantics*.

Context dependence

Easy to explain if we take into account a couple of assumptions:

Context dependence

Easy to explain if we take into account a couple of assumptions:

(i) The definite article introduces a presupposition of unique existence, where uniqueness is with respect to a contextually determined domain of quantification $\exists x \forall y (P(y) \leftrightarrow x = y)$

(ii) What counts as $\mathbf{A}$ depends on a contextually determined comparison class C, where C is a partially ordered set of elements: $\langle D, \geq_{Dim(\mathbf{A})} \rangle$

Context dependence

Easy to explain if we take into account a couple of assumptions:

(i) The definite article introduces a presupposition of unique existence, where uniqueness is with respect to a contextually determined domain of quantification $\exists x \forall y (P(y) \leftrightarrow x = y)$

(ii) What counts as $\mathbf{A}$ depends on a contextually determined comparison class C, where C is a partially ordered set of elements: $\langle D, \geq_{Dim(\mathbf{A})} \rangle$

Constraints on C (Klein 1980):

(C1) $\mathbf{A}$ partitions C into two sets $\mathbf{A}^+$ and $\mathbf{A}^-$ corresponding to the positive and the negative extensions of $\mathbf{A}$, respectively;

(C2) $\mathbf{A}^+$ and $\mathbf{A}^-$ must be non-empty and disjoint, hence $|C| \geq 2$

(C3) $\forall u, u' \in C$ if $u \in \mathbf{A}^+$ and $u' > u$ then $u' \in \mathbf{A}^+$ (and acc. for $\mathbf{A}^-$)

Klein (1980) A semantics for positive and comparative adjectives. *Linguistics & Philosophy* 4: 1–45

Context dependence

- The interpretation of a description like *the big triangle* can be seen as a byproduct of (i) and (ii) above: given a comparison class C, the boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can flexibly shift as long as (i) and (ii) are satisfied

Context dependence

- The interpretation of a description like *the big triangle* can be seen as a byproduct of (i) and (ii) above: given a comparison class C, the boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can flexibly shift as long as (i) and (ii) are satisfied
- Essentially same idea as Kyburg & Morreau, also Syrett et al.

Context dependence

- The interpretation of a description like *the big triangle* can be seen as a byproduct of (i) and (ii) above: given a comparison class C, the boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can flexibly shift as long as (i) and (ii) are satisfied
- Essentially same idea as Kyburg & Morreau, also Syrett et al.

Context dependence

- The interpretation of a description like *the big triangle* can be seen as a byproduct of (i) and (ii) above: given a comparison class C, the boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can flexibly shift as long as (i) and (ii) are satisfied
- Essentially same idea as Kyburg & Morreau, also Syrett et al.

- But this predicts that in a definite description the positive form is equivalent to the superlative. . .

the big triangle = *the biggest triangle*

Kyburg & Marreau (2000) Fitting Words: Vague Language in Context. *Linguistics & Philosophy*, 23: 577–597.

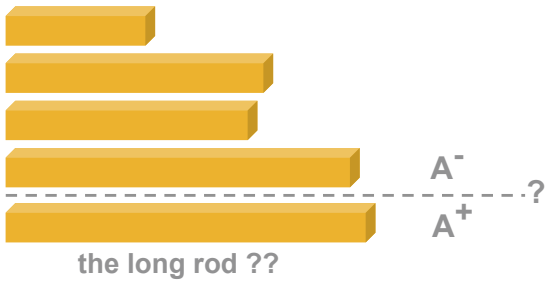
Salience

Salience

- Salience seems to play a role in the use of the positive form

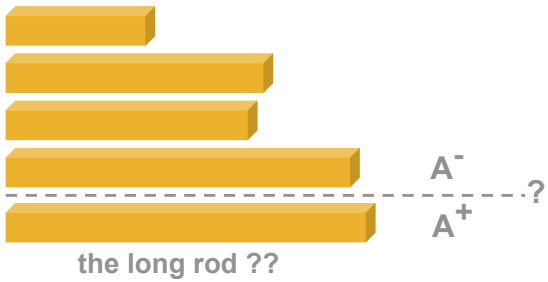
Salience

- Salience seems to play a role in the use of the positive form



Salience

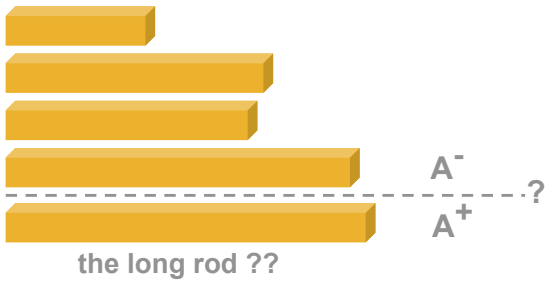
- Salience seems to play a role in the use of the positive form



The boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can only fall between entities that are *saliently dissimilar*

Salience

- Salience seems to play a role in the use of the positive form



The boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can only fall between entities that are *saliently dissimilar*

- Pointed out more or less explicity by Graff 2000, Kennedy 2007 ("the referent needs to stand out"), Rayo 2008, ...
- The gap that sigles out the intended referent needs to be salient with respect to the other gaps: The positive form presupposes the existence of a salient boundary within C

Graff (2000) Shifting Sands: An interest-relative theory of vagueness, *Philosophical Topics*, 20: 45–81.
Kennedy (2007) Vagueness and Grammar, *Linguistics & Philosophy*, 30(1): 1–45.
Rayo (2008) Vague Representation, *Mind*, 117(466): 329–373.

Salience

- How can we account for salience?

Salience

- How can we account for salience?
- We can complement constraints C1–C3 with the following 'similarity contraint':

Salience

- How can we account for salience?
- We can complement constraints C1–C3 with the following 'similarity contraint':

Let $\sim$ be a reflexive, symmetric, and non-transitive relation holding between entities $u \in C$ that satisfies the following conditions:

- $\forall u, u', u'' \in C$ such that $u > u' > u''$, if $u \sim u''$ then $u' \sim u$ and $u' \sim u''$
- If $u' \sim u''$ then $u \in \mathbf{A}^+ \rightarrow u' \in \mathbf{A}^+$ (and accordingly for $\mathbf{A}^-$)

Salience

- How can we account for salience?
- We can complement constraints C1–C3 with the following 'similarity contraint':

Let $\sim$ be a reflexive, symmetric, and non-transitive relation holding between entities $u \in C$ that satisfies the following conditions:

- $\forall u, u', u'' \in C$ such that $u > u' > u''$, if $u \sim u''$ then $u' \sim u$ and $u' \sim u''$
- If $u' \sim u''$ then $u \in \mathbf{A}^+ \rightarrow u' \in \mathbf{A}^+$ (and accordingly for $\mathbf{A}^-$)

- This defines a set of clusters of elements that belong to either $\mathbf{A}^+$ or $\mathbf{A}^-$ and a set of *gaps* that can serve as potential denotational boundaries.

Salience

- How can we account for salience?
- We can complement constraints C1–C3 with the following 'similarity contraint':

Let $\sim$ be a reflexive, symmetric, and non-transitive relation holding between entities $u \in C$ that satisfies the following conditions:

– $\forall u, u', u'' \in C$ such that $u > u' > u''$, if $u \sim u''$ then $u' \sim u$ and $u' \sim u''$

– If $u' \sim u''$ then $u \in \mathbf{A}^+ \rightarrow u' \in \mathbf{A}^+$ (and accordingly for $\mathbf{A}^-$)

- This defines a set of clusters of elements that belong to either $\mathbf{A}^+$ or $\mathbf{A}^-$ and a set of *gaps* that can serve as potential denotational boundaries.
- Related to the semi-order approach of van Rooij: constraints on what constitutes an pragmatically appropriate context

van Rooij (2008) Revealed Preference and Satisficing Behaviour, in *Proc. of the 8th Conf. on Logic and the Foundations of Game and Decision Theory*.

Computing salience using clustering

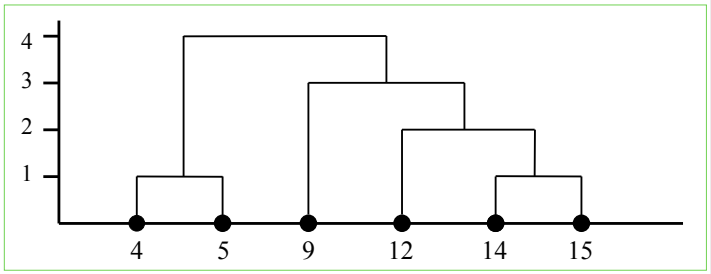
- We explore the use of an agglomerative hierarchical clustering algorithm to model the notions of salience and similarity required

Computing salience using clustering

- We explore the use of an agglomerative hierarchical clustering algorithm to model the notions of salience and similarity required
- The algorithm produces a dendrogram with n nested similarity levels

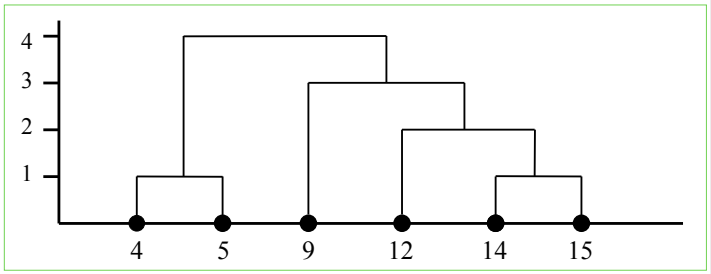
Computing salience using clustering

- We explore the use of an agglomerative hierarchical clustering algorithm to model the notions of salience and similarity required
- The algorithm produces a dendrogram with n nested similarity levels



Computing salience using clustering

- We explore the use of an agglomerative hierarchical clustering algorithm to model the notions of salience and similarity required
- The algorithm produces a dendrogram with n nested similarity levels

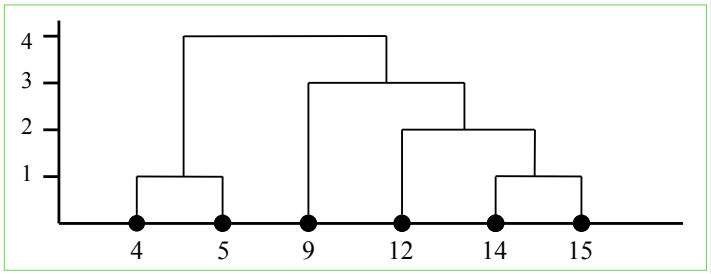


$\delta(14, 15) = 1/4 = 0.25$
$\delta(9, 12) = 3/4 = 0.75$
$\delta(5, 9) = 4/4 = 1$
$\delta(4, 5) = 1/4 = 0.25$
$\ldots$

- We can define the degree of salience δ of a gap between u and u' in C as $\delta(u, u' \in C) = m + 1/n$, where m is the highest level in the dendrogram where u and u' belong to different clusters

Computing salience using clustering

- We explore the use of an agglomerative hierarchical clustering algorithm to model the notions of salience and similarity required
- The algorithm produces a dendrogram with n nested similarity levels



$$\delta(14, 15) = 1/4 = 0.25$$
$$\delta(9, 12) = 3/4 = 0.75$$
$$\delta(5, 9) = 4/4 = 1$$
$$\delta(4, 5) = 1/4 = 0.25$$
$$\ldots$$

- We can define the degree of salience δ of a gap between u and u' in C as $\delta(u, u' \in C) = m + 1/n$, where m is the highest level in the dendrogram where u and u' belong to different clusters
- Since δ is defined as a ratio, it accounts for the *relative salience* of the gaps given the distribution of elements in C

Computing salience using clustering

- We can express restrictions on $\sim$ in terms of heuristic constraints that need to be determined empirically

$$u \sim u' \leftrightarrow \delta(u, u' \in C) < k$$

- The value of k may depend on several factors: the nature of the elements under consideration, the task at hand, the goals of the agents, ...

Computing salience using clustering

- We can express restrictions on $\sim$ in terms of heuristic constraints that need to be determined empirically

$$u \sim u' \leftrightarrow \delta(u, u' \in C) < k$$

- The value of k may depend on several factors: the nature of the elements under consideration, the task at hand, the goals of the agents, ...

  - We can compute the salience of the relevant boundary to decide if the positive form is licenced – but whenever it is licenced, should it always be preferred over the superlative?

Some problematic cases

Some problematic cases

- So far the model predicts that $u \sim u'$ cannot hold for $C = \{u, u'\}$ with $u > u'$ or $u' > u$: the only boundary will always count as salient

Some problematic cases

- So far the model predicts that $u \sim u'$ cannot hold for $C = \{u, u'\}$ with $u > u'$ or $u' > u$: the only boundary will always count as salient
- This is problematic in some contexts: the positive form does not seem to appropriately make what Kennedy calls *crisp judgements*

Some problematic cases

- So far the model predicts that $u \sim u'$ cannot hold for $C = \{u, u'\}$ with $u > u'$ or $u' > u$: the only boundary will always count as salient
- This is problematic in some contexts: the positive form does not seem to appropriately make what Kennedy calls *crisp judgements*

Context: A (99|50)-page novel and a 100-page novel [Kennedy 2007]

(1) The long book. / This book is long compared to that book.

(2) The longer book. / This book is longer than that book.

Some problematic cases

- So far the model predicts that $u \sim u'$ cannot hold for $C = \{u, u'\}$ with $u > u'$ or $u' > u$: the only boundary will always count as salient
- This is problematic in some contexts: the positive form does not seem to appropriately make what Kennedy calls *crisp judgements*

Context: A (99/50)-page novel and a 100-page novel [Kennedy 2007]

(1) The long book. / This book is long compared to that book.

(2) The longer book. / This book is longer than that book.

- Is relative salience not the way to go?

Some problematic cases

- So far the model predicts that $u \sim u'$ cannot hold for $C = \{u, u'\}$ with $u > u'$ or $u' > u$: the only boundary will always count as salient
- This is problematic in some contexts: the positive form does not seem to appropriately make what Kennedy calls *crisp judgements*

Context: A (99/50)-page novel and a 100-page novel [Kennedy 2007]

(1) The long book. / This book is long compared to that book.

(2) The longer book. / This book is longer than that book.

- Is relative salience not the way to go?
- Are there other factors affecting C?

Feature variability

Feature variability

- There seems to be a difference between categories that exhibit different patterns of feature varibility (central tendency vs. uniform frequency distribution)

Feature variability

- There seems to be a difference between categories that exhibit different patterns of feature varibility (central tendency vs. uniform frequency distribution)
- Experiments by van Deemter seem to support this:
  - the positive form is dispreferred with very similar elements
  - but for abstract figures some subjects use the positive form across the board

Feature variability

- There seems to be a difference between categories that exhibit different patterns of feature varibility (central tendency vs. uniform frequency distribution)
- Experiments by van Deemter seem to support this:
  - the positive form is dispreferred with very similar elements
  - but for abstract figures some subjects use the positive form across the board
- A technical "solution": cluster the combined set $\mathcal{C} = C \cup C'$, where the elements in C' are randomly generated from the relevant frequency distribution

Fernández (2009) Salience and Feature Variability in Definite Descriptions with Positive-form Vague Adjectives, in *Proc. of PRE-CogSci, the Annual Meeting of the Cognitive Science Society*.

Summing up

Summing up

- Definite descriptions with relative gradable adjectives

Summing up

- Definite descriptions with relative gradable adjectives
- Pragmatic approach: constraints on context that licence the use of the positive form

Summing up

- Definite descriptions with relative gradable adjectives
- Pragmatic approach: constraints on context that licence the use of the positive form
- Importance of contrast and similarity

Summing up

- Definite descriptions with relative gradable adjectives
- Pragmatic approach: constraints on context that licence the use of the positive form
- Importance of contrast and similarity
- Possible influence of patterns of feature variability

Summing up

- Definite descriptions with relative gradable adjectives
- Pragmatic approach: constraints on context that licence the use of the positive form
- Importance of contrast and similarity
- Possible influence of patterns of feature variability
- Importance of experimental work and computational evaluation