

## Underspecification in anaphoric reference to structured entities



Ruth Filik<sup>1</sup>, Anthony Sanford<sup>1</sup>, Patrick Sturt<sup>1</sup>, & Massimo Poesio<sup>2</sup>

1. University of Glasgow, UK

2. University of Essex, UK



Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Background

Readers and listeners do not seem to always construct detailed semantic interpretations of everything they read or hear (e.g., Ferreira, Bailey, & Ferraro, 2002; Sanford 2002; Sanford & Sturt, 2002).

There is growing empirical evidence which supports the notion of underspecification in language processing:

- **Lexical Underspecification** (e.g., Frazier and Rayner, 1990).
- **Syntactic Underspecification** (e.g., Ferreira 2003; Frazier & Clifton 1996).
- **Semantic anomalies** (e.g., Barton & Sanford, 1993; Erickson & Mattson, 1981).
- **Text change detection** (e.g., Sturt, Sanford, Stewart & Dawydiak, 2004).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Goals of current research

The goal of the current research is to investigate underspecification in relation to anaphoric reference.

This involves:

- Identifying cases in which anaphoric expressions, and especially pronouns are not completely interpreted.
- Determining the interpretation they receive.
- Specifically in relation to *structured reference objects*.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Reference objects and pronoun use

"Jane and Mike were working in the garden. They/He wanted to finish before dark."

Atomic reference object:

- *Mike* is an atom.
- May be referred to with a singular pronoun (*He*).

Complex reference object:

- *Jane and Mike* form a complex reference object.
- Can be thought of as a sum of parts, or as a set of atomic tokens.
- May be referred to with a plural pronoun (*They*).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Formation of Complex Reference Objects

In the example:

*Jane and Mike were working in the garden.*

It is the use of the connective 'and' to create a conjoined NP (e.g., Albrecht & Clifton, 1998) that leads to the formation of a complex reference object. E.g. *Jane and Mike*.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Reference objects and pronoun use

If a complex reference object has been formed:

- Plural pronominal reference will be the preferred means of referring to the Complex Reference Object.
- Singular pronominal reference to the individuals will be difficult, as entities making up the Complex Reference Object have to be individuated (Conjunction Cost).

If a complex reference object has not been formed:

- Singular pronominal reference should be easy.
- Processing of plural pronouns should be difficult because a Complex Reference Object would have to be constructed.

This has been demonstrated empirically (e.g., Moxey, Sanford, Sturt, & Morrow, 2004).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Underspecification in anaphoric reference

Poesio, Reyle, & Stevenson (2001) proposed that underspecification in anaphoric reference may occur if:

- The potential antecedents are part of a single mereological structure (i.e., a sum of parts).
- The structure makes the interpretations equivalent for the purposes of the plan.

Example from TRAINS corpus:

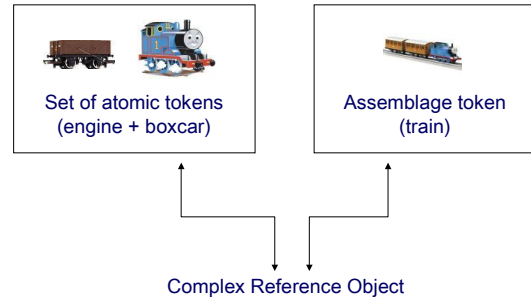
3.1 M: can we .. Kindly hook up  
 3.2 : uh  
 3.3 : engine E2 to the boxcar at .. Elmira  
 4.1 S: ok  
 5.1 M: and send **it** to Corning  
 5.2 : as soon as possible please

It = engine, boxcar, or complex reference object.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Nature of the Complex Reference Object



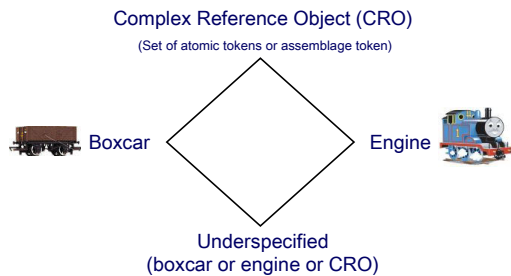
Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Underspecification in anaphoric reference

...hook up the engine and the boxcar and send it to Corning...

In this example, 'it' could refer to:



Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Offline evidence

Poesio et al. used magnitude estimation to compare the acceptability of sentences like:

- (1) a. The engineer hooked up the engine to the boxcar and sent it to London.
- b. The engineer separated the engine from the boxcar and sent it to London.

Examples like (1a) were judged as being more acceptable than (1b), presumably because in (1b) participants have difficulty in resolving the referent of 'it', but in (1a) they would not have this problem if the referent of 'it' is underspecified.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Online evidence

We decided to further test this theory assessing disruption in processing using eye-tracking.

Example:

Mereology Constructing (MC)

The engineer hooked up | the engine and the boxcar, | and sent [it/them] | quickly | to the central station. |

Neutral

The engineer saw | the engine and the boxcar, | and sent [it/them] | quickly | to the central station. |

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Online evidence

We also investigated the nature of the object formed:

'Conventional whole object'

The engineer hooked up | the engine and the boxcar, | and sent [it/them] | quickly | to the central station. |

'No conventional whole object'

The manager stapled together | the letter of complaint and the revised timetable, | and sent [it/them] | quickly | to the central station. |

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Pre-test

13 participants given 96 descriptions.

48 formed a conventional whole object when combined (e.g. 1).

1. *The engineer hooked up the engine and the boxcar.*

48 formed no conventional object (e.g., 2):

2. *The manager stapled together the letter of complaint and the revised timetable.*

Question:

Is the resultant object a conventional whole object?

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Pre-test

Is the resultant object a conventional whole object?	
Object materials	No object materials
86.4% Yes	17.1% Yes

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Predictions

- For neutral sentences there should be a cost in referring to a Complex Reference Object with a singular pronoun (*it*) compared to a plural pronoun (*them*).

The engineer saw the engine and the boxcar, and sent [*it* > *them*] quickly to the central station.

- This cost should be eliminated for MC sentences if the referent of 'it' is underspecified.

The engineer hooked up the engine and the boxcar, and sent [*it* = *them*] quickly to the central station.

- If only the assemblage token (train) is available we might expect there to be a cost in referring to this object with 'them'.

The engineer hooked up the engine and the boxcar, and sent [*it* < *them*] quickly to the central station.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Materials

### MC, conventional whole object

There were many delays. | The engineer hooked up | the engine and the boxcar, | and sent [*it*/*them*] | quickly | to the central station. | He hoped things would improve soon. |

### MC, no conventional whole object

There were many delays. | The manager stapled together | the letter of complaint and the revised timetable, | and sent [*it*/*them*] | quickly | to the central station. | He hoped things would improve soon. |

### Neutral, conventional whole object

There were many delays. | The engineer saw | the engine and the boxcar, | and sent [*it*/*them*] | quickly | to the central station. | He hoped things would improve soon. |

### Neutral, no conventional whole object

There were many delays. | The manager read through | the letter of complaint and the revised timetable, | and sent [*it*/*them*] | quickly | to the central station. | He hoped things would improve soon. |

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Method

- 48 participants
- 48 items
- Participants' eye movements were monitored using a DPI eye-tracker.
- Comprehension questions followed 50% of trials.

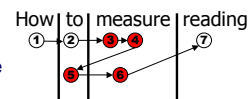
Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Measures of reading behaviour

**Regressions out:** the proportion of trials where readers looked back from the region to an earlier piece of the text between the time when the region was first entered from the left to the time when the region was first exited to the right.

**Regression path:** the sum of fixations from the time that a region was first entered from the left to the time the region was first exited to the right.

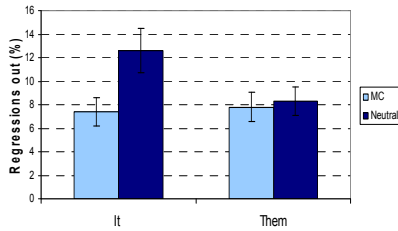


Underspecification in anaphoric reference to structured entities

AMLaP 2005

### Regressions out of the critical region (*and sent it/them*)

Interaction: sentence structure \* pronoun  $F_1 = 3.60, p = .06; F_2 = 4.18, p < .05$ .



For sentences containing *it*: more regressions for neutral than mereology constructing sentences  $F_1 = 8.39, p < .01, F_2 = 10.36, p < .01$ .

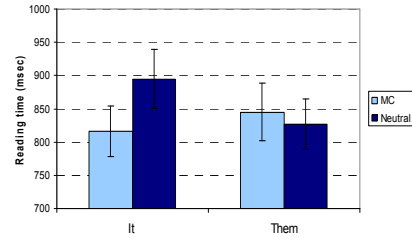
For baseline sentences with *them*: No difference ( $F_s < 1$ ).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

### Regression path final region (*to the central station*)

Interaction: sentence structure \* pronoun ( $F_1 = 8.76, p < .01; F_2 = 5.45, p < .05$ ).



For sentences containing *it*: longer reading times for neutral than mereology constructing sentences ( $F_1 = 8.00, p < .01; F_2 = 5.49, p < .05$ ).

For baseline sentences containing *them*: No difference ( $F_s < 1$ ).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

### Summary

#### For neutral sentences:

There was a processing cost associated with using a singular pronoun (*it*) to refer to the conjoined noun phrase [*it > them*], i.e., a Conjunction Cost.

#### For mereology constructing (MC) sentences:

This cost was eliminated [*it = them*]. Suggesting that readers had access to more than one representation.

- Note there was no cost for 'them' compared to 'it', so it is unlikely that the 'assemblage token' (e.g., train) is the only interpretation available.

**Conclusions:** Compatible with the referent(s) of 'it' being underspecified in some way in MC sentences.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

### Experiment 2: Text change detection

Text change detection recently used to study depth of semantic processing (e.g., Koh, Sanford, Clifton & Dawydiak (in press) Sanford, Sanford, Filik & Molle (in press); Sturt et al. (2004).

- In this task, participants read a piece of text twice, and their task is to detect whether or not a word has changed from the first presentation to the second.
- The likelihood of noticing a change may reflect the depth of processing of that word.

Importantly, Koh et al. demonstrated that changes can be detected on pronouns such as 'he' and 'they'.

Jared and Kathleen went to the mall on Sunday. [They->He] found a parking space near the entrance.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

### Experiment 2: Materials

**MC:** The engineer hooked up the engine and the boxcar and sent [*them -> it*] to London.

**Neutral:** The engineer saw the engine and the boxcar and sent [*them -> it*] to London.

#### Predictions:

**MC sentences:** If the referent(s) of the pronoun is underspecified, participants should detect fewer changes from 'them' to 'it' in this condition.

**Neutral sentences:** More changes from 'them' to 'it' should be noticed as readers try to resolve the referent of 'it'.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

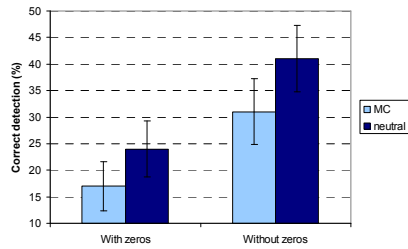
### Experiment 2: Method

- 28 participants.
- 16 items.
- 84 filler items (52 with no change, 32 with random changes).
- Participants read each piece of text twice and verbally indicated if they had noticed a change and what this change was.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Experiment 2: Results



Neutral > MC,  $t_1(27) = 2.15$ ,  $p < .05$ ;  $t_2(15) = 2.18$ ,  $p < .05$

If you remove participants who had scores of zero in both conditions:

Neutral > MC,  $t_1(15) = 2.27$ ,  $p < .05$ ;  $t_2(15) = 2.18$ ,  $p < .05$

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Summary

Results indicate that readers may underspecify the referent(s) of the pronoun 'it' in mereology constructing cases.

### Eye-tracking results:

- For mereology constructing sentences: equally acceptable to refer to the conjoined noun phrase using 'it' or 'them'.
  - Suggesting that more than one representation is available.
- For neutral sentences: cost for using 'it' relative to 'them'.
  - Suggesting a CRO set of atomic tokens representation.

### Change detection results:

- More changes from 'them' to 'it' noticed for neutral than mereology constructing sentences.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## How do people actually represent 'it'?

- Results thus far suggest that readers' representation of what the pronoun 'it' refers to may be underspecified in mereology constructing sentences.
- But what is the nature of this underspecified representation?
- Poesio et al. proposed that it covers all possible representations (i.e., engine or boxcar or CRO).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Paraphrase study

Wanted a task where we could find out how people had represented the referent(s) of 'it', without directly asking them.

- 96 participants were each given 1 sentence from those used in the eye tracking study.
- "The engineer hooked up the engine and the boxcar and sent it to the central station."
- They were asked to rewrite the sentence without using the word 'it', but keeping the meaning of the sentence the same.
- All sentences were mereology constructing.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Classification of responses

**Set of atomic tokens:** 'them', 'both of them', 'the engine and the boxcar' etc.

**Assemblage token:** 'the train', 'the combined object' etc.

**Singular atomic reference object:** one of the objects (i.e., 'engine', 'boxcar').

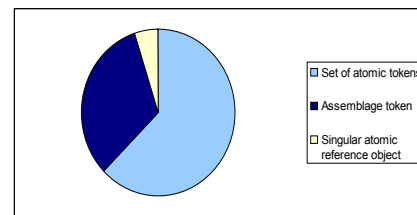
### Predictions:

- According to Poesio et al., the underspecified representation includes 'the engine', 'the boxcar' and the 'Complex Reference Object'.
- If these elements are all present in the representation of the referent(s) of 'it', we would expect all of these elements to be present in participants' responses.

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Results



- All elements (the set of atomic tokens, assemblage token, and singular atomic reference objects) are present in participants' responses.
- Most responses shared between the set of atomic tokens (*them*) and the assemblage token (*train*).

Underspecification in anaphoric reference to structured entities

AMLaP 2005

## Conclusions: nature of the underspecified representation

It is likely that the underspecified representation contains:

- At least the set of atomic tokens and the assemblage token representations.
- May also contain representations of the singular elements (e.g., engine and boxcar).



Engine 



Boxcar 

Underspecification in anaphoric reference to structured entities

AMLP 2005

Thank you



Underspecification in anaphoric reference to structured entities

AMLP 2005