IT–CLEFT SENTENCES DISAMBIGUATION: A LOCAL GRAMMAR PROJECT

Bilyana Ovcharova
University of Forestry, Sofia, Bulgaria

Abstract

The aim of this paper is to write a local grammar to disambiguate it-cleft sentences. In order to do this, the ambiguity of the structure has been established: the pronoun it is used both with referential meaning in sentences with anaphoric or situational reference, and with non-referential meaning in sentences with dummy it subjects denoting time, distance, weather conditions, sentences with extraposition and it-cleft sentences. On the basis of the internal structure of the focused element of it-clefts, disambiguation rules have been written. The environment of it-clefts and, in particular, the sentence elements preceding them, has been studied with the aim of more successful disambiguation. The rules’ precision has been evaluated using examples obtained from the British National Corpus, and the respective disambiguation failures have been analyzed.

Key words: local grammar, disambiguation, it-cleft sentences, English.

Introduction

The aim of this term paper is to write a local grammar to disambiguate it-cleft sentences. It-cleft sentences are defined by Quirk et al (1985) as “devices for giving prominence” comprising the subject pronoun it as an empty theme, followed by the verb be, which introduces the focused element. In example [1] below the dummy it and the focused element are underlined:

[1] In her shock and grief, it was my mother whom Deborah turned to, calling and weeping into the phone.

It-cleft sentences contain the ambiguous construction “it + be [sg]”. The aim of this study is to analyze to what extent the disambiguation will be successful, if possible at all. Therefore, let us first identify the ambiguity of the construction “it + be [sg]”.

1. Ambiguity of it be[sg]

It followed by a verb form of be, in its most straightforward case, is a pronoun which stands for an object in reality as in example [2]:

[2] Desperado. . . . It was a film from the early eighties, and it co-starred Alex Rivers.

In a less obvious case, ‘it’ does not refer to an object in reality, but to a situation, but it still has referential meaning and can be substituted by the demonstrative pronouns this/that. In example [3] such substitution is possible, and is shown in the curly brackets:

[3] As the sun dropped behind the edge of the distant hills, he leaned over and kissed me. It [This] was not the way he had kissed me before—not quiet, not gentle, not testing.

In contrast to the referential it-subjects, there are also syntactically similar structures with dummy it-subjects:

a) in “expressions denoting time, distance, or atmospheric conditions” [Quirk et al, 1985: 348];

[4] It was nearly five in the morning when we arrived, and /…/.

b) in sentences with extraposition;

[5] It was nice to see her thinking of Alex as an ally instead of an enemy, but /…/.

c) in it-cleft sentences.

[6] It was only in his closing that he acknowledged any personal connection between them.

The examples 2-6 all contain “it + be [sg]”, and the task is to write local rules that will disambiguate example [6] as an example of a cleft sentence. Sound local rules are written on the basis of markers such as morphemes, words, POS tags, punctuation marks, sentence borders.

2. Disambiguation rules

2.1. Marker: punctuation

Let us start with the big picture delimiting the borders of a cleft sentence by formulating the following Rule 1: If it be[sg] is followed by a sentence-end punctuation or direct speech punctuation i.e. a comma and inverted commas, then this is not an it-cleft sentence.

IF it be[sg] Punct {sentence-end/direct speech} → not it-cleft

1 The sources of the examples are given in Appendix 1.
2 The abbreviations and symbols used in the rules are given in Appendix 2.
This rule would successfully disambiguate example [7] obtained from the British National Corpus, if the sentence had the expected punctuation.

[7] It wasn’t a repeat oh yeah it was she said I’m going to be sick when, so

2.2. Marker: focused element realized by a personal pronoun or proper noun

The following Rule 2 can be written for it-clefts disambiguation: If it be[sg] is followed by a pronoun in the nominative or the oblique case, or a proper noun followed by who/that, then this is an it-cleft sentence.

IF it be[sg] PronNom/PronObl/NProper, who/that → it-cleft

Examples of Rule 2 disambiguation:
[8] When her younger brother was being bullied it was she who leapt to his defence and up-ended his tormentor in a matter of seconds.
[9] It was she that was responsible for the trashing of so much British manufacturing, with the appreciating pound very much part of this.
[10] Janet shared the stage with the pop group Slade, but next day it was her who made all the headlines.
[11] Her mother was a fighter and had a graceful attitude all throughout her illness, and it was her that gave Jaime the determination to beat cancer.
[12] The story told her unconscious that to follow Hansel’s lead led her back, not forward, and it was also meaningful that although Hansel was the leader at the story’s beginning, it was Gretel who in the end achieved freedom and independence for both, because it was she who defeated the witch.

Quirk et al. (1985) remarks that the use of a personal pronoun in the objective case when it is the subject of the subordinate relative clause is “widely condemned”. Another tendency in the examples provided by Quirk et al. (1985) below is the tendency in familiar English the omission of the relativizer when the focused element is also the subject or an adjunct in the relative subordinate clause [13-14]:
[13] It was the President himself spoke to me.
[14] It is Chelsea (that) he lives in.

The omission of the relativizer requires a new rule. Rule 3: If it be[sg] is followed by a pronoun in the nominative or the oblique case, or a proper noun, then this is an it-cleft sentence, unless it is preceded by what, who, when, where, why, how, whatever, whoever, wherever, whenever.

IF it be[sg] PronNom/PronObl/NProper, unless {what, who, when, where, why, how, whatever, whoever, wherever, whenever} → it-cleft

Examples of Rule 3 disambiguation:
[15] We asked who we talked to and when pressed that it was her we spoke to, she adamantly denied this.
[16] On Monday, Feb. 1, 2010, ECU Joyner Library’s dear colleague Cynthia Jones, who was our Human Resources point person and Administrative Assistant Director, died suddenly on her way to work. It was we who would have sought solace with when this kind of tragedy occurs.
[17] No one ever knew where it was she hoped to go.

It should be noted that Rule 2 will disambiguate sentences as [18] in which the relativizer is not omitted, but is placed at some distance from the focused element:

[18] It was she, it seemed, not Quentin, who might somehow blow hope into hopelessness.

Sentences as in [18] accounted for 14% (7 instances) of the 50 randomly generated examples for the query “it was she” by the British National Corpus. Together Rule 1 (if the sentence had punctuation), Rule 2 and Rule 3 disambiguate 96%, or 48 of the generated sentences. In the two remaining sentences [19, 20] the it subjects can be substituted by this, and the verb be bears its existential meaning.

[19] He accused her of blaming him for his failure to return with Oreste when, if anyone was to blame other than her sister Ellen, it was she herself.
[19a] Обвини я, че хвърли върху него вината, че не се е върнал с Оресте, когато, ако някой изобщо беше виновен, освен сестра й Елън, това беше самата тя.
[20] She had a little mole above her pretty upper lip, so I knew it was she.
[20a] Имаше малка бенка над красивата си горна устна, така че беше сигурен, че това е тя.

Although the sentence translation in Bulgarian shows that the optional use of an emphatic adverb does not change the meaning, sentences [19, 20] are not emphatic it-clefts. In the exchange: “Who is there? It’s me.” it is not possible, without a shift in meaning, to use either a demonstrative or an emphatic adverb when translating it into Bulgarian. Therefore, though accepting 4% disambiguation failure, further re-
search is necessary to study such sentences (the British National Corpus generates only 14 sentences for the query “it was her,” with a full stop at the end).

2.3. Marker: adjunct

a) because of as marker

Rule 4: If it be[sg] is followed by because of, followed by that in position (+1), (+2) or (+3), then this is an it-cleft sentence, unless that is followed by a finite verb:

IF it be[sg] because of that((of+1)/(of+2)/(of+3) → it-cleft

unless

IF it be[sg] because of that((of+1)/(of+2)/(of+3) V[fin]

Examples of Rule 4 disambiguation:

[21] It was because of this that he finally chose medicine over law as his career goal.

[22] It was because of this interest that he was awarded a Red Cross medal for services rendered to the sick and wounded during the Franco-Prussian war.

[23] It was because of that meagre income that she became incensed at her tutor, Maurice Greiffenhagen, who had a habit of sitting down at his students’ drawings and paintings and finishing them off.

[24] If they did, however, it was because of family feuds that might occur.

To establish the disambiguation precision of this rule, a sample of 100 random results for the queries “it is because of” and “it was because of” was obtained from the British National Corpus. The analysis of the sample shows that rule 4 limits the recall in at least two ways:

- that can be distanced from because of by as many as 42 positions as in example 25

[25] It was, then, peace or truce which threatened their existence, and it was because of the termination of hostilities between England and France which resulted from the treaty of Brétigny in 1360, followed as this was in 1364 by the ending of the Navarrese challenge to the royal authority in Normandy and the war of Breton succession, that France, Italy and Spain were to be hosts to the Companies in the 1360s.

- the cleft sentences contain it be[sg] because of can occur without that as in example 26:

[26] She wanted to stay in Harlem and be with Edith for it was because of Edith she was still alive and she finally knew that their relationship was genuine.

In this sample of 100 sentences there are four it-cleft sentences without that. They contain either a pronoun in the oblique case or a proper name. Rule 5 can be written to disambiguate these: If it be[sg] is followed by a pronoun in the oblique case, or a proper noun, then this is an it-cleft sentence.

IF it be[sg] because of PronObl/NProper → it-cleft

Table 1 shows the summary of the results for the sample after analyzing the sentences it contains. As 100 sentences were analyzed, the number and percentages have the same values, but are given for clarity.

<table>
<thead>
<tr>
<th>Table 1. Summary results of the BNC sample for it be[sg] because of</th>
<th>Occurrences</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Disambiguated clefts using rule 4</td>
<td>32</td>
<td>32%</td>
</tr>
<tr>
<td>2 Clefts with no that (after pronoun/name) using rule 5</td>
<td>4</td>
<td>4%</td>
</tr>
<tr>
<td>3 Discarded non-clefts using rule 4</td>
<td>34</td>
<td>34%</td>
</tr>
<tr>
<td>4 Clefts containing that distanced at more than 3 positions</td>
<td>22</td>
<td>22%</td>
</tr>
<tr>
<td>5 Non-clefts containing that distanced at more than 3 positions</td>
<td>8</td>
<td>8%</td>
</tr>
<tr>
<td>Sample total</td>
<td>100</td>
<td>100%</td>
</tr>
</tbody>
</table>

Using Rule 4 and 5 will make possible to correctly identify 70% of the sentences: 36% are correctly disambiguated cleft sentences (lines 1 and 2), and 42% correctly discarded non-clefts (line 3 and 5). The rules fail to disambiguate 22% cleft sentences (lines 3) in which that is place at a distance of up to 42 positions from because of. The recall can be increased, by changing the that position range, to include the 22 sentences which rule 4 fails to disambiguate. However, this increased recall will also include the non-clefts containing that in the new range (line 5), and therefore this needs to be considered by the modified rule.

Rule 4 (modified): If it be[sg] is followed by because of, followed by that in any position from one position after of (of+1) to two position before sentence-end (@@-2), then this is an it-cleft
sentence, unless *that* is followed by a finite verb, or unless in this range *that* occurs as a determiner:

\[
\text{IF } it \ \text{be[sg]} \ \text{because of that(of+1)/...../(@@-2)} \rightarrow \text{it-cleft}\]

unless

\[
\text{IF } it \ \text{be[sg]} \ \text{because of that(of+1)/...../(@@-2)} V[\text{fin}]
\]
or unless

\[
\text{IF } it \ \text{be[sg]} \ \text{because of that(det)(of+1)/...../(@@-2)}
\]

Examples of Rule 4 (modified) disambiguation include all the 22 clefts in line 4 (table 1) as well as two [27, 28] of the 8 non-cleft sentences containing that as determiner.

[27] Probably it was because of what had happened in the play at that point on the previous night.

[28] The first time I murdered it was because of rabbits meeting a fiery death, and meeting that fiery death from the nozzle of a Flamethrower virtually identical to the one I had used to exact my revenge on the warren.

Using Rule 4 (modified) and rule 5, the disambiguation precision is 94%; 6 non-cleft sentences which contain *it be[sg]* followed by *because of, and that* within 2 positions of the sentence end will not be discarded. Solutions may be sought in the environment of *it be[sg]*, and this will be attempted later in the study.

b) *not until* as marker: 100% disambiguation precision of *it*-cleft sentences

It can be expected that the disambiguation of *it*-cleft sentences is not easy considering the other syntactically similar constructions involving “*it be [sg]*”. However, the rule below will disambiguate all *it*-cleft sentences of its type without any error:

\[
\text{IF } it \ \text{be[sg] not until } \rightarrow \text{it-cleft}\]

To test this rule, a search was conducted in the BNC. All the 50 random solutions generated for “it was not until” were *it*-cleft sentences. Of the 36 solutions generated for “it is not until” only one is not a cleft sentence, but it is also not a grammatically correct sentence [29].

[29] Sir Daniel Fleming was certainly interested in the formerly productive mines in his fells around Coniston, but it is not until 1658 is there further note of mines and then not for copper viz

The *it*-cleft sentences under this rule contain either a foregrounded PP [30] or an adjunct clause [31], and in all cases but one the relativizer “*that*” is used; it should be noted that its omission seems to be due to a grammatical error [32].

[30] However it is not until this twentieth century that variable space has become a conceptual reality.

[31] It was not until I wandered back to the harbour that I realised the wind had freshened.

[32] I had been taught not to cry till I was really hurt, and so it was not until about eleven o’clock I really began to be noticed and I went into the theatre.

Conclusions

1. It is possible to establish rules for successful disambiguation of *it*-cleft sentences.
2. The disambiguation precision of the established rules is reliable.

It should be noted that this study is limited in its scope and does not account for the wide diversity of *it*-cleft sentences. Further research is necessary in order to estimate the degree to which successful disambiguation is possible.

References


Appendix 1: Example sources
The British National Corpus, random searches: for examples 7, 8, 10, 12, 17, 18, 19, 20, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32
Lahiri, J., Unaccustomed Earth, Alfred A. Knopf, New York, Toronto, 2008: for examples 1, 6
Picoult, J., Picture Perfect, Penguin Group, 1995: for examples 2, 3, 4, 5,
Appendix 2: Abbreviations and symbols used in the rules

( ) – optional component
/ - or
→ - then
@@ - sentence end

AdjP – adjective phrase
AdvP – adverb phrase
Conj – conjunction
det - determiner
fin - finite
NP – noun phrase
NProper – proper noun
PM – post-modifier
PP – prepositional phrase
PronNom – pronoun (nominative)
PronObl – pronoun (oblique)
Punct - punctuation
sg - singular
VP – verb phrase
XP – any type of phrase