Corpus-Based Analysis of Parenthetical Reporting Clauses

Carla Schelfhout

University of Nijmegen

Abstract

In this paper, the syntactic properties of parenthetical reporting clauses in Dutch are investigated by means of a small corpus. It is shown that an analysis in which the quote is looked upon as the direct object of the reporting verb is inadequate. Therefore, an alternative analysis is proposed, viz. one in which the quote and the reporting clause are taken to be adjoined. Such an analysis, however, is not unproblematic. Here we discuss the two main problems with this analysis.

1 Introduction

A construction that so far has not received much attention in automatic syntactic analysis is the parenthetical reporting clause (rc). The construction has been the subject of a number of theoretical studies including Ross (1973), Luij (1990) and Collins and Branigan (1997), but none of these came near a formalisation, let alone an implementation. For the purpose of this article the term 'parenthetical clause' is defined as a linguistic entity that does not have any syntactic relation to the sentence it interrupts. Parenthetical clauses have restrictions, though: they can only appear at certain positions in the main clause (they usually do not appear within the premodifer of a noun phrase (NP) or between a preposition and its complement, for example) and not every construction is allowed to become parenthetical. An example of a parenthetical clause is given in (1):

(1) He must, I thought, be a wonderful person.

The parenthetical clause 'I thought' does not have any syntactic function in the main clause, nor does one clause appear to be subordinated to the other. Nevertheless, example (1) is a syntactically correct sentence. The fact that the relation between the main clause and the parenthetical clause is unclear implies that this relationship—and therefore these constructions—cannot be implemented in an automatic syntactic parser. In the AMAZON-parser, a surface based parser for Dutch, the parenthetical construction has not been implemented yet. In order to extend AMAZON with at least one variant of the parenthetical construction, a study was made of the variant that seemed easiest to identify: the parenthetical reporting clause. This study consisted of compiling a 50,000-word corpus with a relatively high proportion of these constructions and analysing these real-life observations. Unless stated otherwise, examples present in the paper originate from this corpus. In section 2 I describe the possible forms of reporting constructions, in section 3

1 Examples judged ungrammatical, indicated with *, are either made up entirely or derived from corpus samples. In the latter case, I indicate from which samples.
their analyses will be discussed and the main problems with the proposed analysis will be solved. Section 4 contains some additional remarks and presents a conclusion.

2 Reporting constructions

In Dutch, an author using direct speech can report who said what in basically three ways: he can put the reporting clause in front of the direct speech sentence (2a), he can interrupt the quote by the reporting clause (2b) or he can place the reporting clause after the quote (2c). In all cases, the form of the direct speech sentence remains the same; the form of the reporting clause varies according to its position.

(2) a. Annie zei: "Ik geloof dat je wat moet uitleggen."  
Annie said "I think that you something have to explain"  
'Annie said: "I think that you ought to explain something.""

b. "Ik geloof," zei Annie, "dat je wat moet uitleggen."  
"I think" said Annie "that you something have to explain"  
"I think," Annie said, "that you ought to explain something.""

c. "Ik geloof dat je wat moet uitleggen," zei Annie.  
"I think that you something have to explain" said Annie  
'Annie said: "I think that you ought to explain something.""

The construction exemplified in (2b) situates a reporting clause directly after the finite verb, but reporting clauses can appear in many other positions in the quoted sentence (not at just any position, though). These constructions, in which the Direct speech is Interrupted by a Parenthetical reporting clause (henceforth referred to as DIP-sentences), were the original object of study. During this study, however, it became clear that the construction with the reporting clause positioned directly after the quote, as exemplified in (2c), is in many ways similar to DIP-sentences. The overt similarity that distinguishes constructions like (2b) and (2c) from constructions like (2a), which situate the reporting clause before the quote, is inversion within the reporting clause. Another, covert similarity will be dealt with in section 3.3.2. Therefore, constructions which position the reporting clause after the quote are considered to be variants of DIP-constructions. These constructions can be expected to appear often in fiction dialogue; in real life they are realised in quotes from interviews or political debates. Therefore, the corpus was built from such newspaper articles.

3 What belongs to what?

3.1 Analysis I: the quote as direct object

The main difficulty for the analysis of DIP-constructions is the relationship between the direct speech sentence, or: the quote, and the reporting clause. A com-

2Sentence (2b) is the one that appeared in the corpus; examples (2a) and (2c) were derived from it, to exemplify that all three options are possible with the same sentence.
mon analysis is one in which the reporting clause is the main clause and the reporting verb is the main verb (mv). Accordingly, in this analysis the direct object typically is realised by a discontinuous DIP-sentence; illustrated by example (2b), this analysis can be depicted as in figure 1.

<table>
<thead>
<tr>
<th>Analysis I:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Ik geloof,&quot; zei Annie, &quot;dat je wat moet uitleggen.&quot;</td>
</tr>
<tr>
<td>direct object1 mv subject direct object2</td>
</tr>
</tbody>
</table>

Figure 1: Direct object analysis

This analysis is attractive because the reporting verbs most frequently used (like ‘zeggen’ (to say) and ‘antwoorden’ (to reply)) are transitive verbs in Dutch. As we see in figure 1, with these transitive verbs a direct object is missing in the reporting clause. This problem may be solved by postulating that the quote fulfills this function. When checking this analysis against the corpus data, however, I found major reasons for rejecting this analysis. These are discussed below:

- Not all reporting clauses contain a verb. In Dutch, the *aldus*-construction for reporting clauses in sentence-medial and sentence-final position is quite frequent; for instance:

(3) “Maar,” aldus dokter Hick, “de gemeenschap is hier “but” according to doctor Hick “the community is here sterk.””

“‘But,” according to doctor Hick, ”the community is strong here.””

Whatever category *aldus* may be, it certainly does not allow for a direct object. A construction like

* aldus dokter Hick iets
  according to doctor Hick something

is impossible in Dutch; the *aldus*-construction in non-parenthetical context does not permit any further arguments. This implies that the quote in DIP-constructions cannot be an argument of the *aldus*-construction. As the reporting clause does not contain a verb, the quote could never be its direct object either (nor any other syntactic function).

- If the reporting clause contains a verb, it need not be a transitive one; ‘smalen’ (to revile), ‘grijnzen’ (to grin) and other intransitive verbs are found to occur frequently in reporting clauses.
(4) “Hij is te zwaar,” lacht Huisman, “er moeten nog
he is too heavy” laughs Huisman “there have to another
zeker vijf kilo af.”
至少 five kilos off"
“‘He is too heavy,’” Huisman laughs, “he still has to lose at least
another five kilos.’”

An intransitive verb in a non-parenthetical sentence does not permit a direct object with it;³

* Gisteren lachte Huisman iets.
Yesterday laughed Huisman something

As intransitives do not permit direct objects, it follows that the quote cannot
be a direct object. Nor can it be the subject of the reporting verb, cf. example
(4): here ‘Huisman’ is the subject. Other syntactic functions are impossible
as well; intransitive reporting verbs do not permit any other obligatory roles
and optional roles can only take the form of a prepositional phrase (PP),
e.g. ‘grijnzen naar’ (to grin at) or ‘lachen tegen’ (to laugh to). The quote,
however, does not contain a preposition, nor does the reporting clause con-
tain a provisional prepositional object ‘ertegen’.

- If the reporting verb does take an object, it need not be a direct object. While
verbs such as ‘twijfelen’ (to doubt) and ‘piekeren’ (to worry) do not permit
a direct object, they do permit a prepositional object.

(5) “We weten niet (...),” piekert Tilstra hardop, “en eerlijk
we know not (...)” worried Tilstra out loud “and honestly
gezegd zit er (...)”
said sits there (...)
“We don’t know (...),” Tilstra worried aloud, “and to be honest,
there is (...)

In Dutch, the object of ‘twijfelen’ is a prepositional object with ‘aan’; ‘to
doubt X’ is ‘twijfelen aan X’. Also, ‘to worry about X’ is ‘piekeren over X’.
These verbs do not allow objects without prepositions:

* Hij piekerde zijn probleem.
He worried his problem

* Hij piekerde over zijn probleem iets.
He worried about his problem something

³Instances with excorporated objects like ‘he laughed his loud laugh’ are not discussed here.
Hij piekerde over zijn probleem.
He worried about his problem

In example (5), the quote does not start with the preposition ‘over’; neither does the reporting clause contain a provisional prepositional object ‘erover’. Therefore, it is not very likely that the quote would be either direct object or prepositional object of the reporting verb ‘piekerde’.

- If the reporting clause contains a transitive verb, this transitivity need not be related to the quote. Especially in fiction texts we come across so-called verbs of movement or action verbs in the reporting clause. All syntactic arguments of these verbs are realised within the reporting clause, so there is none that could accommodate the quote.

(6) “Sorry meneer,” duwde Yvette de arm van haar schouder,
“sorry mister” pushed Yvette the arm off her shoulder
“maar wij hoeven even niks.”
“but we need now nothing”
“’Sorry sir,” Yvette pushed his arm off her shoulder, “but we don’t need anything now.”

The reporting clause contains a subject ‘Yvette’, a direct object ‘de arm’ and in this example even an optional prepositional phrase is supplied: ‘van haar schouder’. There is no syntactic function left for the quote. In a non-parenthetical context, the reporting clause could not be extended by a pronominal form like ‘iets’.

* Yvette duwde zijn arm van haar schouder iets.
Yvette pushed his arm off her shoulder something

While a reporting clause in non-parenthetical context could be extended by pronominal forms of place and time, the quote in parenthetical context could not take their position, as these forms can remain in their position in a DIP-construction.

Yvette duwde ooit/ergens zijn arm van haar schouder.
Yvette pushed ever/somewhere his arm off her shoulder

“Sorry meneer,” duwde Yvette ooit/ergens de arm van
“sorry mister” pushed Yvette ever/somewhere the arm off haar schouder, “maar ...
her shoulder “but

These observations lead to the conclusion that action verbs in the reporting clause do not have any argument position left. Henceforth, the quote could not fulfil the direct object role or any other argument role with the reporting clause.
These four arguments together suffice to show that the quote cannot be the direct object of the reporting verb. For reasons similar to the ones mentioned above, the quote cannot occur in any other thematic role of the reporting verb; an analysis of the quote as the subject, indirect object or prepositional object of the reporting verb is not a realistic option. But if the quote does not have a thematic relationship to the reporting clause, what else could the relation be?

3.2 Analysis II: The reporting clause fulfills a thematic role in the quote

In section 3.1 we rejected the idea that the quote would be the direct object of the reporting verb. It was also observed, that the idea of the quote as realizing any other syntactic argument of the reporting verb had to be rejected as well. If the quote does not have a syntactic function in the reporting clause, two options remain: either the reporting clause has a syntactic function within the quote (analysis II, as depicted in figure 2), or there is no syntactic relation between the reporting clause and the quote; in the latter case, one of them would be adjoined to the other (analysis III).

<table>
<thead>
<tr>
<th>Analysis II:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Ik geloof,&quot; zei Annie, &quot;dat je wat moet uitleggen.&quot;</td>
</tr>
<tr>
<td>subj mv argument object-subordinate clause</td>
</tr>
</tbody>
</table>

Figure 2: Reporting clause as part of the citation

There are two good reasons for rejecting analysis II immediately: in the first place, our corpus data show no examples of quotes lacking a syntactic argument. Hence, there is no (argument) role left for the reporting clause. The other point is, that quotes can occur without reporting clauses. As long as the speaker is clear from context, a well-written dialogue does not need reporting clauses. Quotes can also consist of more than one sentence, while a reporting clause always occurs with a single sentence. According to the analysis that the reporting clause would realize a syntactic argument role in the quote, all these sentences would lack an obligatory role and hence be ungrammatical; they obviously are not. These arguments together suffice to reject analysis II.

Of course sentences may consist of more elements than verbs and arguments; they can also contain adverbial clauses, vocatives, interjections, parenthetical clauses and other non-obligatory constituents. Reporting clauses might be comparable to these sentence elements. I will not discuss this possibility here, as these constituents are usually assumed to be adjoined to the main clause, and the possi-

This may be hard to see, as quotes may lack verbs, so that the justification for other sentence elements is unclear; for instance

"Acht woorden," telt Ricardo Aenemaet, "in 35 minuten."
"eight words" counts Ricardo Aenemaet "in 35 minutes."

In these cases, we judged that there was no reason to assume the absence of a syntactic argument.
bility of adjunction of quote and reporting clause is already dealt with in the next section.

3.3 Analysis III: the adjunction-option

If we accept that the quote cannot realise a syntactic argument role with the reporting clause (analysis I) and the reporting clause cannot realise a syntactic argument role with the quote either, the only possible conclusion is, that reporting clause and quote cannot be syntactically related to each other. Therefore, one of them has to be adjoined to the other. This is a more attractive analysis, as it allows us to relate reporting clauses in DIP-sentences to other parenthetical clauses that, according to Ross (1973), are adjoined to the main clause. In line with the analysis of these clauses, and in view of the fact that reporting clauses provide additional information to the quote instead of vice versa, analysis III states that the reporting clause is adjoined to the quote, as illustrated in figure 3.

<table>
<thead>
<tr>
<th>Analysis III:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Ik geloof;&quot; zei Annie, &quot;dat je wat moet uitleggen.&quot;</td>
</tr>
<tr>
<td>quote part 1  adjoined rc quote part 2</td>
</tr>
</tbody>
</table>

Figure 3: Adjoined clauses

While this analysis solves the problems that led to the rejection of analysis I and II, it causes some other, serious difficulties. In the next sections I discuss the most important ones and present solutions.

3.3.1 Two major difficulties with the adjunction-analysis

With the adjunction-analysis none of the problems arise that occur under the assumption that the quote would realise a syntactic argument role of the reporting verb. Instead two other problems come to light: with some reporting verbs syntactic arguments are missing and within the reporting clause inversion may occur.

- The missing of a direct object with obligatorily transitive reporting verbs is exemplified in example (7):

(7) "Een zeker evenwicht (...)," legt Bolle uit, "a kind of harmony (...)" explains Bolle PARTICLE "bevordert de kwaliteit (...)" "improves the quality (...)"

"Some kind of harmony (...)," Bolle explains, "improves the quality (...)"

In Dutch, ‘uitleggen’ (to explain) is obligatorily transitive, optionally ditransitive. In non-parenthetical clauses, it is impossible to leave out the direct object:
Bolle legt (iemand) iets uit
Bolle explains (to someone) something

* Bolle legt (iemand) uit
Bolle explains (to someone)

In example (7), however, the reporting clause only contains the subject ‘Bolle’, while the direct object is missing. This phenomenon occurs in all reporting clauses containing transitive reporting verbs. In analysis I this problem does not occur, because the quote is seen as the direct object. In an adjunction-analysis, the missing role will have to be explained.

- Within the reporting clause inversion may occur. This phenomenon is found to occur with all DIP-sentences in which the reporting clause contains a verb. For example,

(8) a. “Omdat”, zei raadslid J. van Erp, “het voorstel “because” said counsel-member J. van Erp “the proposition niet (...)”
not (...)

“Because,” counsel-member J. van Erp said, “the proposition does not (...)”

Here we notice that the verb ‘zei’ precedes the subject ‘raadslid J. van Erp’. In Dutch, however, the finite verb usually occupies the second position in the sentence. The only possibility for the verb to occur in first position is in some special kinds of subordinate clauses, wishes, imperative clauses or questions; reporting clauses do not belong to any of these. In analysis I, this problem does not arise since under the assumption that the quote is part of the reporting clause, ‘omdat’ occupies the first position and ‘zei’ occurs in second position. But how can this position of the verb be explained in the adjunction-analysis?

3.3.2 A solution for both problems

At this point, all DIP-sentences in the corpus that contain reporting verbs were inspected in order to determine what occurs in the first position in the reporting clause. What we found was that apparently there were only two possibilities: either in first position there was nothing (or: an element that is phonetically empty) or in first position the word ‘zo’ was found—compare for instance example (8a) to example (8b), from which it was derived:

(8) b. “Omdat”, zo zei raadslid J. van Erp, “het voorstel “because” so said counsel-member J. van Erp “the proposition niet (...)”
not (...)

not (...)
"Because," so counsel-member J. van Erp said, "the proposition does not (...)"

The difference between examples (8a) and (8b), the word 'zo', does not make any difference for either the meaning or the acceptability of this sentence. This seems to be the case for all DIP-sentences containing reporting verbs: wherever 'zo' is found in first position it can be left out and wherever zero is found in first position we could insert 'zo', without changing either the meaning or the syntactic acceptability of the sentence.

However, looking through the corpus I find no examples of DIP-sentences with 'zo' appearing in any other position in the reporting clause. 'Zo' seems to be unacceptable in other positions:

(9) a. "Daarom", zo legt Adriaans uit, "is het heel belangrijk that's why" so explains Adriaans PARTICLE "is it very important dat (...)" that (...)"
   "That's why," so Adriaans explains, "it is so important to (...)"
   b. * "Daarom", legt zo Adriaans uit, "is het heel belangrijk dat (...)".
   c. * "Daarom", legt Adriaans zo uit, "is het heel belangrijk dat (...)".
   d. * "Daarom", legt Adriaans uit zo, "is het heel belangrijk dat (...)"

These observations lead to the idea that the first position of the reporting clause in DIP-sentences is always occupied by an abstract ZO, whether ZO is made explicit or not. This explains the inversion: if the quote and the reporting clause are adjunctions to each other, each has its own form. The first position of the reporting clause is not occupied by the first part of the quote, as analysis I states, but is always occupied by ZO—independent of whether ZO is left implicit or is made explicit. The finite verb occupies the second position as usual, the subject follows it and we have inversion.6

But what could be the function of this ZO? In order to establish its nature, I came to divide reporting clauses with a reporting verb and ZO over all possible

5I will use ZO instead of 'zo' to prevent misunderstandings. 'Zo' can fulfil a lot of functions in Dutch; it is used as conjunction, adverb of measure and can fulfil even other functions. ZO is in my opinion a unique variant; by ZO I mean the word 'zo' in the context and function explained above.

6The observation that ZO, when used explicitly, always takes the first position in the reporting clause suggests, that ZO takes an argument position. This idea raises the question, whether ZO or explicit 'zo' somehow takes over the direct object role, as is suggested by the observation, that ZO cannot be made explicit in all contexts; compare for instance:

(6') * "Sorry meneer," zo duwde Yvette de arm van haar schouder, "maar ..."
   "Sorry mister," SO pushed Yvette the arm from her shoulder, "but ..."

(7') "Een zeker evenwicht (...)," zo legt Bolle uit, "bevordert ...
   "A certain balance (...)," SO explains Bolle PARTICLE "improves ..."

Unfortunately, this question could not be dealt with within the scope of this article. It is discussed at length in my doctoral thesis about DIP-constructions.
positions for a reporting clause: preceding the quote, interrupting it or following it (possibility (a), (b) and (c) in section 2). Examples are shown in (10), (11) and (12); (11) and (12) occur with explicit ‘zo’ in the corpus, (10) has been made up.

(10) * zo zei hij: “Het is fijn, jullie te zien.”
    SO said he “it is nice you to see”
    ‘SO he said: “It’s nice to see you.”’

(11) “We hebben,” zo zegt hij, “weer een eigen legioen.”
    “we have” SO says he “again a our-own legion”
    “We have,” SO he says, “a legion of our own again.”

(12) “Kiefer,” zo werd herhaald.
    “Kiefer” SO was repeated
    “Kiefer,” SO was repeated.’

(Readers who know Dutch might observe here, that (10) (and perhaps (9c) as well) is a correct Dutch sentence. This is true, if explicit ‘zo’ is analysed as an adverb, pointing back to an earlier sentence. Zo is not possible here; it has no relation to the quote and it cannot be left out without changing the order of the verb and the subject.)

On the basis of examples (10)-(12) and other examples in the corpus, I want to suggest that zo is only possible when at least a part of the quote has been mentioned. Zo must have a referent; therefore, it probably has an anaphoric function, pointing to the quote. Its function would then be comparable to non-reflexive pronouns like he/him. When we accept an analysis of zo as a free anaphor, sentence (10) is correctly ruled out by principle c of the anaphor-theory, stating that a non-anaphor (the quote) cannot be bound. Contrary to examples (11) and (12), in example (10) zo would c-command the complete quote and bind it. Zo cannot be just a deictic element, as deictic elements are perfectly capable of c-commanding their referent\(^7\); one example would be the (invented) sentence

(13) Hij zei dit/het volgende: “...”
    he said this/the following
    ‘He said this/the following words: “...”’

Thus, if we accept the analysis of zo as a free anaphoric element, it follows that the direct object role is associated with its antecedent via zo. Under these circumstances, the necessity to make this object explicit disappears. This then explains the fact that direct objects appear to be missing in reporting clauses with an obligatory transitive reporting verb.

\(^7\)My attention was drawn to this argument by an anonymous reviewer, for which I would like to thank her/him.
4 Conclusion

In the previous sections we considered three possible analyses of the relation between quotes and reporting clauses in DIP-sentences. The analysis of the quote as the direct object of the reporting verb and that of the reporting verb filling an argument position of the main verb of the quote were rejected on the basis of corpus counterevidence. The final hypothesis, that the reporting clause is adjoined to the quote, was found to be problematic because of the disappearance of obligatory direct objects and inversion in the reporting clause. By postulating the occurrence of an abstract zo in first position in reporting clauses, which may or may not be explicit and stands in an anaphoric relation to the quote, these problems were solved and the adjunction hypothesis became plausible.

By this analysis, not all questions about zo have been answered yet; questions that remain include the following:

- what exactly is the nature of the relation between zo and the quote, and between zo and a possible direct object in the reporting clause;
- are there contexts in which zo could never become explicit, and if so, which are these contexts and why zo cannot become explicit;
- can zo be unified with the quotative operator Collins and Branigan (1997) propose in their analysis of quotative inversion;
- is there a relation between zo and other functions of explicit ‘zo’ in Dutch?

Also with the adjunction-analysis itself there are some questions left unanswered; these include:

- at which positions in the quote can a reporting clause occur, where is it not possible, and why is this so;
- what is the relation of DIP-sentences to reporting constructions where the reporting clause precedes the quote (category (a) in section 2);
- what is the nature of aldus-constructions?

These questions could not be answered within the scope of this article. Some of them are dealt with in the doctoral thesis this article is based on; others remain as yet unsolved. Nevertheless, of the three analyses we discussed here I consider the adjunction-analysis the better option by far. Despite the questions that remain, this analysis is the one that will be implemented in AMAZON; in testing the implementation on the corpus, some answers to the questions above might come to light.

Acknowledgements

I wish to thank Peter-Arno Coppen, Simon van Dreumel, Hans van Halteren and Nelleke Oostdijk for many helpful comments.
References


Collins, C. and Branigan, Ph. (1997), Quotative Inversion, Natural Language and Linguistic Theory 15, 1-41.


Hoek, Th. van den (1971), Woordvolgorde en constituentenstructuur, Linguistics 1, 125-136.


