Syntactic analysis for ellipsis handling in coordinated clauses

Ralph Moreira Maduro*
State University of Campinas, Institute of
Computing, Brazil

Ariadne M. B. R. Carvalho†
State University of Campinas, Institute of
Computing, Brazil

This work is intended as an investigation into elliptical phenomena in natural language. We argue that some types of ellipsis can be resolved at the syntactic level since they are subject to syntactic constraints. We have dealt with four of the major types of ellipsis found in Portuguese, namely: Null VP, Gapping, Stripping and Sluicing. We have used Island Constraints in order to decide on the grammaticality of the sentence. Finally, we have developed and implemented a syntactically-based algorithm that recovers the ellipted constituents and reconstructs the elliptical clause, when applicable. The linguistic data in this work is drawn primarily from Portuguese, but we believe that the results can also be applied to other languages, such as English.

1 Introduction

Ellipsis structures pose an important problem for Natural Language Processing systems designed to provide text understanding, text generation or dialogue handling. Ellipsis is a grammatical phenomenon whereby the structure of the sentence is abbreviated, avoiding redundancy: the sentence, thus, contains a grammatical omission (Quirk et al., 1985). Although ellipsis may in general be regarded in semantic or pragmatic terms as a means of avoiding redundancy of expression, the kinds of reduction which are allowed are largely a matter of syntax. The fundamental problem posed by an elliptical construction is, therefore, to recover the ellided constituent; the actual word(s) whose meaning is understood or implied must be recoverable.

There seems to exist two main approaches to ellipsis resolution (Lappin, 1995). Whereas the first tries to associate an elliptical construction directly with a semantic representation, the latter mediates semantic interpretation through the reconstruction of the syntactic structure of the antecedent. We propose an algorithm which implements the second view of ellipsis. We have dealt with sentences involving ellipsis and coordination simultaneously, because the association between the two phenomena is so close that we cannot understand one without understanding the other. The criteria for ellipsis are (Quirk et al., 1985):

1. The ellipted words are precisely recoverable;

2. The elliptical construction is grammatically defective;

3. The insertion of the missing words results in a grammatical sentence, with the same meaning as the original one;

---

* email: rmaduro@uol.com.br
† email: ariadne@ic.unicamp.br

4.The missing words are recoverable from the neighbouring text; and

5.The missing words are an exact copy of the antecedent.

Ellipsis is typically postulated in order to explain why some normally obligatory element of a grammatical sentence is missing. In a context where no ambiguity of reference arises, there is no doubt as to what words are to be supplied. Consider the following two sentences:

- João tinha dado esses livros ao filho e Maria também tinha [_]. (1)
  [_] = dado esses livros ao filho.
  John had given theses books to his son and Mary had too.

- João gosta de cinema e Pedro [_] de teatro. (2)
  [_] = gosta

  John likes movies and Peter theater.

In (1) the verb complement "dado esses livros ao filho" is missing, which denotes a defective construction. Nevertheless, the ellipted words are precisely recoverable from the neighbouring text and are an exact copy of the antecedent. Therefore, the insertion of the missing words results in a grammatical sentence, with the same meaning as the original one. In (2) only the verb "gosta" is missing, but it is recoverable from the neighbouring text. Since the five criteria for ellipsis apply, the sentence is considered grammatical.

These criteria undoubtedly help to decide on the grammaticality of the sentence through the reconstruction of the elliptical clause. However, when we have a sentence such as

- * João conquistou a confiança de seu chefe e Maria não admite a hipótese de
  que Pedro também [_]. (3)
  *John has gained his boss's confidence and Mary doesn't admit the hypothesis
  that Peter too.*

we cannot reconstruct the elliptical clause using the words "conquistou a confiança de seu chefe" because sentence (3) is ungrammatical at first place. This is due to the fact that some types of ellipsis are subject to syntactic constraints which determine when neighbouring text can be used to fill the gap of the elliptical clause.

In general the kinds of ellipsis vary to some extent from one language to another. Specifically for Portuguese, De Matos (Matos, 1992) has identified five types of ellipsis and studied two of them in detail, namely Null VP and Stripping. She concluded that whereas Stripping is subject to Island Constraints (Haegeman, 1992), Null VP is not. She has shown that besides following the five criteria for ellipsis reconstruction, we must also take these constraints into account when dealing with Stripping and before reconstructing the elliptical clause.

Following De Matos's approach, we have examined two other ellipsis occurrences in Portuguese, namely Gapping and Sluicing regarding syntactic constraints. We have shown that these two types of ellipsis are also subject to Island Constraints and, therefore, during sentence reconstruction these constraints must be respected. We have developed and implemented an algorithm which takes the five criteria for ellipsis and the Island Constraints into account in order to reconstruct the elliptical clause.

The remaining part of this paper is organized as follows. The next section describes other approaches to ellipsis resolution. Section 2 presents the five major types of ellipsis found in Portuguese. Section 3 discusses how syntactic constraints can be used to decide

on the grammaticality of elliptical sentences. Section 4 presents a syntactically-based algorithm to recover the ellipted constituents. Finally, section 5 presents our concluding remarks.

2 Other approaches to ellipsis

There are several other approaches to ellipsis resolution. Dalrymple et al. (Dalrymple, Shieber, and Pereira, 1991) present a generalized semantic approach which employs higher-order unification of property and relation variables to resolve ellipsis. The method presupposes a semantic representation of the antecedent clause and it is argued that the antecedent and the elliptical clause share the same property which, when applied to both clauses, allows for the correct interpretation of the sentence. The strategy is to specify the interpretation of the antecedent clause as an equation between a propositional variable and a predicate argument structure. The arguments of the predicate correspond to the fragments in the ellipsis site, and the ellipsis resolution consists in finding an appropriate value for the predicate variable which can apply to both the sequence of arguments in the interpretation of the antecedent clause, and the sequence of arguments in the ellipsis site.

However, as Lappin (Lappin, 1995) pointed out, it is not clear how higher-order unification can be applied to sentences like

- John sings, and beautifully too. (4)

where there is no corresponding element in the antecedent clause.

Lappin (Lappin, 1995) suggested to point a free manner adverbial function variable in the lexical semantic representation of verbs like "sing". This approach will allow for the correct semantic interpretation of sentence (4), but it still cannot be generalized to sentences such as

- John sang, but not in New York. (5)

Lappin (Lappin, 1995) presented a syntactically-based algorithm to deal with the following types of ellipsis: VP ellipsis, Pseudo-gapping, Stripping and Gapping. The algorithm treats ellipsis resolution as the specification of a relation of correspondence between an unrealized verbal head of an elliptical clause and its arguments and adjuncts as one term of the relation, and the realized head of the antecedent clause and its arguments and adjuncts as the second term. When analysing sentence (4), for example, the algorithm will identify "sings" as the head of the antecedent clause and substitute it for the empty verb. This will produce the following sentence:

- John sings and John sings beautifully too. (6)

Kehler's approach (Kehler, 1995) is based on the following discourse relations: cause-effect and resemblance. He uses these two relations to identify which method should be used for ellipsis resolution. The resemblance relation, for example, requires diverging or converging points between the two involved clauses. He argues that in a resemblance relation the entities present in both clauses share the same property, that is, they act in a similar way in the information context. In a cause-effect relation, on the other hand, the two clauses do not have to share the same property, but there must exist an implication relation between the two clauses, that is, they must be interdependents.

Based on these discourse relations, Kehler proposed a method to identify if either syntactic or semantic analysis should be used to ellipsis resolution. If a resemblance

relation holds between the two clauses, then a syntactic approach to ellipsis must be adopted; if, on the other hand, a cause-effect relation holds between the two clauses, then a semantic approach must be taken. He argues that when the identity between the two clauses is semantic, then a syntactic structure of the antecedent as well as syntactic restrictions are not necessary. When a resemblance relation holds between two clauses, the sentence is subject to syntactic restrictions. If there is not an adequate syntactic structure to recover the elliptical constituent, the sentence is considered ungrammatical.

3 Major types of ellipsis in Portuguese

According to De Matos (Matos, 1992), the major kinds of ellipsis found in Portuguese are: Null VP, Gapping, Stripping, Sluicing and Conjunction Reduction. The difference lies on the type of the structure of the missing constituent.

3.1 Conjunction Reduction

In a Conjunction Reduction occurrence of ellipis a subject noun phrase and, eventually, a verbal constituent are ellided from the sentence. In the following sentence

- João tem comprado muitos livros aos filhos e [_] oferecido muitas flores à mulher. (7)
  [_] = [João tem]

  John has bought many books to his children and ofered many flowers to his wife.

the subject ("João") and the auxiliary verb ("tem") are ellided from the second clause.

3.2 Gapping

In a Gapping occurrence a verb and, optionally, its complements are ellipted, but two other constituents are lexically realized, one of them being usually the subject. In the following sentence

- João deu flores a sua mãe e Pedro [_] chocolates [_]. (8)
  [_] = [deu]
  [_] = [a sua mãe]

  John gave flowers to his mother and Peter chocolates.

verb "deu" and its complement "a sua mãe" are ellipted from the sentence.

3.3 Sluicing

In a Sluicing occurrence an interrogative constituent remains lexically realized as the only representative of a clause. Consider the following sentence:

- Alguém veio lhe procurar, mas eu não sei quem [_] (9).
  [_] = "veio lhe procurar".

  Someone came looking for you, but I don't know who.

The interrogative pronoun "quem" represents the ellided words "veio lhe procurar".

3.4 Null VP

In a Null VP[1] occurrence the verb or an auxiliary verb, when an auxiliary verb is present in the first clause, and an adverb, are lexically realized in the elliptical clause. Consider the following sentence:

---

1 Sometimes also called VP deletion

- Maria atribuiu a culpa do desastre ao motorista e Tereza também atribuiu [_].
  (10)
  [_] = [a culpa do desastre ao motorista].

  Mary blamed the driver for the disaster and so did Theresa.

the verb in both coordinated clauses is identical ("atribuiu"); also, an adverb ("também")
is present in the elliptical clause.

3.5 Stripping

In a Stripping occurrence all constituents, except one and an adverb, are missing. In
Portuguese we can find the adverbs "não", "sim", "também" and "também não", whose
presence in a Stripping ellipsis is compulsory; their function is to recover the constituent
which is the predicate of the elliptical clause. Consider the three sentences below:

- Maria atribuiu a culpa do desastre ao motorista e Teresa também [_]. (11)
  [_] = [atribuiu a culpa do desastre ao motorista].

  Mary blamed the driver for the disaster and Theresa did too.

- Maria atribuiu a culpa do desastre ao motorista e [_] a fuga dos assaltantes
  também. (12)
  [_] = [Maria atribuiu]

  Mary blamed the driver for the disaster and the assailant's escape too.

- Maria ouve sempre o noticiário à hora do almoço e [_] à hora do jantar
  também. (13)
  Mary always hears the news at lunch and at dinner time too.

  [_] = [Maria ouve sempre o noticiário]

  Mary always hears the news.

In (11) a verb phrase ("atribuiu a culpa do desastre ao motorista") is missing; in
(12) a suject followed by a verb ("Maria atribuiu") is ellided from the sentence; in (13)
a noun phrase followed by a verb phrase ("Maria ouve sempre o noticiário") are ellided
from the second clause.

4 Syntactic Constraints on Ellipsis

The fundamental problem of elliptical constructions is to recover the elliptical con-
stituents. De Matos (Matos, 1992) has studied Null VP and Stripping in detail and
observed that, although these types of ellipsis seem to be very similar on the surface,
they are very different when syntactic constraints are concerned. They both require a
linguistic antecedent and a lexically realized adverb in order to be grammatical. Consider
the following two sentences:

- Maria tinha atribuído a culpa do desastre ao motorista e Teresa também tinha
  [_]. (14)
  [_] = [atribuído a culpa do desastre ao motorista].

  Mary had blamed the driver for the disaster and Theresa had too.

- Maria tinha atribuído a culpa do desastre ao motorista e Teresa também [_].
  (15)
  [_] = [tinha atribuído a culpa do desastre ao motorista].

  Mary had blamed the driver for the disaster and Theresa too.

Sentences (14) and (15) present an elliptical predicate and, although both predicates involve a VP, the structure of the two sentences is different. Sentence (14) is an example of Null VP ellipsis because a constituent ("Teresa"), an adverb ("também") and an auxiliary verb ("tinha") are lexically realized. Sentence (15) is an example of Stripping since only one constituent ("Teresa") and an adverb ("também") are realized in the elliptical clause.

De Matos observed that only when attempting to recover elliptical constituents in a Stripping occurrence we must take Island Constraints into account. Therefore, Stripping must follow the Island Constraint which is stated as follows:

When a constituent is moved, it must cross the minimal number of barriers, preferably none (Matos, 1992).

Traditionally this constraint is used to restrict movement of constituents within a sentence (Chomsky, 1986). De Matos has shown that the same principle can be applied, in a similar manner, during the search for an antecedent which can be used to reconstruct the elliptical clause in an ellipsis resolution process.

Consider the following sentences:

- * Que João vá ao cinema hoje é bom, mas $[_{IP}[_{CP}$ que Maria não [_]] é pssimo]. (16)

  That John goes to the movies today is good, but that Mary doen't is awful.

- Que João tenha ido ao cinema é bom, mas $[_{IP}[_{CP}$ que Maria não tenha [VP _]] é péssimo]. (17)

  That John has been to the movies today is good, but that Mary hasn't is awful.

Sentence (16) is an example of Stripping. It is ungrammatical because, since this type of ellipsis is subject to Island Constraints, the antecedent "vá ao cinema hoje" cannot be used to fill the gap in the elliptical clause because in order to do that more than one barrier would have to be crossed. In (17), on the other hand, we have a Null VP which is not sensitive to Island Constraints. Therefore, we can use the antecedent "ido ao cinema" to fill the gap in the elliptical clause and the sentence is considered grammatical.

The following sentences are examples of Stripping and Null VP ellipsis. Whereas the Stripping manifestations are ungrammatical, the Null VP ellipsis are not. In (18) and (19) we have a complex NP in a relative clause. In (18) we cannot use the constituent "falado japonês" to fill the gap in the elliptical clause because this would infringe Island Constraints. Sentence (19) is a Null VP occurrence and, therefore, the words "falado japonês" can be used to fill the gap of the elliptical clause because this type of ellipsis is not sensitive to Island constraints.

- * João fala japonês e eu conheço $[_{NP}$ um aluno $[_{CP}$ que também $[_{VP}$_]]]. (18)

  John speaks Japanese and I know a student who too.

- João tem falado japonês ultimamente e eu conheço $[_{NP}$ um aluno $[_{CP}$ que também tem $[_{VP}$ _]]]. (19)

  John has spoken Japanese lately and I know a student who has too.

In sentence (20) below we have a complex NP, and, again, we cannot use "está doente" to fill the gap of the elliptical clause, because the Island Constraints would be violated.

- * João está doente e Maria não admite [$_{NP}$ a hipótese [$_{CP}$ de que ela também [$_{VP}$-]]]. (20)

  John is ill and Mary doesn't admit the hypothesis that she too.

- João está doente e Maria não admite [$_{NP}$ a hipótese [$_{CP}$ de que ela também esteja [$_{VP}$-]]]. (21)

  John is ill and Mary doesn't admit the hypothesis that she is too.

In sentence (21), on the other hand, the antecedent "doente" is used to fill the gap in the elliptical clause and the sentence is considered grammatical.

Basing our work on De Matos's approach, we have analysed two other types of ellipsis - Gapping and Sluicing - regarding Island Constraints applied to ellipsis resolution. Consider the following example of Gapping ellipsis:

- * João perguntou [$_{CP}$ o que você comeu hoje] e Pedro [$_{VP}$-] ontem. (22)

  John asked what you have eaten today and Peter yesterday.

This example is ungrammatical because, although it is in accordance with the definition of gapping, the elliptical clause cannot be reconstructed with the subconstituents "perguntou o que você comeu hoje" since this would represent a violation of the Island Constraint. Therefore, sentence (22) is considered ungrammatical.

Consider now the following example of Sluicing ellipsis:

- *[$_{IP}$ [$_{CP}$ Que João vá ao cinema] é bom], mas com quem [-].

  That John goes to the movies today is good, but with whom.

In order to fill the gap left in the elliptical clause, we would have to violate the Island Constraint. Therefore, the sentence is considered ungrammatical.

Our conclusion is, therefore, that Gapping and Slucing are also subject to Island Constraints. Therefore, a syntactically-based system to ellipsis resolution must take these constraints into account.

5 An Algorithm for Ellipsis Resolution

We have developed an algorithm which deals only with sentences involving coordination and ellipsis simultaneously and which takes Island Constraints into account in order to reconstruct the ellided material.

The algorithm works in the following way:

1. Decomposing the sentence into syntactic structures;

2. Identifying the type of ellipsis present in the sentence;

3. Checking if this type of ellipsis is subject to syntactic constraints;

4. Identifying the antecedent of the ellided term; and

5. Reconstructing the ellided constituent.

Based on this algorithm we have developed a system, whose architecture is shown in Figure 1.

First, the system decomposes the sentence into syntactic structures using a syntactic parser which deals with elliptical constructions. The grammar, thus, allows ellided constituents wherever the four types of ellipsis treated in this work would. So, for example, the syntactic analyser works on the following sentence
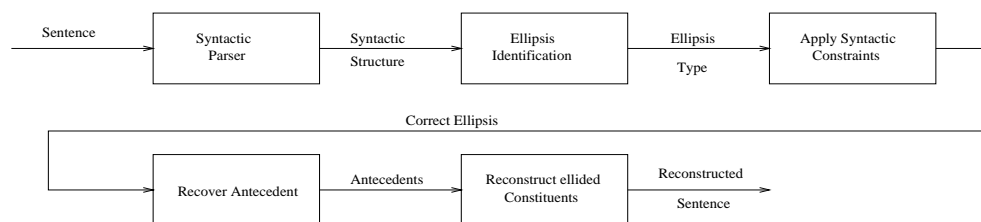
Figure 1
System's architecture for ellipsis handling

- João fala japonês e Carlos também fala []. (24)

  John speaks Japanese and Charles speaks too.

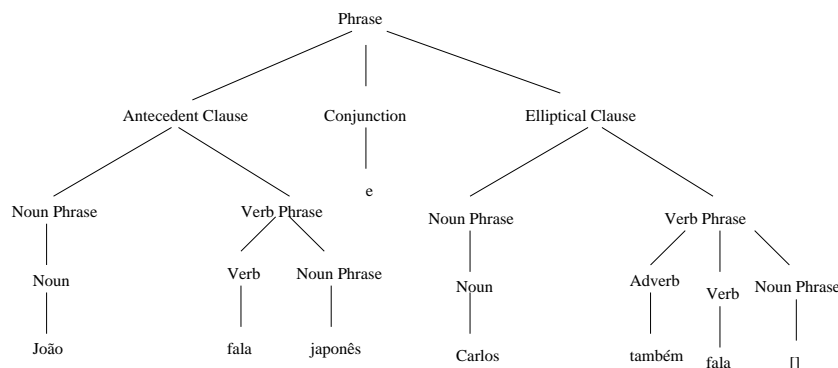producing the syntactic structure shown in Figure 2.

Figure 2
An example of Null VP ellipsis

The next step is to identify the kind of ellipsis present in the sentence. In (24) the system identifies that the constituents which have been lexically realized in the elliptical clause are the noun phrase "Carlos", the adverb "também" and the verb "fala". This configures a Null VP ellipsis. Since this type of ellipsis is not subject to Island Constraints the system recovers the antecedent (that is the noun phrase "japonês") and reconstructs the sentence using it to fill the gap in the elliptical clause.

Sentence (25), on the other hand, is an example of stripping ellipsis, since the lexically realized terms are "quem" and the adverb "também".

- * João vai ao cinema e Maria perguntou [$_{CP}$ quem [$_{IP}$ também [_]]]. (25)

  John is going to the movies and Mary asked who too.

The system identifies this is an Island context because CP constitutes a barrier and, therefore, reconstruction does not take place and the sentence is considered ungrammatical.
Gapping is also subject to Island constraints. Consider now the following sentence, which is an example of Gapping ellipsis:
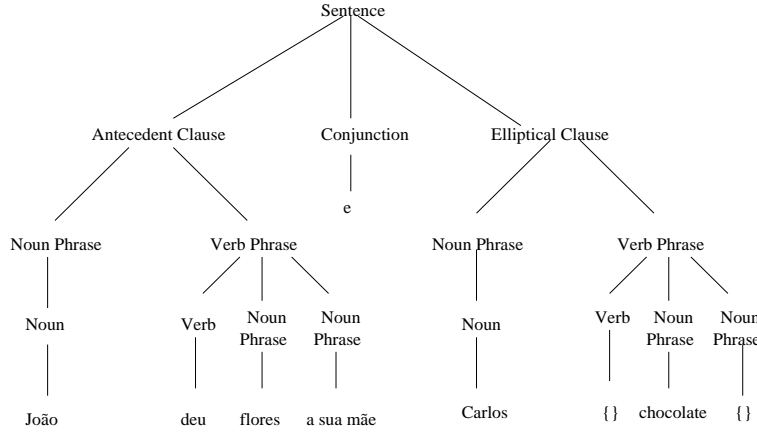
Figure 3
An example of Gapping ellipsis

- João deu flores a sua mãe e Carlos [_] chocolates [_]. (26)

  John gave flowers to his mother and Charles chocolates.

the constituents which are lexically realized are "Carlos" and "chocolates". The correspondent syntactic structure generated by the system is shown in Figure 3.

Verb "dar" and "a sua mãe" are the antecedents present in the first clause. The elliptical sentence is reconstructed because the missing terms are not inside an Island context. Therefore the sentence is reconstructed as shown in figure 4.
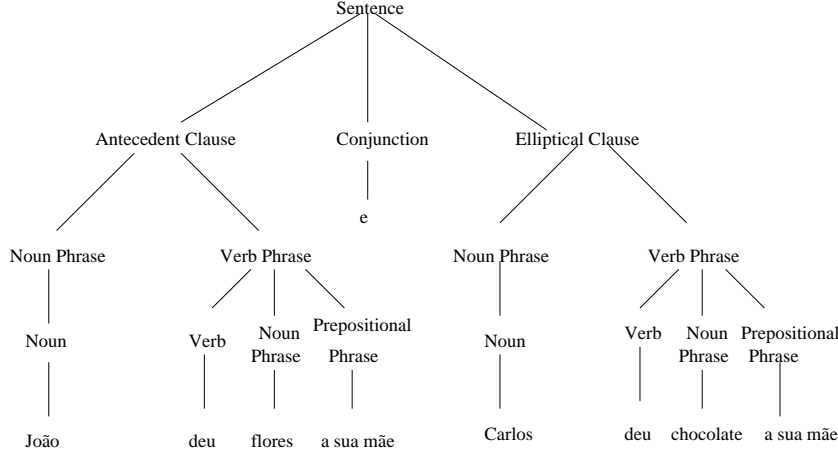

Figure 4
Gapping ellipsis reconstruction

Finally, sentence (27) is an example of Sluicing ellipsis. The pronoun "quando" represents the antecedent clause "os garotos sairão", as it is shown in Figure 5.

- João sabe que os garotos sairão, mas [$_{IP}$ ele não sabe [$_{CP}$ quando [_]]]. (27)

  John knows that the boys will leave, but he doesn't know when.

Since the antecedent can be recovered without violating the Island Constraints, the sentence is reconstructed by the system, as it is shown in Figure 6.
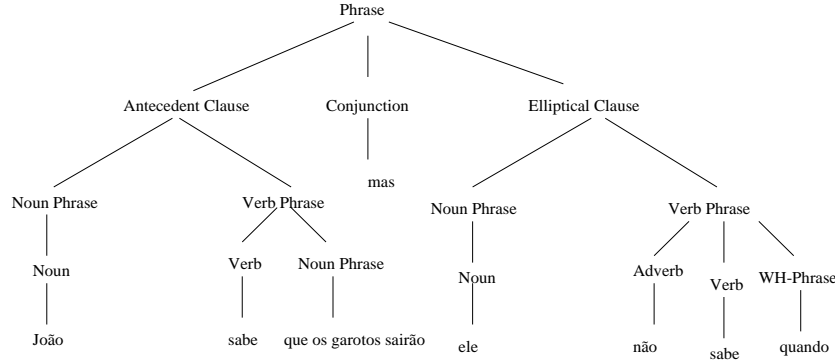
Phrase

Antecedent Clause   Conjunction   Elliptical Clause

mas

Noun Phrase   Verb Phrase   Noun Phrase   Verb Phrase

Noun   Verb   Noun Phrase   Noun   Adverb   Verb   WH-Phrase

João   sabe   que os garotos sairão   ele   não   sabe   quando

Figure 5

An example of Sluicing ellipsis

Phrase

Antecedent Clause   Conjunction   Elliptical Clause

mas

Noun Phrase   Verb Phrase   Noun Phrase   Verb Phrase

Noun   Verb   Noun Phrase   Noun   Adverb   Verb   WH-Phrase

João   sabe   que os garotos sairão   ele   não   sabe   quando os garotos sairão
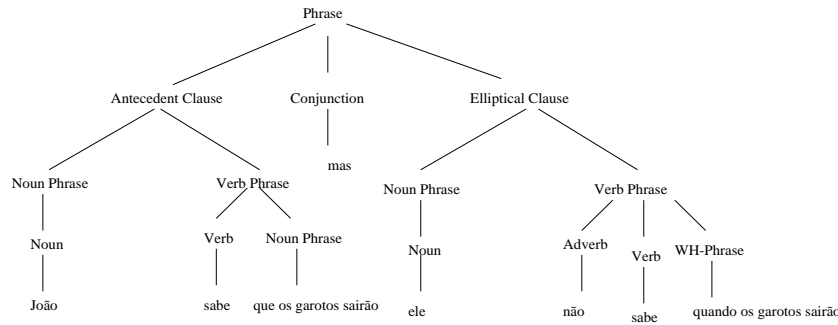
Figure 6

Sluicing ellipsis reconstruction

6 Remarks and Conclusion

We have proposed a syntactically-based algorithm for ellipsis resolution and we have argued that for some types of ellipsis, a syntactic structure is required in order to reconstruct the elliptical clause appropriately. We have not only used the syntactic structure of the antecedent to reconstruct the elliptical clause, but we have also taken syntactic constraints into consideration to check if the elliptical clause can actually be reconstructed. We have used Island Constraints to reconstruct the elliptical clause. Our approach is based on DeMatos' approach to deal with Stripping and Null VP. We have gone one step further dealing with Gapping and Sluicing also. The basic strategy which the algorithm encodes is to reconstruct the ellided clause by (i) decomposing the sentence into syntactic structures; (ii) identifying the type of ellipsis present in the sentence; (iii) checking if this type of ellipsis is subject to syntactic constraints; (iv) identifying the antecedent; and (v) reconstructing the ellided constituent. Future work includes studying other types of ellipsis, such as nominal ellipsis, as well as other syntactic restrictions on ellipsis (Chomsky, 1981). Although the linguistic data in this work is drawn primarily from Portuguese, we believe that the results can also be applied to other languages. Future work also includes investigating how much of the work described here can be applied to other languages, such as English.

Acknowledgments

References

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht:Foris.

Chomsky, Noam. 1986. *Barriers.* The Massachusetts Institute of Technology.

Dalrymple, Mary, Stuart M. Shieber, and Fernando C.N. Pereira. 1991. Ellipsis and higher-order unification. Technical report, Computation and Language E-Print Archive.

Haegeman, Liliane. 1992. *Introduction to Government and Binding Theory.* Blackwell Oxford UK and Cambrigde USA.

Kehler, Andrew. 1995. *Interpreting Cohesive Forms in the Context of Discourse Inference.* Ph.D. Thesis, Harvard University.

Lappin, Shalom. 1995. The interpretation of ellipsis. *The Handbook of Contemporary Semantic Theory.* Blackwell Oxford.

Matos, Maria Gabriela Ardisson Pereira De. 1992. *Construções de Elipse do Predicado em Português - Sv Nulo e Despojamento.* Ph.D. Thesis, Universidade de Lisboa.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Compreensive Grammar of the English Language.* Longman.