The antecedent-likelihood theory: a methodology to analyse and resolve anaphora in dialogues

Marco Antonio Esteves da Rocha

484

February, 1998

ISSN 1350-3162

UNIVERSITY OF SUSSEX AT BRIGHTON

Cognitive Science Research Papers

Acknowledgements

The project was fully funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico under grant no. 200608-92/4.

The antecedent-likelihood theory: a methodology to analyse and resolve anaphora in dialogues

Marco Antonio Esteves da Rocha

Abstract

This paper describes the results of a corpus-based cross-linguistic survey in which over six thousand cases of anaphora were manually analysed in dialogues in English and Portuguese, each language accounting for roughly half of the sample. The analytical tool used was an annotation which classified each case of anaphora according to four properties — namely, type of anaphor; type of antecedent; topical role of the antecedent; and processing strategy — described in the paper. A combination of statistical analysis, observed regular collocations, and specific context features was used to build a theory, called the antecedent-likelihood theory, which organises the information concerning the different types of anaphor in algorithm-like entries. The paper describes the guidelines under which the theory was built, together with the results of subsequent tests carried out for the English and Portuguese versions of the theory in dialogues previously annotated and set apart for the purpose. Possible ways partially or fully to automate the annotation process and the resolution procedures specified in the antecedent-likelihood theory are also discussed.

University of Sussex
School of Cognitive and Computing Sciences
Falmer, Brighton - BN1 9QH
marco@cogs.susx.ac.uk

Chapter 1

The antecedent-likelihood theory: a methodology to analyse and resolve anaphora in dialogues

1.1 Introduction

The antecedent-likelihood theory is the result of a corpus-based study on anaphora in dialogues in English and Portuguese. It was created to systematise and sum up the various types of information found to be relevant for anaphora resolution. The adopted organisation in algorithm-like form allowed manual testing of the theory as a procedure to resolve anaphora in dialogues set apart for the purpose. Although many of its features are evidently quite difficult to automate, the antecedent-likelihood theory (henceforward AL theory) is meant to support an anaphora interpreter in an NLP system. The approach relies on a combination of probabilistic information, regular collocations containing anaphors, and characteristic contextual features, to specify resolution paths for the different types of anaphor.

The concept of anaphora used in the study was as inclusive as possible, comprising all pronominal reference, anaphoric NPs, and all forms of ellipsis, defining anaphoric relations much in the same way as in [Web79]. The corpora used as source of data were the London Lund Corpus, for the dialogues in English, and the Rio de Janeiro Corpus of Clinical Dialogues — which was collected especially for the study — for the dialogues in Brazilian Portuguese. The analytical tool used to gather the information used in the AL theory is an annotation created for the purpose of the research. The annotation is used to analyse each case of anaphora according to four properties, namely, type of anaphor; type of antecedent; topical role of the antecedent; and processing strategy.

The classification used for the types of anaphor consists of categories which are often similar to word classes, with a few additions. The anaphor, in this research, is invariably the phonetically realised word or phrase which triggers the process of searching for an antecedent, thus ensuring the correct semantic interpretation of an utterance. Therefore, concepts like **anaphoric verb** and **anaphoric linking verb** are commonplace and appear as categories in the set of possible classifications for type of anaphor, especially in the Portuguese sample. Notions such as zero pronouns or empty categories are not part of the conceptual framework used in this research.

The classification used for the type of antecedent hinges around the explicit/implicit dichotomy, but there are two other possible categories. The **nonreferential** type is assigned to typically anaphoric words — such as *it* and *that* in English and verbal forms with essential arguments omitted in Portuguese — which do not have an antecedent, such as the *prop it* in weather constructions (see [QGSL85]). There is a fourth type of antecedent, classified as **discourse implicit**, which applies to cases where the antecedent is vague or has to be built out of discontinuous discourse information. The code for this property also includes a number which identifies the referent in a referent list for the dialogue being analysed.

The inclusion of the topical role of the antecedent as a property codified in the annotation is an attempt to capture the effect of topicality upon patterns of anaphora resolution (see [GS86] and [Sid86]). The categories are topical roles assigned to discourse entities prior to the process of annotating the anaphora cases. These topical roles are a result of analytical procedures which select a global topic - called the **discourse topic** - for the dialogue as a whole, and local topics - called **segment topics** - for each stretch of dialogue in which the same topic is thought to be prevailing. Discourse entities related to the discourse topic which are salient throughout the dialogue are assigned the role of **discourse thematic elements**. These typically include animate agents in the discourse, such as the dialogue participants. Discourse entities related to segment topics are called **segment thematic elements**. Their position in the topicality hierarchy is at the local level.

Dialogue segments are often complex enough to be split into subsegments, which are each assigned a **subsegment topic**. This set of topical roles is also used to classify antecedents which are discourse chunks, by singling out the most salient entity in the chunk to create topical roles such as **predicate of the discourse topic**. The procedures for the assignment of topical roles to discourse entities are detailed in [Roc98a]. They rely primarily on notions such as frequency, distribution and position of first occurrence in the dialogue, with a few semantic restrictions for finer distinctions.

The processing strategy was included in the annotation scheme as a way of enriching the classification. The plain assignment of a type of anaphor based on word classes would ignore distinctions in the processing required for the resolution of anaphors of the same type. On the other hand, the inclusion of processing information in the classification used for the type of anaphor would disrupt the intended link of the latter to phonetically realised forms in a strict way. The annotation is entered between brackets. The code for each one of the properties is delimited by semicolons. An example of annotated text is shown below.

(1) **B:** well I think probably what Captain Kay (FNP; ex_222; dthel; LR;)
must have said was a will is legal if it's (SP; ex_224; dthel; FtC;)
witnessed on the back of an envelope

A full description of the annotation scheme, with examples for each category in all properties, can be found in [Roc98a]. A more succint description can be found in [Roc98b]. Using this set of properties and the appropriate categories for the classification, three thousand cases of anaphora were annotated in each language. Techniques of statistical analysis were then used to explore and define the relations between the categories, working towards a model of the anaphora world. The resulting probabilistic model was then combined with regular collocations and contextual information to build the AL theory. This was organised in an algorithm-like form described in the next section. The method and the results of the tests are then described, together with a brief discussion of possible ways to automate the annotation. This is followed by concluding remarks regarding future developments.

1.2 The form of the entries in the AL theory

The first set of statistical procedures used to analyse the annotated sample were frequency counts and cross-tabulations. It became clear that the large number of categories in each variable, particularly types of anaphor and processing strategies, would make most techniques available for the analysis of nominal data unreliable, due to the minimum expected frequencies required for all chi-square-based techniques. The categories were then grouped and chi-square tests used to measure the significance of the relationship between the four variables pairwise. The results showed that the relations between the variables are highly significant in all cases, with $p < 0.00005$ for all pairings. As chi-square results are sensitive to sample size, the relatively large number of cases annotated certainly influenced the high significance obtained.

A measure of association - the Goodman and Kruskal tau - was then chosen to establish how

strongly the distribution of cases in one variable influenced the distribution of cases in another variable. The type of anaphor is the variable which could have its categories mapped out of a POS-tagged dialogue. Thus, it was chosen as the independent variable for the initial association measurements, having in mind a possible automation of the annotation process. The model building technique of loglinear analysis was then used to spot three- and four-way interactions between the variables.

The results, both for the English and the Portuguese samples, showed a strong association between the type of anaphor and the processing strategy, with the proportional reduction of error in predicting the processing strategy, once the type of anaphor is known, rating 44% higher than chance for the English sample and 36% for the Portuguese sample. On the other hand, the association between the type of anaphor and the two variables which classify the antecedent is weak, reducing error prediction in less than 5% for both samples.

The loglinear analysis assessed the four-way interaction between the variables as not significant, but considered all three-way interactions involving the processing strategy as significant. This means that the inclusion of the processing strategy significantly changes the picture of weak association between the type of anaphor and the two variables which classify the antecedent. It became thus clear that the processing strategy plays a special role in the model, acting as a linchpin which holds together the other variables. This also makes psycholinguistic sense, as the processing strategy classifies a pathway used for the resolution of an anaphoric reference, identifying the antecedent on the basis of the information provided by the anaphor and the surrounding context.

The findings of the statistical analysis were used to build probability trees and organise the AL theory. The probability trees use the ungrouped categories of the four properties with one of the types of anaphor as the root. The second level contains the processing strategies used to classify tokens of the type of anaphor at the root. The branches which link the root to each of the categories are assigned a probability calculated on the basis of the data colleced by means of the annotation. The same principle is used to link categories across the four properties, using aggregate totals to calculate the probabilities assigned to the branches. The tree for verbs and adverbials in Portuguese is shown below as an example[1].

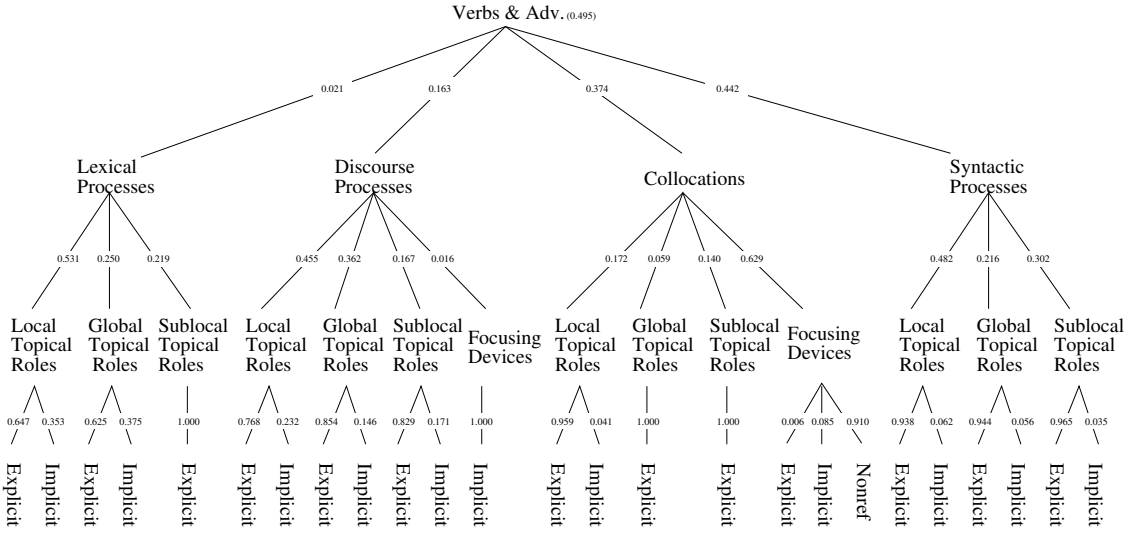

Figure 1.1: Probability tree for verbs and adverbials in Portuguese

Entries in the AL theory are organised as an ordered set of instructions in an algorithm-like form. This set of instructions systematises all the information collected by means of the annotation. It includes the probabilities in the probability trees and contextual information collected by

[1] The umbrella category is used for reasons of space, as trees for the ungrouped categories are rather complex.

means of observation. Each entry in the AL theory has a header which includes: the name of the type of anaphor; a global probability figure, specifying how likely the type of anaphor described in the entry is to occur; and a table, which summarises the information in the probability tree for the type of anaphor. The header for the subject pronoun entry in the version of the AL theory for English is shown as an example below.

Subject pronoun
global probability = 0.247

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh = 0.458	explicit = 0.886	dthel = 0.341
FtC = 0.232	implicit = 0.058	st = 0.188
CK = 0.155	NR = 0.049	sst = 0.156
DK = 0.090	disc.imp = 0.008	dt = 0.055
ScRf = 0.047		thel = 0.110
Pl = 0.012		fdv = 0.056
Dx = 0.003		p_st = 0.038
SK = 0.001		p_sst = 0.024
SetMb = 0.001		p_dthel = 0.014
SetCr = 0.001		p_dt = 0.005
		p_thel = 0.005
		sithel = 0.004
		uthel = 0.003

The header in AL theory entries is followed by a set of instructions. These instructions rely on the taxonomy employed to analyse processing strategies, reflecting the results of the statistical analysis. The typical instruction appears as **check ps**, **ps** being any category included in the list of possible classifications of processing strategy for the type of anaphor. As the probability trees include all category links which have at least one token, but not the ones with no tokens in the sample, the checks are restricted to the processing strategies recorded in the annotation process for the type of anaphor in question.

The typical **check ps** instruction is usually followed by a set of attached probabilities specific to the processing strategy being checked. These probabilities concern categories in the remaining two variables. Other information, such as the probability of predicate topical roles, may be added whenever this is felt to be useful. The subsequent items in a typical **check ps** instruction are **recognition** and **resolution path**. The first item contains information about features of the token itself and the immediate context in which it occurs, based on the observation of corpus data. The purpose is to guide the processing in the attempt to recognise the need for a certain type of processing strategy in order to resolve the anaphoric reference. The second item contains information related to the actual identification of the correct antecedent.

One example of **check ps** instruction is given below. It relates to the **set creation** processing strategy in the **anaphoric demonstratives** entry for Portuguese.

1. check set creation
    - attached probabilities
        - type of antecedent
            - explicit = 0.500
            - implicit = 0.500
        - topical roles

        - thel in all cases
- recognition
    - phrase *só isso*
    - preceding list of entities
- resolution path
    - antecedent is preceding set of entities
    - there may be no clear NP to define the set

The guidelines in the AL theory do not conform to a rigid formalism. Whenever deemed necessary, directions of a very different kind appear as instructions or items within instructions. The sequencing is settled in a way that seems plausible in terms of processing. There are instructions which assume ancillary routines associated to the AL theory. The most important one relates to the instruction:

check collocation list

The instruction, which appears in all entries for types of anaphor with tokens resolved by means of **collocational knowledge**, requires the matching of the move where the token occurs with entries in the collocation list under the type of anaphor to which the token belongs. The collocation list contains a number of entries which attempt to generalise patterns of resolution found to be consistent in specific contexts. The entries are grouped in the list according to the type of anaphor involved. One of the entries for subject pronouns in English is shown below:

it X-be SubjC that-clause

- cataphoric it (Subj)
- antecedent = that-clause

This entry[2] is meant to identify occurrences of cleft sentences in which *it* is systematically resolved as a cataphoric reference to a that-clause following the subject complement in a copular construction. Other regular forms of resolution include, for instance, the *make difference* collocation, which is represented in an entry shown below:

it X-make difference to Obj that-clause

- cataphoric it (Subj)
- antecedent = that-clause

Given an anaphor occurrence such as the first token of *it* in example (2) below, the search in the collocation list would find the entry for the *make difference* collocation and the pathway to resolve the anaphor.

```
(2)
B: I mean what difference could it make
   to the directors of Unilever that
   their shares had got down from say
   eighty to fifty or whatever it is
```

[2]The *X-* symbol means any inflected form of the verb, optionally including tense, aspect and modality. The major structures of the language, such as affirmative, interrogative and negative forms, are also assumed as included in the entry. The other symbols in the entries included as examples in this section stand for subject (Subj); subject complement (SubjC); and object (Obj).

```
A: well in the present circumstances
   not very much because I mean
   everything has gone down but of
   course if they are consistently low
   it makes them more difficult
   it makes it more difficult
   for them to raise money
```

The collocation list comprises thus contexts which contain anaphors regularly resolved in a definable way. The entries in the collocation list are in fact similar to instructions to check other processing strategies in the entries of the AL theory, except that the recognition procedure involves nothing beyond the matching of the move (as defined in [Ste90]) in which the anaphor occurs with a clearly specified pattern. The typical recognition procedure in instructions of the AL theory requires checking elements beyond the move. The pattern specified is also less clear-cut than the plain one-line definition in the collocation list. The collocation list for English anaphors is shown in Appendix B, whereas the one for Portuguese appears in Appendix D.

The second standard routine is associated with the instruction:

select the first appropriate candidate

This instruction specifies the use of a 'naive' algorithm similar to the one specified in [Hob86], but adapted to spoken discourse, in order to select a candidate antecedent. Therefore, **appropriate** means to be an acceptable candidate when syntactic factors only are considered. The candidate may be the antecedent proper or another anaphor, in the case of chains.

Once a candidate is found, it must undergo a good-fit check. The way this check is conducted is specified in the third standard routine below. However, the analysis of corpus data shows that the correct antecedent may be a long way off, making the repeated checking of lexical clues for every possible candidate a cumbersome and possibly expensive pathway to resolution. A search limit of five turns (as defined in [Ste90]) for each participant was defined. The search limit also makes sense from a psycholinguistic point of view. Speech is evanescent. If an adequate candidate is not found within a relatively recent range of utterances, other processing strategies, based on saliency rather than on recency, should be more effective.

The third standard routine is the good-fit check for candidate antecedents. This routine is associated with two distinct instructions which mean roughly the same:

check lexical clues
check if it is a good fit

The good-fit check is a pervasive feature of the AL theory. The procedure applies to any candidate antecedents, no matter how they have been selected. The observation of corpus data suggests two techniques to deal with the problem of checking whether an appropriate selected candidate is a good fit: selectional restrictions and association history.

Selectional restrictions are used here in the usual sense. The second kind of lexical clue is the association history of verbs and other entities with referents available within the context of the discourse. It includes associated noun phrases and adjectives linked to entities by means of copulas in previous occurrences. The absence of previous associations between the verb and the referent signals that it is advisable to consider a resolution which bypasses the first appropriate candidate, as it is unusual — but of course not impossible — that pronoun reference is used in the first instance of an association between a verb and a possible referent.

1.3 Testing and possible automation

Two dialogues — one in the English sample and another in the Portuguese sample — had been annotated but not included in the set of data used to produce the probabilistic model and the AL

theory, as they were reserved for testing purposes from the outset of the project. The testing of the AL theory was carried out manually, using the previously annotated dialogues set aside. The analyst simply browsed through the annotated dialogue and stopped whenever an anaphora case was found. The appropriate entry for the type of anaphor in the AL theory was then used to identify the antecedent and produce the resulting annotation. This annotation was then compared to the pre-existing one in order to assess the accuracy of the annotation produced on the basis of the AL theory.

Two kinds of discrepancy between the pre-existing annotation and the one generated according to the AL theory were marked. The processing strategy and the type of antecedent were considered as crucial in the test. Thus, whenever the information in the AL theory led to the wrong conclusion about the classification of one or both properties, this was counted as a primary error. There were cases in which the correct processing strategy and type of antecedent were selected by using the information in the AL theory, but the data concerning the features of the immediate context were not enough to identify the correct antecedent. Thus, although the guidelines for the recognition of the required processing strategy were good, and the type of antecedent was predicted in the probabilities attached to the processing strategy, the resolution path did not contain adequate information to ensure the identification of the antecedent. This sort of limitation in the AL theory counted as a secondary error.

There were 804 cases of anaphora in the test corpus for English. Only six cases led to primary errors when the appropriate entry in the AL theory was used to resolve them. This means that, in 99.3% of the cases, the guidelines in the AL theory recognised the processing strategy to be used for the resolution correctly, predicting also the correct type of antecedent. However, in nine of these cases, the resolution path did not include information which could successfully handle the anaphoric reference and identify the correct antecedent, counting thus as secondary errors. Therefore, the correct identification of the antecedent was not possible in fifteen cases, lowering the accuracy score to 98.1% of the cases if secondary errors are included. The results of the Portuguese test were slightly worse, with 96.6% accuracy if only primary errors were counted and 90% if secondary errors were counted.

The results are evidently satisfactory, but they should be seen with due caution. Firstly, the assumptions of segmentation and topical role assignment substantially reduce the difficulty of the task. Secondly, the testing was carried out manually by the analyst. It must be said, nevertheless, that the satisfactory results provide strong support for the inclusion of topicality as a crucial element in the resolution of anaphoric references.

The assignment of topical roles is probably the most daunting task for the automatic annotation of dialogues using the scheme described. Nevertheless, work using connectionist networks to model topicality (see [Kar92]) seems to offer some promise, although there is certainly a long way to go before this can be built into an operational system. The task of annotating the anaphora cases is probably a lot more feasible if the assignment of topical roles, together with the segmentation, is assumed to have been accomplished. One approach which bears important similarities to the one described here is the analogical modelling described in [Sko89], as it also uses a combination of probabilities and contextual features extracted from a training corpus to deal with language processing tasks of the same kind in any text.

1.4 Conclusion

In a balanced appraisal of the results above, it seems safe to assert that such high levels of accuracy cannot be due to chance or plainly to the analyst's leniency. Therefore, one might conclude that the AL theory is at least a step in the right direction as a corpus-based approach to anaphora resolution based on a database of annotated cases. It is probably also true that the level of detail included in the annotation, particularly the processing and topicality-related information, is likely to exert a positive influence over the efficiency of an anaphora interpreter in an NLP system. The increased

complexity may therefore be an advantage rather than a limitation.

The entries for each type of anaphor in the version of the AL theory for anaphora in English are shown in Appendix A. The Portuguese counterpart of the theory is shown in Appendix C. Planned future developments for the project which resulted in the AL theory include an attempt to explore the possibilities of annotating anaphoric relations automatically on the basis of the information organised in the entries for each type of anaphor. In spite of the limitations mentioned above, it seems worthwhile to establish how effective the approach could be once the assumptions of POS tagging — probably demanding partial parsing — and segmentation on the basis of topicality, with the assignment of topical roles, were in place.

Bibliography

[GS86] Barbara Grosz and Candace Sidner. Attentions, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

[Hob86] Jerry Hobbs. Resolving pronoun references. In B.L. Webber, Barbara Grosz, and K. Jones, editors, *Readings in Natural Language Processing*. Morgan Kaufmann, Palo Alto, CA, 1986.

[Kar92] Lorraine Karen. Identification of topical entities in discourse: a connectionist approach to attentional mechanisms in language. In Noel Sharkey, editor, *Connectionist natural language processing*, chapter 13, pages 275–94. Intellect, Oxford, 1992.

[QGSL85] Randolph Quirk, Sidney Greenbaum, Jan Svartvik, and Geoffrey Leech. *A comprehensive grammar of the English language*. Longman, 1985.

[Roc98a] Marco Rocha. A description of an annotation scheme to analyse anaphora in dialogues. Technical Report 427, University of Sussex - School of Cognitive and Computing Sciences, Brighton, 1998.

[Roc98b] Marco Rocha. A probabilistic approach to anaphora resolution in dialogues in english. In Magnus Ljung, editor, *Corpus-based studies in English*, pages 261–278. Rodopi, Amsterdam, 1998.

[Sid86] Candace Sidner. Focusing in the comprehension of definite anaphora. In Karen Jones Barbara Grosz and Bonnie Webber, editors, *Readings in natural language processing*. Morgan Kaufman, Palo Alto, CA, 1986.

[Sko89] Royal Skousen. *Analogical Modelling of Language*. Kluwer, Dordrecht, 1989.

[Ste90] Anna-Brita Stenström. Lexical items peculiar to spoken discourse. In Jan Svartvik, editor, *The London-Lund corpus of spoken English: description and research*. Lund University Press, Lund, 1990.

[Web79] Bonnie Lynn Webber. *A Formal Approach to Discourse Anaphora*. Garland, New York, 1979.

Appendix A

The AL theory for anaphors in the English sample

A.1 Pronouns

A.1.1 Subject pronouns

global probability = 0.247

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh=0.458	explicit=0.886	dthel=0.341
FtC=0.232	implicit=0.058	st=0.188
CK=0.155	NR=0.049	sst=0.156
DK=0.087	disc=0.008	thel=0.110
ScRf=0.047		fdv=0.056
Pl=0.012		dt=0.055
DsAn=0.003		p_st=0.038
Dx=0.003		p_sst=0.024
SK=0.001		p_dthel=0.014
SetMb=0.001		p_dt=0.005
SetCr=0.001		p_thel=0.005
		sithel=0.004
		uthel=0.003

1. check if POS tag is Q-tag item
   - if not, go to instruction 2; if yes
   - go to tag-question entry in collocation list
   - follow resolution path in entry
2. identify pronoun
   (a) pronoun is *he*, *she* or *they*
      - go to instruction 5

   (b) pronoun is *it*
      - go to instruction 4

   (c) pronoun is first or second person

- go to instruction 3

3. check secondary reference

- attached probabilities
  - type of antecedent
    - explicit = 0.889
    - implicit = 0.111
    - implicit antecedents are in a chain
    - ultimate resolution by shared knowledge
  - topical roles
    - dthel = 0.750
    - st = 0.250
- recognition
  - separate from endophoric usage
  - previous move
    - verbs *say*; *ask*; *answer*; *explain*
    - subject a third person pronoun or personal name
    - simultaneous tense and person shift between utterances
    - typically past to present tense; third person to first
  - if it is a second person pronoun
  - check identifying vocative in the utterance
- resolution path
  - select first human candidate searching backwards
  - check lexical clues
  - if there is an identifying vocative
    - select it as the antecedent

4. check collocation list

- if no match found
  - go to instruction 5
- if a match is found
  - follow resolution pathway in entry

5. select first appropriate candidate

- if no appropriate candidate found
  - go to instruction 6
- if an appropriate candidate is found
  - go to instruction 7

6. check shared knowledge

- attached probabilities
  - type of antecedent: implicit (1.000)

  - topical role
    - discourse thematic element: dthel = 1.000
- recognition
  - beginning of dialogue (up to 40 tone units)
    - if not, go to instruction 7
  - no appropriate candidates
  - it often starts a chain
- resolution path
  - check dt and dthel in history of previous interactions
  - track chain till a definite description occurs

7. check lexical clues

- if it is a good fit
  - go to instruction 10
- if it is not a good fit
  - go to instruction 8

8. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.758
    - implicit = 0.182
    - discourse implicit = 0.061
  - topical roles
    - dthel = 0.273 p_st = 0.106
    - st = 0.197 sst = 0.076
    - thel = 0.136 fdv = 0.061
    - dt = 0.121 p_sst = 0.030
    - saliency more useful for specific types
- recognition and resolution divided in types
- if pronoun is *he* or *she*
  - go to instruction 8(g)
- if pronoun is *it* or *they*

  (a) shift *it-they*
    - recognition
      - first-candidate chain *it-they*
      - vice-versa seems possible
      - ultimate antecedent is an institution
    - resolution path
      - ignore agreement conflict
      - select institutional referent
      - check lexical clues

(b) plural NP antecedent of *it*
- variation of type (a)
- recognition
  - first-candidate strategy selects plural NP
  - plural NP in previous move or turn
- resolution path
  - ignore agreement conflict
  - select plural NP as antecedent
  - check lexical clues
  - encyclopedic knowledge may be needed

(c) antecedent extraction from complex NP
- recognition
  - first candidate is a complex noun phrase
  - poor fit ascertained
  - it may require complex semantic processing
- resolution path
  - select internal constituent as antecedent
  - check syntax and lexical clues
- if no recognition in 8(a),(b),(c)
  - pronoun *they*
    - go to instruction 8(g)
  - pronoun is *it*
    - go to instruction 8(d)

(d) strained anaphora
- recognition
  - finite or non-finite verb in previous turn or move
  - nominalised form fits as antecedent
- resolution path
  - nominalise verb in previous turn or move
  - check lexical clues

(e) discourse-chunk antecedent
- recognition
  - *it* interchangeable with *this/that*
  - frequent collocations
    - *it was because...*
    - *it* + Copula + Adjective as in:
    - *it was necessary*
- resolution path
  - check previous move or turn
  - check clausal constituents within
  - begin with that-clauses and non-finites

(f) altered reiteration
- recognition
  - previous move or turn is a sentential *it* collocation

        - reiteration with the non-finite verb and complement missing
      - resolution path
        - antecedent is the missing verb and complement
      - if poor fit persists
        - go to item **candidate bypass**
  (g) candidate bypass
      - recognition
        - lexical clues
        - discourse-marker clues
        - prosodic clues
      - resolution
        - bypass first candidate
        - if pronoun is *it*
          - select dt and st
          - check syntax and lexical clues
        - if any other subject pronoun
          - select dthel and st
          - check syntax and lexical clues
  (h) return pop
      - recognition
        - token in segment resumption boundary
        - lexical clues
        - discourse-marker clues
        - prosodic clues
      - resolution path
        - select st of resumptive segment
        - check appropriateness and lexical clues
        - select dthel(s) in resumptive segment
        - check appropriateness and lexical clues
        - invert order if pronoun is not *it*
        - select dt and thel(s) in resumptive segment
        - check appropriateness and lexical clues
  (i) sequence of questions
      - recognition
        - two (or more) questions in sequence
        - set of answers follows set of questions
        - anaphors in the answers or in both
      - resolution path
        - for anaphors in first answer
          - bypass antecedent candidates in second question
          - select antecedent candidates in first question
        - for anaphors in second answer
          - bypass antecedent candidates in first answer
          - select antecedent candidates in second question

(j) sum-up utterances
- recognition
  - *it* followed by link verb and complement
  - a judgement on the state of things described
  - the utterance may sum up long stretches
  - complex processing required
- resolution
  - discourse-implicit antecedent
  - a NP of vague generic meaning
  - *the problem*; *the matter*;
  - *the question*; *the situation*;
  - *the thing*; *the stuff*; *the issue*

(k) set/set member oscillation
- recognition
  - chain of *it* tokens (two or more)
  - collocation signalling reference to set (unstable)
    - *you know how (the way) it is*
  - aspect clue: habitual-progressive
- resolution path
  - interpret antecedent as set or set member as appropriate
  - the real problem lies in recognition

9. check set creation

- recognition
  - pronoun *they*
  - lexical clue in utterance
    - the word *simultaneous* in tone unit
- resolution path
  - search backwards for NPs with the same head
  - search backwards for NPs in a semantic set
  - select conjoined NPs as created set

10. check deixis

- recognition type one
  - pronouns *he*, *she* or *they*
  - dialogue with three or more participants
  - two participants refer to the other(s) one(s)
- resolution type one
  - select other(s) participant(s) as antecedent(s)
- recognition type two
  - first-candidate chain selects a demonstrative
  - search backwards within the limit

- resolution type two
  - visual input needed for identification

11. check parallel

- recognition type one
  - two appropriate candidates in previous move or turn
  - sharply distinct syntactic functions
    - typically subject and object
  - verb often identical
- resolution path type one
  - ignore first candidate
  - select candidate with the same syntactic function
- recognition type two
  - pronoun *it*
  - cleft sentence in previous move or turn
  - complement omitted in current move
- resolution type two
  - antecedent is that-clause in previous turn or move
- recognition type three
  - pronoun *it*
  - alternative interrogative clause in previous turn or move
- resolution path type three
  - antecedent is alternative clause
  - if there is embedment of alternative clauses
    - antecedent is the last one

12. set member

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical role
    - universal thematic element in all cases
- recognition
  - first-candidate strategy selects a compound indefinite pronoun
  - antecedent in the same clause
- resolution path
  - meaning constrained by dialogue context
  - check lexical clues
  - select according to verbal semantics
- if 11 and 12 don't apply

13. accept first candidate

A.1.2 Object pronoun

global probability = 0.095

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh = 0.432	explicit = 0.915	dthel = 0.259
FtC = 0.293	NR = 0.048	st = 0.235
DK = 0.126	implicit = 0.031	dt = 0.228
CK = 0.082	disc.imp = 0.007	sst = 0.088
Pl = 0.027		thel = 0.088
ScRf = 0.017		fdv = 0.051
SK = 0.010		p_st = 0.027
Dx = 0.007		p_sst = 0.010
WK = 0.003		p_dt = 0.003
SetMb = 0.003		p_thel = 0.003
		p_dthel = 0.003
		sithel = 0.003

1. identify pronoun
   - pronoun is *him*, *her* or *them*
     - go to instruction 4
   - if pronoun is *it*
     - go to instruction 3
   - if pronoun is first or second person
     - go to instruction 2
2. check secondary reference
   - attached probabilities
     - SR tokens by all pronoun tokens
       - us = 0.166
       - me = 0.032
       - you (SP + OP) = 0.003
     - type of antecedent
       - explicit in all cases
     - topical roles
       - dthel = 0.600
       - st = 0.200
       - thel = 0.200
   - recognition
     - separate from endophoric usage
     - previous move or turn
       - verbs *say*; *ask*; *answer*; *explain*
       - subject a third person pronoun or personal name
       - tense switch

  - identifying vocative in the utterance
- resolution path
  - if there is an identifying vocative
    - select the entity as the antecedent
  - if there is no vocative
    - select first human candidate searching backwards
    - check lexical clues

3. check collocation list

- if no match is found
  - go to instruction 4
- if a match is found
  - follow resolution pathway in entry

4. select first appropriate candidate

- if no appropriate candidate found
  - go to instruction 5
- if an appropriate candidate is found
  - go to instruction 6

5. check shared knowledge

- attached probabilities
  - type of antecedent: implicit (1.000)
  - topical role
    - discourse topic: dt = 0.750
    - discourse thematic element: dthel = 0.250
- recognition
  - beginning of dialogue
  - no appropriate candidates
  - it often starts a chain
- resolution path
  - check dt and dthel in history of previous interactions
  - track chain till a definite description occurs

6. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.973
    - discourse implicit = 0.027
  - topical roles
    - dt = 0.444 p_dthel = 0.028

    - st = 0.250 sst = 0.028
    - dthel = 0.083 fdv = 0.028
    - p_st = 0.083 p_dt = 0.028
    - thel = 0.028 sithel = 0.028
    - saliency effect stronger than for SP cases

- recognition type one
  - discourse-chunk antecedent (0.139 in DK)
  - *it* interchangeable with *this/that*
  - frequent collocation: *I'm glad to hear it*
  - *remember*; *mention*; *read*; *move*; or
  - subsequent *this* or *that* in the same turn
  - especially if object of *say* and
  - subject of link verb with a clausal complement
    - verb *say* in clause as in example below
    - *she didn't remember it but that's what mother said*
- resolution path type one
  - antecedent in the same or previous turn
  - often that-clause as verb object
  - list of possible verbs in CGE 16.30ff
  - nominal wh-interrogative clauses
    - subject complement in copulas
- recognition type two
  - candidate bypass
  - lexical clues
  - discourse-marker clues
  - prosodic clues
- resolution type two
  - bypass first candidate
  - if pronoun is *it*
    - select dt and st
    - check syntax and lexical clues
  - if any other object pronoun
    - select dthel and st
    - check syntax and lexical clues
- recognition type three
  - sum-up utterances
  - lexical clues
    - *look at it*
    - no association history
  - discourse-marker clues: *so*
  - prosodic clues
- resolution type three
  - discourse-implicit antecedent

  - a NP of vague generic meaning
  - *the problem*; *the matter*;
  - *the question*; *the situation*;
  - *the thing*; *the stuff*; *the issue*

7. check deixis

   - recognition type one
     - pronouns *him*, *her* or *them*
     - dialogue with three or more participants
     - two participants refer to the other(s) one(s)
   - resolution type one
     - select other(s) participant(s) as antecedent(s)
   - recognition type two
     - first-candidate chain selects a demonstrative
     - search backwards within the limit
   - resolution type two
     - visual input needed for identification

8. check syntactic parallelism

   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical role
       - st = 0.375
       - dthel = 0.375
       - dt = 0.125
       - thel = 0.125
   - recognition type one
     - sequence of identical or very similar moves
       - same verb; same arguments;
   - resolution path type one
     - assign the same antecedents for the anaphors
       - bypass identical pronouns in different syntactic functions
   - recognition type two
     - sequence with identical core (especially verb)
       - intervening adjunct (such as a PP)
       - the PP contains an acceptable candidate
   - resolution path type two
     - assign the same antecedents for the anaphors
       - ignore alternative candidate
   - recognition type three

  - sequence with semantically related verbs *advice-argue*
- resolution path type three
  - ignore identical pronouns with different syntactic functions
  - bypass alternative candidates with different syntactic functions

9. check world knowledge

- attached probailities
  - type of antecedent
    - implicit in all cases
  - topical roles
    - thematic element in all cases
- recognition
  - semantically inadequate first candidate
  - complex semantic processing required
- resolution path
  - information concerning subject in utterance
  - information involves experiential knowledge

10. check set member

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - subsegment topic in all cases
    - unreliable sample size
- recognition
  - only pronoun *it*
  - *them* seems acceptable
  - first candidate a noun phrase
    - a "no matter which" determiner
    - *every*; *any*
- resolution path
  - no specific referent for the anaphor
  - "no matter which" member of the set

11. accept first candidate

A.1.3 Demonstratives

global probability = 0.107

Category probabilities

processing strategy	type of antecedent	topical role
DK = 0.380	explicit = 0.830	st = 0.210
CK = 0.185	implicit = 0.103	p_st = 0.188
FtCCh = 0.149	disc.imp = 0.040	p_sst = 0.149
FtC = 0.134	NR = 0.027	dt = 0.122
Dx = 0.134		sst = 0.091
Pl = 0.006		thel = 0.079
DsAnCh = 0.006		fdv = 0.055
SetCr = 0.003		p_dthel = 0.043
DsAn = 0.003		dthel = 0.033
		p_thel = 0.018
		p_dt = 0.012

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution path in entry
2. select first appropriate candidate
   - if no appropriate candidate found
     - go to instruction 3
   - if an appropriate candidate is found
     - go to instruction 4
3. check distant anaphora
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - dt = 0.666
       - sst = 0.333
         - *these* token
         - in fact it refers to dt and a dthel
   - recognition
     - pronouns *this* and *these*
     - no appropriate candidate within search limit
   - resolution path
     - if pronoun is *this*
       - select dt and dthel(s)
       - check syntax and lexical clues
       - select best candidate
     - if pronoun is *these*

        - select dt and dthel(s) in pairs
        - check syntax and lexical clues for the pairs
        - select best pair

4. check lexical clues

    - if it is a good fit
        - go to instruction 10
    - if it is not a good fit
        - go to instruction 5

5. check discourse knowledge

    - attached probabilities
        - type of antecedent
            - explicit = 0.847
            - implicit = 0.081
            - disc. implicit = 0.073
        - topical roles
            - p_st = 0.293 fdv = 0.048
            - p_sst = 0.250 sst = 0.032
            - st = 0.153 dthel = 0.032
            - thel = 0.056 p_thel = 0.024
            - p_dthel = 0.056 p_dt = 0.008
            - dt = 0.048
    - if anaphor is subject of non-copular verb
        - if verb is *explain* type
            - select previous move or turn as candidate
            - typical verbs
                - *show*; *account*; *explain*;
        - if verb is not *explain* type
        - identify pronoun
            - pronoun is *this* or *these*
                - select segment topic
                - check lexical clue
                - if it is a good fit, accept it
                - if it is not a good fit
                - select discourse topic
                - check lexical clues
                - if it is a good fit, accept it
                - if it is not a good fit
                - select subsegment topic
                - check lexical clues
                - if it is a good fit, accept it
                - if it is not a good fit

- select previous move
- make standard adjustments to check good fit
- *the fact that X*; *what X*;
- if selected move is a complex utterance
- break it up in clausal constituents
- check clausal constituents in separate
- expand move
- up to next discourse-unit boundary
- up to a delimiting discourse marker
- use lexical clues in anaphor utterance
- use lookahead for other clues
- if it is a good fit, accept it
- if it is not a good fit
- select first move which contains st (if different)
- if it is a good fit, accept it
- if it is not a good fit
- repeat first-move check with sst and dt
- if it is a good fit, accept it
- if it is not a good fit
- check thel(s) and first move with thel
- if it is a good fit, accept it
- if it is not a good fit
- nominalise verb in previous turn or move
- check lexical clues

* pronoun is *that*
  - follow the same procedure
  - check moves before noun phrases
* if a good fit is found, accept it
* if it is not, check discourse implicit
  - check NPs of generic meaning
  - *this X-be where*
  - check at this point
  - broad reference to current discourse unit
  - broad reference to previous discourse unit
  - broad reference to discourse unit before the previous
  - make standard adjustments
  - recorded collocations
  - *does that mean...*
  - *how does that strike you*
  - *this is because*
  - *Subj (that) this X-do*

- if anaphor is subject of link verb
  - follow the same procedure
  - check moves before noun phrases for singular pronouns
  - make standard adjustments to check good fit

        - if pronoun is *this*
            - check cataphoric reference as well (next move)
- if a good fit is found, accept it
- if it is not, check discourse implicit
    - *all* as a clue
- if anaphor is object of verb
    - pronoun is *this*
        - if verb is not in clausal complementation list
            - check noun phrases until solution is found
        - if verb is in the clausal complementation list
            - check previous move and turn first
            - make standard adjustments to check good fit
        - most frequent verbs in the list
            - *say*; *answer*; *ask*; *learn*;
            - *imagine*; *avoid*; *suggest*;
        - if a good fit is found, accept it
        - if it is not, check discourse implicit
    - pronoun is *that*
        - follow the same procedure
        - check moves before noun phrases
        - make standard adjustments to check good fit
        - if a good fit is found, accept it
        - if it is not, check discourse implicit
    - anaphor is complement of copular verb
        - follow the same procedure for all pronouns
        - check moves before noun phrases
        - make standard adjustments to check good fit
        - if a good fit is found, accept it
        - if it is not, check discourse implicit
    - anaphor is object of preposition
        - follow the same procedure for all pronouns
        - check moves before noun phrases
            - except if anaphor follows CopAdj + *than*
            - use adjective as a clue to search NP antecedent
        - make standard adjustments to check good fit
        - if a good fit is found, accept it
        - if it is not, check discourse implicit

6. check deixis

- attached probabilities
    - type of antecedent
        - explicit = 0.523
        - implicit = 0.477

  - topical roles
    - st = 0.500 sst = 0.091
    - dt = 0.182 dthel = 0.068
    - thel = 0.159
- recognition
  - prosodic pattern
  - identifiable entity in physical environment
- resolution path
  - visual input needed for identification

7. check syntactic parallelism

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - st = 0.500
    - p_sst = 0.500
- recognition
  - repetition or near repetition
  - alternative candidates in different syntactic functions
- resolution path
  - select candidate with same syntactic function
  - ignore alternative candidates

8. check set creation

- attached probabilities
  - type of antecedent
    - implicit in all cases
  - topical roles
    - segment topic in all cases
- recognition
  - pronoun it these or *those*
  - no appropriate candidate within search limit
- resolution path
  - search backwards for NPs with the same head
  - search backwards for NPs in a semantic set
  - select conjoined NPs as created set

9. if 5, 6, 7 and 8 don't apply

10. accept candidate

A.1.4 Determinative possessives

global probability = 0.032

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh = 0.469	explicit = 0.980	dthel = 0.500
FtC = 0.439	implicit = 0.020	thel = 0.194
DK = 0.051		sst = 0.143
ScRf = 0.041		st = 0.092
		dt = 0.021
		uthel = 0.010

1. check secondary reference
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - dthel = 0.750
            - thel = 0.250
    - recognition
        - separation from endophoric usage
        - first-person subject pronoun in chain in the same move
        - previous move or turn
            - verbs *say*; *ask*; *answer*; *explain*
            - subject a third person pronoun or personal name
            - tense switch in relation to anaphor move
        - identifying vocative in the utterance
    - resolution path
        - if there is a personal pronoun in chain
            - select the same antecedent
        - if there is an identifying vocative
            - select the entity as the antecedent
        - if there is no vocative nor chain
            - select first human candidate searching backwards
            - check lexical clues
2. select first appropriate candidate
    - check lexical clues
        - if it is a good fit and
        - candidate does not appear in reported speech
            - go to instruction 4
        - if candidate appears in reported speech
            - go to instruction 3 type one
        - if it is not a good fit

        - go to instruction 3 type two

3. check discourse knowledge

    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - thel = 0.600
            - dthel = 0.200
            - dt = 0.200
    - recognition type one: return to direct reference
        - first candidate appears in reported speech
        - speaker of reported speech is also a candidate
    - resolution path
        - check lexical clues, especially possessed object
        - use lookahead if still ambiguous
        - select best candidate
    - recognition type two: candidate bypass
        - lexical clues rule out first candidate
    - resolution path type two
        - bypass first candidate
        - check alternatives within search limit
        - select best candidate using lexical clues

4. accept first candidate

A.1.5 Independent possessives

global probability = 0.001

Category probabilities

processing strategy	type of antecedent	topical role
SetMb = 0.500	explicit = 1.000	dt = 0.500
FtC = 0.250		st = 0.250
FtCCh = 0.250		thel = 0.250

1. check set member

    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - discourse topic in all cases
            - sample size very small: unreliable
    - recognition type one

    - first or second person pronoun
  - resolution path type one
    - antecedent is possessed object
    - select first appropriate set
      - lexical clues, especially
      - association history: possessor-possessed
    - antecedent is the set member identified by the anaphor
  - recognition type two
    - third person pronoun
  - resolution path type two
    - double reference
    - possessor and possessed object
    - resolve possessor as determinative
    - resolve possessed as type one

2. disconsider first-candidate-strategy tokens

   - accidental omission of possessed object

A.1.6 One-anaphora

global probability = 0.006

Category probabilities

processing strategy	type of antecedent	topical role
SetMb = 0.650	explicit = 0.700	sst = 0.300
Dx = 0.200	implicit = 0.300	thel = 0.300
FtC = 0.100		st = 0.150
AM = 0.050		dthel = 0.150
		dt = 0.100

1. check deixis

   - attached probabilities
   - small sample size: unreliable
     - type of antecedent
       - explicit = 0.500
       - implicit = 0.500
     - topical roles
       - thel = 0.500
       - st = 0.250
       - sst = 0.250
   - recognition
     - demonstrative a typical modifier
     - prosodic pattern
     - often pointing gesture
     - identifying entity in physical environment

- resolution path
  - visual input needed for identification

2. select first appropriate candidate

- if it is a NP modified by *only*
  - accept head as the antecedent
- if it is an indefinite description
  - check anaphor utterance for lexical clues
    - anaphor takes a demonstrative determinant
    - anaphor has no other or no identifying modifiers
  - accept description as antecedent
- anaphor has distinguishing modifiers
  - go to instruction 4
- if it is neither of those
  - go to instruction 3

3. check modified antecedent

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - thel = 1.000
    - small sample size: unreliable
- recognition
  - there are one or more modifiers attached to anaphor
  - same modifier appears in previous NP
  - other NPs may appear with same head and different modifier
  - the other NP or NPs may occur between anaphor and antecedent
- resolution path
  - select head and identical modifier as antecedent

4. check set member

- attached probabilities
  - type of antecedent
    - explicit = 0.692
    - implicit = 0.308
  - topical roles
    - sst = 0.385
    - dthel = 0.231
    - dt = 0.154
    - thel = 0.154
    - st = 0.077

- recognition and resolution divided in types
- recognition type one
  - anaphor determiner is an indefinite article or
  - anaphor is plural
  - modifier selects from a previously introduced set
  - first candidate is a set member introducing the set
- resolution path type one
  - select head of NP candidate as antecedent
  - analyse antecedent as an implicit set
- recognition type two
  - as type one except that
  - first candidate defines a set
- resolution type two
  - select head and lexical modifiers as antecedent
- recognition type three
  - anaphor determiner is a definite article
  - anaphor is not a subject complement in a copula
  - subject complement in a *there X-be* construction included
  - there is a modifier attached to the anaphor or
  - there is a modifying clause attached to anaphor
- resolution path type three
  - select NP candidates sequentially
  - check fit with modifier
  - accept the best fit
  - if there is doubt, prefer the subsegment topic
- recognition type four
  - as type three but
  - anaphor is a subject complement in a copula
- resolution path type four
  - the subject of the copula is the antecedent
  - if subject is a first or second person pronoun
    - the participant referred to is the antecedent
    - analyse the antecedent as implicit
- recognition type five
  - anaphor determiner is a definite article
  - modifier is a weak selector (such as an ordinal)
    - not enough to characterise a modified antecedent
- resolution path type five
  - select NP candidates sequentially
  - check lexical clues in utterance
  - accept the best fit
- recognition type six

  - anaphor is *the other one(s)*
- resolution type six
  - if anaphor is singular
    - search for two members of the same set
  - if anaphor is plural
    - search for two subsets in the set
  - check lexical clues
  - select the appropriate option
  - complex discourse processing may be required
- recognition type seven
  - anaphor is *another one*
- resolution path type seven
  - select head of the first NP candidate
  - check lexical clues

A.1.7 Numerals

global probability = 0.007

Category probabilities

processing strategy	type of antecedent	topical role
SetMb = 0.522	explicit = 0.609	thel = 0.435
Dx = 0.130	implicit = 0.391	dt = 0.304
LR = 0.087		sst = 0.217
VMm = 0.087		
LS = 0.043		
AM = 0.043		

1. check set member

- attached probabilities
  - type of antecedent
    - explicit = 0.750
    - implicit = 0.250
  - topical roles
    - dt = 0.500
    - sst = 0.250
    - thel = 0.167
    - dthel = 0.083
- recognition and resolution divided in types
- recognition type one
  - NUM + *of which* + VERB + NP
- resolution path type one
  - lookahead in anaphor utterance
  - select NP head in verb complement

  - search backwards for plural NP with same head
  - antecedent is always explicit
- recognition type two
  - DET + (Adj) + NUM
  - numeral may occur in its own
- resolution path type two
  - search backwards for either
    - an indefinite NP
      - check if dt; st; dthel; sst;
      - check lexical clues
    - if it is a good fit
      - accept as antecedent
    - if it is not a good fit
      - move on to next candidate
    - or if it is a NP with a numeral postdeterminer
      - check if numeral postdeterminer matches anaphor
      - if yes, NP head is antecedent
      - if not, check numeral postdeterminer
      - if it is an ordinal postdeterminer
      - NP head is antecedent
      - check synonims, e.g., *front/first*
      - if it is a cardinal postdeterminer
      - full NP is antecedent
      - *of-the* relation with anaphor
      - embeddings possible
      - (three X (two Xy [one Xya and one Xyb) and one Xz)
      - (three X (two Xy and one Xz; two Xy (one Xya and one Xyb))
      - X is omitted and should be identified as antecedent
      - y, ya, yb and z are modifiers
      - typically PPs or relative clauses

2. check lexical repetition

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - thematic element in all cases
- recognition
  - numeral pronoun
- resolution path
  - search for repeated numeral
  - current and previous turn

3. check lexical signalling
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - thematc element in all cases
            - small sample size
    - recognition
        - numeral pronoun
    - resolution path
        - search backwards
        - current and previous turn
            - NP with numeral postdeterminer
            - NP with *of-the* structure
4. check modified antecedent
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - thematic element in all cases
    - recognition
        - numeral pronoun occurs with a modifier
        - modifier is typically a PP or a rel. clause
    - resolution path
        - search current and previous turn
        - antecedent is almost identical to anaphor
        - slightly different modifier

A.1.8 Reflexives

global probability = 0.005

Category probabilities		
processing strategy	type of antecedent	topical role
FtC = 0.500	explicit = 1.000	thel = 0.375
FtCCh = 0.500		dthel = 0.250
		uthel = 0.188
		sst = 0.063
		st = 0.063
		dt = 0.063

1. select the first appropriate candidate
2. accept it as the antecedent

A.1.9 Indefinite pronouns

global probability = 0.003

Category probabilities

processing strategy	type of antecedent	topical role
Pl = 0.778	explicit = 1.000	sst = 0.556
SetMb = 0.111		st = 0.222
VMm = 0.111		thel = 0.222

1. check parallel
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - sst = 0.571
            - thel = 0.286
            - st = 0.143
    - recognition
        - *several*; *none*; *many*
    - resolution path
        - select first NP as a candidate
        - check parallel with IP
        - adjust as required
2. check verbatim memory
    - explicit and sst in all cases
    - recognition
        - pronoun *some*
    - resolution path
        - select first NP as candidate
        - adjust as required
3. check set member
    - explicit and st in all cases
    - recognition
        - pronoun *each (of which)*
    - resolution path
        - select first NP as candidate
        - adjust as required

A.2 Verbs and adverbials

A.2.1 Adverbs of place

global probability = 0.005

Category probabilities

processing strategy	type of antecedent	topical role
DK = 0.355	explicit = 0.677	thel = 0.320
Dx = 0.226	implicit = 0.290	dthel = 0.161
FtC = 0.194	NR = 0.032	st = 0.129
FtCCh = 0.129		dt = 0.065
CK = 0.097		sst = 0.065
		p_dthel = 0.065
		sithel = 0.032
		fdv = 0.032
		p_sst = 0.032

1. check collocation list
   - if a match is found
     - follow resolution path in entry
   - if a match is not found
     - go to instruction 2
2. select first appropriate candidate
   - if none is found within search limit
     - go to instruction 3
   - if one is found
     - check lexical clues
     - if it is a good fit, accept candidate
     - if it is not a good fit, repeat search
3. check discourse knowledge
   - attached probabilities
     - type of antecedent
       - explicit = 0.666
       - implicit = 0.333
     - topical roles
       - thel = 0.555
       - dt = 0.111
       - st = 0.111
       - sst = 0.111
       - dthel = 0.111
   - recognition type one
     - anaphor is adverb *there*
     - first candidate is not a good fit

- resolution path
  - use information in the utterance, especially
    - subject and object association history
    - resumptive segment
  - bypass first candidate
  - select the one that fits with lexical clues
- recognition type two
  - adverb refers to situation
  - typical collocation: *go from there*
  - first candidate is not a good fit
- resolution path type two
  - select the segment or subsegment topic
  - check lexical clues
- recognition type three
  - *there* as a place assumed to exist
  - location of an event being talked about
  - antecedent is implicit
- resolution path type three
  - assume a location where the event took place
  - construe it according to available information
- recognition type four
  - adverb refers to object
  - *here* in non-deictic usage
- resolution type four
  - analyse the utterance for clues
  - use lexical clues to select candidate
  - check previous move or turn if required

4. check deixis

- attached probabilities
  - type of antecedent
    - explicit = 0.571
    - implicit = 0.429
  - topical roles
    - thel = 0.571
    - st = 143
    - dthel = 0.143
    - sithel = 0.143
- recognition
  - *here* in broader sense
  - *there* with no appropriate antecedent
- resolution path
  - visual input is needed if *there*
  - *here* refers to location of dialogue
    - town, state or country

A.2.2 Adverbs of response

global probability = 0.036

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 1.000	explicit = 0.982	p_sst = 0.464
	implicit = 0.009	p_st = 0.348
		p_dthel = 0.089
		p_dt = 0.054
		p_thel = 0.036
		p_fdv = 0.009

1. explicit in all cases except
   - if antecedent contains another anaphor and
   - anaphor has an implicit antecedent
2. recognition
   - *yes* or *no* as responses
3. resolution path
   - select query as antecedent
   - adjust as required

A.2.3 Wh-word

global probability = 0.001

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 1.000	explicit = 1.000	p_st = 0.333
		p_sst = 0.333
		p_dt = 0.333

1. select previous move or turn as antecedent
2. if there is an interrogative clause
   - select clause as antecedent
3. if not, select the whole move
4. adjust as required

A.2.4 Prepositional Phrases

global probability = 0.003

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 1.000	explicit = 1.000	p_st = 0.625
		p_sst = 0.250
		p_dthel = 0.125

1. select previous move or turn as antecedent
2. if there is a prepositional phrase
   - select PP as antecedent
3. if not, select the whole move
4. adjust as required

A.2.5 *So* anaphora

global probability = 0.002

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 0.857	explicit = 1.000	p_st = 0.714
Pl = 0.143		p_sst = 0.143
		p_thel = 0.143

1. check verbatim memory
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - discourse chunk in all cases
       - p_st = 0.833
       - p_sst = 0.167
     - recognition type one
       - collocations *think so*; *suppose so*;
       - collocations *say so*; *tell so*;
       - list in CGE 12.47
     - resolution path
       - check previous turn
         - if there is a subordinate that-clause
         - select that-clause as antecedent
         - check if it is a good fit; if not
         - select the whole move as antecedent
         - if there are coordinate clauses
         - select last clause

- · check if it is a good fit
- · if previous turn is not a good fit
- · check if the main clause is missing
- · if it is, check previous move
- · if it is a clarification question
- · select the previous turn
- · join it with the move before the anaphor utterance
- · make the necessary adjustments
- · select the move pieced together as antecedent
- · check lexical clues

2. check parallel

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_thel in all cases
    - sample size small
- recognition
  - collocation *even more so*
- resolution path
  - search for the main verb
  - select it as antecedent
  - make adjustments as required
    - adjectives in verbal complement may be made comparative

A.2.6 *Do* phrase anaphora

global probability = 0.010

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 0.438	explicit = 0.843	p_sst = 0.406
DK = 0.281	implicit = 0.157	p_st = 0.281
FtCCh = 0.250		p_dt = 0.094
Pl = 0.031		dthel = 0.063
		p_dthel = 0.063
		p_thel = 0.031
		st = 0.031
		thel = 0.031

1. check verbatim memory

- attached probabilities
  - type of antecedent
    - explicit in all cases

  - topical roles
    * p_sst = 0.643
    * p_st = 0.143
    * p_dthel = 0.143
    * p_dt = 0.071

- recognition
  - according to general definition
- resolution path
  - select predicate of previous move as candidate
    * if previous move is a short response
      · select fully reconstructed move
    * if previous move is a De + X-mean + NF-clause
      · select NF-clause as antecedent
      · check if it is a good fit
      · if it is, accept the candidate
    * if previous move contains a DPA
      · go to instruction 3
    * if previous move matches the structure of anaphor utterance
    * and if DPA is the only significant difference
      · go to instruction 4
  - remove operators if there are any
  - remove semiauxiliary and catenative verbs
  - check if it is a good fit
    * if it is, accept the candidate
  - if it is not a good fit
    * select next move with the same subject as DPA move
    * check if it is a good fit
      · if it is, accept the candidate
      · if it is not, check if it is a subordination structure
      · if it is, select main clause as candidate
      · check if it is a good fit
      · if it is, accept the candidate
      · if it is not, check subordinate clause
      · if it is a good fit, accept the candidate
      · more reliable if subordinate clause is direct object
      · if it is not, check if it is a coordination structure
      · if it is, check if clause subject is the same as DPA utterance
      · if both do, select the last clause as candidate
      · check if it is a good fit
      · if it is, accept the candidate
      · if it is not, check if previous move is a copula
      · if it is, check if subject is clausal
      · if it is, select as candidate
      · check if it is a good fit

            - if it is, accept the candidate
            - if none of these conditions apply or
            - if none yields a good fit
            - check if DPA utterance is a *whether* DPA
            - if it is, select only the main verb in previous move
            - if none of these conditions apply or
            - if none yields a good fit
            - go to instruction number 2

2. check discourse knowledge

    - attached probabilities
        - type of antecedent
            - explicit = 0.667
            - implicit = 0.333
        - topical roles
            - p_st = 0.333
            - p_dt = 0.222
            - p_sst = 0.111
            - p_thel = 0.111
            - st = 0.111
            - thel = 0.111
    - recognition type one
        - complex parallel
            - DPA utterance is a complex sentence
            - antecedent utterance is a similar complex sentence
            - typical case
                - sequence of subordination structures
                - if-clauses or that-clauses in both utterances
                - appositions and intervening moves are possible
                - they may involve similar clauses as well
                - that-clause antecedents may be omitted in parallel utterance
                - they must be retrieved in previous discourse
                - especially if there is affirmative-negative contrast
                - *they know*/*they don't know* that-clause
    - resolution path type one
        - rearrange utterances to reveal parallel
        - retrieve omitted elements from previous discourse
        - match DPA with antecedent
    - recognition type two
        - frequency word in DPA utterance
        - previous move is an example of event referred to
    - resolution path type two
        - select previous move as candidate

  - compare move with stretch up to discourse unit boundary
  - if a subsegment, extend comparison to segment boundary
  - test a *this kind of thing* referent
  - complex semantic processing required
- recognition type three
  - phrase is X-do + demonstrative
  - *advise* type verb within search limit
- resolution type three
  - postulate a hypothetical action
  - delimit scope within discourse context
- if none of these conditions apply or
- if none yields a good fit
  - go to instruction 4

3. classify resolution as FtCCh
   - tense adjustments may be required
   - classify TA and TR as the anaphor in chain
4. classify resolution as Pl
   - explicit in all cases

A.2.7 Linking verb

global probability = 0.009

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 1.000	explicit = 1.000	p_st = 0.259
		sst = 0.222
		p_sst = 0.185
		p_dt = 0.148
		p_dthel = 0.074
		p_thel = 0.037
		st = 0.037
		uthel = 0.037

1. search for previous LV token
2. antecedent is subject complement
   - polarity adjustment may be required
   - agreement is not always accurate
     - in question tags especially

A.2.8 Operator

global probability = 0.029

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 0.966	explicit = 0.989	p_sst = 0.427
DK = 0.023	dim = 0.011	p_st = 0.371
FtCCh = 0.011		p_thel = 0.079
		p_dt = 0.067
		p_dthel = 0.056

1. if operator is *do* or *does*
   - search for previous token of present tense
   - select predication of utterance as candidate
   - check if it is a good fit
   - if it is not, go to instruction 4
2. if operator is *did*
   - search for previous token of present tense
   - select predication of utterance as candidate
   - check if it is a good fit
   - if it is not, go to instruction 4
3. if it is any other operator
   - search for previous token of same operator
   - select predication of utterance as candidate
   - check if it is a good fit
   - if it is not, go to instruction 4
4. select a *this kind of thing* type of antecedent
5. if it is a good fit
   - accept it as antecedent
   - classify antecedent as **dim**

A.2.9 Anaphoric verbs

global probability = 0.009

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 0.448	explicit = 0.931	p_st = 0.242
FtCCh = 0.276	implicit = 0.069	dthel = 0.208
Pl = 0.138		p_dthel = 0.172
FtC = 0.138		st = 0.103
		thel = 0.103
		p_sst = 0.103
		sst = 0.069

1. check verbatim memory

    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - clausal antecedent in all cases
            - p_st = 0.545
            - p_dthel = 0.364
            - p_sst = 0.091
    - recognition type one
        - anaphor is a semiauxiliary or catenative verb
        - complement is missing
    - resolution path
        - analyse previous move
            - if there is a NF-clause
            - select it as candidate
            - check lexical clues
            - if it is a good fit
            - accept it as antecedent
            - if it is not a good fit
            - search for move with the same subject
            - repeat analysis
    - if there isn't a NF clause
        - select predication
            - check lexical clues
            - if it is a good fit
            - accept it as antecedent
            - if it is not a good fit
            - search for move with the same subject
            - repeat analysis
    - recognition type two
        - form of verb *to know*
        - clausal object omitted
    - resolution path type two
        - analyse previous move
            - if there is an interrogative clause or question
                - select it as antecedent
                - especially if object of verb *to know*
                - questions need adjustments
            - if there is a conditional clause
                - select it as antecedent
            - if the conditions do not apply

· go to instruction 2

- recognition type three
  - anaphor is non-finite verbal form
  - anaphor is in a coordination structure
  - anaphor is in the second clause
  - retrieval of subject and auxiliary(ies) required
- resolution path type three
  - retrieve subject and auxiliary(ies) in first clause
  - accept them as antecedent

2. check parallel strategy

- attached probabilities
  - type of antecedent
    - explicit = 0.750
    - implicit = 0.250
  - topical roles
    - p_sst = 0.250
    - p_st = 0.250
    - st = 0.250
    - dthel = 0.250
- recognition type one
  - transitive verb with omitted object
  - typical verbs: *think*; *tell*; *agree*;
  - special case: *go*
    - select object of previous token of *go*
- resolution path
  - analyse previous move or turn
  - select object of transitive verb in move or turn
- recognition type two
  - verb with omitted subject
  - not in coordination structure
  - first appropriate candidate is in a subordinate clause
- resolution path two
  - bypass candidate in subordinate clause
  - select subject of main clause in previous move

3. check first candidate strategies

- verb with omitted subject
- coordination structure
  - in most cases the second clause
- select first appropriate candidate
  - in most cases subject of first clause

A.2.10 Copula-FNP

global probability ≈ 0.000

Category probabilities

processing strategy	type of antecedent	topical role
FtC = 1.000	explicit = 1.000	dt = 1.000

1. first candidate is antecedent

A.2.11 Copula-Adj

global probability ≈ 0.000

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh = 1.000	explicit = 1.000	dthel = 1.000

1. first candidate is antecedent

A.2.12 Copula-PP

global probability ≈ 0.000

Category probabilities

processing strategy	type of antecedent	topical role
FtCCh = 1.000	explicit = 1.000	st = 1.000

1. first candidate is antecedent

A.2.13 Copula-Clause

global probability ≈ 0.000

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 1.000	explicit = 1.000	p_sst = 1.000

1. recognition

   - according to definition

2. resolution

   - antecedent is subject of previous move

A.2.14 Non-finite clauses

global probability = 0.001

Category probabilities

processing strategy	type of antecedent	topical role
SetMb = 1.000	explicit = 1.000	p_sst = 1.000

1. recognition

- ordinal precedes anaphor
- if there are two anaphors in sequence
  - the second may be preceded by *and then*

2. resolution

   - antecedent is NP like: NUM + NOUN
   - low semantic content possible: *two things*

A.3 Nominals

A.3.1 Nonpronominal noun phrases

global probability = 0.380

Category probabilities

processing strategy	type of antecedent	topical role
LR = 0.457	explicit = 0.776	thel = 0.276
AM = 0.167	implicit = 0.217	dthel = 0.235
LS = 0.165	disc.imp. = 0.008	st = 0.194
SK = 0.072		sst = 0.155
WK = 0.056		dt = 0.105
CK = 0.032		p_st = 0.010
Dx = 0.023		sithel = 0.007
DK = 0.015		p_sst = 0.006
SetCr = 0.007		fdv = 0.005
SetMb = 0.003		uthel = 0.004
VMm = 0.003		p_dthel = 0.002

1. check collocation list
2. if a match is found

   - follow resolution path in entry

3. if no match is found

   - go to instruction 4

4. check lexical repetition

   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - dthel = 0.340
       - thel = 0.223
       - st = 0.191
       - dt = 0.128
       - sst = 0.117
   - recognition

  - any kind of NP
- resolution path
  - search history list
  - if a precise match is found
    - select it as antecedent
  - if a partial match is found
    - go to instruction 4

5. check modified antecedent

- attached probabilities
  - type of antecedent
    - explicit = 0.897
    - implicit = 0.103
  - topical roles
    - st = 0.299
    - thel = 0.273
    - sst = 0.211
    - dthel = 0.119
    - dt = 0.098
- recognition
  - partial match with entity in history list
- resolution path
  - check the full history list
    - if there is a previous precise match
      - accept ot as the antecedent
    - if there is not a previous precise match
  - if partial match involves the NP head
    - select partial match as antecedent
    - check lexical clues
    - if it is a good match
      - accept it as antecedent
    - if it is not a good match
      - classify antecedent as new and implicit
      - typical case: anaphor is a plural noun
  - if partial match is a modifier
    - if it is a proper noun, check usage
    - if it isn't, check lexical clues

6. check lexical signalling

- attached probabilities
  - type of antecedent
    - explicit = 0.552

        * implicit = 0.448
      - topical roles
        * thel = 0.309
        * dthel = 0.211
        * sst = 0.175
        * st = 0.175
        * dt = 0.082
        * p_sst = 0.015
        * uthel = 0.015
        * p_st = 0.010
        * sithel = 0.005
    - recognition
      - definite description
      - no entry in history list
    - resolution path
      - check dictionary entries of entities
        * if a match is found
          · accept anaphor as an implicit antecedent
        * if no match is found
          · go to instruction 5

7. check world knowledge

    - attached probabilities
      - type of antecedent
        * implicit = 0.742
        * explicit = 0.258
      - topical roles
        * thel = 0.485
        * sst = 0.258
        * st = 0.167
        * dthel = 0.061
        * dt = 0.015
        * sithel = 0.015
    - recognition
      - definite description
      - no match in history list
      - no match in dictionary entries of entities
    - resolution path
      - complex semantic processing
      - domain information
      - select an antecedent based on lexical clues
        * if no good fit is found
        * go to instruction 6

8. check shared knowledge
    - attached probabilities
        - type of antecedent
            - implicit = 0.726
            - explicit = 0.274
        - topical roles
            - thel = 0.417
            - sst = 0.167
            - dthel = 0.143
            - st = 0.131
            - dt = 0.107
            - sithel = 0.024
            - uthel = 0.012
    - recognition
        - definite description
        - no match in history list
        - no match in dictionary entries of entities
        - no useful world knowledge information
9. resolution path
    - search entities in record of previous interactions
        - if no match is found
    - go to instruction 7
10. check deixis
    - attached probabilities
        - type of antecedent
            - implicit = 0.815
            - explicit = 0.185
        - topical roles
            - thel = 0.407
            - sst = 0.259
            - sithel = 0.111
            - dthel = 0.074
            - st = 0.074
            - dt = 0.037
            - uthel = 0.037
    - recognition
        - demonstrative a typical modifier
        - prosodic pattern
        - often pointing gesture
        - identifying entity in physical environment

- resolution path
  - visual input needed for identification

11. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.500
    - implicit = 0.333
    - disc.impl. = 0.167
  - topical roles
    - dthel = 0.333
    - st = 0.278
    - thel = 0.167
    - dt = 0.111
    - sst = 0.056
    - p_st = 0.056
- recognition type one
  - definite descriptions
  - no match in lexical processing based on the anaphor
- resolution path
  - analyse modifiers in search of lexical clues
  - analyse anaphor move in search of lexical clues
  - use association history to select candidate
  - check salient topical roles first
- recognition type two
  - special expressions
    - *people*; *such venture*;
- resolution path type two
  - search for lexical clues in anaphor move
  - establish restrictions to select candidate
  - if no acceptable candidate is selected
    - check discourse implicit
    - dim associated with dt; dthel; st;
- recognition type three
  - NUM + *kinds*
- resolution path type three
  - cataphoric reference
  - antecedents are enumerated after anaphor
- recognition type four
  - *sort(s) of organisations*;
  - *in that connexion*;
  - *from the ADJ point of view*

- resolution type four
  - implicit antecedent
  - postulate a referent according to NP head
  - specify according to lexical clues in anaphor move

12. check set creation

- attached probabilities
  - type of antecedent
    - explicit = 0.875
    - implicit = 0.125
  - topical roles
    - dthel = 0.250
    - sst = 0.250
    - thel = 0.125
    - dt = 0.125
    - st = 0.0125
    - p_sst = 0.125
- recognition type one
  - generic NP
  - indefinite article as determiner
- resolution path type one
  - search for set members using NP head
  - create set as a new referent
- recognition type two
  - plural NP modified by a numeral
  - generic or low semantic content
    - *maths, philosophy-two possible subjects*
    - *whatever1, whatever2-those two things*
- resolution path type two
  - check lexical clues in NP head
  - check lexical clues in anaphor move
  - select candidate on good fit
- recogniton type three
  - generic NP + PP (*people with money*)
  - reference to set members in previous discourse
    - often scattered throughout; difficult
- resolution path type three
  - check lexical clues in PP
  - inference may be required
    - relation *money-support*

13. check set member

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - thel in all cases
- recognition
  - two or more NPs with same head
  - distinguished by modifiers
    - often adjectives with *one*/*other*
    - *one at* Adj1 X-noun *and the other at* Adj2 X-noun
    - antecedent = two Xs
    - antecedent is generally recent
- resolution path
  - select plural head as antecedent
  - search previous discourse

14. check verbatim memory

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_dthel = 0.667
    - p_st = 0.333
- recognition
  - NP is response to question
  - full response ellipted
- resolution path
  - select question minus wh-word as candidate
  - check lexical clues

A.3.2 Anaphoric adjectives

global probability = 0.004

Category probabilities

processing strategy	type of antecedent	topical role
VMm = 0.692	explicit = 0.846	sst = 0.308
LR = 0.077	implicit = 0.154	p_sst = 0.308
LS = 0.077		p_sst = 0.231
FtCCh = 0.077		dt = 0.154
AM = 0.077		

1. check verbatim memory

- attached probabilities

  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_sst = 0.444
    - p_st = 0.333
    - sst = 0.222
- recognition type one
  - adjective takes clausal complementation
  - complement is omitted
  - typically a subject complement in copulas
  - typical anaphoric adjectives
  - *sure*; *certain*; *positive*
- resolution path type one
  - search for token of adjective within search limit
  - if found, complement is antecedent
  - if complement is again omitted, continue search
  - if no other token is found within search limit
    - check previous move or turn
    - attach it as a that-clause to anaphor
    - if it is a good fit, accept it as antecedent
    - if previous move or turn is a response adjective
      - retrieve ellipted response
      - repeat operation above
- recognition type two
  - adjective modifies an omitted NP head
  - this includes comparative and superlative forms
  - this includes NPs used as adjectives
- resolution path
  - search for NP within search limit
  - check previous move or turn
  - select salient NP head as antecedent
  - if it is a good fit, accept it as antecedent
  - if another adjective anaphor is found
    - go to instruction 2
  - if an identical adjective anaphor is found
    - go to instruction 3
  - if selected NP is modified by same adjective
    - go to instruction 4
  - if selected NP is modified by semantically related adjective
    - go to instruction 5

2. classify former token as FtCCh strategy

- repeat VMm procedure for new token

3. classify former token as LR strategy
    - repeat VMm procedure for new token
4. classify former token as AM strategy
    - repeat VMm procedure for new token
5. classify former token as LS strategy
6. classify antecedent as implicit
    - repeat VMm procedure for new token

Appendix B

The collocation list for the English sample

B.1 Collocations with subject pronouns

it + X-be + time expression

- time of the clock; day; month; year;
- after/before phrases
- resolution path
  - nonreferential

it + X-go + wrong

- summary of extended explanations
- *the fact that...* NP
- resolution path
  - check segment topic
  - check predicate of segment topic
  - repeat with subsegment topic

it's not because

- summary of extended explanations
- *the fact that...* NP
- resolution path
  - check segment topic
  - check predicate of segment topic
  - repeat with subsegment topic

it X-be a question of

- nonreferential

it X-be like that

- collocation reiterates discourse chunk
- resolution path
  - pronouns are nonreferential

WH-word it X-be that-clause

- emphasis for the subsequent that-clause
  - nonreferential

pleonastic it

- subject of utterance should be easily identifiable
- resolution path
  - nonreferential or reiteration

cleft sentences

- the most common collocation
- sentential *it*
- recognition
  - it + X-be + time and/or place expression + that-clause
  - it + X-be + (modif) NP (modif) + that-clause
  - it + X-be + (modif) ADJ (modif) + that-clause
  - variants of the that-clause possible
    - who-clause; which-clause; *that* omitted;
- resolution path
  - the that-clause is the antecedent
- special cases
  - subject of that-clause is *it*
    - antecedent is first candidate
    - subject of cleft sentence disconsidered
  - adjustments required
    - agreement; nominalisation
    - *it's not the Indologists you want to convince*
    - *THOSE you want to convince ARE not...*
  - that-clause antecedent must be retrieved
    - syntactic parallelism often present
    - that-clause may not exist as such

            - the whole previous move is the antecedent
        - whether + it X-be + Comp
            - that-clause missing
            - opposite polarity in previous move
            - antecedent is predicate with opposite polarity
    - that-clause with required *that* omitted
        - a pause substitutes for *that*
        - *it is stated* + (pause) + clause
        - antecedent is still that-clause

it + X-be + NUM + of those + (NUM) + things

- nonreferential

it + LV + ADJ (AdvP) (PP) (FOR ObjP) + NF-clause

- recognition
    - list of adjectives requiring complementation
        - *frustrating*; *fascinating*;
        - *kind of you*; comparatives;
- resolution path
    - the antecedent is the NF-clause
    - the nonfinite clause may be omitted
        - it must be retrieved from a previous utterance

it + LV + Comparative + as it is

- the antecedent is *as it is*

it + LV + of-phrase (modifiers) + NF-clause

- nonfinite clause is the antecedent

it + X-be + (modif) ADJ

- recognition
    - generic characterisation
    - list of adjectives
        - *(so) strange*; *(so) curious*; *(too) late*
- resolution path
    - nonreferential

tag questions

- recognition

  - Q-tag

- resolution path

  - check utterance to which the tag question is attached
  - syntactic parallel
  - agreement is not always respected

it + X-be + IndArt + (modif) ADJ + NP

- recognition

  - subject complement is an indefinite description

- resolution path

  - antecedent is a definite description
  - NP head is head of definite description
  - *it was a very hurried interview =*
  - *the interview was a very hurried one*
  - *it's a very difficult time =*
  - *this time is a very difficult one*

it + X-be (not) a matter of + NF-clause

- nonreferential

it + X-be + Pos + fault + if-clause

- cataphoric
- antecedent is if-clause

NP or whatever it is
NP or whoever it was

- resolution path

  - analyse NP
  - select a suitable generic for canidate antecedent
  - *two years or whatever it is*; years=time;
  - *the assistant or whoever it was*; assistant=person

it + LV(X-be) + a long time

- mostly nonreferential; however,
- *nine years is a long time*
- *yes, it is a long time*

it + LV + worthwhile + NF-clause

- antecedent is the NF-clause
- NF-clause is usually gerund

it means that

- resolution path
  - select previous turn
  - it may be a long chunk
  - check good fit
  - nominalise as *the fact that...*
  - check good fit
  - adjustments may be required

what is it

- recognition
  - discourse marker
  - often preceded by *now*
  - a rhetorical question
- resolution path
  - nonreferential

there it X-BE

- recognition
  - *there* must not refer to a location
- resolution path
  - nonreferential

that's the way (how) it goes

- nonreferential

it + X-be + a pleasure

- nonreferential

it + X-be + a pleasure + NF-clause

- antecedent is NF-clause

it + X-take + time expression

- response to questions with *how long*
- nonfinite clause complement

it X-be (not) a question of

- nonreferential

it + X-be + the same + Noun

- the antecedent is *the* + Noun

it + X-be + Pos + Noun

- the antecedent is *the* + Noun

it X-be a lie

- followed or preceded by a subordinate conditional clause
- antecedent is the conditional clause

how + ADJ + it + X-be

- followed or preceded by a subordinate conditional clause
- antecedent is the conditional clause

it + X-be + SubjC + De

- nonreferential
- demonstrative must be resolved

it X-make DET difference

- check preceding question by other participant
- antecedent is subordinate alternative clause or that-clause
- resolution may involve discourse knowledge
  - such as building the antecedent out of an X-or-X NP

it X-make DET difference to Obj that-clause

- cataphoric
- antecedent = that-clause or alternative clause

it X-vary

- check preceding question by other participant
- antecedent is subordinate alternative clause or that-clause
- resolution may involve discourse knowledge
  - such as building the antecedent out of an X-or-X NP

it X-matter a damn (PHRASE) to Obj whether-clause

- cataphoric
- antecedent = whether-clause

it X-be no good NF-clause

- typically present participle
- cataphoric
- antecedent = NP-clause

it X-be (not) on the basis of

- check previous move
- antecedent = nominalised sentence *the fact that*

it + X-be + for + ObjP + NF-clause

- cataphoric
- antecedent = NF-clause

it + X-seem (+ to + ObjP + ADJ) + that-clause

- object pronoun is often *me*
- cataphoric
- antecedent = that-clause

it + X-look + to + ObjP + like-clause

- nonreferential

it + X-turn out + that-clause

- cataphoric
- that-clause is antecedent

it X-stand

- nonreferential

B.2 Collocations with object pronouns

Subj + X-mean + it

- sense of *X-be sincere*
- nonreferential

Subj + X-take + it + that-clause

- sense of *would have thought*

- first person subject frequent
- often apposition or subordinate adverbial clause
- that-clause is antecedent
- *that* may be omitted

that is it

- it may appear in question form
- nonreferential

X-put it mildly/this way

- often non-finite
- nonreferential

as I X-understand it

- sense of *in my opinion*
- nonreferential

X-stick it out

- in the sense of *hold out*
- nonreferential

X-be pleased about it

- discourse-chunk antecedent: *the fact that...*
- subordinate clause in preceding move
- syntax may be broken up

and all the rest of it

- verbless syntactically disconnected phrase
- prosodic pattern significant
- compare with next entry
- nonreferential

(and) all the rest of it

- subject or object of clause
- distinct prosodic pattern
- antecedent is NP in preceding move

do it

- non-finite form
- not a DPA; no verb-phrase antecedent
- discourse-implicit antecedent
- discourse topic

there X-be no question about it
there X-be no doubt about it
Subj + X-have + no doubt about it

- cataphoric
- subsequent move
- it could be anaphoric

X-put it about

- sense of *saying openly*
- cataphoric
- subsequent that-clause

X-find + it + (modif) + Adj

- sense of *X-think it is*
- cataphoric
- subsequent that-clause or
- subsequent NF-clause

for the sheer tickle of it

- verb and object in preceding move
- adjustments may be needed

Subj + X-think + it + Adj

- token in conditional clause
- main clause is antecedent

X-cope with it

- often discourse-chunk antecedents
- often preceding move
- nominalisation with *the fact that...*

it makes it + Adj + for + ObjP + Inf

- the antecedent is the PP after the Adj
- arguably nonreferential as an option

X-Verb it all-PRONOUN

- nonreferential
- sense of *whatever there is to* + Verb

B.3 Collocations with demonstratives

Subj-be + (that) + Subj + X-want + to ask ObjP + X-be + this
Subj-be + (that) + Subj + X-want + to say + X-be + this

- both anaphoric and cataphoric (but often cataphoric)
- antecedent = discourse chunk (question or request)
- intervening moves possible
  - preamble to actual question or request
  - lexical clues to identify antecedent
    - repeated X-ask or X-say
- Subj-be: *what*; *the last thing*; *the other thing*;
- the embedded clause may be omitted

there X-be something in this

- anaphoric; discourse-chunk antecedent;
- previous move: Subj X-tell that-clause
- antecedent = *the fact that* + that-clause

X-understand this that-clause

- simultaneously anaphoric and cataphoric
- implicit antecedent made explicit as subsequent move
- subsequent move is that-clause
- anaphor is pleonastic; common in spoken language

that is to say

- anaphoric
- antecedent = discourse chunk; previous move;
- selection out of move may be required
  - lexical clue in same or subsequent move
  - partial repetition
  - syntactic parallelism

anything/something/nothing like that

- anaphoric
- antecedent = discourse chunk
- previous or same move

- connecting *or* frequent

NP like that

- anaphoric
- antecedent = discourse chunk
- it may be a description
- lexical clue: preceding NP

it + X-be + like that

- pronouns are nonreferential
- collocation reiterates discourse chunk

that + X-be + the + *n* time + that-clause

- cataphoric
- antecedent = *this time*

that + X-be (not) + (Adv) right/wrong

- anaphoric
- antecedent = discourse chunk; previous move;

this X-be right/wrong

- anaphoric
- antecedent = discourse chunk;
- previous move *the fact that...*
  - if Subj X-say that-clause
  - Subj X-say may or may not be included

that's (not) true

- anaphoric
- antecedent = discourse chunk;
- previous move *the fact that*
  - if coordinated clauses
  - it may be the second one only

this X-be all right

- anaphoric
- plain NP; first candidate;

this means

- anaphoric
- antecedent = discourse chunk
- previous move

Subj/main sentence X-say that

- anaphoric
- antecedent = previous move;
- selection on lexical clue;
- previous *say* token
- I/you correspondences
- complex discourse processing may be needed
  - sequential tokens with different referents (S02.01.01-0143)
  - clue in adverbs: *never/absolutely right*
- if previous move *tell* X (adjuncts) that-clause
- that-clause is antecedent

that's it

- nonreferential

that is one thing

- anaphoric
- discourse implicit: long stretches
- sum-up phrase
- if followed by clause
- it should not be treated as a collocation (DK)

that's/this is why

- anaphoric
- discourse chunk
  - previous move
  - coordination with clausal antecedent possible
  - X-ant *and this is why*

that X-BE time of the clock

- nonreferential

- special case: narratives
  - *that's about half past two*
  - the meaning is: *that happened about...*
  - antecedent is previous move

sure about that

- anaphoric
- discourse chunk;
- previous move *the fact that...*

X doubt about this

- cataphoric
- discourse chunk
- following move; same participant

that is so

- anaphoric
- discourse chunk
- previous move

apart from that

- anaphoric
- discourse chunk; previous move;
- lexical clue () may appear
- *apart from* repeated

there X-be no question about that

- anaphoric
- previous move

that's all

- anaphoric
- discourse implicit
- sum-up phrase
- a discourse-chunk antecedent may sometimes be identified
- if followed by that-clause (DK)
- it should not be treated as a collocation

that X-be the reason for X

- anaphoric
- discourse implicit;
- previous subsegment(s)
- agreement may be imperfect
- more than one *reason* as antecedent

that's all right/OK

- nonreferential

know (about) that

- anaphoric
- antecedent: discourse chunk; previous move
- selection may be required
- coordination: the second clause

B.4 Collocations with adverbs of time

wrong/right there

- anaphoric
- discourse chunk
- previous move
- *wrong or right* may occur in questions

B.5 Collocations with nonpronominal noun phrases

anything/something/nothing of this kind

- explicit NP antecedent
- NP + *or* + COL
- antecedent is NP

this sort of line

- there may be a subsequent explanation
- *when I say COL, I mean ANT*

this sort of thing

- anaphoric
- discourse-chunk antecedent

- demonstrative anaphor in previous move
- it refers to the antecedent
- *the fact that...*
- predicate of segment topic
- it sums up a description

the sort of thing

- anaphoric
- segment topic
- NP explicit antecedent
- *is this the sort of thing*

this thing

- NP explicit antecedent
- lexical clue: verb
- syntactic parallel
- the first appropriate NP candidate
- digressions may occur in between

the whole point

- discourse implicit
- focusing device
- it sums up a description but vaguely

the whole thing

- pronoun-like resolution
- use *it* as a guide
- check OP or SP entry as appropriate
- discourse-chunk antecedents possible
- implicit antecedents possible

the thing

- pronoun-like resolution
- use *it* as a guide
- check OP or SP entry as appropriate

- no discourse-chunk antecedents found
- no implicit antecedents found

the kind of things

- anaphoric
- previous turn of same participant
- segment or subsegment topic

that kind of thing

- anaphoric
- NP antecedent
- it sums up a description detailing the antecedent
- it may appear in coordination with a description antecedent
- segment or subsegment topic

the only kind of thing

- as above

and things

- end of utterance
- nonreferential
- focusing device
- like an ornament or way of speaking

the other thing

- explicit NP antecedent
- same turn
- several moves backwards

the X-Adj* thing

- subject complement in a copula
- subject is clue of implicit antecedent

there X-be one other thing Subj X-want to say

- cataphoric
- introducing topic of a new segment
- boundary move

this fact

- anaphoric
- discourse chunk related to segment topic
- previous exchange nominalised as *the fact that*
- minor adjustments are likely to be needed

the stuff

- anaphoric
- explicit NP antecedent
- high-saliency entity
- verb as a lexical clue

this stuff

- as above

at this stage

- implicit; thematic element
- the actual *stage* is not clearly specified

put it this/that way

- cataphoric
- discourse-chunk antecedent
- predicate of segment or subsegment topic
- subsequent move

words to that effect

- anaphoric
- discourse-chuk explicit antecedent
- previous move
- chains may occur
- previous move: *say that*

in that sense

- anaphoric
- antecedent is implicit adjective or qualifier;
- inside current turn; immediately preceding moves;

- move is a boundary move

the way X-clause

- anaphoric
- previous move
- segment or subsegment topic

in this way

- anaphoric
- PP attached to verb
- meaning like X-verb *as described above*
- discourse implicit; focusing device
- it sums up a narrative or description

X-verb (Obj) that way

- anaphoric
- discourse-chunk antecedent; implicit
- previous moves
- sum-up phrase

this/that point of view

- anaphoric; discourse implicit; focusing device;
- sum-up phrase
- subsequent move may contain explanation
- introduced by *I mean*

on that occasion

- explicit;
- event
- segment topic

in that case

- explicit; discourse chunk
- previous move or turn
- chain with anaphoric demonstrative possible
- a fact or event

under those terms

- explicit
- discourse-chunk antecedent
- previous turn or move
- selection may be required

Appendix C

The AL theory for anaphors in the Portuguese sample

C.1 Pronouns

C.1.1 Subject pronouns

global probability = 0.079

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtCCh=0.647	explicit=0.900	dthel=0.618
FtC=0.158	implicit=0.100	thel=0.315
DK=0.154		sst=0.046
ScRf=0.029		st=0.017
SK=0.008		sithel=0.004
Dx=0.004		

1. identify pronoun
   (a) pronoun is third person
      - go to instruction 3
   (b) pronoun is first or second person
      - go to instruction 2
2. check secondary reference
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - dthel = 0.714
       - thel = 0.286
   - recognition
     - separate from endophoric usage
     - analyse previous move
       - verbs *dizer*; *falar*

        - subject a third person pronoun or personal name
        - human-denoting NPs also possible: *a colega*
      - simultaneous tense and person shift
      - tense shift may be also `perfeito` to `imperfeito`
      - tense shift may be the only clue in some cases
  - resolution path
      - select first human candidate searching backwards
      - check lexical clues
      - if there is an identifying vocative
        - select it as the antecedent

3. select first appropriate candidate

  - if no appropriate candidate found
      - go to instruction 4
  - if an appropriate candidate is found
      - go to instruction 5

4. check shared knowledge

  - attached probabilities
      - type of antecedent
        - explicit = 0.500
        - implicit = 0.500
      - topical roles
        - thel in all cases
  - recognition
      - usually beginning of dialogue (up to 40 tone units)
      - no appropriate candidates
      - it often starts a chain
        - one exception to the features above
        - reference occurs late in the dialogue
        - lexical clues signal absence of fit candidates
        - association history in dialogue
        - still a particularly difficult case
  - resolution path
      - check dt and dthel in history of previous interactions
      - track chain till a definite description occurs
        - this would solve the tough case above

5. check lexical clues

  - if it is a good fit
      - go to instruction 7
  - if it is not a good fit

- go to instruction 6

6. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.528
    - implicit = 0.472
  - topical roles
    - thel = 0.528
    - dthel = 0.472
- recognition type one
  - indeterminate reference
  - pronoun *eles*
  - presence of location adjunct in turn
  - *lá* + PP
  - PP contains institutional referent
- resolution path type one
  - implicit antecedent
  - people in the institution
- recognition type two
  - candidate bypass
  - lexical clues
  - discourse-marker clues
  - prosodic clues
- resolution path type two
  - bypass first candidate
  - select dthel and st
  - check syntax and lexical clues
  - implicit antecedents possible
    - chains on DK, not FtCCh
    - ultimate resolution on SK
    - intervening NP possible, then:
    - antecedent made explicit in chain
- recognition type three
  - shift secondary to primary reference
  - preceding token of ScRf
  - DK token may occur between two ScRf tokens
  - *no meu entender, ele disse, eu não sou...*
  - shift back to primary reference is often unsignalled
- resolution type three
  - select antecedent of ScRf token as candidate
  - check lexical clues

- recognition type four
  - shift deictic to discourse reference
  - preceding token of Dx
  - identical pronoun
- resolution type four
  - bypass pronoun resolved by deixis
  - select next appropriate candidate
  - check lexical clues

7. check deixis
   - attached probabilities
     - type of antecedent
       - implicit in all cases
     - topical roles
       - dthel in all cases
   - recognition
     - dialogue with three or more participants
     - two participants refer to the other(s) one(s)
   - resolution path
     - select other(s) participant(s) as antecedent(s)
8. if 6 and 7 don't apply
9. accept first candidate

C.1.2 Object pronouns

global probability = 0.019

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtCCh=0.754	explicit=0.982	dthel=0.825
DK=0.141	implicit=0.018	thel=0.105
FtC=0.088		st=0.035
Dx=0.018		sst=0.035

1. select first appropriate candidate
2. check lexical clues
   - if it is a good fit
     - go to instruction 3
   - if it is not a good fit
     - go to instruction 2
3. check discourse knowledge
   - attached probabilities

  - type of antecedent
    - explicit in all cases
  - topical roles
    - dthel = 0.571
    - thel = 0.286
    - st = 0.143
- recognition type one
  - candidate bypass
  - lexical clues
  - discourse-marker clues: *tanto que*
  - prosodic clues
- resolution path type one
  - inanimate antecedent
  - bypass first candidate
  - check st and dt first

4. check deixis
   - attached probabilities
     - type of antecedent
       - implicit in all cases (only one)
     - topical roles
       - dthel in all cases
   - recognition
     - dialogue with three or more participants
     - two participants refer to the other(s) one(s)
   - resolution path
     - select other(s) participant(s) as antecedent(s)
5. if 3 does not apply
6. accept first candidate

C.1.3 Demonstratives

global probability = 0.056

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
CK=0.323	explicit=0.689	st=0.216
DK=0.269	NR=0.144	fdv=0.198
FtC=0.132	implicit=0.114	sst=0.168
FtCCh=0.108	disc. impl.=0.054	p_sst=0.168
Dx=0.072		thel=0.132
SetMb=0.054		p_st=0.060
LS=0.012		dthel=0.024
SetCr=0.012		p_dthel=0.018
AM=0.006		dt=0.012
Pl=0.006		p_uthel=0.006
WK=0.006		

1. check collocation list
    - if no match found
        - go to instruction 2
    - if a match is found
        - follow resolution pathway in entry
2. check world knowledge
    - attached probabilities
        - type of antecedent
            - implicit in all cases (one)
        - topical roles
            - thel in all cases
    - recognition
        - pronoun *o* in neutral sense
        - *piorar mesmo mais do que o que eu estou*
    - resolution path
        - nominalise verb in the modifier relative clause
        - check lexical clues for good fit
        - antecedents like *estado* out of *X-estar*;
3. check lexical signalling
    - attached probabilities
        - type of antecedent
            - implicit in all cases
        - topical roles
            - st = 0.500
            - dthel = 0.500
    - recognition
        - presence of modifier or modifying clause
        - cataphoric in questions
    - resolution path
        - use lexical information in modifier
        - select antecedent within search limit
4. check modified antecedent
    - attached probabilities
        - type of antecedent
            - explicit in all cases (one)
        - topical roles
            - thel in all cases
    - recognition

  - presence of modifier or modifying clause
  - lexical information does not lead to resolution
- resolution path
  - detach modifier
  - match modifier with discourse entities
  - select best match if close enough
  - if no good match is found
  - go to instruction 4

5. select first appropriate candidate
6. check lexical clues
   - if it is a good fit
     - go to instruction 10
   - if it is not a good fit
     - go to instruction 6
7. check discourse knowledge
   - attached probabilities
     - type of antecedent
       - explicit = 0.844
       - implicit = 0.156
     - topical roles
       - st = 0.244 p_sst = 0.333
       - sst = 0.111 p_st = 0.133
       - thel = 0.111 p_dthel = 0.022
       - dthel = 0.022 dt = 0.022
       - discourse-chunk antecedents = 0.488
   - recognition type one
     - pronoun is followed by modifying clause
     - discourse-specific information needed
     - *o* and *aquilo* are typical
   - resolution path
     - select candidates within search limit
     - check lexical clues
     - if none is a good fit
     - select st; dt; dthel
     - check lexical clues
   - recognition type two
     - pronouns *isso* and *aquilo*
   - resolution type two
     - check previous move or turn
     - event described is the antecedent

  - it is often implicit
  - if implicit, previous move is clue
  - if a good fit is not found
  - check st, dt and dthel(s)
- recognition type three
  - pronoun *isso*
- resolution type three
  - antecedent is a discourse chunk
  - select previous move or turn
  - check if it is a good fit
  - make adjustments
    - nominalisation with *the fact that...*
    - string moves together to form a complete utterance
    - especially if the previous turn includes several moves
    - comment moves may have to be left out
    - semantic conflicts have to be handled
  - check next move for clarifying questions
- recognition type four
  - pronoun *isso*
  - pronoun is object of verbs *pedir*; *falar*
  - pronoun may be preposed and linked by *que*
  - especially if next move begins with conjunction *que*
- resolution path type four
  - cataphoric reference
  - select subsequent move
  - check if it is a good fit
- recognition type five
  - pronoun *isso*
- resolution path type five
  - candidate bypass
  - check st, dt and dthel
  - select best fit
- recognition type six
  - pronoun *isso*
  - return pop
  - anaphor is in a boundary move
  - move introduces resumptive segment or subsegment
- resolution path type six
  - check topic of resumptive unit
  - if it is a good fit, accept it

8. check set member

- attached probabilities
  - type of antecedent
    - explicit = 0.889
    - implicit = 0.111
  - topical roles
    - sst = 0.444
    - st = 0.333
    - thel = 0.222
- recognition
  - pronouns *esse(s)*; *essa(s)*
  - pronouns *o(s)*; *a(s)*
  - modifiers always present with the latter
- resolution path
  - select first set-defining candidate
  - strip of modifiers
  - in doubt prefer the segment topic
  - combine head with anaphor modifiers
  - identify antecedent

9. check set creation

- attached probabilities
  - type of antecedent
    - explicit = 0.500
    - implicit = 0.500
  - topical roles
    - thel in all cases
- recognition
  - phrase *só isso*
  - preceding list of entities
- resolution path
  - antecedent is preceding set of entities
  - there may be no clear NP to define the set

10. check deixis

- attached probabilities
  - type of antecedent
    - explicit = 0.417
    - implicit = 0.583
  - topical roles
    - sst = 0.417
    - thel = 0.333
    - st = 0.083

        - dthel = 0.167
    - recognition
        - prosodic pattern
        - identifiable entity in physical environment
        - often followed by *aqui* or *aí*
    - resolution path
        - visual input needed for identification

11. check parallel

    - attached probabilities
        - type of antecedent
            - explicit in all cases (one)
        - topical roles
            - sst in all cases
    - recognition
        - pronouns *o*, *a*, *os*, *as*
        - pronouns *aquele*, *aquela*, *aquilo* seem possible
        - relative clause or PP modifiers
        - syntactic structure is similar to previous move
        - differences are the anaphors
    - resolution path
        - match modifiers with identical bits of similar move
        - fill the gaps to reconstruct move
        - select antecedent
        - Subj X-verb NP-Obj RelCl
        - Subj X-verb De-Obj RelCl

12. if it doesn't apply

13. accept selected candidate

C.1.4 Determinative possessives

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
ScRf=0.667	explicit=1.000	st=0.333
FtCCh=0.333		thel=0.333
		dthel=0.333

1. check secondary reference

    - attached probabilities
        - type of antecedent
            - explicit in all cases

  - topical roles
    - thel = 0.500
    - dthel = 0.500
- recognition
  - separation from endophoric usage
  - previous move or turn
    - verbs *dizer*, *falar*, *explicar*
    - subject a third person pronoun or personal name
    - tense switch in relation to anaphor move
- resolution path
  - if there is a personal pronoun in chain
    - select the same antecedent
  - if there is an identifying vocative
    - select the entity as the antecedent
  - if there is no vocative nor chain
    - select first human candidate searching backwards
    - check lexical clues

2. if it is a third-person pronoun
3. select the first appropriate candidate
4. accept it as the antecedent

C.1.5 Independent possessives

global probability = 0.000

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtCCh=1.000	explicit=1.000	st=1.000

1. select the first appropriate candidate
2. accept it as the antecedent

C.1.6 Numerals

global probability = 0.011

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
SetMb=0.353	explicit=0.912	sst=0.294
LS=0.176	implicit=0.088	st=0.235
LR=0.176		dt=0.176
AM=0.118		dthel=0.147
VMm=0.059		thel=0.088
DK=0.059		p_st=0.059
FtC=0.029		
WK=0.029		

1. check lexical repetition
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - dt = 0.167
            - sst = 0.666
            - dthel = 0.167
    - recognition
        - numeral is NP head
        - identifying modifier present
    - resolution path
        - search backwards
        - if an identical NP is found
        - accept it as antecedent
2. check modified antecedent
    - attached probabilities
        - type of antecedent
            - explicit = 0.750
            - implicit = 0.250
        - topical roles
            - st = 0.250
            - sst = 0.250
            - dthel = 0.500
    - recognition type one
        - numeral is NP head
        - a similar NP is within search limit
    - resolution path type one
        - search backwards
        - select similar NP as candidate
        - check if it is a good fit
    - recognition type two
        - numeral is an addition to NP head
        - typically linked by preposition *e*
        - *setenta quilos e duzentos*
        - it follows very closely a previous reading
        - the previous reading is very similar
        - for instance, *setenta quilos*
    - resolution type two
        - typically implicit antecedent
        - subdivision of measurement

  - accept it as antecedent if it fits

3. check lexical signalling

   - attached probabilities
     - type of antecedent
       - explicit = 0.833
       - implicit = 0.167
     - topical roles
       - st = 0.333
       - sst = 0.333
       - thel = 0.167
       - dthel = 0.167
   - recognition type one
     - numeral is NP head
     - it specifies a measurement reading
     - or a type of subset based on a measurement
   - resolution path type one
     - search backwards
     - if a NP is found which
       - with a measure or a type of subset as head
       - a numeral as modifier specifying the reading
     - accept the head as antecedent
   - recognition type two
     - numeral is NP head
     - it is a time of the clock
     - a time of the clock has been mentioned before
   - resolution type two
     - search backwards for a time of the clock
     - link it to the anaphor if it rephrases it

4. check world knowledge

   - attached probabilities
     - type of antecedent
       - implicit in all cases
     - topical roles
       - thel in all cases
   - recognition
     - numeral is NP head
     - PP modifier
   - resolution path
     - modifier contains an uncountable noun
     - a unit is used to measure quantities

  - antecedent is this typical unit
  - ex.: *duzentos e cinquenta de carne*

5. check set member

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - dt = 0.417
    - st = 0.250
    - sst = 0.250
    - thel = 0.083
- recognition type one
  - numeral is NP head
  - no determiner
  - adverb modifier *só* possible
  - contraction with preposition possible: *num(a)*
- resolution path type one
  - search backwards
  - select NP head as candidate if
  - it is a NP with a numeral postdeterminer
    - especially if NP head is *vezes*
  - it is a NP with a numeral as head
    - thus an anaphor as well
    - select the same antecedent as candidate
  - it is a set-defining description
    - indefinite description
    - set member which introduces the set
      - definite article or demonstrative determiner
      - adjectival and clausal modifiers possible
    - generic with no determiner
  - check lexical clues and topical role
    - use modifiers for check
    - adjectives and clauses attached to the anaphor
  - if it is a good fit, accept as antecedent
    - if numeral anaphor is *um(a)*
    - adjustment from plural to singular is required
  - if it is not a good fit
  - select next candidate within search limit
- recognition type two
  - numeral is NP head
  - determiner is a demonstrative
- resolution path type two

  - search backwards
  - select plural NP as candidate
  - determiner is indefinite pronoun
  - antecedent is NP head
- recognition type three
  - numeral is NP head
  - adjective as modifier
- resolution type three
  - search backwards
  - use adjective as clue
  - antecedent is head of (N + Adj) NP
  - Adj is semantically related to anaphor modifier
  - often either synonymy or antonymy

6. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - st = 0.500
    - dthel = 0.500
- recognition type one
  - numeral is NP head
  - set member check yields an uncertain fit
  - typically two candidates in previous move
- resolution path type one
  - analyse previous move in search of clues
  - see if move implicitly defines a quantity
  - check which candidate the quantity is assigned to
  - check if it improves acceptance of candidate
  - if yes, accept it as antecedent
- recognition type two
  - numeral is NP head
  - there is no appropriate candidate within search limit
- resolution path type two
  - analyse anaphor move and previous move for clues
  - check dt, st and dthel(s)
  - if a good fit is found
  - accept it as antecedent

7. check verbatim memory

- attached probabilities

  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_st in all cases
- recognition
  - numeral is NP head
  - numeral is response to *quanto* question
  - or restates part of previous move
  - typically subject complement of copula in previous move
- resolution path
  - select previous move with subject and verb
  - if it is good fit, accept it as antecedent
  - if it is a question, build antecedent

8. accept first candidate
9. typically a preposed noun phrase

C.1.7 Indefinite pronoun

global probability = 0.003

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
SetMb=0.555	explicit=1.000	sst=0.333
Pl=0.222		st=0.222
DK=0.111		p_st=0.222
CK=0.111		dthel=0.111
		sithel=0.111

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. check set member
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - sst = 0.600
       - st = 0.200
       - dthel = 0.200
   - recognition

  - pronouns *outro(a)(s)*; *algum(a)(s)*
- resolution path
  - search backwards
  - select set-defining NP within search limit
  - check lexical clues
  - if it is a good fit, accept it
  - if not, select next candidate within search limit

3. check parallel

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - st = 0.500
    - p_st = 0.500
- recognition
  - pronouns *nenhum((a)*; *muito(a)(s)*
- resolution path
  - analyse last move
  - select NP in same function as anaphor as candidate
  - check lexical clues
  - if anaphor is elliptical response to question
  - match IP with possible antecedent
  - check lexical clues

4. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - sithel in all cases (one)
- recognition
  - pronouns *todo(a)(s)*; *nenhum(a)*
  - no syntactic parallel in previous move
  - especially if syntax is broken in previous move
- resolution path
  - similar to SetMb resolution
  - however, reference is to full set
  - consider candidates introduced in false starts

C.2 Verbs and adverbials

C.2.1 Adverbs of Place

global probability = 0.044

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
Dx=0.396	explicit=0.856	dt=0.295
FtCCh=0.237	implicit=0.122	dthel=0.288
DK=0.158	NR=0.022	st=0.144
FtC=0.158		thel=0.129
CK=0.022		sst=0.101
SK=0.007		fdv=0.022
LR=0.007		p_dthel=0.014
		uthel=0.007

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. select first candidate
3. check lexical clues
   - if it is a good fit
   - go to instruction 8
   - if it is not a good fit
   - go to instruction 4
4. check shared knowledge
   - attached probabilities
     - type of antecedent
       - implicit in all cases
     - topical roles
       - thel in all cases (one)
   - recognition
     - no candidate fits within search limit
     - a distinct adverb is used
     - no chain with adverbs within search limit
   - resolution path
     - use lexical clues in move to identify location
     - *eu vi ali escrito*
     - use the writing in question to identify the location

5. check discourse knowledge

    - attached probabilities
        - type of antecedent
            - explicit = 0.864
            - implicit = 0.136
        - topical roles
            - dthel = 0.364
            - dt = 0.318
            - st = 0.136
            - sst = 0.091
            - thel = 0.045
            - uthel = 0.045
    - recognition type one
        - candidate bypass
        - two competing locations within search limit
    - resolution path type one
        - check lexical clues
            - especially if anaphor is object of verb *mandar*
            - search for a previous token of *mandar*
        - check next move for reiteration
        - if the more distant candidate is a better fit
        - accept it as antecedent
    - recognition type two
        - no candidate within search limit
        - anaphor is object of verb *voltar*
    - resolution path type two
        - check association history of subject
        - select possible locations
        - check dt, dthel(s), st, sst
        - check next move for confirmation
    - recognition type three
        - return pop
        - anaphor occurs in boundary move
        - boundary moves starts a resumptive discourse unit
    - resolution type three
        - analyse boundary move
        - identify resumptive discourse unit
        - check dt, st, sst, dthel(s)
        - use lexical clues and next move for confirmation
    - recognition type four
        - adverb *lá*
        - no candidate within search limit or

  - no good fit when lexical clues are used
- resolution type four
  - implicit antecedent
  - assumed location where a fact described occurs
  - create a referent for the assumed location
  - accept it as antecedent to the anaphor

6. check deixis

- attached probabilities
  - type of antecedent
    - explicit = 0.782
    - implicit = 0.218
- topical roles
  - dthel = 0.473
  - st = 0.236
  - thel = 0.236
  - sst = 0.055

7. recognition

- recognition
  - pronouns *aqui*; *aí*; *ali*; *cá*
  - contractions with preposition *de*
  - prosodic pattern
  - identifiable entity in physical environment
  - chains are frequent
- resolution path
  - antecedent usually an object and not a location
    - however, *aqui* in a restrictive sense refers to locations
    - it is only annotated if it means an actual location
    - not simply the place where the conversation takes place
  - visual input needed for identification
  - domain specific: parts of the body
  - pointing gesture or touch

8. check verbatim memory

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_dthel = 1.000 (one case)
- recognition
  - adverbs of place requiring complement

- resolution path
  - previous move is antecedent

9. check lexical repetition
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - p_dthel = 1.000 (one case)
   - recognition
     - anaphor is a repetition
     - previous token is within search limit
   - resolution path
     - select same antecedent as previous token
     - check if it is a good fit
     - accept it as antecedent
10. accept selected candidate

C.2.2 Adverbs of response

global probability = 0.023

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=0.956	explicit=0.985	p_sst=0.456
CK=0.044	NR=0.015	p_st=0.353
		p_dthel=0.088
		p_thel=0.074
		fdv=0.015
		p_dt=0.015

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. recognition
   - *não* and *uhum*
   - response to questions
3. verbatim memory in all cases
4. explicit in all cases
5. discourse-chunk antecedents in all cases
6. previous move in all cases
7. adjustments may be required

C.2.3 Adverbs of Time

global probability = 0.003

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_st=0.667
		p_sst=0.333

1. *ainda*; *já*; *logo*
2. adverbial phrases were included: *todo dia*
3. previous move is antecedent
4. adverb is often part of previous move
5. adjustments may be required

C.2.4 Adverbs of Manner

global probability = 0.005

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
DK=0.500	explicit=0.929	p_st=0.357
CK=0.429	disc.impl.=0.071	p_sst=0.214
VMm=0.071		p_thel=0.143
		sst=0.071
		fdv=0.071
		dthel=0.071
		p_dt=0.071

- all tokens are the adverb *assim*
- discourse-chunk antecedents: 0.785

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. check discourse knowledge
   - attached probabilities
     - type of antecedent
       - explicit = 0.857
       - dim = 0.143
     - topical roles
       - p_sst = 0.286
       - p_st = 0.143

    - sst = 0.143
    - fdv = 0.143
    - dthel = 0.143
    - p_thel = 0.143
- recognition type one
    - return pop
    - anaphor occurs in boundary turn
    - preceding move signals change
    - *aí é aquele negócio que eu falei...*
- resolution path type one
    - identify resumptive discourse unit
    - check st, dt, dthel(s)
    - if anaphor is attached to a NP
    - antecedent is NP
    - contiguity is not a sure factor
- recognition type two
    - anaphor attached to NP
    - attachment may not be evident
    - NP in question appears in previous move or moves
- resolution path type two
    - discourse-chunk antecedent
    - select previous move containing NP as antecedent
    - check if it is a good fit
    - if not, select move before the previous
    - check if it is a good fit
- recognition type three
    - anaphor attached to NP
    - no token of NP in previous discourse
    - NP is *alteração*; *mudança*;
    - NP is *situação* or similar
- resolution type three
    - select previous move as candidate
    - if it is a good fit, accept it
    - if it is not, break in into clauses
    - check clauses separately
    - make required adjustments
- recognition type four
    - anaphor attached to verb
    - attachment may not be evident
    - verb in question appeared in previous discourse
- resolution path type four
    - search backwards for previous token of verb

- antecedent is adjunct to previous token

3. check verbatim memory

   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - p_st in all cases
   - recognition
     - token of identical anaphor in previous move
     - both anaphors have the same syntactic function
   - resolution path
     - select the same antecedent as previous anaphor
     - check if it is a good fit
     - if it is, accept it as antecedent

C.2.5 Adverbs of Frequency

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=0.500	explicit=0.500	p_st=0.500
DK=0.500	implicit=0.500	p_sst=0.500

1. select previous move as candidate
2. if it is a good fit
3. accpet it as antecedent
4. classify the token as VMm resolution
5. if it is not a good fit
6. check discourse knowledge

   - recognition
     - anaphor is *nunca*
     - previous move contains a coordination
     - the coordinating conjunction is *or*
   - resolution path
     - antecedent is implicit
     - change coordinating conjunction to *and*
     - make required adjustments
     - accept the move as antecedent

C.2.6 Adverbs of Intensity

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_st=0.333
		p_sst=0.333
		thel=0.333

- all anaphor tokens are *mais*
- all antecedents are discourse-chunks
    1. select previous move as candidate
    2. adjust according to repetitions in the anaphor move
    3. accept it as antecedent

C.2.7 Adverbs of Exclusion

global probability = 0.000

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_sst=1.000

- one token: *só*
- discourse-chunk antecedent
    1. select previous move or turn as antecedent
    2. make required adjustments

C.2.8 Wh-word

global probability = 0.003

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=0.778	explicit=0.889	p_st=0.333
FtCCh=0.111	implicit=0.111	p_sst=0.333
DK=0.111		st=0.111
		p_thel=0.111
		fdv=0.111

1. check first candidate
    - recognition
        - wh-word is *qual*
        - determinative function
        - NP head missing
        - often question is inverted
        - *qual* at the end of question and stressed

- resolution path
  - select first candidate
  - if it is an anaphor, resolve as chain
  - if it is a good fit, accept it as antecedent

2. check discourse knowledge

   - recognition
     - wh-word is *porque*
     - previous move is not a good fit
   - resolution path
     - try to interpret concealed meanings in previous move
     - for instance, laughs and ironic intonation
     - *porque* **você está rindo?**

3. check verbatim memory

   - recogntion
     - any wh-word requiring complement
     - typical cases: *sabe porque?*;
     - *como? não sei*; *para que?*
   - resolution path
     - discourse-chunk antecedents
     - select previous move as candidate
     - make required adjustments
     - check if it is a good fit
     - if it is, accept it as antecedent

C.2.9 Prepositional phrase

global probability = 0.013

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=0.263	explicit=1.000	p_st=0.316
AM=0.211		sst=0.184
LR=0.211		st=0.132
DK=0.132		thel=0.105
WK=0.105		p_sst=0.105
FtC=0.026		p_dthel=0.079
CK=0.026		dthel=0.053
FtCCh=0.026		dt=0.026

1. check collocation list

   - if no match found
     - go to instruction 2
   - if a match is found

  - follow resolution pathway in entry

2. check world knowledge

   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - sst = 0.666
       - thel = 0.333
   - recognition
     - anaphor is not a repetition
     - anaphor attaches to an NP in previous move
     - attachment relies on knowledge about the world
     - world knowledge is retrieved by lexical information
     - lexical information required is conveyed by the anaphor
   - resolution path
     - check NPs in previous move
     - select appropriate head as antecedent
     - check lexical clues reconstructing move

3. check lexical repetition

   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - p_st = 0.500
       - sst = 0.125
       - thel = 0.125
       - p_sst = 0.125
       - p_dthel = 0.125
   - recognition
     - anaphor is a repetition of a PP in previous move
     - anaphor is often an answer to a question
   - resolution path
     - select previous move as antecedent
     - check if it is a good fit
     - if it is, accept it as antecedent

4. check modified antecedent

   - attached probabilities
     - type of antecedent
       - explicit in all cases

  - topical roles
    - sst = 0.375
    - st = 0.250
    - p_dthel = 0.250
    - thel = 0.125
- recognition
  - anaphor is a partial repetition of NP in previous move
  - PP is often attached to NP head in previous move
- resolution path
  - analyse move where partial repetition occurs
  - usually but not always the previous move
  - select head to which partial repetition is attached
  - check if it is a good fit
  - if it is, accept it as antecedent

5. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_st = 0.400
    - st = 0.200
    - sst = 0.200
    - thel = 0.200
- recognition type one
  - anaphor is a repetition
  - previous token is beyond search limit
- resolution path type one
  - analyse previous move
  - check attachment to NPs in previous move
  - use information in previous token
  - check if it is a good fit
  - if it is, accept it as antecedent
- recognition type two
  - anaphor is not a repetition
  - anaphor is a PP requiring complement
  - *abaixo*; *no lugar*
  - complement has to be retrieved from previous discourse
  - retrieval requires full processing of moves
- resolution path two
  - analyse previous move or moves
  - select NP which fits best

  - check lexical clues in anaphor and anaphor move
  - if a good fit is found
  - accept it as antecedent
- recognition type three
  - anaphor is not a repetition
  - previous move is a candidate
  - lexical clues do not rule it out
  - the misfit is slight
  - move has to be bypassed
  - antecedent chunk is further back
- resolution path type two
  - complex discourse processing required
  - if a slight misfit is detected
  - check first candidate move where st appears
  - check dt; dhtel
  - if a good fit is found
  - consider bypassing
  - if anaphor is in a boundary move of new subsegment
  - stronger possibility of required bypass

6. check first candidate

- recognition
  - PP is *no meio*; *diferente da*
  - the complement is missing
- resolution path
  - if PP is *no meio*
    - search for plural NP
    - especially if a numeral postdeterminer is present
    - check if it is a good fit
    - if yes, accept it as antecedent
- if PP is *diferente da*
  - if it is a copula and PP is complement
  - antecedent is subject
  - if it is not a copula and PP is adjunct
  - antecedent is the whole preceding clause

7. check verbatim memory

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_st = 0.600
    - p_sst = 0.300

        - st = 0.100
    - recognition
        - anaphor is in a move that is a verbless or subordinate clause
        - anaphor requires attachment to a previous move for interpretation
        - preposition is often a surface clue to attachment
        - lexical semantics is not a definite clue to attachment
        - anaphor may be a response or suggested response to a query
    - resolution path
        - mostly discourse-chunk antecedents
        - analyse previous move
            - if previous move is also an anaphor
            - and antecedent is a discourse chunk
            - analyse reconstructed move
        - attach it to anaphor to reconstruct move
        - previous move is usually the antecedent
        - anaphor is usually attached to the end of antecedent
        - adjustments may be required
        - overlapping may require deletion of words in antecedent

C.2.10 Do-phrase anaphora

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_st=0.500 p_dthel=0.250 p_sst=0.250

1. recognition
    - X-fazer *isso*
    - reference to a verb phrase
    - it may appear inverted
    - *isso eu faço*
2. resolution path
    - verb phrase in previous move is antecedent

C.2.11 Linking verbs

global probability = 0.130

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
CK=0.859	NR=0.470	fdv=0.475
FtCCh=0.078	explicit=0.511	p_sst=0.186
FtC=0.051	implicit=0.016	p_st=0.138
DK=0.009	disc.impl.=0.002	p_dthel=0.041
WK=0.002		p_thel=0.034
		thel=0.028
		st=0.028
		sst=0.025
		p_dt=0.018
		uthel=0.005
		p_uthel=0.005
		dt=0.002
		p_fdv=0.002

1. identify linking verb
2. if it is *é*, *né* or *(es)tá*
3. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
4. select first candidate for subject complement
5. it is a good fit, accept it as antecedent
6. if it is not a good fit, go to instruction 13
7. if there is no overt subject as well
8. use parallelism information
9. select first candidate for subject as well
10. if it is not a good fit and it is *né*
11. go to instruction 12
12. check world knowledge
    - attached probabilities
      - type of antecedent
        - implicit in all cases (one)
      - topical roles
        - uthel in all cases
    - recognition
      - Q-tag to verbless clause
      - NP only in clause; uthel inferrable

- resolution path
  - analyse NP in clause
  - if NP is a proper noun
  - anaphor refers to, for instance, *seu nome*
  - other forms of uthel possible

13. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.750
    - implicit = 0.250
  - topical roles
    - st = 0.250
    - thel = 0.250
    - dthel = 0.250
    - p_dt = 0.250
    - dthel
- recognition type one
  - simple candidate bypass
  - *vê o que que é*
  - previous move has no candidates that fit well
  - typically entities are in chain with main clause subject
  - antecedent is further back, but usually within search limit
  - reference may cross segment boundaries
  - *pode ser*
  - omitted antecedent - broken syntax
  - *era o que ?*
- resolution path
  - bypass first candidate
  - select a non-human entity
  - check lexical clues
  - if it is a good fit, accept as antecedent
- recognition type two
  - *pode ser* as an uncomplemented question
  - previous move has a good fit for SubjC
  - not a good fit for SUBJ
- resolution path type two
  - nominalise verb in previous move
  - check lexical clues
  - if it is a good fit, accept as antecedent
- recognition type three
  - plain linking verb
  - previous move has a good fit for subject

  - syntax in previous move is broken
  - no SubjC to retrieve; implicit
- resolution type three
  - use the information in the previous move
  - postulate a SubjC
  - check lexical clues
  - if it is a good fit, accept as antecedent
- recognition type four
  - *era o que* question
  - typically followed by *era* + SubjC
  - the next move is also a question
- resolution type four
  - check next move
  - use SubjC in next move to infer the SUBJ
  - typically a part-whole or set-subset relation
  - if SubjC is a complex NP, check head as antecedent
  - if it is a good fit, accept as antecedent

C.2.12 Operators

global probability = 0.012

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_sst=0.500
		p_st=0.333
		p_thel=0.167

1. search previous token of operator
   - found in previous turn in all cases
   - response to question
   - agreement adjustment for search
2. antecedent is discourse chunk
3. chunk begins at word after operator
4. in all cases the main verb
5. chunk ends at end of utterance
6. treatment expressions excluded

C.2.13 Anaphoric verbs

global probability = 0.181

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtCCh=0.339	explicit=0.770	st=0.232
FtC=0.270	NR=0.105	sst=0.156
CK=0.200	implicit=0.087	thel=0.147
DK=0.172	disc=0.038	fdv=0.143
ScRf=0.011		dthel=0.140
SK=0.005		p_st=0.064
Pl=0.002		p_sst=0.049
		dt=0.036
		p_dthel=0.015
		p_dt=0.005
		p_thel=0.009
		sithel=0.002
		uthel=0.002

1. check argument structure of main verb
2. compare with main verb of token
3. if essential arguments are missing
4. select token as anaphoric
5. check person of verb form
   - verb form is third person
     - go to instruction 7
   - verb form is first or second person
     - go to instruction 6
6. check secondary reference
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - dthel = 0.833
       - sst = 0.167
   - recognition
     - separate from endophoric use
     - check previous move
     - verb *dizer* in previous move
       - third person form with human subject
       - tense shift from past to present
     - first or second person ScRf pronoun in previous move

  - first or second person ScRf anaphoric verb form in previous move
  - change in prosodic pattern
- resolution path
  - select candidate in previous move on the basis of argument structure
  - check lexical clues
  - if it is a good fit, accept it as antecedent
  - no tokens in different argument positions

7. check collocation list
   - if no match found
     - go to instruction 8
   - if a match is found
     - follow resolution pathway in entry
8. examine candidates within search limit
9. use argument structure to select candidate
10. check if it is a good fit
    - if it is a good fit
    - go to instruction 13
    - if it is not a good fit
    - go to instruction 11
11. check shared knowledge
    - attached probabilities
      - type of antecedent
        - implicit in all cases
      - topical roles
        - st in all cases
    - recognition
      - the referent has not been introduced
      - domain knowledge strongly suggests a referent
      - first candidate backwards may be acceptable but not perfect
    - resolution path
      - use domain knowledge information
      - establish association history of verb
      - select the best candidate
12. check discourse knowledge
    - attached probabilities
      - type of antecedent
        - explicit = 0.823

    - implicit = 0.165
    - disc. impl. = 0.013
  - topical roles
    - st = 0.342 p_st = 0.051
    - dthel = 0.215 thel = 0.051
    - sst = 0.127 p_dthel = 0.013
    - dt = 0.114 fdv = 0.013
    - p_sst = 0.063 sithel = 0.013

- recognition type one
  - missing verb phrase head and object
  - verb *precisar*
  - move introduced by *mas* or *aí*
- resolution path type one
  - bypass verb phrase in previous move
  - select first verb phrase in preceding subsegment
  - check if it is a good fit
  - look ahead may reveal a reiteration in next move
- recognition type two
  - missing object of verb
  - long-distance retrieval
  - association history may contain alternative candidate
  - or it may be the first token of verb in discourse
- resolution path type two
  - if association history of verb is conclusive
  - select associated candidate in spite of distance
  - accept it as antecedent if saliency supports choice
  - if association history is inconclusive or none
  - local topic should be a lexical clue
  - check association history of local topic
  - if it is a subsegment, check st
  - if st is a good fit, accept it as antecedent
  - if it is not or if the discourse unit is a segment
  - select dt; and then dthel(s) within association history of st
  - check and accept best fit as antecedent
  - if none is found, check all elements in association history of st
  - lookahead may detect a reiteration
- recognition type three
  - indirect object of verb missing
  - explicit direct object
  - verb is a causative verb
  - *forçar*; *obrigar*;
  - or direct object is missing
  - verb is *querer*; *poder*; etc.

- resolution path type three
  - antecedent is an infinitive form
  - search discourse unit for infinitive form
  - the antecedent may be part of a noun phrase
  - *vontade de comer*; *interesse em fazer...*
- recognition type four
  - verb object is missing
  - often first token of verb in discourse
  - jarred use of first candidate as a clue
- resolution path type four
  - antecedent may be implicit or
  - verb may be a weak selector
  - complex discourse processing required
  - selectional restriction of verb
  - world knowledge
  - information in discourse unit
  - lexical semantics of first candidate
  - *dente* leading to:
  - *extração; rote; local da extração*
  - *exame* leading to *resultado do exame*
- recognition type five
  - subject and object omitted
  - verb can also be intransitive
  - *continuar*; *modificar*
  - agreement is flawed
- resolution path type five
  - use object of previous token as subject of second
  - resume to previous structure if required
- recognition type six
  - simple candidate bypass
  - subject or object
  - first candidate may appear in a clarifying question
  - lexical clues yield a poor fit for first candidate, but:
  - poor fit may not be so clear in weak selectors: (*fazer*)
  - recognition of poor fit may depend on world knowledge
- resolution path type six
  - bypass first candidate
  - check if the same verb occurs within segment
  - if yes, use argument structure to select antecedent
  - check fit in anaphor move
  - if it is a good fit, accept it as antecedent
  - if it isn't, check st, dt, and dthel(s)
  - accept best fit as antecedent

- recognition type seven
  - verb *ir*
  - complement omitted
- resolution path type seven
  - analyse previous turn
  - if it is a question with the same subject
  - transform the predication in an infinitive verb phrase
  - select as candidate and check if it is a good fit
  - if it is, accept it as antecedent
- recognition type eight
  - verb is third person plural
  - subject is omitted
- resolution path type eight
  - antecedent is indetermined *eles*
  - a specific referent is not required
  - it may refer to a hypothetical agent
- recognition type nine
  - subject or object omitted
  - first candidate is a complex NP
- resolution type nine
  - check if a NP component is highly salient
  - especially if it is st or sst; if yes:
  - extract candidate from NP (typically the object of a PP)
  - check if it is a good fit
  - if it is, accept it as antecedent
- recognition type ten
  - subject or object omitted
  - no adequate candidates within search limits
- resolution path type ten
  - nominalise verb in previous move
  - if it is a LV, go to move before the previous
  - if move is part of a complex turn
  - check all verbs in turn
  - check if it is a good fit
  - if it is, accept it as antecedent
- recognition type eleven
  - verb *saber*
  - object missing
  - no appropriate candidate within search limit
- resolution path type eleven
  - antecedent is implicit
  - it is the subject of previous conversation

  - it is too specific to be discourse implicit
  - *a causa*; *o motivo*;
- recognition type twelve
  - mismatch between verb and object
  - selectional restrictions would disallow the object
- resolution path type twelve
  - implicit antecedent
  - lexical clues in inadequate object
  - lookahead for reiteration of verb in next move or turn
- recognition type thirteen
  - return pop
  - anaphor is in a boundary move
  - prosodic clues
  - discourse-marker clues (*mas*)
  - lexical clues: candidates within search limit are a poor fit
- resolution path type thirteen
  - check topic of resumptive discourse unit
  - if it is a good fit
  - accept it as antecedent
- recognition type fourteen
  - verb is in subordinate clause
  - first candidate in main sentence is not a good fit
- resolution path type fourteen
  - bypass first candidate
  - select first candidate in previous turn or move
  - if it is a good fit, accept it as antecedent
  - if it is not, follow type six path
- recognition type fifteen
  - verb *achar*
  - omitted object
- resolution path type fifteen
  - typically discourse-chunk antecedent
  - previous move or nominative clause in previous move
- recognition type sixteen
  - sequence of questions
  - two question moves in sequence
  - typically different verbs in each move
  - two response moves containing anaphors follow
  - verbs in response moves match verbs in questions
- resolution path type sixteen
  - match each question with each response
  - use verbs as signal

  - select antecedents as in first-candidate strategies
- recognition type seventeen
  - sequence of moves with the same verb
  - anaphoric verb in second move
  - different senses of verb
  - typically verb *ter*
  - the sense of *haver* and of *possuir*
- resolution path type seventeen
  - bypass first candidate
  - correct antecedent within search limit
  - select from available candidates
  - check lexical clues
  - accept best fit as antecedent

13. check parallel

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - sst in all cases
- recognition
  - same verb within search limit
  - intervening token of different verb
- resolution path
  - select identical verb token arguments
  - check lexical clues
  - if it is a good fit, accept it as antecedent

14. accept candidate as antecedent

C.2.14 Copula-FNP

global probability = 0.028

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtC=0.366	explicit=0.805	sst=0.280
FtCCh=0.329	implicit=0.134	thel=0.244
DK=0.195	NR=0.037	st=0.159
CK=0.073	disc.impl.=0.024	fdv=0.061
WK=0.037		dthel=0.061
		dt=0.061
		p_sst=0.049
		p_thel=0.037
		p_st=0.024
		p_dthel=0.012
		p_dt=0.012

1. check collocation list
    - if no match found
        - go to instruction 2
    - if a match is found
        - follow resolution pathway in entry
2. select first appropriate candidate within search limit
3. check if it is a good fit
4. if it is, go to instruction 8
5. if it is not, go to instruction 6
6. check world knowledge
    - attached probabilities
        - type of antecedent
            - implicit in all cases
        - topical roles
            - thel in all cases
    - recognition
        - no candidate is a good fit within search limit
    - resolution path
        - analise NP in anaphor
        - if the NP defines a member of a specialised set S
        - antecedent is *o nome de* + DefArt + S
        - if the NP is of the form N + modifiers
        - antecedent is DefArt + N
7. check discourse knowledge
    - attached probabilities
        - type of antecedent
            - explicit = 0.800
            - implicit = 0.150
            - disc.impl. = 0.050
        - topical roles
            - thel = 0.312 st = 0.062
            - p_sst = 0.188 dt = 0.062
            - sst = 0.188 p_dt = 0.062
            - fdv = 0.062 p_st = 0.062
    - recognition type one
        - candidate bypass
        - first candidate is not a good fit
        - or previous move is an apposition

- resolution path type one
  - bypass first candidate in previous move
  - if the move before the previous is a copular construction
  - select candidate and check if it is st or sst
  - if it is, check lexical clues in anaphor move
  - if it is not, check st or sst first
  - if either is a good fit, accept it as antecedent
  - if NP in anaphor is st or sst
  - check thel(s) until a plausible candidate is found
- recognition type two
  - first candidate is inappropriate or not a good fit
  - previous move is not a copular structure
- resolution type two
  - nominalise verb in previous move
  - select resulting NP as candidate
  - check lexical clues in anaphor move
  - if it is a good fit, accept it as antecedent
  - if it is not a good fit
  - search for a token of the discourse unit topic
  - nominalise verb and check if it is a good fit
  - if it is a subsegment, repeat process with st
- recognition type three
  - generic NP in anaphor
  - judgement on the state of things
  - *não é preferência*
- resolution path type three
  - discourse implicit antecedent
  - *o problema*; *a questão*
- recognition type four
  - NP in anaphor is an item in an enumeration
  - the enumeration is being carried out
- resolution path type four
  - select an explicit thel in the segment
  - if it is a plural definite description
  - the singular form is the antecedent
- recognition type five
  - *eu não sei se* + anaphor
  - there may be a sequence of tokens
  - *se* + anaphor with main clause omitted
- resolution path type five
  - antecedent is implicit
  - lookahead for an indefinite description

  - if one is found, accept it as antecedent
  - if not, postulate an appropriate generic phrase
  - as specific as context allows
- recognition type six
  - omitted clausal subject
  - polarity contrast with previous move typical
- resolution path type six
  - implicit discourse-chunk antecedent
  - it must be built out of discourse information
  - this information is typically scattered in discourse unit
  - procedure is uncertain but:
  - invert polarity of previous move
  - use verb to build a relative clause
  - NP in anaphor is the subject (relative pronoun)
  - object is local topic or clause containing local topic
  - adjustments are often required
- recognition type seven
  - omitted clausal subject
  - previous move is an apposition
  - or previous move reiterates the preceding one
  - typically two questions which mean the same
- resolution path type seven
  - bypass apposition move or repeated question
  - select the preceding move as candidate antecedent
  - check if it is a good fit as relative clause
  - if it is, accept it as antecedent

8. accept it as antecedent

C.2.15 Copula-Adj

global probability = 0.032

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
CK=0.515	explicit=0.485	fdv=0.495
DK=0.172	NR=0.465	st=0.202
FtCCh=0.162	disc.impl.=0.030	sst=0.152
FtC=0.151	implicit=0.020	p_sst=0.061
		p_st=0.040
		dthel=0.020
		thel=0.010
		p_dt=0.010
		dt=0.010

1. check collocation list

- if no match found
  - go to instruction 2
- if a match is found
  - follow resolution pathway in entry

2. select first appropriate candidate within search limit
3. check if it is a good fit
4. if it is, go to instruction 7
5. if it is not, go to instruction 6
6. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.941
    - implicit = 0.059
  - topical roles
    - p_sst = 0.294 thel = 0.059
    - sst = 0.294 dthel = 0.059
    - st = 0.235 p_dt = 0.059
- recognition type one
  - omitted clausal subject
  - anaphor is often in nominative clause
  - main clause verb is usually *saber*
  - plain tokens as questions possible
  - plain tokens like *é melhor* also possible
- resolution path type one
  - if adjective accepts infinitive clauses as complement
    - make predicate of previous move an infinitive clause
    - strip of time expressions
    - select it as candidate antecedent
    - check if it is a good fit
    - adjust agreement morphemes of anaphor if required
    - if it is, accept it as antecedent
    - if it is not, bypass it and repeat the process
    - especially if previous move is a short apposition or comment
    - if verb in previous move is a complex verb phrase
    - and if verb phrase contains an infinitive clause
    - select infinitive clause only as antecedent
    - repeat the process
    - if subject in any of the moves is a clause
    - check if it is a good fit
    - especially if anaphor is similar to subject complement
  - if adjective only accepts nominative clauses as complement

        - select the whole previous move as antecedent
        - check if it is a good fit
        - if it is, accept it as antecedent
- recognition type two
    - anaphor is in a subordinate clause
    - clause is a correction of a previous move
    - previous move is partially repeated without the subject
    - partial reiteration is usually followed by *não*
    - subsequent move is the anaphor move
    - main clause is usually not reiterated
- resolution path type two
    - subject of corrected clause is antecedent
- recognition type three
    - candidate bypass
    - first candidate is not a good fit
- resolution path type three
    - bypass first candidate
    - if anaphor is in a subsegment
    - select st as candidate
    - check if it is a good fit
    - if it is, accept it as antecedent
    - if it is not or if it is not a subsegment
    - check dt and dthel(s)
- recognition type four
    - copular verb is in the present subjunctive
    - previous move is also in the present subjunctive
    - previous move is not a copular structure
- resolution type four
    - nominalise verb phrase in previous move
    - select it as candidate
    - check if it is a good fit
    - if it is, accept it as antecedent
- recognition type five
    - conditional subordinate clause
    - *se (não) for muito difícil*
    - sequence within segment with inverted polarity possible
    - the same referent contrasted
- resolution type five
    - both anaphoric and cataphoric cases
    - if main clause is in the conditional tense
    - or in the imperfect past
        - cataphoric reference

        - antecedent is a nominalisation of main clause
        - nominalise main clause as an infinitive clause
    - if main clause is in the imperative
        - anaphoric reference
        - antecedent is a discourse chunk
        - if there is a similar clause in segment
        - typically with the inverted polarity
        - antecedent is the same as for the previous anaphor
        - if there isn't a similar clause in segment
        - antecedent must be built out of segment moves
        - adjective in anaphor should be a clue

7. accept the candidate as antecedent

C.2.16 Copula-PP

global probability = 0.013

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtC=0.300	explicit=0.800	st=0.375
DK=0.300	implicit=0.100	sst=0.125
FtCCh=0.275	NR=0.075	fdv=0.100
CK=0.100	disc=0.025	dthel=0.100
LS=0.025		p_st=0.075
		p_sst=0.075
		dt=0.075
		thel=0.075

1. check collocation list
    - if no match found
        - go to instruction 2
    - if a match is found
        - follow resolution pathway in entry
2. select first appropriate candidate within search limit
3. check if it is a good fit
4. if it is, go to instruction 7
5. if it is not, go to instruction 6
6. check lexical signalling
    - attached probabilities
        - type of antecedent
            - explicit in all cases (one)
        - topical roles

        - p_st in all cases
    - recognition
        - anaphor is a question
        - previous move has a PP adjunct as well
    - resolution path
        - check if anaphor contrasts with PP in previous move
        - *de dia-à noite*
        - if yes, remove PP; remaining move is antecedent

7. check discourse knowledge

    - attached probabilities
        - type of antecedent
            - explicit = 0.750
            - implicit = 0.250
        - topical roles
            - st = 0.500
            - dthel = 0.167
            - p_st = 0.167
            - st = 0.083
            - p_sst = 0.083
    - recognition type one
        - anaphor is in nominal subordinate clause
        - typically a clausal direct object
        - typically introduced by *que* or *se*
    - resolution path type one
        - nominalise verb in previous move
        - if previous move is verbless or a short answer
        - retrieve first move with a verb
        - nominalise verb and make required adjustments
        - check if resulting NP is a good fit
        - if it is, accept it as antecedent
    - recognition type two
        - anaphor reiterates previous move as response
        - omitted subject
        - agreement flawed for first candidate
        - first candidate is one of the participants
    - resolution type two
        - implicit antecedent
        - analyse PP in anaphor semantically
        - *é pelo olho = mede pelo olho*
        - check first candidate adjusting agreement
        - if it is a good fit, accept it as antecedent

- recognition type three
  - first candidate strategy would work
  - it depends on the resolution of a preceding anaphor
  - preceding anaphor requires a discourse-chunk antecedent
  - current anaphor requires a NP antecedent
  - NP antecedent is part of discourse-chunk antecedent
- resolution path type three
  - check reconstructed previous move
  - select first candidate using anaphor lexical clues
  - if it is a good fit, accept it as antecedent
- recognition type four
  - missing clausal subject
  - reconstruction typically in inverted form
- resolution path type four
  - select previous move as candidate
  - make required adjustments
  - typically polarity and agreement
  - PPs may have to be replaced by anaphor PP
  - especially in cases of evident contrast
  - *de dia*/*à noite*
- recognition type five
  - anaphor fits uneasily as subject complement
  - PP does not link to copular verb as expected
- resolution path type five
  - bypass first candidate in previous move
  - select candidate in first copular structure
  - check if it is a good fit
  - in a subsegment, check if candidate is st
  - if both are true, accept it as antecedent
- recognition type six
  - first candidate fits uneasily as antecedent
  - previous move contains mental process verb
  - *entender*; *perceber*; etc.
  - *não é só na hora em que...aí...*
- resolution path type six
  - analyse previous move
  - substitute anaphor for direct object
  - add integrating conjunction if necessary
  - if it is a good fit, accept it as antecedent
- recognition type seven
  - PP contains current topic
  - presupposition is strongly salient

  - *é pelo INPS ?*
  - antecedent can only be *tratamento* in context
- resolution type seven
  - domain knowledge associated to discourse knowledge
  - token is part of a reported conversation
  - complex discourse processing required

8. accept the candidate as antecedent

C.2.17 Copula-Clause

global probability = 0.002

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
CK=0.400	NR=0.400	fdv=0.400
DK=0.400	implicit=0.400	p_dthel=0.400
FtC=0.200	explicit=0.200	thel=0.200

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. select first appropriate candidate within search limit
3. check if it is a good fit
4. if it is, go to instruction 7
5. if it is not, go to instruction 6
6. check discourse knowledge
   - attached probabilities
     - type of antecedent
       - implicit in all cases (two)
     - topical roles
       - p_dthel in all cases
   - recognition
     - previous move is often a question
     - typical anaphors: X-ser *porque* or X-ser *que*
   - resolution path
     - response to question is the antecedent
     - precise response has to be inferred from PP
     - reiterate the move in both affirmative and negative form
     - check which fits best

  - make required adjustments
  - if previous move is not a question
  - reiterate move and check if it is a good fit
  - if it is, accept it as antecedent
  - if it is not, break move into clauses
  - check each clausal constituent, particularly the last
  - check best fit until an accpetable antecedent is found

7. accept candidate as antecedent

C.2.18 Copula-Adv

global probability = 0.006

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
CK=0.429	explicit=0.500	fdv=0.429
FtCCh=0.286	disc.impl.=0.375	thel=0.286
FtC=0.214	implicit=0.071	st=0.143
DK=0.071	NR=0.071	dthel=0.143

1. check collocation list
   - if no match found
     - go to instruction 2
   - if a match is found
     - follow resolution pathway in entry
2. select first appropriate candidate within search limit
3. check if it is a good fit
4. if it is, go to instruction 7
5. if it is not, go to instruction 6
6. check discourse knowledge
   - explicit is all cases (one)
   - thel in all cases
   - recognition
     - simple bypass
     - intervening comment or short apposition
   - resolution path
     - bypass first candidate
     - check preceding move
     - select a candidate
     - typically the subject
     - check if it a good fit
     - if it is, accept it as antecedent
7. accept the candidate as antecedent

C.2.19 Copula-NUM

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtCCh=0.500	explicit=1.000	st=1.000
DK=0.500		

1. select first appropriate candidate within search limit
2. check if it is a good fit
3. if it is, accept it as antecedent
4. if it is not, go to instruction 5
5. check discourse knowledge
   - recognition as defined above
   - resolution path
     - search for identical or similar move in segment
     - select candidate in similar argument position
     - it is likely to be the segment topic
     - thus, move can be found by searching for st as well
     - check if it is a good fit
     - if it is, accept it as antecedent

C.2.20 Non-finite clause

global probability = 0.000

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=1.000	explicit=1.000	p_thel=1.000

1. analyse previous move
2. select supporting verb in VP
3. if anaphor is preceded by *só*
4. invert polarity of supporting verb
5. check if it is a good fit when added to anaphor
6. if it is, accept it as antecedent

C.2.21 That-clause

global probability = 0.001

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
FtC=0.500	explicit=1.000	dthel=1.000
VMm=0.500		

1. recognition
   - sequence of clauses introduced by *que*
   - main clause has to be retrieved for interpretation
2. resolution path
   - search initial clause introduced by *que*
   - if *que* is a relative pronoun
   - relative pronoun antecedent is the antecedent
   - if *que* is an integrating conjunction
   - preceding main clause is the antecedent
   - verbatim memory is required in second case

C.3 Nominals

C.3.1 Nonpronominal NP

global probability = 0.319

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
LR=0.477	explicit=0.748	sst=0.289
LS=0.161	implicit=0.245	thel=0.249
AM=0.126	disc.impl.=0.006	st=0.239
WK=0.104		dthel=0.091
SK=0.037		dt=0.091
CK=0.032		p_st=0.011
DK=0.031		p_sst=0.009
Dx=0.014		fdv=0.007
VMm=0.008		uthel=0.005
SetCr=0.007		sithel=0.003
FtCCh=0.001		p_dt=0.001
ScRf=0.001		p_thel=0.001
		p_dthel=0.001

1. check collocation list
   - if a match is found
     - follow resolution path in entry
   - if no match is found
     - go to instruction 2

2. check secondary reference
    - attached probabilities
        - type of antecedent
            - explicit in all cases (one)
        - topical roles
            - dthel in all cases
    - recognition
        - NP is *a gente*
        - separation from endophoric usage
        - check previous move
        - verb *dizer* in previous move
            - third person form with human subject
            - tense shift from past to present
        - first or second person ScRf pronoun in previous move
        - first or second person ScRf anaphoric verb form in previous move
        - change in prosodic pattern
    - resolution path
        - if there is a personal pronoun in chain
            - select the same antecedent
        - if there is an identifying vocative
            - select the entity as the antecedent
        - if there is no vocative nor chain
            - select first human candidate searching backwards
            - check lexical clues
3. check lexical repetition
    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - sst = 0.294
            - st = 0.253
            - thel = 0.199
            - dt = 0.130
            - dthel = 0.123
            - uthel = 0.002
    - recognition
        - any kind of NP
    - resolution path
        - search history list
        - if a precise match is found
            - select it as antecedent

  - if a partial match is found
    - go to instruction 3

4. check modified antecedent

- attached probabilities
  - type of antecedent
    - explicit = 0.877
    - implicit = 0.123
  - topical roles
    - st = 0.336
    - thel = 0.197
    - sst = 0.279
    - dthel = 0.090
    - dt = 0.098
- recognition
  - partial match with entity in history list
- resolution path
  - check the full history list
    - if there is a precise match preceding the partial match
      - accept it as the antecedent
    - if there is not a precise match preceding the partial match
  - if partial match involves the NP head
    - select partial match as antecedent
    - check lexical clues
    - if it is a good match
      - accept it as antecedent
      - typical cases: diminutives or
      - modifier is *mesmo(a)(s)* or *mesmo tipo de*
      - anaphor agrees in number with previous token
      - and has no modifiers except for determiners
    - if it is not a good match
      - classify antecedent as new and implicit
      - typical cases: anaphor is a plural NP or
      - anaphor is modified by a distinctive adjective or
      - anaphor contains a relative clause
      - relative clause modifies a NP head
      - NP head matches previous token
      - generic-specific contrast
  - if partial match is a modifier
    - if modifier is a proper noun, check usage
      - if partial match is an integral part of a full name
      - ex.: *PAM Marechal Rondon/Marechal Rondon*
      - accept it as antecedent

            - if it is not, classify anetecedent as new and implicit
        - if it isn't, check lexical clues in anaphor move

5. check lexical signalling

    - attached probabilities
        - type of antecedent
            - implicit = 0.583
            - explicit = 0.417
        - topical roles
            - st = 0.250
            - sst = 0.276
            - thel = 0.385
            - dt = 0.038
            - dthel = 0.045
            - p_sst = 0.006
    - recognition
        - definite description
        - no entry in history list
    - resolution path
        - check dictionary entries of entities
            - if a match is found
                - accept anaphor as an implicit antecedent
            - if no match is found
                - go to instruction 5

6. check world knowledge

    - attached probabilities
        - type of antecedent
            - implicit = 0.723
            - explicit = 0.277
        - topical roles
            - thel = 0.406
            - sst = 0.376
            - st = 0.109
            - uthel = 0.040
            - dthel = 0.030
            - dt = 0.020
            - sithel = 0.010
    - recognition
        - definite description
        - no match in history list
        - no match in dictionary entries of entities
    - resolution path

    - complex semantic processing
    - domain information
    - select an antecedent based on lexical clues
        * if no good fit is found
        * go to instruction 6

7. check shared knowledge

    - attached probabilities
        - type of antecedent
            * implicit = 0.778
            * explicit = 0.222
        - topical roles
            * st = 0.278
            * sst = 0.250
            * thel = 0.250
            * dthel = 0.139
            * dt = 0.083
    - recognition
        - definite description
        - no match in history list
        - no match in dictionary entries of entities
        - no useful world knowledge information available

8. resolution path

    - search entities in record of previous interactions
        - if no match is found
    - go to instruction 7

9. check deixis

    - attached probabilities
        - type of antecedent
            * implicit = 0.786
            * explicit = 0.214
        - topical roles
            * sst = 0.429
            * thel = 0.214
            * sithel = 0.143
            * st = 0.143
            * dt = 0.071
    - recognition
        - demonstrative a typical modifier
        - *aqui* + de + Art + anaphor
        - prosodic pattern

  - often pointing gesture
  - location where dialogue occurs
  - identifying entity in physical environment
- resolution path
  - visual input needed for identification

10. check discourse knowledge

- attached probabilities
  - type of antecedent
    - explicit = 0.633
    - implicit = 0.367
  - topical roles
    - thel = 0.167
    - sst = 0.400
    - st = 0.233
    - dt = 0.100
    - p_sst = 0.067
    - p_dt = 0.033
- recognition type one
  - X-ter (IP(such as *algum*) + *problema*
  - first candidate in previous move is the direct object of verb
- resolution path
  - if first candidate is a complex NP
  - typically an NP of the type N + *de* + N
    - select PP in NP as candidate
    - check if it is a good fit
    - if it is, accept it as antecedent
  - if first candidate is a simple NP (only N)
    - nominalise full verb phrase
    - check if it is a good fit
    - if it is, accept it as antecedent
- recognition type two
  - cataphoric Det + *negócio*
  - ex.: *é aquele negócio que eu falei...*
  - ex.: *não tinha esse negócio não*
  - ex.: *negócio de* + N
- resolution path type two
  - if it is a negative declarative clause
    - analyse next move
    - if the move is or contains a conditional clause
    - check subsequent move using path below
    - if move begins with *mas*
    - check if object direct of verb is a good fit

        - if it is, accept it as antecedent
    - if it is an affirmative declarative clause
        - check subsequent move
        - if it is a good fit, accept it as antecedent
    - it it is a *negócio de* + N
        - check association history of N
        - select any NPs of the N + N form
        - accept the best fit as antecedent
- recognition type three
    - verbless clause
    - either NP on its own or with adverbials
- resolution path type three
    - search backwards for first move with a verb
    - select move as candidate
    - if it is a good fit, accept it as antecedent
    - if it is not a good fit
        - analyse direct object
        - if it is a complex NP
        - check fit using NP head only as antecedent
        - make adjustments as required
        - adjustments may involve world knowledge
- recognition type four
    - NP is *de* + N
    - head is missing
    - often direct object of verb
    - resolution path type four
    - check st or sst
    - token may occur in a boundary move
    - use lexical information in N to check fit
- recognition type five
    - NP meaning is incomplete
    - a modifier must be retrieved from previous discourse
    - typically a *de* + N antecedent to link with an N
- resolution path type five
    - select st, sst, dt and dthel(s)
    - check fit as *de* + N modifiers of anaphor
    - make required adjustments
    - an inferred adjective modifier may be required
        - typically *ideal* or *adequado*
        - prosodic and discourse-marker clues possible
        - *né* with a special kind of intonation
        - lookahead within search limit may reveal adjective token
    - accept the best fit as antecedent

- recognition type six
  - token is a past participle in feminine gender (V-ada)
  - there is no token of V in history list
  - only acceptable in spoken language but very common
  - typically it appears in *X-dar uma NP*
- resolution type six
  - analyse previous discourse
  - check synonyms and antonyms in search of lexical signals
  - include variations such as *engordar=aumentar de peso*
  - the statement can also be taken face value
  - the phenomenon could be analysed as a word-formation rule
- recognition type seven
  - restricted interpretation of NPs
  - restriction is given by discourse context
  - typically NPs like *dia-a-dia*; *área*; *chance*
- resolution type seven
  - if NP is part of a complex NP
    - check information in NP head
    - ex.: *quantidades do dia-a-dia*
    - as *quantidades* refers to food, *dia-a-dia* of food
  - if NP is simple, i.e., Det + N
    - check for a solution within search limit
    - if anaphor is argument of a repeated verb
    - check if entity in same argument position is a good fit
    - domain knowledge combined with discourse context
    - *essa área=cardiologia*
    - complex comprehensive discourse processing may be required
    - very long distance retrieval may also be required
- recognition type eight
  - *o máximo* + LV + NUM
- resolution type eight
  - search for entity being measured
  - use lexical clues
  - *quanto* questions; *pouco*; *muito*
  - it is usually found within search limit
- recognition type nine
  - anaphoric nature of NP is hard to notice
  - indefinite description
  - it relates to the gist of a previous passage
  - anaphoric link may be unnecessary for processing
- resolution type nine
  - full complex discourse processing required

    - X *se dá muito com* Y
    - X *me disse* : *um amigo nosso*(reported speech)
  - recognition type nine
    - history list contains two senses of the same word
    - the one closest to the anaphor is incorrect
  - resolution path type ten
    - check both senses of the word
    - select the best fit as antecedent

11. check set creation

  - attached probabilities
    - type of antecedent
      - explicit = 0.571
      - implicit = 0.429
    - topical roles
      - thel = 0.429
      - st = 0.286
      - dthel = 0.286
  - recognition type one
    - generic NP
    - no determiner
  - resolution path type one
    - search for set members using NP head
    - domain knowledge may be crucial
    - create set as a new referent
  - recognition type two
    - complex NP with one new and one repeated constituent
    - partly lexical repetition
    - complex NP is hence referred to as one entity
      - *verduras-verduras e legumes*
      - *whatever1-whatever1 e whatever2*
  - resolution path type two
    - create a new referent defined by complex NP
    - record creation of new set which comprehends previous one
  - recognition type three
    - IP + plural NP like: *vários médicos*
    - or *a questão de* + Det + plural NP
    - head is in history list as distinct singular tokens
    - it may require domain knowledge: *raio-x-exames*
  - resolution path type three
    - create new referent defined by plural NP
    - record creation of set which comprehends previous tokens

- recognition type four
  - Det + NUM + *junto(as)*
  - typically NUM is *dois*
- resolution type four
  - tokens of entities grouped into set within search limit
  - tokens may be pronouns, especially if entities are humans

12. check verbatim memory

- attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p_st = 0.625
    - sst = 0.250
    - p_thel = 0.125
- recognition type one
  - NP is an addition to a previous move
  - move is verbless
  - *pedaço*; NUM + N;
  - repetition of NP in previous move
  - confirmation question
- resolution path type one
  - check previous move
  - if an identical NP is found
  - accept the whole move as antecedent
  - if no identical NP is found
  - check attachment to NPs in move
  - typically attached to nonhuman NP
- recognition type two
  - NP is response to question
  - full response ellipted
  - often but not always a verbless move
  - PPs with omitted prepositions
  - sequence of responses possible
- resolution path type two
  - select question minus wh-word as candidate
  - check lexical clues
  - if move is not verbless
    - select a NP antecedent
    - typically the quantified entity in a *quanto* question

C.3.2 Anaphoric adjectives

global probability = 0.013

Category Probabilities

Processing Strategy	Type of Antecedent	Topical Role
VMm=0.395	explicit=0.868	st=0.211
CK=0.184	NR=0.079	dt=0.132
FtCCh=0.158	implicit=0.053	sst=0.158
SetMb=0.079		p_dthel=0.132
LR=0.053		p_st=0.079
DK=0.053		p_dt=0.079
AM=0.053		fdv=0.079
Pl=0.053		dthel=0.079
LS=0.026		p_sst=0.026
		p_thel=0.026

1. check collocation list
   - if a match is found
     - follow resolution path in entry
   - if no match is found
     - go to instruction 2
2. check verbatim memory
   - attached probabilities
     - type of antecedent
       - explicit in all cases
     - topical roles
       - p_dthel = 0.333
       - sst = 0.200
       - p_st = 0.133
       - p_sst = 0.067
       - st = 0.067
       - p_dt = 0.067
       - p_thel = 0.067
       - dthel = 0.067
   - recognition type one
     - adjective takes clausal complementation
     - complement is omitted
     - typically a subject complement in copulas
     - typical anaphoric adjectives
     - *sure*; *certain*; *positive*
   - resolution path type one
     - search for token of adjective within search limit
     - if found, complement is antecedent
     - if complement is again omitted, continue search
     - if no other token is found within search limit
       - check previous move or turn
       - attach it as a that-clause to anaphor

        - if it is a good fit, accept it as antecedent
        - if previous move or turn is a response adjective
            - retrieve ellipted response
            - repeat operation above
    - recognition type two
        - adjective has an adverb function in previous move
        - anaphor is a repetition in a verbless utterance
        - response to question or question based on previous move
        - it may be an addition to previous move
        - the adverb form may be used in the previous move
        - partial repetition: *bem-bom*
    - resolution path type two
        - check previous move as antecedent
        - if it is a good fit, accept it as anecedent
    - recognition type three
        - adjective modifies an omitted NP head
        - this includes comparative and superlative forms
        - *o principal*; *a mesma*
    - resolution path
        - search for NP within search limit
        - check previous move or turn
        - select salient NP as antecedent (typically sst)
        - if it is a good fit, accept it as antecedent
        - if another adjective anaphor is found
            - go to instruction 2
        - if an identical adjective anaphor is found
            - go to instruction 3
        - if selected NP is modified by same adjective
            - go to instruction 4
        - if selected NP is modified by semantically related adjective
            - go to instruction 5
        - if selected NP is modified by a PP or unrelated adjective
            - typically an adjective like *o principal*
            - go to instruction 6

3. classify token as FtCCh strategy
    - accept NP head as antecedent
4. classify token as LR strategy
    - accept NP head as antecedent
5. classify token as AM strategy
    - accept NP head as antecedent

6. classify token as LS strategy
    - accept NP head as antecedent
    - classify antecedent as implicit
7. classify token as SetMb strategy
    - accet NP head as antecedent
    - classify antecedent as implicit

Appendix D

The collocation list for the Portuguese sample

D.1 Collocations with demonstratives

isso

- nonreferential
- previous move an information statement
- no confirmation value
- discourse marker signalling:
  - the hearer's attention and encouragement
  - the hearer's familiarity with the procedure described
  - the hearer's acknowledgment of normality
  - hearer's assumptions confirmed
- prosodic pattern is likely to be crucial
- differentiation from utterance confirmation *isso*

isso

- discourse-chunk antecedent
- previous move or turn
- adjustments often needed

isso

- discourse implicit; focusing device
- discourse marker
- confirmation or approval of a description
- a way of doing things
- difficult to pinpoint a definite antecedent

- degree of referentiality is enough to rule out NR
- some cases are close to definite reference
- distinction from explicit discourse-chunk antecedent blurred

isso

- confirmation or approval
- equivalent to *that's it*

isso isso

- discourse implicit; focusing device
- variation of the description approval entry

X-verb isso, X-verb aquilo

- identical verbs
- nonreferential; focusing device
- two tokens

por isso ou por aquilo

- nonreferential; focusing device

então, isso aí

- nonreferential; focusing device
- assertive unstressed
- the speaker confirms what has just been said

isso aí

- nonreferential; focusing device
- assertive stressed
- the speaker confirms emphatically what has just been said

isso aí (SUBJ) X-verb Compl

- discourse implicit; focusing device
- what was being talked about before
- it requires full discourse processing
- no identifiable antecedent

isso mesmo

- nonreferential; focusing device

- whatever has been said just before

é isso mesmo

- discourse-chunk antecedent
- previous move
- chain with Q-tag possible
- chain helps definiteness of reference

por isso que (SP) X-verb Compl

- discourse-chunk antecedent
- previous move or moves is antecedent
- check next move as well

por isso que (SP) X-verb Compl

- discourse implicit; focusing device
- uncomplemented X-verb
- interruption or broken syntax

pelo menos isso

- discourse-chunk antecedent
- previous move or moves is antecedent

não é isso (Q-tag)

- discourse-chunk antecedent
- previous move

não é isso (not a Q-tag)

- implicit built discourse-chunk antecedent
- constituents are explicit
- sequence of moves and turns
- boundary case between implicit and explicit

é isso (Q-tag)

- discourse-chunk antecedent
- previous move
- selection may be required
- X-dizer que-clause: antecedent is que-clause

é isso

- nonreferential; focusing device
- confirmation and sum-up

isso é que (SP) X-verb

- discourse implicit; focusing device
- complex discourse processing
- connecting and analysing of several turns
- explicit discourse-chunk antecedent: boundary

X-verb Obj, X-verb isso-Obj

- nonreferential; focusing device
- indeterminate antecedent within a set such as *comida*

X-verb isso-Obj

- discourse implicit; focusing device
- sum-up utterance
- X-verb semantic group *checar*
- including *ver* in the sense of *checar*

X-verb Obj nisso

- discourse implicit; focusing device
- sum-up utterance
- Obj semantic group *checada*

o que tiver (X-verb future subjunctive)

- nonreferential; focusing device
- no matter what; whatever;

sabe o que que é sabe qual é o problema

- nonreferential; focusing device
- discourse marker
- introductory phrase for an explanation

vê o que que é

- nonreferential; focusing device
- *see what the matter is*

o que (SP) X-verb (Adv+) X-be SUBJ

- SUBJ is antecedent
- or *a coisa*; *o problema*
- inverted copula with relative clause attached to anaphoric De
- there may be a preceding clue: *entre todas essas coisas*

(não) X-saber o que X-be

- explicit antecedent
- but a little vague
- *um problema de...*

é o que eu (OP) X-verb COL para OP or personal name or treatment expression

- X-verb in *falar* semantic group
- *dizer*; *explicar*;
- discourse-chunk antecedent
- next move is antecedent

se isso X-be o que X-verb

- nonreferential; focusing device
- expletive *o*

mais do que o que (Adv) X-verb

- nonreferential; focusing device
- expletive *o*

D.2 Collocations with indefinite pronouns

X-clause nem nada

- discourse-chunk antecedent
- antecedent is X-clause main verb and complements
- it fits well as infinitive form with auxiliaries out

D.3 Collocations with prepositional phrases

é (em função) de + Art (contraction) + NP

- analyse NP to determine semantics of *de*
- test possible solutions

D.4 Collocations with adverbs of place

sei lá

- nonreferential; focusing device
- a substitute for *não sei*

vá lá

- nonreferential; focusing device
- a substitute for 'vá em frente'
- if no location is found within search space
- if referential, not a COL

D.5 Collocations with adverbs of response

não (Q-tag)

- explicit discourse-chunk antecedent
- the utterance to which it is attached

uhum

- nonreferential; focusing device
- response to *tá* when *tá* is NR as well

D.6 Collocations with adverbs of manner

X-dizer assim

- cataphoric; discourse-chunk antecedent
- next move is antecedent
- speech reported verbatim follows
- a token of ScRf very likely in next move
- adjustments may be required

X-ser/estar assim

- both anaphoric and cataphoric tokens
- antecedent is previous or next move
- move may involve speech verbatim
- adjustments may be required

X-fazer assim

- cataphoric prosodic pattern

- discourse-chunk antecedent
- antecedent is next move or moves
- it may involve a long enumeration
- selections may be required
- if a move is not a good fit
  - break it into clauses
  - check clauses separately as candidates

X-verb assim

- cataphoric reference
- cataphoric prosodic pattern
- especially *ver*; *entender*
- next move is often a relative clause
- clause is attached to anaphor
- especially *que* as connector
- or next move is a non-finite clause

D.7 Collocations with linking verbs

né/não é (Q-tag)

- verb in clause is *ser*
  - confirmation
  - antecedent is preceding clause
  - adjustments may be required
  - if move is *acho que* + nominative clause
  - nominative clause is antecedent
- implicit verb *ser*
  - prosodic pattern
  - isolated NP before Q-tag
  - antecedent is reconstructed copular construction
  - *e eu corri lá para aquele hospital de lá, Miguel Couto, né*
- verb in clause is not *ser*
  - check if preceding clause fits confirmation pattern
  - semantics and prosodic pattern
  - *sempre tem probabilidade de sentir melhor, né ?*
  - if it does, antecedent is preceding clause

  - if it doesn't, collocation is nonreferential
  - information statement; no confirmation expected
  - *peixe raramente eu como, não é?*

tudo bem, né

- nonreferential; focusing device

é (response)

- verb in previous move is not *ser*
- reaction signal function; typically referential
- antecedent is previous move
- there are NR cases though
- check semantics and prosody of confirmation

é (Q-tag)

- not verb *ser* in preceding clause
- confirmation
- antecedent is previous clause
- if verb *ser* in previous clause
- same tense; not a COL
- different tense; confirmation as above

X-ser EVENT

- *acontecer* sense
- subject is postponed
- no subject complement
- nonreferential; focusing device

ah é ?

- response; request for confirmation
- antecedent is previous turn

como é (a coisa)?

- implicit; first candidate
- *ele tem dificuldade assim de andar ? como é (a locomoção dele)?*

pois é

- nonreferential; focusing device

(você) sabe o que que é

- cataphoric
- discourse-chunk antecedent
- next move

é que talvez seja

- nonreferential; focusing device

X-ser para ObjP X-verb-INF

- nonreferential; focusing device

aí é aquele negócio

- discourse implicit; focusing device
- vaguely cataphoric

será que

- nonreferential; focusing device
- tentative statement
- similar to perhaps

(es)tá (Q-tag)

- nonreferential; focusing device
- approval or request for approval
- verb in clause may be *estar*
- still a NR interpretation
- prosodic pattern

(es)tá (response)

- not verb *estar* in clause
- nonreferential; focusing device
- request for approval

tá que tá

- nonreferential; focusing device
- emphasis
- relies on intonation strongly

D.8 Collocations with anaphoric verbs

entendeu (Q-tag)

- explicit discourse-chunk antecedent
- preceding clause is antecedent

entendeu (Q-tag)

- nonreferential; focusing device
- preceding clause is an information statement
- there is nothing to be understood

entendeu (Q-tag)

- discourse implicit; focusing device
- connection of explanatory moves required
- distinction from a large chunk may be difficult

sei

- response to statement
- sense of *entendi*
- if no verb *saber* in previous move
  - check previous move
  - if move or constituent fits as object of *sei*
  - explicit discourse-chunk antecedent
  - if clue in move, like *e assim por diante*
  - discourse implicit; focusing device
  - if none of the above applies
  - nonreferential; focusing device
- sense of *conheço*
- explicit NP antecedent
- if verb *saber* in previous move
  - check prosody and semantics
  - if it fits as a lexical response, not a COL
  - if it does not, nonreferential
- if negative form of verb *saber* in previous move
  - nonreferential; focusing device
  - but, if overt stressed pronoun as subject, not a COL

acontece

- explicit discourse-chunk antecedent
- subsequent move is antecedent
- postposed subject connected by conjunction *que*

X-pegar

- nonreferential; focusing device
- meaning next action connected by *e*
- it is expletive

sabe (Q-tag)

- check semantic and prosody
- if confirmation pattern present
- explicit discourse-chunk antecedent
- clause to which Q-tag is attached

quer dizer

- apposition
- nonreferential; focusing device
- distinct from *quer dizer que*
- prosodic pattern and syntax

não X-dar

- nonreferential; focusing device
- sense of *impossible*; *doesn't work*
- no identifiable *impossible* fact or event

X-verb (active)

- nonreferential; focusing device
- indeterminate subject; passive voice function;
- especially *fazer*; *dizer*

X-ter que X-verb-INF

- nonreferential; focusing device
- passive function: *it must be* + X-verb-PastP

X-verb (medial) passivising particle (*se*)

- nonreferential; focusing device
- indeterminate subject

X-complicar

- discourse implicit; focusing device
- *a situação*; *o quadro*
- sum-up utterance

D.9 Collocations with copular constructions

D.9.1 Copula + FNP

não é isso (Q-tag)

- nonreferential; focusing device

X-ser o que eu expliquei para NP

- nonreferential; focusing device
- expletive *o*

mas não é dizer que...

- discourse-chunk antecedent
- explicit
- typically the previous move or turn

é uma droga

- discourse-chunk antecedent
- often implicit but identifiable
- built by linking elements in previous moves

é o seguinte

- discourse implicit; focusing device
- *o problema*; *a questão*; *o negócio*
- discourse marker; subsequent description

se for o caso

- nonreferential; focusing device
- usually removable without loss of meaning

D.9.2 Copula + Adj

Subj X-be (Adv) Adj (Adv) (Prep) Subj X-INF Compl

- discourse-chunk antecedent
- explicit
- antecedent is Subj X-INF Compl

é lógico

- discourse implicit; focusing device
- discourse marker
- approval-continuity

não é possível

- apposition (*no way*)
- nonreferential; focusing device
- referential tokens possible

'tá oquei

- nonreferential; focusing device
- discourse marker
- agreement

'tá certo

- nonreferential; focusing device
- discourse marker
- agreement

'tá legal

- nonreferential; focusing device
- discourse marker
- agreement

'tá bom (Q-tag or not)

- nonreferential; focusing device
- discourse marker
- request of agreement sign
- if combined with a PP, not a COL
- *pra mim 'tá bom*

'tá ótimo

- nonreferential; focusing device
- discourse marker
- emphatic agreement

D.9.3 Copula + PP

(não) X-ser para X-INF

- nonreferential; focusing device
- meaning *one must (not)* X-INF
- prosodic pattern

X-ser (often conditional) para X-INF

- discourse implicit; focusing device
- presupposition
- antecedent is the general aim of the conversation
- it is previously known by participants
- it is unspecific

D.9.4 Copula + Clause

é que X-clause

- nonreferential; focusing device
- expletive

(não) é para SUBJ X-INF

- nonreferential; focusing device
- meaning SUBJ *must (not)* X-INF

D.9.5 Copula + Adv

é (X-ser) assim

- no further constituents in predicative
- discourse implicit; focusing device
- *a coisa*; *as coisas*
- *that's the way it is* or *it's like that*
- there is no need to identify a specific referent
- it would often be a fairly long discourse chunk
- many parts would be superfluous
- not worth the processing

'tá (tudo) bem

- nonreferential; focusing device
- *it's all right*

D.10 Collocations with anaphoric noun phrases

DET problema

- explicit NP
- search *problema*; *questão*; etc. within search limit
- if a good fit is found, not a COL
- if not, discourse implicit; focusing device
- antecedent is hard to define in one entity
- not necessarily a recent topic
- distance is likely to add to vagueness

(dar) (ter) vontade

- no complement
- explicit discourse-chunk antecedent
- typically an infinitive clause
- resolution path
  - select previous move or turn
  - change it into an infinitive clause
  - check if it is a good fit
  - if it is, accept it as antecedent
  - if it is not, go further back to next move
  - if no good fit is found within search limit (unlikely)
  - check previous move containing st, dt, dthel(s)
- *feel like*

o seguinte/a seguinte coisa

- cataphoric
- explicit discourse-chunk antecedent
- immediately subsequent move
- no clear equivalent in English
- discourse marker to draw the hearer's attention

essa coisa toda

- discourse implicit; focusing device
- it sums up a stretch of discourse of variable size
- it may be closer to nonreferential or closer to implicit

- some cases have an acceptable explicit interpretation
- discourse chunk
- a possible antecedent would combine implicit and explicit elements

NP (pause) essa coisa toda

- antecedent is a set
- it may or may not have been introduced
- resolution path
  - select preceding NP as antecedent
  - check if set that contains it is a better fit

NP (pause) outras coisas

- similar to entry above

NP (Art + N + Adj) (pause) uma coisa Adj

- antecedent is N
- anaphor is a reinforcemen

as coisas + relative clause

- equivalent to: *o que eu posso comer*
- the relative clause describes the antecedent
- in this case, *alimentos*
- check history list and association list for explicit

grandes coisas

- particularly good results of Y
- Y is the antecedent
- explicit NP
- description in previous moves should provide complement

a tendência X-ser

- missing complement is antecedent
- antecedent is explicit and probably local topic
- check st and sst

esse tipo de coisa assim

- both anaphoric and cataphoric tokens

- implicit antecedent (*como* ANT) if cataphoric
- thematic element
- it may appear within a description
- difficult to identify
- resolution path
  - check verb in anaphor move
  - if it appears in a move within search limit
  - check NPs in move
  - if there is a good fit, accept it as antecedent

essas coisas

- discourse implicit; focusing device
- preceding description
- usually limited by a discourse unit (subsegment or segment)
- general feeling or attitude described
- *things like that*

NP essas coisas

- implicit antecedent
- coisas como NP

entre todas essas coisas o que mais...

- set of antecedents
- generalise on the one given in the move

certas coisas

- explicit set of antecedents
- anaphor is direct object of verb
- resolution path
  - check association history of verb
  - antecedent should be one of the associated entities
  - if not, check association history of nominalised verb
- antecedent is usually near
- there may be several intervening turns though

uma coisa Adj

- implicit antecedent
- clues in previous moves
- likely to be a description
- antecedent sums up or is the process or fact described

(essa) história (toda) (Adj)

- explicit antecedent
- often segment topic
- it may be discourse implicit as well
- resolution path
  - if there is an Adj, nominalise the adjective
  - check if it fits as an antecedent
  - if it does, accept it as antecedent
- if there is no adjective, discourse implicit

esse meio caminho/no meio do caminho

- discourse implicit; focusing device
- antecedent is a situation described in discourse
- it usually involves distinct outcomes

o resto

- previous move contains a NP like *não-sei-que* PN
- implicit antecedent
- *o resto do nome de* + definite description
- definite description derives from PN

D.11 Collocations with anaphoric adjectives

o ideal

- explicit NP antecedent
- omitted head of NP
- strong topicality effect
- resolution path
  - select st; sst;
  - check lexical clues
- *ideally*

o máximo

- explicit NP antecedent
- *o número máximo* de X
- X is typically in previous move
- resolution path
  - check NPs in previous move
  - accept the best fit
- *as many as*

ótimo

- nonreferential; focusing device
- discourse marker
- agreement signal
- *fine*