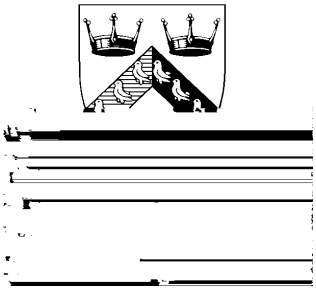
An

UNIVERSITY OF



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

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**

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

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

    - thel in all cases

- recognition
  - phrase *so 'sso*
  - 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 entrr. e mns

theory, as they were reserved for testing purposes from the outset of the project. The testing of

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

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 , *s* or *t* *y*
      - go to instruction 5

   (b) pronoun is *t*
      - 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*; *as* ; *answ r*; *xp a n*
    - subject a third person pronou[(;)-6.93181]TJ/R31 10.k5rt.3778.7821(r)-5.52048]TJ/R19 10.95 Tf3

  - 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 or *s*
  - go to instruction 8(g)
- if pronoun is *t'* or *t y*

(a) shift *t'-t y*

  - recognition
    - first-candidate chain *t'-t y*
    - 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 *'t*
- 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 *t y*
    - go to instruction 8(g)
  - pronoun is *'t*
    - 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
  - *'t* interchangeable with *t 's t at*
  - frequent collocations
    - *'t was b caus*
    - *'t* + Copula + Adjective as in:
    - *'t was n c ssary*
- resolution path
  - check previous move or turn
  - check clausal constituents within
  - begin with that-clauses and non-finites

(f) altered reiteration
- recognition
  -

    - 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 *'t*
      - 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 *'t*
    - select dt and thel(s) in resumptive segment
    - check appropriateness and lexical clues

(i) sequence of questions

(j) sum-up utterances

- recognition
  - *’t* 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
  - *t prob* ; *t att r*;
  - *t qu st’on*; *t s’tuat’on*;
  - *t t’n* ; *t stu* ; *t ’ssu*

(k) set/set member oscillation

- recognition
  - chain of *’t* tokens (two or more)
  - collocation signalling reference to set (unstable)
    - *you now ow t way ’t ’s*
  - 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 *t y*
  - lexical clue in utterance
    - the word *s’ u tan ous* in tone unit
- resolution path
  -

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 *y* , *r* or *t*
     - go to instruction 4
   - if pronoun is *y t*
     - 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

  - 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
  -

- a NP of vague generic meaning
- *t prob ' ; t att r;*
- *t qu st' on; t s' tuat' on;*
- *t t' n ; t stu ; t ' ssu*

7. check deixis

- recognition type one
  - pronouns ' , *r* or *t*
  - 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 secm BT/R19 10.95 Tf238.11026(l)-6.934311(s)-268.535(a)5.64311(n)-4.1

–

        - 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
        - to27.26 Tm(fl)Tj/R19 1(o27.26 Tm(fl)Tj/R19 1(o27.26 Tm(fl)Tj/R19 1(o27.26 Tm(fl)Tj7(d)-4.10914(t

- select previous move
- make standard adjustments to check good fit
- *t act t at X*; *w at 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 *t at*
  - 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
  - *t ' s X b w r*
  - 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
  - *do s t at an*
  - *ow do s t at str' you*
  - *t ' s ' s b caus*
  - *ub t at t ' s 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 *t ' s*
        - check cataphoric reference as well (next move)
- if a good fit is found, accept it
- if it is not, check discourse implicit
    -

- topical roles
  - st = 0.500 sst = 0.091
  - dt =

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*; *as* ; *answ r*; *xp a n*
    - subject a third person pronoun or personal name
    - tense switch in relation to anaphor move
  - identifying vocative in the utterance
- resolution path
  -

- go to instruction 3 type two

3. check discourse knowledge

- attached probabilities
  - type of antecedent
    -

  - 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

-

- resolution path
  - visual input needed for identification

2. select first appropriate candidate

- if it is a NP modified by *on y*
  - 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 *t r X b* 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
  -

  - anaphor is *t ot r on 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 *anot r on*
- resolution path type seven
  - select head of the first NP candidate
  - check lexical clues

  - 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., *ront rst*
      - if it is a cardinal postdeterminer
      - full NP is antecedent
      - *o t* 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 *o t* 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.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

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 *o* 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 *t' n so*; *suppos so*;
       - collocations *say so*; *t 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 *v n or 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
    *

- 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 *J t a J I s*

  - compare move with stretch up to discourse unit boundary
  - if a subsegment, extend comparison to segment boundary
  - test a *t' s' nd o t' n* referent
  - complex semantic processing required
- recognition type three
  - phrase is X-do + demonstrative
  - *adv' s* 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

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 *do s*
   - 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 *d' d*
   - 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 *t ' s ' nd o t ' n* type of antecedent
5. if it is a good fit
   - accept it as antecedent
   - classify antecedent as **dim**

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 now*
  - 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 now*
      - 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: *t' n* ; *t* ; *a r* ;
  - special case: *o*
    - select object of previous token of *o*
- 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

  - 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

    -

- 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
    - *p op* ; *suc v ntur* ;
- 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 + *'nds*
- resolution path type three
  - cataphoric reference
  - antecedents are enumerated after anaphor
- recognition type four
  - *sort s o or an'sat'ons*;
  - *'n t at conn x'on*;
  - *ro t ADJ po'nt o v' w*

- 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

- attached probabilities
  - type of antecedent
    *

  - 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
  - *sur ; c rta n; pos t v*
- 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
- *t act t at* NP
- resolution path
  - check segment topic
  - check predicate of segment topic
  - repeat with subsegment topic

it's not because 0 308 1 T7T7T7

- Q-tag

- resolution path

-

- 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 *t act t at*
  - 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
  - *t r* 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 *ow on*
- nonfinite clause complement

it X-be (not) a question of

- nonreferential

- 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 *t act t at*

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

- cataphoric
- antecedent = NF-clause

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

- object pronoun is often
- 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 b s nc r*
- nonreferential

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

- 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 *say'n op n y*
- cataphoric
- subsequent that-clause

X-find + it + (modif) + Adj

- sense of *X t 'n 't's*
- 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 *t act t at*

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 *w at v r t r 's 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: *w at*; *t ast t' n* ; *t ot r t' n* ;
- the embedded clause may be omitted

there X-be something in this

- anaphoric; discourse-chunk antecedent;
- previous move: Subj X-tell that-clause
- antecedent = *t act t at* + 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 = *t s t*

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

- anaphoric
- antecedent = discourse chunk; previous move;

this X-be right/wrong

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

this X-be all right

- anaphoric
- plain NP; first candidate;

- special case: narratives
  - *t at s about a past two*
  - the meaning is:

that X-be the reason for X

- anaphoric
- discourse implicit;
- previous subsegment(s)
- agreement may be imperfect
- more than one *r ason* 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
- *wron or r' t* 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
- *w n I say C L I an AN*

this sort of thing

- anaphoric
- discourse-chunk antecedent

- demonstrative anaphor in previous move
- it refers to the antecedent
- t act t at
- predicate of segment topic
- it sums up a description

the sort of thing

- anaphoric
- segment topic
- NP explicit antecedent
- 's t 's t sort o t 'n

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 't as a guide
- check OP or SP entry as appropriate
- discourse-chunk antecedents possible
- implicit antecedents possible

the thing

- pronoun-like resolution
- use 't as a guide
- check OP or SP entry as appropriate

this fact

- anaphoric
- discourse chunk related to segment topic
- previous exchange nominalised as *t act t at*
- 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 *sta* 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 t at*

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 d scr' b d abov*
- 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 an*

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

        - subject a third person pronoun or personal name
        - human-denoting NPs also possible: *a co a*
      - 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 *s*
  - presence of location adjunct in turn
  - *a* + 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 u nt nd r d' ss u nao 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

- type of antecedent
  - explicit in all cases
- topical roles
  -

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
        - *p' orar s o a' s do qu o qu u stou*
    - resolution path
        - nominalise verb in the modifier relative clause
        - check lexical clues for good fit
        - antecedents like *stado* out of *X star*;
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 *aqu'o* 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 *'sso* and *aqu'o*
   - 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)

•

- attached probabilities
  - type of antecedent
    - explicit = 0.889
    - implicit = 0.111
  - topical roles
    - sst = 0.444
    - st = 0.333
    - thel = 0.222
- recognition
  - pronouns

        - dthel = 0.167
    - recognition
        - prosodic pattern
        - identifiable entity in physical environment
        - often followed by *aquí* or *a*
    - 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 *aqu , aqu a, aquí o* 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 *d'z r, a ar, xp'car*
    - 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	h i5.64311(a)5.64534(l)-248.02(R)9.46686(o)-4.111373181(s)-246.624	

4. accepcs a t as the antecedent

ee , **C.1.5 Independri**

Caegory Probabilities

Procesning Strategy	Type of Antecedent	Topical Role
1(n)-81(o)-4.109661(p)-4.10914(r)4.2)n1(a(o)-4.11137.1598	FtCCh=1.20648s	

1. check lexical repetition

    - attached probabilities
        - type of antecedent
            - explicit in all cases
        - topical roles
            - dt = 0.167
            -

- 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

- search backwards
-

  - 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

.213(m)10.8307(o)-4.11021(S)87.49307(n)-4.11026(g)-2232048(t)-6.9307(i)-6.9289(r)-4.1451 Tf3.(e)5.9318.7207.36 Tc(7)-4.10914(d)-4

  - sype of antecedent

(e)5.64422(c)5.64.11(r)4.2.1-.967(i)-4.10914(.)8.90389(0)-4.11137(2)-4.10914(0)-4.10914(0)-3097 0.24 Tf1 0 0 -1 111.12 682.54 4.060)Tj/R1

  - topical roles

  - pronouns *outro a s ; a u 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 *n n u a ; u'to 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 ; n n u a*
  - no syntactic parallel in previous move
  - especially if syntax is broken in previous move
- resolution path
  -
  -

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
     - u v’a’ scr’to
     - 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 *andar*
            - search for a previous token of *andar*
        - 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 *vo tar*
    - resolution path type two
        - check associ4.23177(o)-4.10914(r)-236.875(o)-4.11137(c)5.64Tf-10.8m6.62 Tm(fl)Tj/R19 10.95

- 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 *aqu'* ; *a* ; *a'* ; *ca*
  - contractions with preposition *d*
  - prosodic pattern
  - identifiable entity in physical environment
  - chains are frequent
- resolution path
  - antecedent usually an object and not a location
    - however, *aqu'* 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 verbast377(.)-495.293181(o)-411137]TJ((–)-4.10914]TJ/R1538.38 Tm()T20914(a)5.68 0 Td[(h20914(a)5.

– needed for identification

- 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

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. *a' nda*; *a*; *o o*
2. adverbial phrases were included: *todo d' a*
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 *ass'*
- 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
  - *a aqu n oc'o qu u a'*
- 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 *a t racao*; *udanca*;
  - NP is *s'tuacao* 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 *a's*
- 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: *so*
- discourse-chunk antecedent

  1. select previous move or turn as antecedent
  2. make required adjustments

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

3. check verbatim memory

- recogntion
  - any wh-word requiring complement
  - typical cases: *sab porqu* ↯

- 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

- (a) attached probabilities
  - type of antecedent
    - explicit in all cases
  - topical roles
    - p

  - 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 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
  -

    * 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 anteced

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 *, n* or *s ta*
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 *n*
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, *s u no*
  - 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
  - *v o qu qu*
  - 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
  -

  - 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
  - *ra o qu* question
  - typically followed by *ra* + 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

  - 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 associationf e–

    - 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 *pr c' sar*
  - move introduced by *as* or *a*
- 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
  - *orcar*; *obr' ar*;
  - or direct object is missing
  - verb is *qu r r*; *pod r*; etc.

  - it is too specific to be discourse implicit
  - *a causa*; *o ot' vo*;
- 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 (*as*)
  - lexical clues: candidates witliutdrkh biopdm
    - raeoothn
    - lov
    - sooootis ms (
    - shseial
      - mi[illegible] e2380914(u)-4.10914(a)5.64311(s)-5.52271(:)-2913181(i)-6.93181(o)-4.109

– select antecedents as in first-candidate strategies

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 no d* + 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
  - *nao pr r nc'a*
- resolution path type three
  - discourse implicit antecedent
  - *o prob a; a qu stao*
- 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
  - *u nao s's* + anaphor
  - there may be a sequence of tokens
  - *s* + 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
  -

•

    - ∗ 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 *nao*
    - – 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
    - – *s nao or uʼtodʼ cʼ*
    - – 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

*

- 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
  -

- make required adjustments
- if previous move is not a question
-

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 *so*
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

2. check secondary reference
    - attached probabilities
        - type of antecedent
            - explicit in all cases (one)
        - topical roles
            - dthel in all cases
    - recognition
        - NP is *a nt*
        - separation from endophoric usage
        - check previous move
        - verb *d' z r* 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 *s o a s* or *s o t’po d*
      - 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.: *AM Mar c a ondon Mar c a ondon*
      - 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 lexica5.213(a)56 21

– complex semantic processing
– domain information
– select an antecedent based on lexical clues

*

  - 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

- 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 u a N*
- resolution type six
  - analyse previous discourse
  - check synonyms and antonyms in search of lexical signals
  - include variations such as *n ordar au ntar d p so*
  - 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 *d'a a d'a*; *ar a*; *c anc*
- resolution type seven
  - if NP is part of a complex NP
    - check information in NP head
    - ex.: *quant'dad s do d'a a d'a*
    - as *quant'dad s* refers to food, *d'a a d'a* 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
    - *ssa ar a=card'o o 'a*
    - complex comprehensive discourse processing may be required
    - very long distance retrieval may also be required
- recognition type eight
  - *o ax' o* + LV + NUM
- resolution type eight
  - search for entity being measured
  - use lexical clues
  - *quanto* questions; *pouco*; *u'to*
  - 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 *s da u’to co* Y
– X *d’ss* : *u a’ o nosso*(reported speech)

- recognition type four
  - Det + NUM + *unto as*
  - typically NUM is *do's*
- 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
  - *p daco*

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

*

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 *'sso*

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 *t at s't*

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
-

- 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 ´

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

- SUBJ is antecedent
- or *a co'sa*; *o prob a*
- inverted copula with relative clause attached to anaphoric De
- there may be a preceding clue: *ntr todas ssas co'sas*

(não) X-saber o que X-be

- explicit antecedent
- but a little vague
- *u prob a d*

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

- X-verb in *a ar*

D.4 Collocations with adverbs of place

sei lá

- nonreferential; focusing device
- a substitute for *nao s'*

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 *ta* when *ta* is NR as well

D.6 Collocations with adverbs of

- 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 *v r*; *nt nd r*
- next move is often a relative clause
- clause is attached to anaphor
- especially *qu* 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 *s r*
  - confirmation
  - antecedent is preceding clause
  - adjustments may be required
  - if move is *ac o qu* + nominative clause
  - nominative clause is antecedent
- implicit verb *s r*
  - prosodic pattern
  -

- if it doesn't, collocation is nonreferential
- information statement; no confirmation expected
- p x rara nt u co o nao

tudo bem, né

- nonreferential; focusing device

é (response)

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

X-ser EVENT

- *acont c r* 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
- t d cu dad ass d andar co o a oco ocao d

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

ser´

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 *nt nd'*
- if no verb *sab r* in previous move
  - check previous move
  - if move or constituent fits as object of *s'*
  - explicit discourse-chunk antecedent
  - if clue in move, like *ass' por d'ant*
  - discourse implicit; focusing device
  - if none of the above applies
  - nonreferential; focusing device
- sense of *con co*
- explicit NP antecedent
- if verb *sab r* 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 *sab r* 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 *qu*

X-pegar

- nonreferential; focusing device
- meaning next action connected by
- 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 *qu r d'z r qu*
- prosodic pattern and syntax

não X-dar

- nonreferential; focusing device
- sense of *' poss'b* ; *do sn t wor*
- no identifiable *' poss'b* fact or event

X-verb (active)

- nonreferential; focusing device
- indeterminate subject; passive voice function;
- especially *az r*; *d'z r*

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' ta bo*

'tá ótimo

- nonreferential; focusing device
-

D.9.3 Copula + PP
(n˜

D.10 Collocations with anaphoric noun phrases

DET problema

- explicit NP
- search *prob a*; *qu stao*; 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 74v.8011(v)17.8011(e)-257.137elife 74v.801; fo4.10914(e)5.643C12i7(e)Td[(;)

- 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 qu u posso co r*
- the relative clause describes the antecedent
- in this case,,elto97 0.24 Tf81(n)-245.213(t..6)-6.93181.2latn4.124.8(9911(e)-23[(e)5.64422(q)-4.109[-25.4402 Td

- 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 ngmipelisetheedvaci
  - ifitdis as31246.624(a)5.64422(n)-a10914(e)5.64422(n)-4.10914(t)-6.93181(a)5.64311(c)5.10914(e)5.64311

itbinncinncaim

o máximo

- explicit NP antecedent
- *o nu ro ax' o* de X
- X is typically in previous move
- resolution path
  - check NPs in previous move
  - accept the best fit
- *as any as*

ótimo

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