III. Le problème des ambiguïtés pour l'analyse de donnés textuelles

1. Définitions

Interprétation

L'**interprétation** consiste à attribuer un sens à un énoncé.
Ce méchanisme fait intervenir plusieurs paramêtres:

- des connaissances linguistiques (i.e. les méchanismes qui régissent le système de la langue utilisée - cf II. Méthodes et connaissances de la langue pour l'analyse de données textuelles)
- la forme de l'énoncé (i.e. la séquence de graphie ou de sons)
- le contexte* d'énonciation (*Qu'est-ce qu'il fait chaud!* peut signifier le contraire de son sens propre si les interlocuteurs se trouvent dans un environnemement pariculièrement froid)
- des connaissances du monde

  Exemple:
  Mary hit the vase against the sink, it broke.
  Mary hit her bag against the vase, it broke.
  connaissance du monde: le vase est plus fragile que l'évier et que le sac.

  protons et éléctrons chargés négativement
  → * negatively charged protons and electrons
  → protons and negatively charged electrons
  connaissance du monde: les électrons sont chargés négativement

- de plus, l'usage répété de certains énoncé au sein d'une communauté linguistique donnée (région, entreprise, ...) entraine la création d'**interprétations spécifiques à cette communauté linguistique donnée**. De ce fait, l'interprétation est également dépendante de la communauté linguistique des personnes qui communiquent.
- enfin, l'expérience personnelle de chaque individu (ses priorités , ses préoccupation, ses désirs, ...) intervient dans son interprétation des énoncés qu'il perçoit. C'est la difficulté liée à l'**interprétation**: sa **subjectivité**.
  On est confronté à cette difficulté, dès que l'on essaye d'évaluer un système d'analyse qui doit reproduire un mécanisme d'interprétation (cf étiquetage sémantique, recherche d'information, ...)

* le terme **contexte** signifie "ce qui entoure un élément". Il peut désigner suivant les cas, le contexte temporelle, géographique, culturel, social (*comme c'est le cas ici*) ou un contexte linguistique comme par exemple le contexte syntaxique (les catégories syntaxiques qui entourent un élément) ou le contexte lexical (les mots qui entourent un élément - dans ce cas, parfois appelé *co-texte*.)

Ambiguïté

L'**ambiguité** est la propriété de certains énoncés qui présentent plusieurs sens. (i.e. à une même forme, on peut associer plusieurs interprétations). Cette propriété est due à l'**absence de certaines informations** qui permettent normalement de **désambiguiser.**

2. Types d'ambiguités

⟹ On peut distinguer différents types d'ambiguité en fonction du niveau d'analyse linguistique où l'ambiguité est observée:

- /lə potoʀoz/, /ʀɑ̃dʀlapaʀɛj/, /setœ̃dezastʀ/, /lasiʀi/

- Exemple: *le, bois*
  au **niveau morphologique:**
  le → det ou pron
  bois → N ou V

  Exemple: *le bois*
  au **niveau syntaxique:**
  → det N ou pron V
  mais det V et pron N impossibles

  Exemple: *le bois était sec*
  en contexte: non ambigu

Exemples d'énoncés ambigus

Ambiguités syntaxique

Ambiguïté de rattachement

- rattachement des adjectifs

  1. a. "A céder: [$_{GN}$ commerce [$_{GP}$ de dame [$_{GAdj}$ facile]]]" ***(petite annonce)***
  b. "A céder: [$_{GN}$ commerce [$_{GP}$ de dame] [$_{GAdj}$ facile]]"

  2. a. J'ai rencontré [$_{GN}$ un professeur [de football] américain]
  b. J'ai rencontré [$_{GN}$ un professeur [de football américain]]

  3. J'ai rencontré [$_{GN}$ un professeur [$_{GP}$ de judo] [$_{GAdj}$ italien]]

  4. J'ai rencontré [$_{GN}$ un professeur [$_{GP}$ de droit international]]

  5. a. Les [$_{GN}$ tartes [$_{GP}$ aux amandes [$_{GAdj}$ fraîches]]].
  b. Les [$_{GN}$ tartes [$_{GP}$ aux amandes] [$_{GAdj}$ fraîches]].

- rattachement prépositionnel

  Important : Amb syntaxique toujours par rapport à une grammaire donnée (éventuellement implicite)

  6. a. Paul emmène [$_{GN}$ les clefs [$_{GP}$ de la maison]] [$_{GP}$ au garage].

  b. Paul emmène [$_{GN}$ les clefs] [$_{GP}$ de la maison] [$_{GP}$ au garage].

  7. a. J'ai reçu [$_{NP}$ un vase [$_{PP}$ de Chine $_{PP}$] $_{NP}$].
  b. J'ai reçu [$_{NP}$ un vase $_{NP}$] [$_{PP}$ de Chine $_{PP}$].

8. a. J'ai reçu [$_{NP}$ un bonnet [$_{PP}$ de coton $_{PP}$] $_{NP}$].
   b. * J'ai reçu [$_{NP}$ un bonnet $_{NP}$] [$_{PP}$ de coton $_{PP}$].

9. Jean regarde l'homme avec le télescope.

- rattachement d'adverviaux

10. "vends tricycle pour infirme en bon état" ***(petite annonce)***

11. Deux conducteurs étaient interpellés par les gendarmes en état d'ivresse. (*Var Matin*, 13/07/1994)

- GN sujet / objet

12. J'ignore quels ennemis redoutaient les soldats.
cf. J'ignore quel danger redoutaient les soldats.
J'ignore quelle population redoutait les soldats.

Ambiguïté de catégorie syntaxique

13. Il répugne à accepter ce savant compromis
→ savant: N, compromis: Adj
→ savant: Adj, compromis: N

14. la belle ferme le voile.

15. Oser le savoir

16. la bonne cuisine avec des navets.

Ambiguité de prédication

Les noms prédicatifs

17. la crainte de l'ennemi
→ X craint l'ennemi
→ l'ennemi craint X

Comparer avec:

18. J'ai lu la critique de Chomsky
→ X a critiqué Chomsky
→ Chomsky a critiqué X

19. la production de lait

Constructions elliptiques

20. Il a souvent vu manger des poulets *(infinitives)*
→ Il a souvent vu des poulets manger *qqch*: *poulet* = arg1 de manger
→ Il a souvent vu *des gens* manger des poulets: *poulet* = arg2 de manger

21. Il m'a reçu comme un prince
→ Il m'a reçu comme un prince *reçoit*
→ Il m'a reçu comme *si j'étais* un prince.

22. Elle aime son fils plus que son mari
→ elle aime son fils plus que son mari *aime son fils*
→ elle aime son fils plus qu'*elle aime* son mari

23. J'ai fait porter les fleurs à Marie
→ J'ai fait que X porte des fleurs à Marie.
→ J'ai fait que Marie porte les fleurs

Ambiguité lexicale (morphologique)

homonymie / polysémie

- au niveau de la forme graphique ET phonique

  24. le vol de l'oiseau
  → l'oiseau a volé qqch
  → l'oiseau vole dans les airs

  25. les jumelles grossissent
  → les deux soeurs prennent du poids
  → les lunettes augementent la puissance de vision

  26. Comme il faisait très chaud le jour de l'enterrement, on sortit la bière.
  → on sortit le cercueil
  → on sortit les boissons

  27. "Il y a quinze jour que je suis au lit avec le Dr Machin et il ne m'a encore rien fait" *(lettre à la sécurité sociale)*

  28. Les kines se sont masses contre les grilles de la préfecture.

- au niveau de la forme graphique uniquement (la forme phonique lève l'ambiguïté)

  29. portion, couvent, ...

  30. Elle a perdu ses fils.

  NB: Ce dernier type d'ambigïté est en fait rare en français:

Précédence des opérateurs

Portée des quantifieurs

31. les régiments défilent drapeau en tête
Si l'on considère que chaque régiment a un drapeau
→ les *n* drapeaux des *n* régiments sont portés en groupe devant l'ensemble des régiments,
→ chaque drapeau est porté en tête de son régiment

32. Tous les hommes aiment une femme
→ Il existe une femme que tous les hommes aiment
→ Pour chaque homme il existe une femme telle que cet homme aime cette femme

Interaction avec la négation

33. Toutes les victimes n'avaient pas été vaccinées
→ Certaines victimes n'avaient pas été vaccinées
→ Aucune victime n'avait été vaccinée.

34. Elle ne pleure pas parce qu'il est parti
→ Il est faut que (il partir implique elle pleurer) i.e. Ce n'est pas parce qu'il est parti qu'elle pleure, mais elle pleure pour une autre raison
→ Il partir implique (elle ne pas pleurer)

ambiguïté référentielle

pb de coréférence

Anaphores

Il y a anaphore lorsqu'**une unité ne désigne pas un référent par elle-même, mais seulement par le biais d'un élément antérieur**, qui est donc, en termes traditionnels, son "antécédent". Dans le cas prototypique de l'anaphore (avec le pronom *il* par exemple), on observe une relation de coréférence entre l'expression anaphorique et son antécedent (i.e. ils désignent le même référent):

ex. : *Pierre est content : **il** a été reçu à son examen.*

Cependant, **une anaphore n'est pas nécessairement coréférentielle**. Le référent d'une expression anaphorique n'est pas toujours dénoté explicitement par un terme antérieur, mais peut s'appuyer sur son *contexte*. On parle alors d'anaphore indirecte ou *in abstentia*. C'est le cas de l'*anaphore associative*, de l'*anaphore lexicale* et de l'*anaphore collective*:

L'*anaphore associative* un phénomène pragmatique, qui fait qu'un élément peut très bien se trouver en relation anaphorique avec un autre élément par simple association d'idées, sans qu'on puisse considérer que le premier serait l'"antécédent" de l'autre.
ex. Un village apparaissait à l'horizon. Le clocher se dressait droit dans le ciel.
Dans cet énoncé, clocher, bien qu'apparaissant pour la première fois dans la chaîne, est employé avec l'article LE, avec une valeur spécifique. Cela tient au fait que le stéréotype de village inclut pour un francophone la notion d'église, et donc celle de clocher.

Le pronom *en* met en place une *anaphore lexicale*:
ex.: *Elle jeta sa cigarette dans le jardin, **en** alluma **une autre***
Dans cet exemple, le pronom *en* reprend seulement la matière notionnelle de *cigarette*, mais construit un référend différent de celui du groupe nominal antérieur.

De même pour l'*anaphore collective*
ex.: *A Paris, **ils** roulent comme des fous.*
le pronom *ils* a un référent collectif et anonyme inféré à partir de la ville indiquée (Paris).

Les phénomènes d'anaphores concernent le plus souvent la reprise de groupes nominaux, mais ils peuvent aussi concerner des adverbes, des verbes ou des adjectifs (cf RIEGEL94 p615-616)

→ pb pour calculer la résolution des anaphores

→ pb d'ambiguïté avec les anaphores:
- *The monkey ate the banana because it was hungry / because it was ripe*
- *The soldiers shot at the women and some of them fell / missed*

Les soldats ont tiré sur les femmes et qques unes sont tombées / qques uns ont raté

- **antécédent d'un pronom relatif**

  35. "elle montra les photos de ses deux enfants qu'elle avait fait avec un appareil minuscule"
  → elle montra les photos de [ses deux enfants$_i$] [qu']$_i$ elle avait fait [O$_i$] avec un appareil minuscule
  → elle montra [les photos$_i$ de ses deux enfants] [qu']$_i$ elle avait fait [O$_i$] avec un appareil minuscule

- **pronom personnel**

  36. Pierre demande à Jean s'il peut partir
  → Pierre demande à Jean si *Jean* peut partir
  → Pierre demande à Jean si *Pierre* peut partir

  37. Vous me demandez de renvoyer le certificat sur l'honneur mais je l'ai perdu depuis longtemps *(perles dans des courriers aux assurances)*
  →Vous me demandez de renvoyer [le certificat$_i$ sur l'honneur] mais je l$_i$'ai perdu depuis longtemps
  → Vous me demandez de renvoyer le certificat sur [l'honneur$_i$] mais je l'ai perdu depuis longtemps

  Un mot rapide sur les P. Labyrinthes : Percues à tort comme syntaxiquement mal formées, à cause d'une amb locale non percues

  Le lac que l'écrivain décrit dans ce livre contemple est le Lac de Come

Pour information

Ambiguïté à l'oral

Homophones

au niveau de la forme phonique uniquement (la forme graphique lève l'ambiguïté)

- /so/
  → sceau: "cachet, empreinte"
  → saut: "bond, bondissement"
  → sot: "bête, idiot, borné"
  → seau: "recipient cylindrique"
- /si/
  (si six scies scient six cyprès, six cent scies scient six cent cyprès
)
- /v«nu, ma{S/
  → venu, venue, venues, venus
  → marche, marchent, marches

problème de segmentation de la chaîne parlée

- /k«poto{oz/
  → le poteau rose
  → le pot aux roses

- /{A)d{«lapa{Ej/
  → rendre la pareille
  → rendre l'appareil
- /estEpE{sevE{/
  → Esther perd ses verres
  → Esther perd ses vers
  → Esther persévère
- /lekals)sES/
  → les caleçons sèchent
  → les cales sont sèches

Les cinq exemples suivants sont tirés de *Approches de la langue parlée en français* Claire Blanche Benveniste

- /lasiri/
  → la Syrie
  → l'Assyrie
- /¿)dezast{/
  → désastre
  → des astres
- /lapatrisek«peiul)nE/
  → la patrie, c'est le pays où l'on nait
  → la patrie, c'est le pays où l'on n'est
- /seyndemisj)d«liEnE{pe/
  → c'est une démission de l'INRP
  → c'est une des missions de l'INRP
- /ils)d«v«nydezE{kylty{El/
  → ils sont devenus des aires culturelles ...
  → ils sont devenus déserts culturels

- "Gall, amant de la reine, alla, tour magnanime,
  galamment de l'arêne à la tour Magne à Nimes." (*Victor Hugo*)
- Castro, c'est une sorte de mythe errant (/ Mitterrand)

  Lu dans un journal satirique lors de la réception de Fidel Castro à l'Elysée. Cité dans *Les ambiguités du français* Catherine FUCHS

Les **formes linguistiques** sont **très fortement ambigües** → il faut un des informations complémentaires pour en déterminer le sens exact.

3. Ambiguïtés " réelle " / " virtuelle "

On peut appeler **ambiguïté réelle**, les ambiguïtés qui subsitent pour l'humain.

En plus de ces ambiguïtés réelles, les sytèmes de traitement automatique des langues ajoutent des **ambiguÏtés virtuelles**: d'une part parce qu'ils n'ont pas accès à autant d'informations que l'humain (connaissances extra-linguistiques) et d'autre part, parce qu'ils traitent généralement les différents niveaux d'analyse (syntaxe, sémantique, ...) séparement.

Distinguer *structure ambiguë* et *phrase ambiguë*

- J'ai rencontré [un professeur [de judo] italien]
- J'ai rencontré [un professeur [de droit international]]
- Il voit la maison verte (// Il voit M. heureuse).
- Il a acheté 1 costume à carreaux (// à crédit).
- La période de reconstruction de l'après-guerre (// de la ville) fut pénible.

Désambiguisation syntaxique:

création d'ambiguïtés artificielles

> *Exemple:*
>
> *I put the bouquet of flowers that you gave me for Mother's day in the vase that you gave me for my birthday on the chest of drawers that you gave me for armistice day*
>
> J'ai mis le bouquet de fleurs que tu m'as offert pour la fete des mères dans le vase que tu m'as offert pour mon anniversaire sur le dessus de la commode que tu m'as offerte pour le jour de l'armistice

8 PP → 4862 arbres syntaxiques (Church & Patil 82), et l'ambiguité n'est pas percue du tout !! (Et elle serait conservée lors de la traduction).

4. Désambiguation

La **désambiguation** consiste à spécifier les informations supplémentaires qui permettent de selectionner l'une des interprétations possibles.

L'**humain utilise un très grand nombre d'information pour éviter l'ambiguité**, mais ces informations ne sont pour la plupart **difficilement implémentable**.

Lorsque les informations provenant de **tous** les niveaux d'analyse **linguistiques** et **extra-linguistiques**, les énoncés ambigüs sont rares (l'ambigüité relève alors de l'interprétation personnelle du locuteur et de son interlocuteur. il est clair que l'analyse automatique des langue est encore très loin de ce genre de problèmatique ...).

Désambiguïser un énoncé revient donc à fournir les connaissances nécessaires.

Connaissances linguistiques

- intonation

  cf *J'ai rencontré un professeur de football américain*

  pauses, courbe intonative, ...

- influence de la taille du contexte

  > *Exemple:*
  > Je comparais [$_{GN}$ mes maîtresses [$_{GP}$ au visage radieux]]
  > Je comparais [$_{GN}$ mes maîtresses] [$_{GP}$ aux femmes de ma famille]
  > Je comparais mes maîtresses au visage maussade des femmes de ma famille.

Connaissances extra-linguistiques

- contexte (extra-lingusitique), situation d'énonciation, connaissances du monde:

  En dehors de tout contexte, de nombreux énoncés sont ambigus (en l'absence de contexte, l'être humain a tendance à en créer un, afin de désambiguiser).

  - J'ai reçu un vase de Chine.
  - J'ai reçu un bonnet de coton.
  - Les croque-morts déposèrent la bière dans le corbillard
  - La soif les poussa à commander deux bières bien fraîches
  - Pour marquer l'anniversaire de Marie, je suis allé chez un fleuriste du quartier chez qui j'ai choisis un très beau bouquet et j'ai fait porter les fleurs à Marie.
  - Pour que la cérémonie soit réussie, j'ai fait porter les fleurs à Marie
  - J'ai vu le cheval avec le télescope

    Exemple:
    1. Zoé a mangé un avocat
    2. Zoé a rendez-vous avec un avocat
    3. Zoé aime l'avocat

    Information nécessaire pour désambiguiser 3.:

    - catégorisation sémantique des cplt de V
      Zoé a mangé un avocat
      + HUMAIN manger COMESTIBLE
      → Zoé ate an avocado.

      Cette information permet aussi d'identifier des tournures figées:
      Zoé a mangé de la vache enragée.
      ≠ HUMAIN manger COMESTIBLE
      + HUMAIN manger de la vache enragée
      Zoé went through hard times.

      HUMAIN manger à LIEU (*Zoe a mangé au resto)*
      → Zoé had a meal out.
      APPAREIL manger combustible (*le poele mange beaucoup de charbon*)
      → This stove uses a lot of coal.
      ORAGAINSATION manger ORAGANISATION (Les grosses entreprises mangent les petites)
      → The big firms swallow up the smaller ones.
      ...

    - classification sémantique des noms

    MAIS
    4. Zoé a mangé un avocat
    → pour désambiguiser *avocat* il faut la catégorie sémantique de *aimer* et vice versa ...

  → explosion combinatoire des analyses possibles

- d'où la nécessité d'avoir une représentation des connaissances du monde (cf IV. Représentation conceptuelle)
- L'anayse automatique de texte se heurte non seulement à la **quantité d'information nécessaire à la désambiguation** mais aussi à la **difficulté de définir un moyen implémentable de représentation de ces informations**