



Sur la complexité logique et expérimentale du langage naturel

Christian Retoré LIRMM Texte

Journée du LIRMM 1er juillet 2015

1. Plusieurs notion de sens

Etant donné un texte, une conversation,...

De quoi ça parle ?

Qu'est ce qui est affirmé ? réfuté ?

Geach est-il l'élève de Wittgenstein ?

(Wikipedia) En 1941, IL épousa la philosophe Elizabeth Anscombe, grâce à LAQUELLE IL entra en contact avec Ludwig Wittgenstein. Bien qu'IL N'ait JAMAIS suivi l'enseignement académique de CE DERNIER, cependant IL EN éprouva fortement l'influence.



2. Exemples de difficultés sémantiques

- (1) Il observe la fille avec des lunettes noires.
- (2) Les enfants prendront une pizza.
- (3) Tous les étudiants ne sont pas autorisés à se garer sur le parking.
- (4)
 - a. James Bond pense qu'un chercheur du labo est un espion.
 - b. Il soupçonne le directeur. $\exists x. pense(JB, espion(x))$
 - c. Il a trouvé un micro collé sous un bureau. $pense(JB, \exists x. espion(x))$
- (5) The Brits love France.
- (6)
 - a. Ma fille est grande. (pour son âge).
 - b. Ma fille est petite. (et doit se placer devant pour la photo)
- (7) J'ai fini mon livre. Lire, écrire, imprimer, relier,...
- (8) Pierre aime sa femme et Paul aussi.
- (9)
 - a. A ce qu'il paraît Marie conduit habituellement une Twingo.
 - b. (1) A ce qu'il paraît > (2) > habituellement > (3) le \exists est en (1) ou (2) ou (3)

- 
- (10) a. Marie n'a pas vu la dernière Palme d'Or ce week-end.
b. chaque ingrédient peut être nié
- (11) Paul est tombé. Pierre l'a poussé.
- (12) Dimanche, nous sommes allés visiter l'arrière pays. Nous sommes partis vers 10h, nous avons déjeuné à St. Guilhem, puis nous sommes redescendus nous baigner au Pont du Diable, avant de rentrer à Montpellier.

3. Sémantique logique

Spécificité de ce style d'analyse :

- phrase(s) \rightarrow formule(s) logique(s), ce qui est affirmé et relation entre ces propositions (discours)
- analyse profonde (de quelques phrases)

interface homme/machine

compréhension/simulation du comportement humain

Un exemple actuel, l'inférence textuelle :

- texte – analyse \rightarrow (conjonction de) formules H
- énoncé – analyse \rightarrow formule C
- démonstrateur automatique (prover) $H \Rightarrow C ?$

4. Compositionnalité du sens

Selon Frege (fin XIX) le sens d'une expression complexe (une phrase) est fonction du sens des éléments qui la composent et de la structure syntaxique :

- (13) a. Le chat qui dort sous ma voiture est celui du voisin.
b. Le voisin qui dort sous ma voiture est celui du chat.
- (14) a. Il a aimé trois livres que Chomsky a écrits.
b. Combien de livres que Chomsky a écrits a-t-il aimé ?
c. Il= Chomsky... possible ou pas ?



5. Cadre général : syntaxe (phrase) \longrightarrow sémantique (formule)

Analyse catégorielle (= preuve = λ -terme) + λ -termes lexicaux
 \rightarrow formule logique exprimant « qui fait quoi »

mot	type sémantique u^* semantics : λ -terme de type u^* x^v la variable ou constante x est de type v
certains	$(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t)$ $\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\&^{t \rightarrow (t \rightarrow t)} (P x)(Q x))))$
énoncés	$e \rightarrow t$ $\lambda x^e (\text{statement}^{e \rightarrow t} x)$
parlent_de	$e \rightarrow (e \rightarrow t)$ $\lambda y^e \lambda x^e ((\text{speack_about}^{e \rightarrow (e \rightarrow t)} x)y)$
eux-mêmes	$(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t)$ $\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x)$



Pour la phrase :

Some statements speak about themselves.

La syntaxe fournit un λ -terme de ce genre :

((some statements) (themselves speak_about))

de type $S^* = t$

On remplace les mots par les lambda termes correspondant,
et on réduit....

$$\begin{aligned} & \left((\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\&(P x)(Q x)))) \right. \\ & \quad \left. (\lambda x^e (\text{statement}^{e \rightarrow t} x)) \right) \\ & \left((\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x)) \right. \\ & \quad \left. (\lambda y^e \lambda x^e ((\text{speak_about}^{e \rightarrow (e \rightarrow t)} x)y)) \right) \end{aligned}$$

$$\begin{aligned} & \downarrow \beta \\ & (\lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\&^{t \rightarrow (t \rightarrow t)} (\text{statement}^{e \rightarrow t} x)(Q x)))) \\ & \quad (\lambda x^e ((\text{speak_about}^{e \rightarrow (e \rightarrow t)} x)x)) \end{aligned}$$

$$\begin{aligned} & \downarrow \beta \\ & (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\&(\text{statement}^{e \rightarrow t} x)((\text{speak_about}^{e \rightarrow (e \rightarrow t)} x)x)))) \end{aligned}$$

le résultat est y plus lisible ainsi :

$$\exists x : e (\text{statement}(x) \wedge \text{speak_about}(x, x))$$

6. Intégration du sens lexical

Analyse habituelle avec deux types :

- **e** entités (tout, par ex : "le chat qui dort sous ma voiture" "un amour de vacances" ...)
- **t** ou **prop** propositions logiques (vraies ou fausse)

réification avec un type événement (**v**) pour éviter l'ordre supérieur :

- (15) a. Hussein court vite
b. $vite(court(Hussein))$ — ordre supérieur
c. $court(e, Hussein) \& vite(e)$ — premier ordre, e : événement

Beaucoup de types : restrictions de sélection, gérer la coprédication (possible ou non) et quantification sur les types (CaML) pour factoriser les expressions qui agissent uniformément sur les types (par ex. conjonction de deux prédicats qui s'applique à un objet ayant deux types différent : "ce livre est lourd mais intéressant")

7. Exemples de problèmes lexicaux

- (16)
 - a. la chaise aboie
 - b. la surveillante a aboyé après un élève très pénible
- (17)
 - a. Barcelone refuse un statut transfrontalier à la Cerdagne
 - b. Barcelone a remporté la ligue des champions
 - c. ? Barcelone refuse un statut transfrontalier à la Cerdagne et a remporté la ligue des champions
 - d. Barcelone a gagné 5 ligues des champions et organisé les Jeux Olympiques.
- (18) Tout jeune Napoléon était très maigre et officier d' artillerie plus tard il devint empereur alors il prit du ventre et beaucoup de pays.
(Prévert)



8. Questions classiques

Portée

Quantification, négation

Pluriels

Nom massif

Glissement de sens

Relations lexicales privilégiées

(lire un livre plutôt que peindre un livre)



9. Ambiguités et difficulté d'analyse

Ambiguités beaucoup de choix pour interpréter

- sous spécification (formulation complexe, réalité cognitive)
- calculs de toutes les interprétations (facile, mais long)



10. Changement de point de vue : du traitement automatique des langues aux sciences cognitives

Modélisation, algorithmes :

phrases -> formules logiques

Ce modèle peut-il prédire nos difficultés de compréhension ?

Ambiguïtés, complexité du calcul des formules logiques

Beaucoup de paramètres, de mesures :

par ex alternance de quantificateurs, glissement de sens enchaînés,

- (19) Une secrétaire présente chaque étudiant à un professeur.
- (20) a. Barcelone a viré son entraîneur
b. ? Barcelone a viré ses jardiniers.
(ceux qui tondent le Campnou)



11. Comment vérifier les prédictions ? quel(s) comportement(s) standard ?

Inobservable sur un corpus standard. (sauf peut-être : certaines formulations absentes le sont en raison de leur complexité)

Dialogues spontanés.

Dialogues guidés (par exemple la quantification universelle est rare)



12. Expériences spécifiques

dialogue fortement guidés

temps de réaction

suivi oculaire

Electroencéphalogramme potentiels évoqués ondes N400 (langage)



13. Exemple plus détaillé : la quantification

Interprétation usuelle des quantificateurs

(21) Keith a joué **une** chanson des Beatles.

une (\exists) appliqué à deux prédicats :

- B=chanson_des_Beatles (x)
- J=Keith a joué (x)
- "une" est appliqué aux deux prédicats : $\exists x. B(x) \& J(x)$

Quelques problèmes :

- le calcul de la formule ne colle pas à la structure syntaxique
- le groupe nominal n'est pas un individu
- symétrique

(22) Certains politiciens sont des menteurs.

(23) ?? Certains menteurs sont des politiciens.

14. Epsilon de Hilbert

(cf. workshop epsilon, Montpellier 10-12 juin 2015,
merci à Elisabeth Greverie)

Formule à une variable libre, propriété \rightarrow terme (\sim témoin de Henkin)

$$F(\epsilon_x F) \equiv \exists x. F(x)$$

Règles :

- de $P(t)$ on infère $P(\epsilon_x P(x)) \equiv \exists x P(x)$
- de $P(x)$ avec un x générique (quelconque)

$$P(\epsilon_x. \neg P(x)) \equiv P(\tau_x. P(x)) \equiv \forall x P(x)$$

- quelqu'un dort \rightarrow $dort(\epsilon_x dort(x))$
- un chat dort \rightarrow $dort(\epsilon_x chat(x))$

Evite les problèmes mentionnés précédemment :

suit la structure syntaxique, asymétrique, groupe nominal = individu

15. Etude de la portée des quantificateurs

PEPS CNRS HUMAIN HUmanités MATHématiques INformatique Complexité et Langage : une étude formelle et expérimentale des mécanismes de compréhension 2013 et 2014 LABRI LIRMM LPL I2M

Exemple très spécifique :

- (24) a. Un rond est relié à tous les carrés
- b. Un rond est relié à chaque carré.
- (25) a. Un gardien est placé devant chaque entrée du musée.
- b. ? Un gardien est placé devant toutes les entrées du musée.

Idée : les interprétations $\exists\forall$ et $\forall\exists$ existent toutes les deux dans les phrases

un ... verbe (tous_les / chaque) ...

mais $\forall\exists$ est plus probable avec "chaque" qu'avec "tous les"

16. Expérience (LPL, Aix)

- une trentaine de volontaires
- phrase entendue (rond et carré, haut bas, plus neutre)
- 4 images sont présentées
le sujet en donner une qui correspond à sa compréhension
- mesure du temps de réaction
- suivi oculaire enregistré



17. Complexité de compréhension du langage

Nous proposons d'étudier quelques phénomènes de syntaxe et de sémantique, en analyse et en compréhension.

- Comment interprétons nous ?
- Quel temps cela prend-il ?
- Construisons nous toutes les interprétations ?
- Y a-t-il/quelles sont les interprétations privilégiées ?
- Etendre et affiner le modèle logique pour traiter simultanément des différents phénomènes.
- Définir des mesures de complexité au sein du modèle.
- Définir les expériences pour confirmer ou infirmer les prédictions.
- Les réaliser.
- Y a t il un ou des comportements standard ?
- Dépendent ils des pathologies langagière des sujets ?



18. Quelques questions à étudier

- La quantification : portée respective.
- Quantificateurs généralisés, éléments génériques, classe de comparaison.
 - (26) Ma fille est grande.
 - (27) Le lion a tendance à s'échapper. (générique/particulier)
 - (28) Un représentant de chaque labo connaît un conseiller de chaque commune. (très complexe à vérifier)
- Polysémie logique, glissements de sens,....
 - (29) Ma voiture est crevée.
 - (30) ?? Ma voiture est bouchée.
 - (31) J'ai fini mon livre



19. Populations spécifiques

Différence ou pas ?

Indication de diagnostic ?

Mise en oeuvre difficile (problèmes éthiques)



20. Expérience : négation et enfants dyslexiques

Delfitto & Vender, Vérone

Hypothèse : mémoire à court terme plus limitée :

- (32) Une baleine n'est pas un poisson.
- (33) Une baleine n'est pas un oiseau. (plus difficile)

Classement de complexité

affirmatives vraies > affirmatives fausses > négatives fausses >> négatives vraies

Enfants dyslexiques : très forte amplification du phénomène
(temps de réaction, erreurs)



21. Autre expérience : structure du discours de patients schizophrènes

Amblard, PEPS Humain Schizophrénie et Langage SLAM

Pas d'expériences spécifiques.

Corpus entretiens patient/psychologue totalement anonymisé.

Particularités dégagées :

- Arbres des relations discursives : élaboration (vertical) narration (horizontal) Normalement on ne peut repartir que des positions les plus à droites.
- Peut repartir sur un autre sens d'un mot précédemment utilisé.

- (34) a. V87 Vous savez ça arrive à tout le monde d'avoir des moments biens et des moments où on est perdu
- b. G88 Oui j'ai peur de perdre tout le monde

22. Autre expérience : relations lexicales

Bruno Gaumé (Toulouse)

Association de mots (à la JeuxDeMots)

Beaucoup plus riches sur les enfants autistes, surdoués, très limités chez les patients de Alzheimer



23. Projets passés, en cours, futurs

PEPS CNRS HUMAIN HUmanités MATHématiques INformatique Com-
plexité et Langage : une étude formelle et expérimentale des méca-
nismes de compréhension 2013 et 2014 LABRI LIRMM LPL I2M

ANR Contenus numériques et interaction. POLYMNIE Analyse et syn-
thèse dans les grammaires catégorielles abstraites du lexique au dis-
cours. durée sept 2012 – février 2016 LORIA LABRI IRIT INRIA-Paris
LIRMM

Modèle logique / expériences sur des populations dites saines ou spéci-
fiques

Projet de projet FET RIA THUMP Type theoretical study of the semantic
HUMAN Processing : logical modelling, linguistic experiments and de-
sign of human-inspired automated semantic analysis Montpellier, Vé-
rone, Londres, Tuebingen, Tilburg, Amsterdam (merci à Marie-Christine
Pothier)