

Logique, Linguistique et Informatique

Christian Retoré, Université Bordeaux 1

Équipe Signes linguistiques, grammaire et sens :
algorithmique logique de la langue

INRIA-Futurs, LaBRI -C.N.R.S. et Département des Sciences du Langage Université Bordeaux 3

Linguistique et informatique :une longue histoire	2
Exemple d'outils de linguistique computationnelle	3
Linguistique computationnelle	5
Niveaux d'analyse de la langue	7
Modèles informatiques et mathématiques en linguistique	10
Un classique:grammaire générative et langages formels	13
Logique et grammaire : un lien naturel et traditionnel.....	19
Syntaxe et sémantique des langues en théorie des types	24
Réseaux de démonstration	45
Conclusion	46

Linguistique et informatique : une longue histoire

- 1949 Machine Translation → aide à la traduction,

- The flesh is weak but the spirit is willing
- The meat is rotten but the vodka is strong
- La chair est faible mais l'esprit est fort.

aujourd'hui réparti dans les deux disciplines suivantes

- 1960 **Computational Linguistics**
Structuration du précédent (théories mathématiques, linguistiques)
- 1965 Automatic / Natural Language Processing
Focalisé sur les outils et plus particulièrement:
 - ▶ Analyse syntaxique
 - ▶ Méthodes statistiques
- 1970 Natural Language Understanding (AI) → approches cognitives

Exemple d'outils de linguistique computationnelle

- ▶ Le Graal: la traduction automatique (il faut savoir tout traiter pour y parvenir)
- ▶ Aide à la traduction:
 - domaine spécifique
 - repère les expressions idiomatiques (aller bon train)
 - propose pour chaque mot ou expression des traductions
 - les assemble avec les choix du lecteur
 - (éviter au maximum la représentation des connaissances)
- ▶ L'interface homme/machine en langue naturelle
 - par exemple: interrogation de BD en langage naturel
 - Quels sont les films des années cinquante qui passent actuellement à Bordeaux?

- ▶ Correcteurs orthographiques (pas simple):
Quels livres crois-tu qu'il sait que je pense que tu as lus?
- ▶ Génération automatique de bulletins météo, de comptes-rendus,...
- ▶ Résumé automatique
- ▶ Recherche d'information (notamment sur le web)
production laitière / production de lait
production minière / production de mine(s)???
- ▶ reconnaissance de la parole (par ex. pour sous-titrage)
nécessite une analyse morpho-syntaxique pour fonctionner en temps réel

Linguistique computationnelle

(ou linguistique informatique)

Un domaine interdisciplinaire :

- mathématiques
 - logique
 - théorie des langages
 - probabilités
- informatique
 - algorithmique
 - génie logiciel
- linguistique
 - paramètres phonologiques
 - grammaire générative
 - syntaxe structurale
 - philosophie du langage

Des objectifs variés :

- ▶ Réalisation d'outils de traitement des langues
- ▶ Formalisation des théories linguistiques
vérification ou réfutation d'hypothèses linguistiques
surtout si celles-ci sont de nature algorithmique
ex. les modèles syntaxiques doivent admettre des algorithmes
d'apprentissage efficaces à partir d'exemples positifs
- ▶ Développement des théories informatiques et mathématiques,
pour elles-mêmes,
éventuellement pour d'autres objectifs

Niveaux d'analyse de la langue

La langue étant un système complexe un des apports de la linguistique a été d'organiser son étude suivant divers modules.

phonétique étude des sons concrets d'une langue
Accoustique - Système phonatoire/auditif

phonologie Les sons abstraits: système discret (dans un continu)

Bali / Paris indistincts pour un japonais

prosodie Structure du phrasé et de leur enchainement: pauses, intonation

"Je serai très heureux de venir parler au LaBRI, laboratoire auquel je dois ma formation initiale en informatique, par exemple sur la lambda-DRT."

"Je serai très heureux de venir parler au LaBRI — laboratoire auquel je dois ma formation initiale en informatique — par exemple sur la lambda-DRT."

morphologie structure des mots

morphologie dérivationnelle [formation des mots]

préfixes, suffixes, nom composés, etc.

changement de catégorie possible

noble → noblesse petit → petitesse

maison → maisonnette camion → camionnette carpe → carpette?

morphologie flexionnelle [déclinaisons, conjugaisons]

en général pas de changement de catégorie (sauf except, part present)

arriver → arriv[er][ons]

cheval → chevaux

syntaxe étude de la structure de la phrase

*Je fais la réparer

Je la fais réparer

* [[Pierre [mange une]] pomme]

Pierre [mange [une pomme]]

sémantique étude du sens des mots, des phrases hors contexte

sémantique lexicale sens des mots et relation entre eux

livre, imprimer (objet concret), lire (contenu abstrait)

sémantique logique 2 aspects indépendants

sémantique vériconditionnelle déterminer les conditions dans lesquelles un énoncé est Vrai ou Faux ou ...

(sens = formule logique)

sémantique compositionnelle calcul du sens d'un constituant à partir du sens des parties

(λ -calcul pour gérer la composition du sens et les substitutions)

pragmatique utilisation de la langue pour communiquer dans un contexte énonciatif

Référence des indexicaux: 1 et 2 personnes (je, nous, vous), ici, maintenant, démonstratifs, ...

Allons plutôt dans ce restaurant.

Modèles informatiques et mathématiques en linguistique

- probabilités, statistiques

marquage des parties du discours sur un texte en fonction des n mots précédents : si les mots précédents sont *article*, *nom*, *adjectif* il est peu probable que le mot suivant soit *article* sauf apposition

■ grammaires formelles

- morphologie: automates d'états finis, transducteurs
nombres, dates
chanterons → chanter 1 pers. pl. , futur

- syntaxe générative, théorie des langages formels

[La [petite brise]] [la glace]

[La petite] [[brise [la glace]]]

Il [regarde [une passante]] [avec des lunettes noires].

Il [regarde [une [passante [avec des lunettes noires]]]]

Elle [[a trouvé] [son [parapluie bizarre]]]

Elle [[[a trouvé] [son parapluie]]] [bizarre]

Syntaxe étendue vers la sémantique:

* Il_i a revendu un livre que Pierre_i a lu.

Combien de livres que Pierre_i a lus a-t-il_i revendu?

■ logique

- logique pour la sémantique

tous les médecins sont des conducteurs

(donc) tous les médecins français sont des conducteurs français

*(donc) tous les bons médecins sont des bons conducteurs

J'avais trois pièces, j'en ai perdu une. Je la cherche.

*J'avais trois pièces, j'en ai perdu deux. Je la range dans ma poche.

- une particularité dans la famille logique linéaire

logique pour la syntaxe

structure d'analyse = déduction formelle = graphe

Un classique: grammaire générative et langages formels

En informatique, mathématique, biologie,... développements conséquents: compilation, parallélisme, groupes profinis, génômique,...

grammaires formelles: notion linguistique due à Noam Chomsky (+ contribution mathématique de Marcel-Paul Schützenberger)
première notion de grammaire hors-contexte Paṇini, 5e s. av. J.C.

- Une langue n'est pas l'ensemble des énoncés de ses locuteurs :
on peut toujours produire des phrases nouvelles
qui sont identifiées comme des phrases par les locuteurs
par ex. $E_{i+1} = \text{Il croit que } E_i$
- → Hypothèse: La langue est un ensemble de règles (inconscientes) :
les enfants surgénéralisent lorsqu'ils ont acquis une règle: "vous faites"

■ d'où grammaires formelles, hiérarchie de Chosmky

- Distinction **compétence** /performance : la **grammaire** / ce qu'on utilise

Le loup a dévoré la chèvre.

La chèvre que le loup a dévoré avait mangé le chou.

? Le chou que la chèvre que le loup a dévoré avait mangé appartenait au passeur.

?? Le passeur auquel le chou que la chèvre que le loup a dévoré avait mangé appartenait possède plusieurs bateaux.

??? Les bateaux que le passeur auquel le chou que la chèvre que le loup a dévoré avait mangé appartenait possède sont des barges.

(néanmoins correct, en prenant son temps et un crayon)

- quelle est l'allure des règles (exprimant la compétence des locuteurs), c.-à-d. la situation des langues dans cette hiérarchie abstraite?

2 principes:

- phrase analysables (compréhensibles) mais en un temps raisonnable (polynomial?)
- grammaire apprenable à partir d'exemple positifs exclusivement

■ phrases analysables mais en temps polynomial

- ▶ les langages réguliers ne suffisent pas:
(ex. précédent) Sujet1 Sujet2 Sujet3 ... Verbe3 Verbe2 Verbe1

- ▶ les langages hors-contextes non plus:
(complétives NL) Sujet1 Sujet2 Sujet3 ... Verbe1 Verbe2 Verbe3

- ▶ un peu plus que hors-contexte, mais avec analyse polynomiale :
TAG ou grammaires hors-contexte avec mouvements

■ Quelques principes de la grammaire universelle:

- Tout groupe nominal doit recevoir un cas, et seul un verbe **conjugué** donne un cas.

Il semble que l'été arrive.

L'été semble arriver.

* Il semble que l'été arriver.

* Il semble l'été arriver.

- Un pronom doit être gouverné par son antécédent
(position relative dans l'arbre d'analyse)

Le chien de Carlotta pense que il ne l'aime pas.

il \neq l' tout les autres égalités sont possibles

Le chien de Carlotta pense que il ne s'aime pas.

il=s' tout les autres égalités sont possibles

* Il_i a aimé deux livres que Chomsky_i a écrit.

Combien de livres que Chomsky a écrit a-t-il_i aimés?

Logique et grammaire : un lien naturel et traditionnel

Depuis l'antiquité (Aristote, Denis de Thrace)
puis au Moyen-Âge (scholastique), au 18(Port-Royal)...

La phrase a une structure logique.

Les enfants prendront une pizza.

$$\forall x \ (enfant(x) \Rightarrow (\exists y \ pizza(y) \wedge prendre-futur(x,y)))$$
$$\exists y \ (pizza(y) \wedge \forall x \ (enfant(x) \Rightarrow prendre-futur(x,y)))$$

Une pizza sera servie aux enfants.

$$\exists y \ (pizza(y) \wedge \forall x \ (enfant(x) \Rightarrow prendre-futur(x,y)))$$
$$\forall x \ (enfant(x) \Rightarrow (\exists y \ pizza(y) \wedge prendre-futur(x,y)))$$

Subtiles différences:
chaque, tout, tous les
un, des, certains

Quantificateurs généralisés

La langue présente d'autres quantificateurs:
la plupart, les, un grand nombre de, peu de,

La plupart des politiciens ont lu un livre d'économie.

les **nombre**s sont aussi des sortes de quantificateurs:

Mettre huit gouttes dans trois cuillères à soupe d'eau.

$3 \times 8 = 24$ gouttes ?

8 gouttes ?

Interprétation, mondes possibles, intentionnalité

Sémantique vériconditionnelle:

le sens d'un énoncé est l'ensemble des conditions qui le rendent vrai
il s'identifie avec la classe des mondes possibles dans lesquels il est vrai.

Cet étudiant croit que Chomsky est informaticien.

Dans tous les mondes possibles compatibles avec les croyances de cet étudiant, Chomsky est un informaticien.

Lectures [de re](#) et [de dicto](#)

James Bond croit que l'un des chercheurs du laboratoire est un espion.

James Bond pense que Blofeld est un espion.

James Bond a trouvé un microfilm dans le laboratoire.

Compositionnalité

Frege le sens du tout est construit à partir du sens des parties.

Paul, que je connais, n'est pas arrivé.

Limites de la compositionnalité:

Si un paysan possède un âne, alors il_i le_j bâte.

Si $(\exists p \exists a \text{ Ane}(a) \wedge \text{Paysan}(p) \wedge \text{Poss}(p,a))$ alors $\text{Bât}(p,a)$.

2a et 2e p LIBRES??

Limites d'une approche purement logique

J'avais trois trombones dans ma poche, je les ai tous perdus sauf un. Je le range dans un tiroir.

* J'avais trois trombones dans ma poche, j'en ai perdu deux. Je le range dans un tiroir.

Aspects sémantiques des catégories syntaxiques

Les catégories ou parties du discours ont une contre partie logique.
Parallèle entre les catégories fondamentales **Verbes** et **Noms**
les **Prédicats** et les **Individus** de la logique.

- ▶ Groupes nominaux: individus (individus ou variables d'individus quantifiables)
- ▶ Verbes, groupes verbaux: prédicats
- ▶ Adjectifs partage les caractères avec les noms (accord, déclinaisons) et avec les verbes (expriment un prédicat)
- ▶ Groupes prépositionnels : ni des prédicats, ni des individus

Syntaxe et sémantique des langues en théorie des types

On s'intéresse plus particulièrement aux niveaux en rouge

phonétique

phonologie

prosodie

morphologie dérivationnelle

morphologie inflexionnelle

syntaxe

sémantique

sémantique prédicative

sémantique lexicale

pragmatique

**techniques logiques:
grammaires catégorielles
systèmes déductifs
lambda calcul
logique d'ordre supérieur**

Objectifs concrets

- Analyse d'un énoncé produisant une représentation de son sens.
Interrogation en langue naturelle,
traduction.
- Produire un énoncé à partir d'une représentation sémantique.
Génération de texte,
réponse en langue naturelle à une recherche d'information,
traduction.

Point d'articulation: syntaxe, sémantique compositionnelle

Point fort des grammaires catégorielles (qui sont aussi un calcul logique).

Ingrédients

Formalisme syntaxique tourné vers la sémantique: grammaires catégorielles

Formalisme naturel pour la sémantique compositionnelle :
logique d'ordre supérieur vue dans le lambda calcul simplement typé.

Morphologie:
catégories d'accord
ou unification de traits
ou traitement préalable et catégories "accordées"
ou module avec des transducteurs

Coté syntaxe on pourrait compiler la grammaire en une grammaire de réécriture plus efficace (RCG)

Schéma général

Coté syntaxique:

- ▶ On associe aux mots une formule décrivant leur comportement syntaxique.
- ▶ Un calcul logique permet d'attribuer une catégorie à un composé à partir de la catégorie syntaxique des constituants.
(Les règles ne dépendent pas de la langue considérée.)
- ▶ Si la catégorie d'une suite de mots est S , on a affaire à une phrase.

Coté sémantique:

- ▶ À chaque mot correspond un type sémantique dérivé du type sémantique et un terme qui décrit son comportement sémantique.
- ▶ À chaque assemblage syntaxique correspond une opération sur les termes sémantiques
- ▶ À l'objet phrase on associe ainsi une formule logique.

Grammaires de Lambek (1958)

Une grammaire de Lambek = un lexique Lex

$Lex(mot)$ = un ensemble fini de types (comportement syntaxique de mot)

Types ou formules $L ::= P = \{S, sn, n, \dots\} \quad | \quad L \multimap L \quad | \quad L \circ - L$

$m_1 \cdots m_n \in Langage(Lex)$ ssi $\forall i \exists t_i \in Lex(m_i) \quad t_1, \dots, t_n \vdash S$

Grammaires lexicalisées = seul le lexique diffère d'un langage à un autre
les règles sont les mêmes pour tous les langages

Coté sémantique, on utilise deux types pour représenter les formules logiques.

- ▶ e individus
- ▶ t valeurs de vérités

Dormir: $e \rightarrow t$ fonction qui pour un individu dit si le fait qu'il dorme est vrai ou faux.

Aimer: $e \rightarrow (e \rightarrow t)$ fonction qui pour un individu dit si le fait qu'il dorme est vrai ou faux.

Notation $\lambda x.u$ la fonction qui à la variable x associe le terme u

Pour calculer $(\lambda x.u) t$ on remplace x par t dans u .

Exemple $(\lambda x \lambda y.((\text{aime } y)x))\text{MariePierre}$

$= \lambda y.((\text{aime } y)\text{Marie})$

$= \text{aimePierreMarie}$

$\text{aime}: e \rightarrow (e \rightarrow t)$

$\text{Pierre, Marie}: e$

Quelques exemples de typages dans le lexique

- syntagmes nominaux, noms propres: sn
- nom communs: pomme, voiture : n
- le, la, les, une, des, ... : $sn \circ- n$
- dort: $sn \circ- S$
- mange: $(sn \circ- S) \circ- sn, sn \circ- S$
- rouge: $n \circ- n$
- qui (relatif): $(n \circ- n) \circ- (sn \circ- S)$
- oui: S, n ("Oui.", "Un oui franc et massif")

Quelques exemples de typage attendus du système

- voiture qui vient de passer: n
- la voiture qui vient de passer: sn
- mange une pomme: $sn \circ- S$

Quelques exemples de types sémantiques

- syntagmes nominaux: noms propres e
- nom communs: pomme, voiture : $e \rightarrow t$
- le, la, les, une, des, ... : $(e \rightarrow t) \rightarrow e$
- dort : $e \rightarrow t$
- mange : $e \rightarrow (e \rightarrow t), e \rightarrow t$
- rouge : $(e \rightarrow t) \rightarrow (e \rightarrow t)$
- qui (relatif) : $(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow (e \rightarrow t))$
- oui : $t, (e \rightarrow t)$ ("Oui.", "Un oui franc et massif")

Règles du calcul syntaxique

<p>$A = \text{hyp.libre la plus à gauche}$</p> <p>$\dots [A] \dots$</p> $\frac{B}{A \multimap B} \multimap_i \text{ qui lie } A$	$\frac{\Delta \quad \Gamma}{A \quad A \multimap B} \multimap_e$
<p>$A = \text{hyp.libre la plus à droite}$</p> <p>$\dots [A] \dots$</p> $\frac{B}{B \multimap A} \multimap_i \text{ qui lie } A$	$\frac{\Gamma \quad \Delta}{B \multimap A \quad A} \multimap_e$

Sens des connecteurs et des règles

$u : B \circ\text{-} A$ signifie que si u est suivi de $v : A$ alors $uv : B$

$u : B \text{-}\circ A$ signifie que si u est précédé de $v : A$ alors $vu : B$

exemple: $la : sn \circ\text{-} n$ voiture : n permet d'obtenir $la\ voiture : sn$

(modus ponens ou élimination)

les règles d'introduction permettent de faire
apparaître puis disparaître des constituants fictifs:

si $u : A \circ\text{-} B$ et $v : B \circ\text{-} C$ on a $uv : A \circ\text{-} C$

en effet on a avec $x : C$ $uvx : C$ avec deux modus ponens
et donc $uv : A \circ\text{-} C$ avec une règle d'introduction

exemple: $très : (n \text{-}\circ n) \circ\text{-} (n \text{-}\circ n)$ donne $très\ très : (n \text{-}\circ n) \circ\text{-} (n \text{-}\circ n)$

Lien CFG \leftrightarrow grammaires de Lambek

Mots : terminaux

Types et sous-types : non-terminaux

- CFG \rightarrow grammaire de Lambek
 - ▶ CFG \rightarrow CFG en forme normale de Greibach
 - ▶ si $X \rightarrow aTUV$ ajouter au lexique $a : ((X \multimap V) \multimap U) \multimap T$
 - ▶ La grammaire de Lambek obtenue est faiblement équivalente à la CFG (assez facile)
- grammaire de Lambek \rightarrow CFG
 - ▶ Grammaire de Lambek, taille max d'un type: k
 - ▶ pour chaque séquent démontrable $A, B \vdash C$ où A, B, C sont de taille $\leq k$ prendre une règle $C \rightarrow AB$
 - ▶ pour chaque $a : A$ du lexique prendre une règle $A \rightarrow a$
 - ▶ La CFG (en forme normale de Chomsky) obtenue est faiblement équivalente (conjecture de Chomsky 63, preuve de Pentus 93)

Types syntaxiques, types sémantiques (Montague 1970)

Formules logiques écrites en λ calcul simplement typé avec 2 types: individus e , valeurs de vérités t .

prédicat à n places : $e \rightarrow (e \rightarrow (e \rightarrow (\dots \rightarrow t)))$
fonction à n arguments : $e \rightarrow (e \rightarrow (e \rightarrow (\dots \rightarrow e)))$

Constantes logiques

$\wedge, \vee, \Rightarrow$: $t \rightarrow (t \rightarrow t)$
 \exists, \forall : $(e \rightarrow t) \rightarrow t$

Correspondance types syntaxiques / types sémantiques

S^* = t phrases: valeurs de vérités
 sn^* = e individus
 n^* = $e \rightarrow t$ prédicats à un argument

$(A \multimap B)^* = (B \multimap A)^* = A^* \rightarrow B^*$ extension à toutes les formules

Un exemple de mini lexique On va construire un lexique analysant la (!) phrase:
« Certains énoncés parlent d'eux-mêmes. »

mot **Type syntaxique** u
Type sémantique u^*
Représentation sémantique : λ -terme de type u^*
 x^v **variable ou constante** x de type v

énoncés $n = E$

"énoncés" est un nom commun

$$e \rightarrow t = E^*$$

du point de vue sémantique c'est un prédicat à une place

$$\lambda x^e (\text{enonce}^{e \rightarrow t} x)$$

ce prédicat est la fonction qui a tout individu x associe la valeur de vérité de " x est un énoncé"

mot

Type syntaxique u

Type sémantique u^*

Représentation sémantique : λ -terme de type u^*

x^v **variable ou constante** x **de type** v

parlent_de $(sn \multimap S) \multimap sn = P$

parler_de attend à sa droite un groupe nominal, et à sa gauche un groupe nominal pour produire une phrase

$e \rightarrow (e \rightarrow t) = P^*$

du point de vue sémantique "parler_de" est un prédicat à deux places

$\lambda y^e \lambda x^e ((\text{parler_de}^{e \rightarrow (e \rightarrow t)} x)y)$

c'est-à-dire une fonction qui prend deux individus et rend vrai si et seulement si le second argument (le sujet) parle du premier (l'objet)

mot

Type syntaxique u

Type sémantique u^*

Représentation sémantique : λ -terme de type u^*

x^v **variable ou constante** x de type v

eux-mêmes $((sn \multimap S) \multimap sn) \multimap (sn \multimap S) = X$

"eux-mêmes" (objet) attend à sa gauche un verbe transitif, pour produire une phrase à la quelle il manque un sujet

$(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t) = X^*$

du point de vue sémantique, "eux-mêmes" prend un prédicat à deux places $P(x,y)$ (le verbe transitif) et rend un prédicat à une place

$\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x)$

le prédicat fabriqué par "eux-mêmes" à partir de $P(x,y)$ est $P(x,x)$

mot **Type syntaxique** u
Type sémantique u^*
Représentation sémantique : λ -terme de type u^*
 x^v **variable ou constante** x **de type** v

certain $(S \circ - (sn \circ S)) \circ - n = C$

"certains" (sujet) attend à droite un nom puis une phrase à laquelle il manque un sujet pour donner une phrase.

$$(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t) = C^*$$

étant donné un prédicat à une place P (nom commun) et un prédicat à une place Q (groupe verbal) "certains" fabrique une formule close

$$\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists (e \rightarrow t) \rightarrow t (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} (P x)(Q x))))$$

la formule fabriquée par "certains" est $\exists x P(x) \wedge Q(x)$

mot	Type syntaxique u Type sémantique u^* Représentation sémantique : λ -terme de type u^* x^v variable ou constante x de type v
certains	$(S \circ - (sn \circ S)) \circ - n = C$ $(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t) = C^*$ $\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} (P x)(Q x))))$
énoncés	$n = E$ $e \rightarrow t = E^*$ $\lambda x^e (\text{enonce}^{e \rightarrow t} x)$
parlent_de	$(sn \circ S) \circ - sn = P$ $e \rightarrow (e \rightarrow t) = P^*$ $\lambda y^e \lambda x^e ((\text{parler_de}^{e \rightarrow (e \rightarrow t)} x)y)$
eux-mêmes	$((sn \circ S) \circ - sn) \circ - (sn \circ S) = X$ $(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t) = X^*$ $\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x)$

Analyse syntaxique

$(S \circ (sn \circ S)) \circ n, n, (sn \circ S) \circ sn, ((sn \circ S) \circ sn) \circ (sn \circ S) \vdash S$?

$$\frac{\frac{C \vdash (S \circ (sn \circ S)) \circ n \quad E \vdash n}{C, E \vdash (S \circ (sn \circ S))} \circ_e \quad \frac{P \vdash (sn \circ S) \circ sn \quad X \vdash ((sn \circ S) \circ sn) \circ (sn \circ S)}{P, X \vdash (sn \circ S)} \circ_e}{C, E, P, X \vdash S} \circ_e$$

Ossature sémantique de l'énoncé

$$\frac{\frac{C^* \vdash (e \rightarrow t) \rightarrow (e \rightarrow t) \rightarrow t \quad E^* \vdash e \rightarrow t}{C^*, E^* \vdash (e \rightarrow t) \rightarrow t} \rightarrow_e \quad \frac{P^* \vdash e \rightarrow e \rightarrow t \quad X^* \vdash (e \rightarrow e \rightarrow t) \rightarrow e \rightarrow t}{P^*, X^* \vdash e \rightarrow t} \rightarrow_e}{C^*, E^*, P^*, X^* \vdash t} \rightarrow_e$$

λ -terme correspondant :

$$((c^{C^*} e^{E^*})(x^{X^*} p^{P^*}))^t$$

Calcul de la sémantique

variable := λ -termes sémantiques (de mêmes types)

$$\left((\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge (P x)(Q x)))))(\lambda x^e (\text{enonce}^{e \rightarrow t} x)) \right) \\ \left((\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x))(\lambda y^e \lambda x^e ((\text{parler_de}^{e \rightarrow (e \rightarrow t)} x)y)) \right)$$

$\downarrow \beta$

$$(\lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} (\text{enonce}^{e \rightarrow t} x)(Q x)))))) \\ (\lambda x^e ((\text{parler_de}^{e \rightarrow (e \rightarrow t)} x)x))$$

$\downarrow \beta$

$$(\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge (\text{enonce}^{e \rightarrow t} x)((\text{parler_de}^{e \rightarrow (e \rightarrow t)} x)x))))$$

en d'autres termes :

$$\exists x : e (\text{enonce}(x) \wedge \text{parler_de}(x,x))$$

Résumé de la méthode

- Grammaire définie comme un lexique qui à chaque mot associe des types.
- Analyse syntaxique = typage de la phrase par le type S
= démonstration de S dans une logique sensible aux ressources
- Correspondance types syntaxiques / types sémantiques
analyse syntaxique / démonstration \rightarrow démonstration intuitionniste
= lambda terme (linéaire) qui exprime la structure compositionnelle
- Mettre les lambda termes sémantiques (non linéaires) à la place des mots
beta réduction \rightarrow la formule correspondant au "sens" de l'énoncé
- Ici système très pauvre (surtout syntaxiquement)
 \rightarrow logiques plus sophistiquées
par ex. logique multimodale Moortgat avec postulats en plus des règles.

Réseaux de démonstration

Si analyse = démonstration.... il ne faut pas avoir plusieurs démonstrations pour une même analyse.

Alors on arrive à rendre compte de la complexité instantanée de compréhension d'une phrase pour les relatives imbriquées, la portée des quantificateurs,

....

Désolé, pas de transparents, voir:

Mark E. Johnson. Proof nets and the complexity of processing center-embedded constructions. In C. Retoré, editor. Special Issue on Recent Advances in Logical and Algebraic Approaches to Grammar, volume 7(4) of Journal of Logic Language and Information. Kluwer, 1998, pages 433–447.

Glyn Morrill. Incremental processing and acceptability. Computational Linguistics, 26(3):319– 338, 2000.

Christian Retoré, The logic of categorial grammars – lecture notes. INRIA Research Report 5703 2005.

Conclusion

- ▶ Pour un **petit fragment**, on sait établir une relation calculable entre certains aspects du sens (*grosso modo* qui fait quoi) et la forme d'un énoncé.
- ▶ D'un point de vue **syntactique** le fragment est bien trop restreint:
 - constituants discontinus
Je **ne** sais **pas** .
 - extraction médiane
La personne que_i j'ai vu *t_i* hier rappellera.
 - pronoms clitiques des langues romanes:
Je la fais réparer.
Je sais la réparer.

- ▶ D'un point de vue sémantique aussi, le modèle est trop restreint:
 - Le temps, le lieu, ... ne sont pas pris en compte.
 - Pour rester compositionnel, il faut raffiner la présentation de la logique.
 - La coréférence entre pronoms et référents n'est pas bien prise en compte.
 - L'ajout de paramètres complique le modèle.
 - La sémantique des mondes possibles est très peu réaliste.
 - Lien avec la sémantique lexicale difficile (logiquement, celle-ci peut se voir comme des relations entre prédicats d'ordre supérieur).
- ▶ D'un point de vue pratique, ces techniques manquent de ressources (sauf Grail pour le néerlandais)
 - Absence de lexiques suffisants (acquisition sur corpus annoté, Grail)
 - Complexité des algorithmes d'analyse (supertagging, minimization dans les réseaux, Grail)
 - Mauvaise gestion des phrases incorrectes, pourtant fréquentes.

Moralité: il reste bien des questions intéressantes.