



journée "Prédication" du CRISCO



Comment discriminer automatiquement
les formes verbales
des formes nominales ?

Jacques Vergne

GREYC - Université de Caen

<http://www.info.uni-caen.fr/~jvergne>



journée "Prédication" du CRISCO



Comment modéliser la proposition ?

Jacques Vergne

GREYC - Université de Caen

<http://www.info.unicaen.fr/~jvergne>



journée "Prédication" du CRISCO



*Comment modéliser la proposition ?
Et comment utiliser ce modèle
en analyse automatique ?*

Jacques Vergne

GREYC - Université de Caen

<http://www.info.uni-caen.fr/~jvergne>

plan de l'exposé

- 1. *Question de fond* : choix d'un modèle de la proposition
 - . 1.1 Rappel sur les modèles
 - . 1.2 Modèle de quel objet ?
 - . 1.3 Critères de comparaison entre modèles
 - . 1.4 Le modèle scolaire traditionnel
 - . 1.5 Le modèle chomskien importé en TAL
 - . 1.6 Le modèle de Tesnière importé en TAL
- 2. Le modèle du Groupe Syntaxe du GREYC
 - . 2.1 Le concept de "chunk" (Abney 91, Déjean 98)
 - . 2.2 Le modèle de proposition constituée de chunks
 - . 2.3 L'analyse automatique comme calcul dans la proposition
 - . 2.4 Comment d'abstraire de la langue ?

1. *Question de fond* : choix d'un modèle de la proposition

- . 1.1 Rappel sur les modèles : pluralité des modèles
- . 1.2 Modèle de quel objet ? pluralité des objets
- . 1.3 Critères de comparaison entre modèles
- . 1.4 Le modèle scolaire traditionnel
- . 1.5 Le modèle chomskien vu du TAL
- . 1.6 Le modèle de Tesnière vu du TAL

1.1 Rappel sur les modèles

- modèle : représentation mentale d'un objet,
explicitée, partagée, construite
et non pas : description du réel, ou vérité sur le réel
- modèle : outil à penser, à comprendre l'objet
- modèle : outil à agir sur l'objet
- **pluralité des modèles** pour un même objet

1.2 Modèles de quel objet ?

- Quels **grains** ? (au moins deux)

mot dans la phrase

groupe dans la proposition

mot dans le groupe, syntagme, "chunk"

paragraphe dans le texte entier

pluralité
des objets

- Avec quels regards sur ce grain ?

parties du discours, relations de constituance

fonctions, relations de dépendance

sujet - prédicat, SN - SV, thème - rhème

verbe (actants, circonstants)

- Parole ou écrit ? ou les deux ?

1.3 Critères de comparaison entre modèles

- quel objectif ? pour les concepteurs, pour les utilisateurs
- quels **constituants**, quelle catégorisation, hiérarchisés, ou récursifs ?
- relations entre **constituants** : constituance, dépendance, ...
- représentation graphique
- ordre linéaire <--> représentation graphique
- liens avec le matériau attesté

1.4 Le modèle scolaire traditionnel

- objectif : prescrire, enseigner le bien écrire des **phrases**
- pas d'intérêt pour le matériau attesté
- **phrase** = **proposition** principale (+ **propositions** subordonnées)
- **modèle centré explicitement sur le constituant proposition**
- **proposition** : sujet - verbe - compléments du verbe (fonctions)
- catégories des **mots**, dites "parties du discours"
- fonction des **mots** (groupes ? récursifs ?) dans la proposition

1.5 Le modèle chomskien vu du TAL

- comment une **phrase** est **constituée** de **mots** et de **syntagmes**
- modèle de la **phrase** ? en fait modèle **implicite** de la **proposition**
- modèle dégradé dérivé du modèle classique de la prédication

prédication \longrightarrow prédicat

relation \longrightarrow constituant

proposition = sujet - prédicat (fonctions des 2 constituants)



proposition = SN - SV (catégorie du mot-tête du constituant)

1.5 Le modèle chomskien vu du TAL

- le syntagme récursif ---> mot tête + syntagme récursif complément
un tout **de même type** que certaines de ses parties
constituant fondé sur la complémentation
dépendance incluse, cachée dans la constituance
=> pas de hiérarchie des constituants

1.5 Le modèle chomskien vu du TAL

- représentation graphique : arbre des constituants

racine = constituant phrase (P)

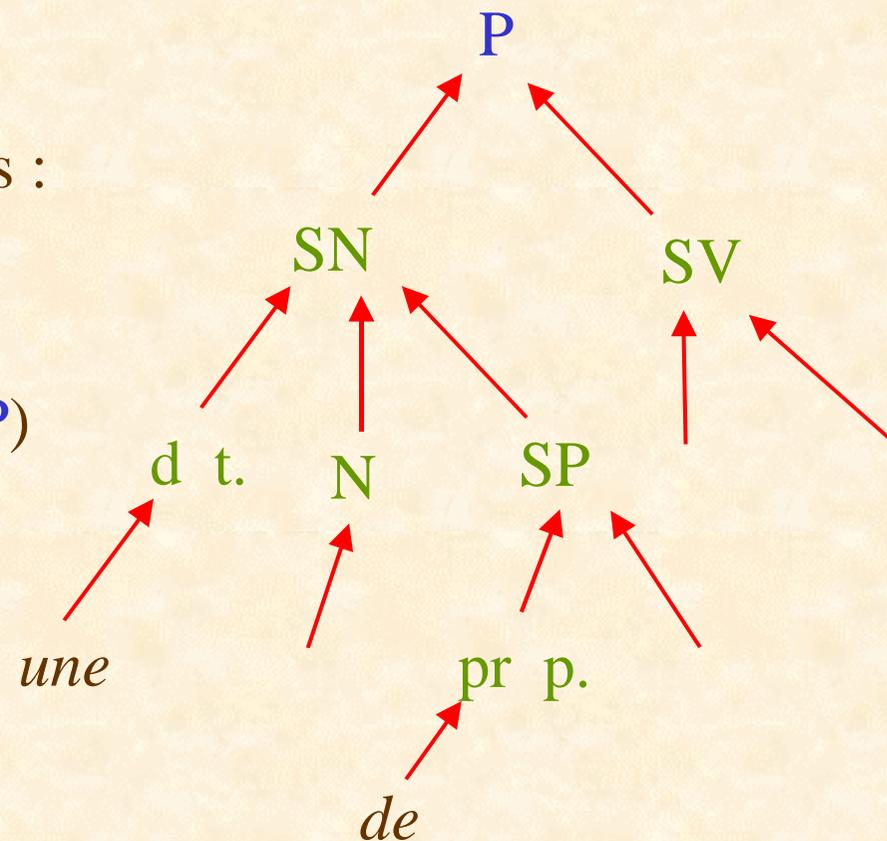
noeuds = catégories des constituants :

syntagmes, mots

arcs = relation de constituance

(SN est constituant de P)

feuilles = graphies des mots

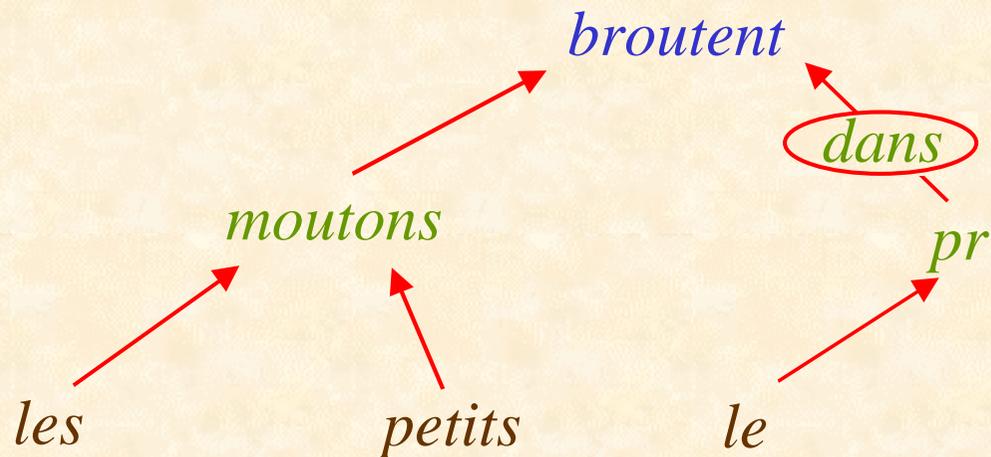


1.6 Le modèle de Tesnière vu du TAL

- comment les mots **dépendent** d'autres mots dans la phrase
- modèle de la phrase ? surtout modèle **implicite** de la proposition
- **proposition** = verbe + ses dépendants (actants et circonstants)
- sujet : un actant parmi les autres
- hiérarchie des constituants minimale : mots dans une phrase
- la translation enracine la fonction dans la catégorie

1.6 Le modèle de Tesnière vu du TAL

- représentation graphique : arbre des dépendances entre mots
 - racine* = graphie du verbe de la proposition
 - nœuds* = graphies des mots
 - arcs* = relation de dépendance (*tel mot* dépend de *tel autre mot*)
 - feuilles* = graphies des mots qui n'ont pas de dépendant



plan de l'exposé

- 1. *Question de fond* : choix d'un modèle de la proposition
 - . 1.1 Rappel sur les modèles
 - . 1.2 Modèle de quel objet ?
 - . 1.3 Critères de comparaison entre modèles
 - . 1.4 Le modèle scolaire traditionnel
 - . 1.5 Le modèle chomskien vu du TAL
 - . 1.6 Le modèle de Tesnière vu du TAL

- 2. Le modèle du Groupe Syntaxe du GREYC
 - . 2.0 Caractéristiques générales
 - . 2.1 Le concept de "chunk" (Abney 91, Déjean 98)
 - . 2.2 Le modèle de proposition constituée de chunks
 - . 2.3 L'analyse automatique comme calcul dans la proposition
 - . 2.4 Comment d'abstraire de la langue ?

2. Le modèle du Groupe Syntaxe

2.0 Caractéristiques générales

2.1 Le concept de "chunk" (Abney 91, Déjean 98)

2.2 Le modèle de proposition constituée de chunks

2.3 L'analyse automatique comme calcul dans la proposition

2.4 Comment d'abstraire de la langue ?

2.0 Le modèle du Groupe Syntaxe

- unifie constituance et dépendance
- comment les **constituants X** sont **constitués** de **constituants Y \neq X**
- comment les **constituants X** **dépendent** de **constituants Y= ou \neq X**
- représentation en machine : graphe des constituants

nœuds = constituants hiérarchisés avec leurs attributs

arcs = relations de constituance, dépendance, coordination, ...

feuilles = mots avec leurs attributs

2.0 Le modèle du Groupe Syntaxe

- double hiérarchie en miroir de constituants non récursifs :

les constituants physiques :

document, paragraphes, mots

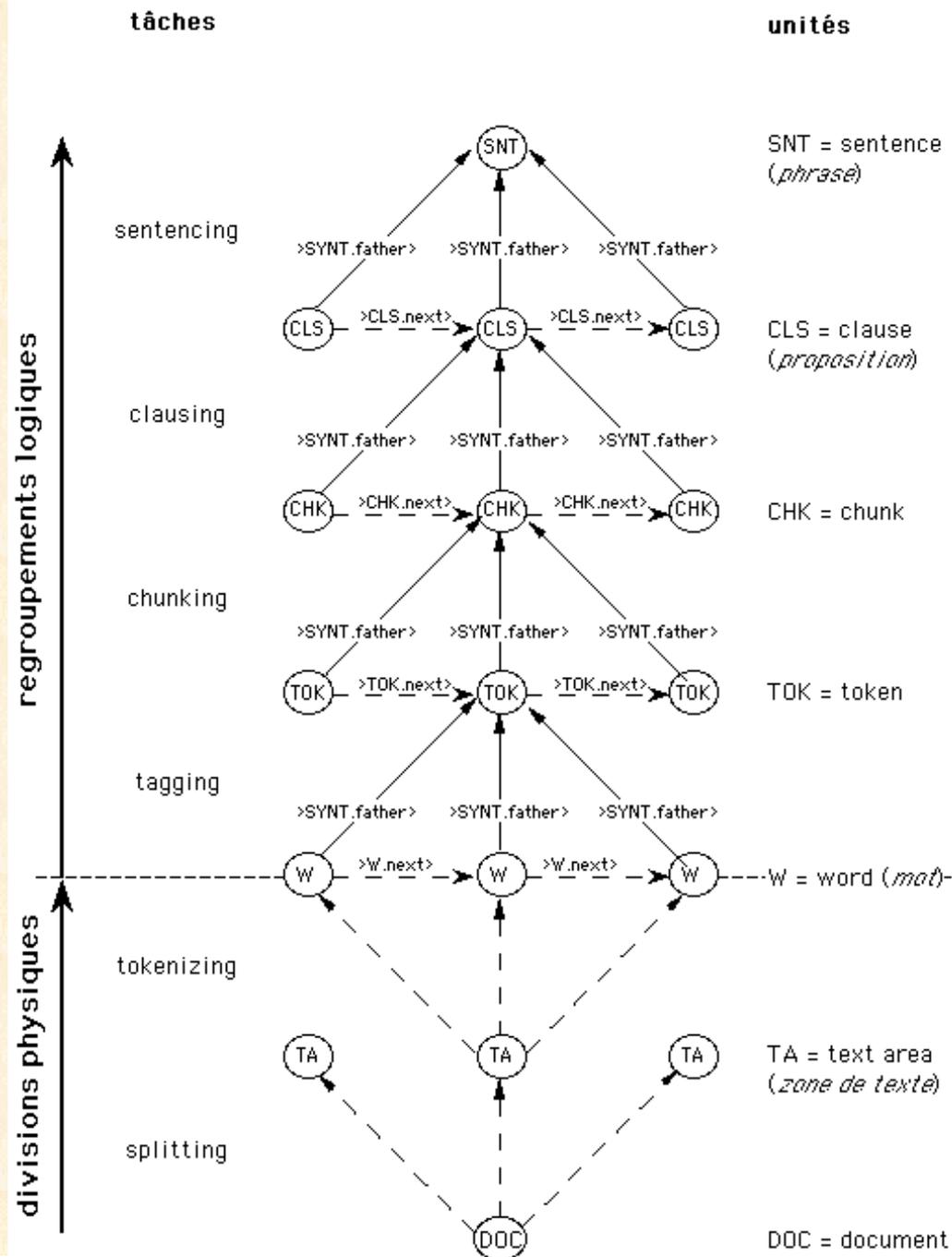
les constituants linguistiques:

mots, tokens, chunks, propositions, phrases

- la proposition est composée de "chunks" nominaux ou verbaux
- opposition verbo-nominale au niveau des chunks

2.0 Le modèle du Groupe Syntaxe

Hiérarchie des unités et relations entre unités



- double hiérarchie en miroir de constituants non récursifs

2. Le modèle du Groupe Syntaxe

2.0 Caractéristiques générales

2.1 Le concept de "chunk" (Abney 91, Déjean 98)

2.2 Le modèle de proposition constituée de chunks

2.3 L'analyse automatique comme calcul dans la proposition

2.4 Comment d'abstraire de la langue ?

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

[I begin] [with an intuition] : [when I read] [a sentence] ,
[I read it] [a chunk] [at a time] .

un segment prosodique : (un groupe accentuel)

These chunks correspond in some way to **prosodic patterns**.

[...] the strongest stresses in the sentence fall one to a chunk,
and pauses are most likely to fall between chunks.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

[I begin] [with an intuition] : [when I read] [a sentence] ,
[I read it] [a chunk] [at a time] .

un segment prosodique : (un groupe accentuel)

These chunks correspond in some way to **prosodic patterns**.

[...] the strongest stresses in the sentence fall one to a chunk,
and pauses are most likely to fall between chunks.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

● [I begin] [with an intuition] : [when I read] [a sentence] ,
● [I read it] [a chunk] [at a time] .

un segment prosodique : (un groupe accentuel)

These chunks correspond in some way to **prosodic patterns**.

[...] the strongest stresses in the sentence fall one to a chunk,
and **pauses** are most likely to fall between chunks.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

[I begin] [with an intuition] : [when I read] [a sentence] ,
[I read it] [a chunk] [at a time].

structure interne :

The typical chunk consists of a single content word surrounded by a constellation of function words, matching a fixed template.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

[I **begin**] [with an **intuition**] : [when I **read**] [a **sentence**] ,
[I **read** it] [a **chunk**] [at a **time**].

structure interne :

The typical chunk consists of **a single content word** surrounded by a constellation of function words, matching a fixed template.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

un exemple :

[I begin] [with an intuition] : [when I read] [a sentence] ,
[I read it] [a chunk] [at a time].

structure interne :

The typical chunk consists of **a single content word** surrounded by **a constellation of function words**, matching a fixed template.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

l'ordre des mots à l'intérieur du chunk :

The typical chunk consists of a single content word surrounded by a constellation of function words, matching **a fixed template**.

A simple context-free grammar is quite adequate to describe the structure of chunks.

2.1 Le concept de chunk d'Abney 91 dans "Parsing by Chunks"

l'ordre des mots à l'intérieur du chunk :

The typical chunk consists of a single content word surrounded by a constellation of function words, matching **a fixed template**.

A simple context-free grammar is quite adequate to describe the structure of chunks.

l'ordre des chunks à l'intérieur d'une phrase :

By contrast, the relationships between chunks are mediated more by lexical selection than by rigid templates.

[...] the **order** in which chunks occur is **much more flexible** than the order of words within chunks.

2.1 Le concept de chunk illustré par Molière

dans Le Bourgeois Gentilhomme :

[Belle marquise] , [vos beaux yeux] [me font] [mourir] [d'amour] .

[d'amour] [mourir] [me font] , [Belle marquise] , [vos beaux yeux] .

[vos beaux yeux] [d'amour] [me font] , [Belle marquise] , [mourir] .

Molière fait permuer les chunks (pas les mots)

2.1 Le concept de chunk d'après Hervé Déjean

morphèmes
de début
de chunk

morphèmes
de milieu
de chunk

morphèmes
de fin
de chunk

de la
of the

-ation
-ancy

- structure acquise semi-automatiquement
sur corpus de langues très variées
- morphèmes de début et de fin : petites listes closes

2. Le modèle du Groupe Syntaxe

2.0 Caractéristiques générales

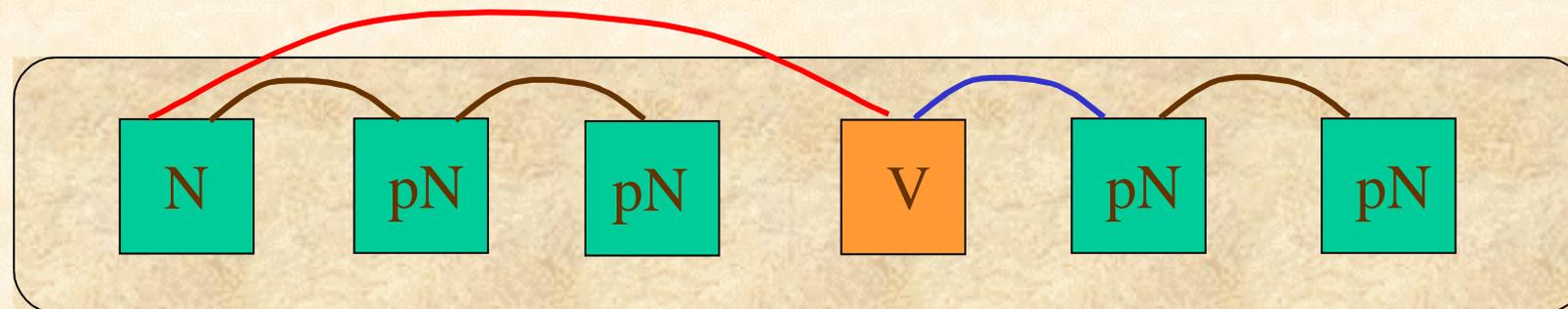
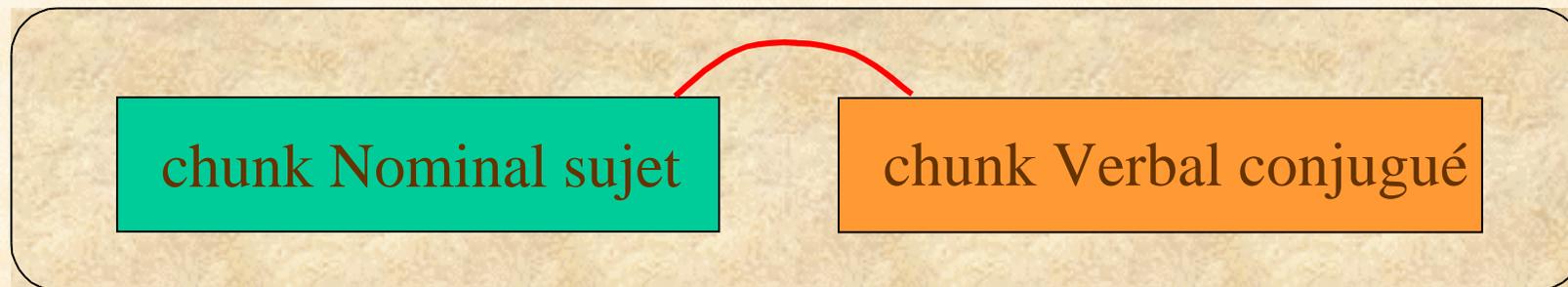
2.1 Le concept de "chunk" (Abney 91, Déjean 98)

2.2 Le modèle de proposition constituée de chunks

2.3 L'analyse automatique comme calcul dans la proposition

2.4 Comment d'abstraire de la langue ?

2.2 Modèle de proposition constituée de chunks



2. Le modèle du Groupe Syntaxe

2.0 Caractéristiques générales

2.1 Le concept de "chunk" (Abney 91, Déjean 98)

2.2 Le modèle de proposition constituée de chunks

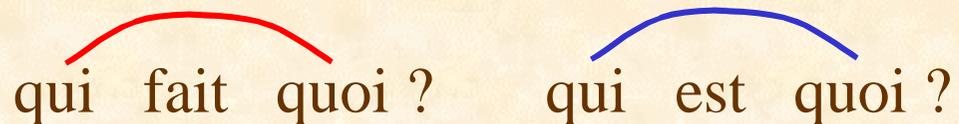
2.3 L'analyse automatique comme calcul dans la proposition

2.4 Comment d'abstraire de la langue ?

2.3 L'analyse syntaxique comme calcul

- discriminer le nominal du verbal, catégoriser :
une étape intermédiaire du calcul
pour calculer des fonctions (ou des relations)
- la catégorisation a un intérêt linguistique (\Rightarrow métalangage)
mais calculer des fonctions (ou des relations)
rend possible l'interprétation par un récepteur
(sans métalangage) :

qui fait quoi ? qui est quoi ?



2.3 L'analyse syntaxique comme calcul

- le matériau vu de manière dynamique,
comme un **flux** qui passe entre 2 humains qui communiquent
=> intérêt pour les *processus* de production et de réception
ces *processus* contraignent les formes du flux (dont l'ordre)
l'étude des processus est à inclure dans l'étude des formes du flux

2.3 L'analyse syntaxique comme calcul

- rupture de paradigme :
 - ne pas choisir parmi des opérandes pré-énumérés
 - > calculer par des opérateurs sur des opérandes inconnus
- les **graphies** des mots :
 - ne pas les énumérer dans un dictionnaire exhaustif en entrée
 - > mots grammaticaux et suffixes suffisent
- les **structures** des constituants :
 - ne pas les expliciter dans une grammaire formelle en entrée
 - > les calculer et les produire en sortie
- le **processus** d'analyse :
 - non pas un processus combinatoire (ou arborescent)
 - > un processus de complexité linéaire,
explicité par des règles appliquées aux grains du flux

2.3 L'analyse syntaxique comme calcul

- l'analyse syntaxique est une simulation du **processus** de **réception** :
 - ou la transformation d'une **chaîne (le flux)** en un **graphe**
 - ou la transformation : **ordre linéaire** --> **ordre structural** (Tesnière)
 - ou la reconstruction d'un graphe à partir de ses nœuds énumérés

chaîne = flux
= éléments reçus
dans un certain ordre



graphe = éléments
avec leurs liens
calculés, restitués

- quel traitement du flux ?
le découper, le segmenter,
et relier les segments

2.3 L'analyse syntaxique comme calcul

- regrouper les mots en chunks, en utilisant leurs frontières
--> chunks délimités (pas tous) et typés (pas tous)
- regrouper les chunks en propositions, en utilisant leurs frontières
--> propositions délimitées et typées
- calcul forme-position des fonctions dans la proposition :
le premier chunk (pouvant être) nominal est le sujet
le premier chunk (pouvant être) verbal est le verbe
- s'il n'y a pas de relation sujet-verbe, un chunk est coupé en deux
--> sujet (complément du sujet) | verbe

2.3 L'analyse syntaxique comme calcul

FRANCE : AIRBUS DÉFIE BOEING DANS LES AVIONS DE 100 PLACES . (TBNDF)

Airbus a décidé de lancer l' A318 , version raccourcie de la famille A320 .

Le consortium démarre avec un portefeuille de 109 commandes , émanant de plusieurs compagnies .

L' A318 , dont l' industrialisation a été décidée hier , sera le quatrième appareil de la famille A320 .

Sa capacité d' une centaine de sièges complète par le bas une famille d' appareils « monocouloir » .

Le site allemand de Dasa à Hambourg devra assembler ce nouvel avion .

Airbus sera ainsi en concurrence directe avec Boeing , qui propose déjà une version de cette capacité .

Air France passe commande de 25 appareils , dont 15 achats fermes .

Airbus table sur plus de 1.300 ventes d' A318 d' ici à 2020 .

La Tribune

2.4 Comment s'abstraire de la langue ?

- certaines propriétés du flux entrant sont indépendantes de sa langue
 - le flux entrant est toujours unidimensionnel
 - ce sont toujours 2 humains qui communiquent (processus) :
 - . même appareil vocal
 - . même système cognitif
 - . même recherche du moindre effort (optimisation)
- le chunk est un constituant fondé sur l'oral, contraint par l'appareil vocal
hypothèse : c'est un concept indépendant de la langue

2.4 Comment s'abstraire de la langue ?

- dans l'analyse automatique,
comment s'abstraire de la langue du flux entrant ?
 - premier paquet de règles : graphies --> attributs des unités
 - paquets suivants : calculs sur les attributs,
indépendants de la langue du flux
- dans l'analyseur du GREYC, opérations communes anglais - français :
 - segmentation en propositions
 - mise en relation des chunks dans les propositions
 - segmentation en phrases
 - mise au point des règles sur corpus anglais - français

fin de l'exposé

- vous trouverez cette présentation en .pdf sur <http://www.info.unicaen.fr/~jvergne>



vos questions ?