

IFT-17586

Intelligence artificielle I



Traitement Automatique du Langage Naturel

Benoît Potvin

Département d'informatique et de génie logiciel
Faculté des sciences et de génie, Université Laval

Été 2003

© Tourigny, Potvin et Capus (2003)

1

TALN



◆ Référence

- G.F. Luger, Artificial Intelligence. Structures and Strategies for Complex Problem Solving, Addison Wesley, (Mass.), 4e éd., 2002, ch. 13, pp. 553-595.

© Tourigny, Potvin et Capus (2003)

2

TALN

*Quelles sont les ÉTAPES
pour COMPRENDRE
un texte écrit en LN ?*

© Tourigny, Potvin et Capus (2003)

3

TALN

◆ Plan

- **Niveaux d'analyse du langage**
- Traitement du langage
- Étapes de l'analyse du langage
- Grammaire, grammaire attribuée
- Analyse et synthèse

© Tourigny, Potvin et Capus (2003)

4

TALN : niveaux d'analyse du langage

◆ PROSODIE

- {règles} relatives au **rythme** et à l'**intonation** dans la langue. (Ex.: analyse des émotions dans le discours d'un enfant)

« *L'instituteur, dit le directeur, est un incapable.* »

« *L'instituteur dit : 'le directeur est un incapable'.* »

« **Je vais terminer** » (par opposition à quelqu'un d'autre)

« **Je vais terminer** » (par opposition à une action déjà accomplie)

« **Je vais terminer** » (par opposition à une autre action)

© Tourigny, Potvin et Capus (2003)

5

TALN : niveaux d'analyse du langage

◆ PHONÉTIQUE

- Étude des **sons** du langage par rapport à leur articulation ou leur réception auditive. (sons d'une langue)

« *La phonétique est utile pour savoir prononcer les mots nouveaux* »

[la fɔnɛtik ɛ tytil pur savwar prɔnɔ̃sɛ lɛ mo nuvo]

© Tourigny, Potvin et Capus (2003)

6

TALN : niveaux d'analyse du langage

◆ MORPHOLOGIE

- Étude des **formes des mots** (racines, préfixes, suffixes, nombre, genre, temps, séparation, etc.)

grand / grande; gros / grosse

écouter, écoute, écoutes, écoutons, écoutez, écoutent

« *Jean a mangé des pommes.* »

nom propre, masculin, singulier ← *manger*, verbe, passé composé, indicatif, 3^{ème} personne, singulier, constructions : transitif, ...

© Tourigny, Potvin et Capus (2003)

7

TALN : niveaux d'analyse du langage

◆ SYNTAXE

- Étude de la **structure des phrases**

« *Marie aime les fraises.* »

« *Les fraises que Marie a achetées sont délicieuses.* »

« *Le boucher sale la tranche.* »

© Tourigny, Potvin et Capus (2003)

8

TALN : niveaux d'analyse du langage

◆ SÉMANTIQUE

– Étude du **sens** des mots, phrases.

« *Les idées vertes incolores crient* »

« *Elle a mangé du poisson avec des amis.* »

« *Pierre est parti à la piscine* »

Où est Pierre ?

Que fait Pierre ?

TALN : niveaux d'analyse du langage

◆ PRAGMATIQUE

– Étude de l'**utilisation** du discours et de son **effet** concret sur l'auditeur.

« *Savez-vous quelle heure il est ?* »

« *Il a demandé l'addition à la serveuse et l'a payée.* »

TALN : niveaux d'analyse du langage

◆ CONNAISSANCE DU MONDE

- Intégration du discours dans son **contexte**

« Marie s'approche de la vendeuse. Elle lui demande si elle a besoin d'aide. »

TALN

◆ Plan

- Niveaux d'analyse du langage
- **Traitement du langage**
- Étapes de l'analyse du langage
- Grammaire, grammaire attribuée
- Analyse et synthèse

TALN : Traitement morphologique

- ◆ Décomposition du texte en mots
- ◆ Séparation des mots ?
 - Discours écrit (espaces, signes de ponctuation, tirets dans les mots composés)
 - Discours parlé (silence, euh, etc.)
- ◆ Difficultés (Discours écrit)
 - Usage du tiret (ex : tire-laine, a-t-il, etc.)

TALN : Traitement morphologique

- ◆ Quels mots faut-il conserver ?
 - Tous les mots
 - être: suis, es, est, sommes, êtes, sont, étais, était, étions, étiez, étaient, fus, serais, serez, etc.
 - aimer : aime, aimes, aimons, aimez, aiment, etc.
 - beau : beaux (masc. plur.), belle (fém. sing.), belles (f. p.)
 - Les formes canoniques des mots
 - racines et finales
 - aimer : aime, aimes, aimons, aimiez, aiment
 - finir : finis, finit, finissons, finissez, finissent
 - règles de formation des mots
 - formation des mots au pluriel, au féminin, ...
 - Ambiguïtés possibles : ex. publions : nom (créations), adjectif (abscons), verbe conjugué (mangions)

TALN : Traitement lexical

- ◆ Recherche des mots dans un dictionnaire
- ◆ But : lever certaines ambiguïtés sur les mots
 - Ressource : dictionnaire

TALN : Traitement lexical

- ◆ Exemples
 - camions : nom pluriel seulement
 - « Nous savons que Marie sait que les savons que Louise a achetés sont excellents. »
 - verbe savoir à la première personne du pluriel (nous savons)
 - nom au pluriel (les savons que Louise a achetés)
 - « Les poules du couvent couvent. »
 - nom : du couvent
 - verbe couvrir à la troisième personne du pluriel : les poules couvent

=> Nécessité de faire une **ANALYSE SYNTAXIQUE**

TALN : Traitement syntaxique

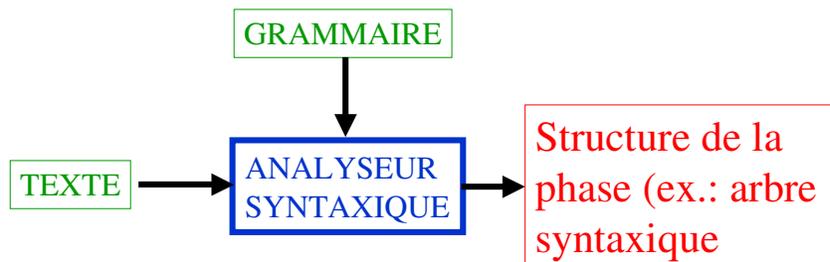
- ◆ Détermination de la structure de la phrase (sujet, verbe, positions, ...)
- ◆ Ressources :
 - Grammaire
 - Programme : analyseur syntaxique (**parser**)

© Tourigny, Potvin et Capus (2003)

17

TALN : Traitement syntaxique

- ◆ Architecture d'un analyseur syntaxique (**parser**) :



© Tourigny, Potvin et Capus (2003)

18

TALN : Traitement pragmatique

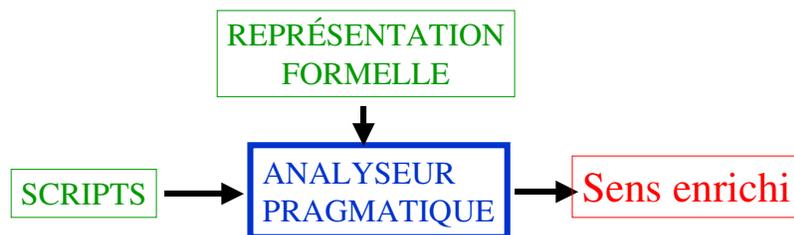
- ◆ But : enrichir, interpréter la représentation formelle d'une BC
- ◆ Ressources :
 - Scripts (schémas)
 - Programme : analyseur pragmatique
 - Exemples
 - Avez-vous l'heure ? : **Quelle heure est-il ?**
 - Avez-vous de l'argent ? : **Me prêteriez-vous \$?**

© Tourigny, Potvin et Capus (2003)

21

TALN : Traitement pragmatique

◆ Architecture



© Tourigny, Potvin et Capus (2003)

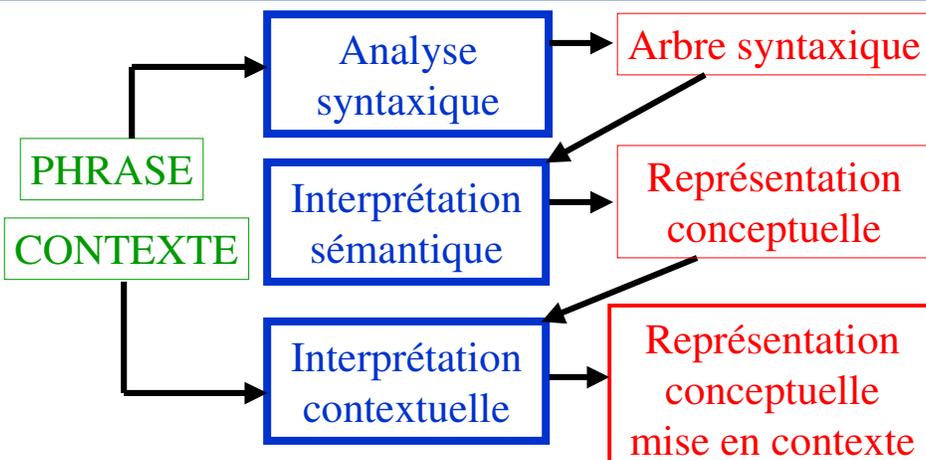
22

TALN

◆ Plan

- Niveaux d'analyse du langage
- Traitement du langage
- ***Étapes de l'analyse du langage***
- Grammaire, grammaire attribuée
- Analyse et synthèse

TALN : Étapes de l'analyse du langage

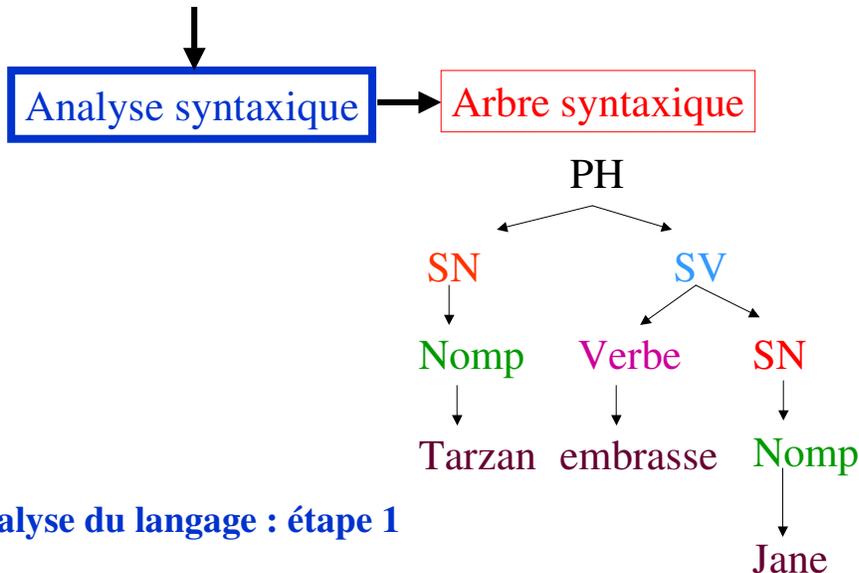


TALN : Étapes de l'analyse du langage

◆ ANALYSE SYNTAXIQUE

- BUT : identifier les structures syntaxiques des phrases
- Relations sujet-verbe verbe-objet, etc.
- Base pour l'interprétation sémantique
- Connaissances utilisées : syntaxe, morphologie, lexicologie, quelques connaissances sémantiques

Tarzan embrasse Jane



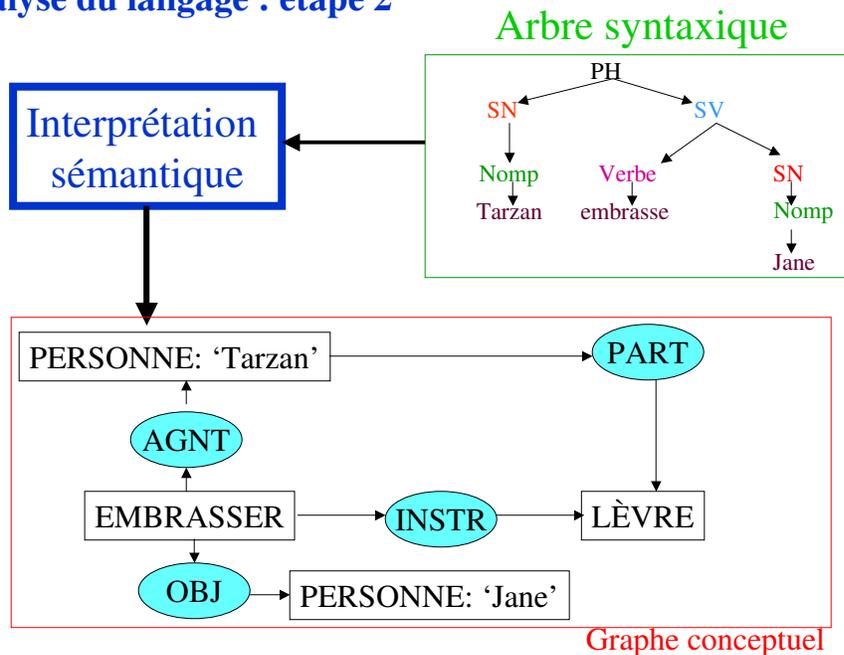
Analyse du langage : étape 1

TALN : Étapes de l'analyse du langage

◆ INTERPRÉTATION SÉMANTIQUE

- BUT : produire une représentation conceptuelle
- Modes de représentation des connaissances
 - ex.: graphe conceptuel (GC), schémas, scripts, etc.

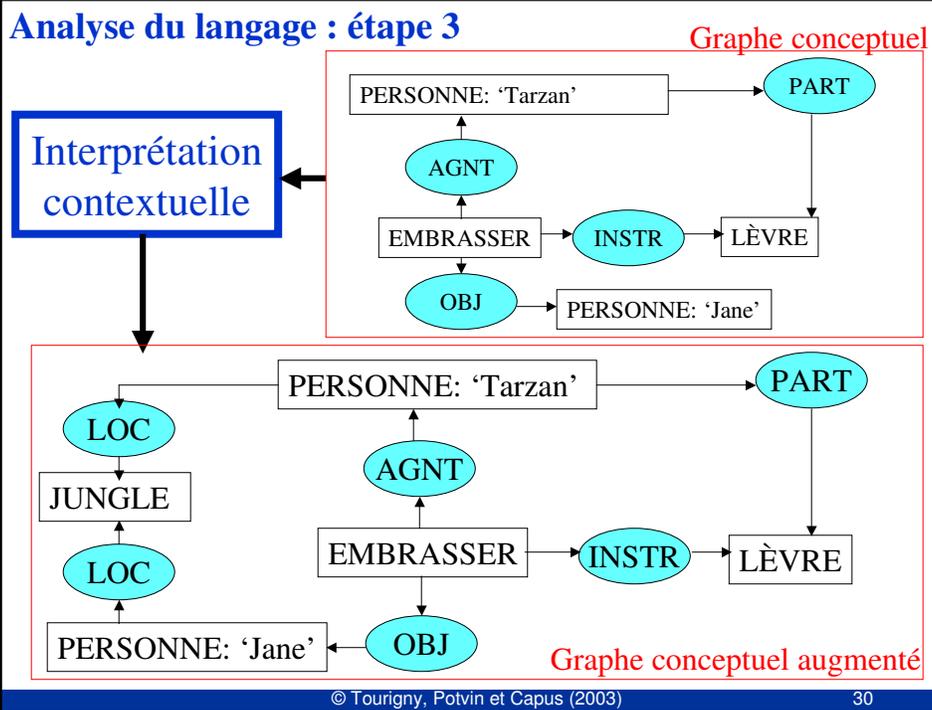
Analyse du langage : étape 2



TALN : Étapes de l'analyse du langage

◆ INTERPRÉTATION CONTEXTUELLE

- Enrichissement de la représentation contextuelle à l'aide de la BC contextuelle



TALN

◆ Plan

- Niveaux d'analyse du langage
- Traitement du langage
- Étapes de l'analyse du langage
- **Grammaire, grammaire attribuée**
- Analyse et synthèse

GRAMMAIRE

V_T : {symboles **Terminaux**} **non dérivables**

V_N : {symboles Non terminaux **dérivables**
(catégories syntaxiques) }

axiome S

R : {RÈGLES DE RÉÉCRITURE}

<phrase > → <syntagme_nominal >, <syntagme_verbal >

<syntagme_nominal > → <nom >

<syntagme_verbal > → <verbe >, <syntagme_nominal >

<nom > → 'Tarzan' | 'Jane'

GRAMMAIRE

{Règles de réécriture}

```

<phrase > → <syntagme_nominal >,
             <syntagme_verbal >
<syntagme_nominal > → <nomp >
<syntagme_verbal > → <verbe >,
                    <syntagme_nominal >
<verbe > → embrasse
<nomp > → 'Tarzan' | 'Jane'
    
```

Symboles

non terminaux :

dérivables

terminaux :

non dérivables

© Tourigny, Potvin et Capus (2003)

33

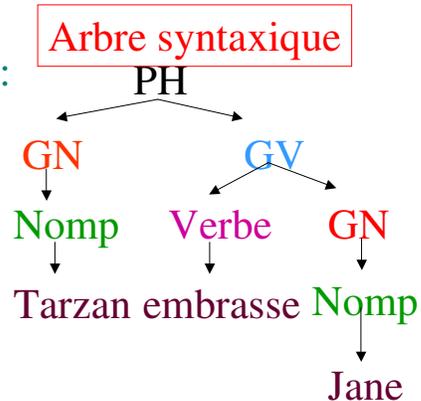
GRAMMAIRE

```

<phrase > → <syntagme_nominal >, <syntagme_verbal >
<syntagme_nominal > → <nomp >
<syntagme_verbal > → <verbe >, <syntagme_nominal >
<nomp > → 'Tarzan' | 'Jane'
    
```

GÉNÉRATION de phrases :

Tarzan embrasse Jane
Jane embrasse Tarzan



© Tourigny, Potvin et Capus (2003)

34

GRAMMAIRE

<phrase > → <syntagme_nominal >, <syntagme_verbal >
<syntagme_nominal > → <nomp >
<syntagme_verbal > → <verbe >, <syntagme_nominal >
<nomp > → 'Tarzan' | 'Jane'

<syntagme_nominal > → <article >, <nom >
<nom > → femmel homme
<article > → un | une

Tarzan embrasse Jane

Jane embrasse Tarzan

un homme embrasse une femme

une femme embrasse un homme

© Tourigny, Potvin et Capus (2003)

35

GRAMMAIRE ATTRIBUÉE

<syntagme_verbal(SV) > → <verbe(V) >, <syntagme_nominal(SN) >
{SV= gv(V,SN)}

- ATTRIBUTS : paramètres

SV V SN

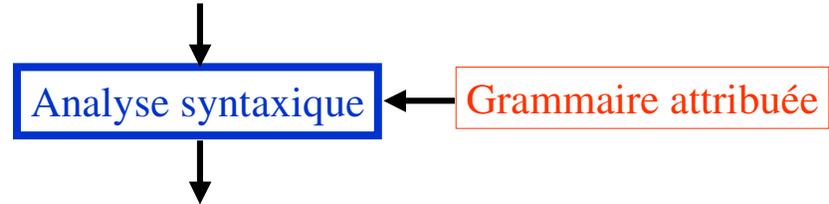
- ACTIONS ou procédures

© Tourigny, Potvin et Capus (2003)

36

GRAMMAIRE ATTRIBUÉE

« *Tarzan embrasse Jane* »



$gv(gn(np(Tarzan)), gv(v(embrasse), gn(np(Jane))))$

Analyse du langage : étape 1

© Tourigny, Potvin et Capus (2003)

37

TALN

◆ Plan

- Niveaux d'analyse du langage
- Traitement du langage
- Étapes de l'analyse du langage
- Grammaire, grammaire attribuée
- **Analyse et synthèse**

© Tourigny, Potvin et Capus (2003)

38

TALN : Analyse et synthèse

