

# Informatique linguistique

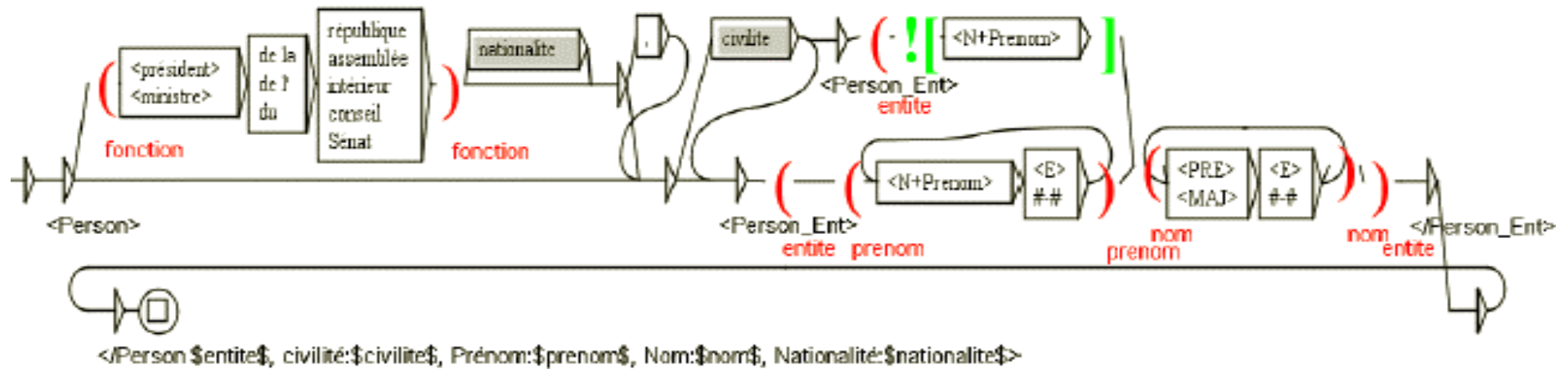
De la description linguistique  
à l'ingénierie des langues



# Activités appliquées

- Objectifs principaux : recherche et extraction d'informations
- Financement : projets subventionnés, contrats avec des PME
- 2 projets de startups à Louvain-la-Neuve : recherche d'informations, correction orthographique
- Méthodes
  - Pour la robustesse, statistiques lexicales
  - Pour le contrôle sur les performances, ressources linguistiques maintenues manuellement : grammaires locales (Gross, 1997), lexiques, qui sont les résultats de recherches fondamentales sur plusieurs dizaines d'années

# Extrait d'une grammaire locale

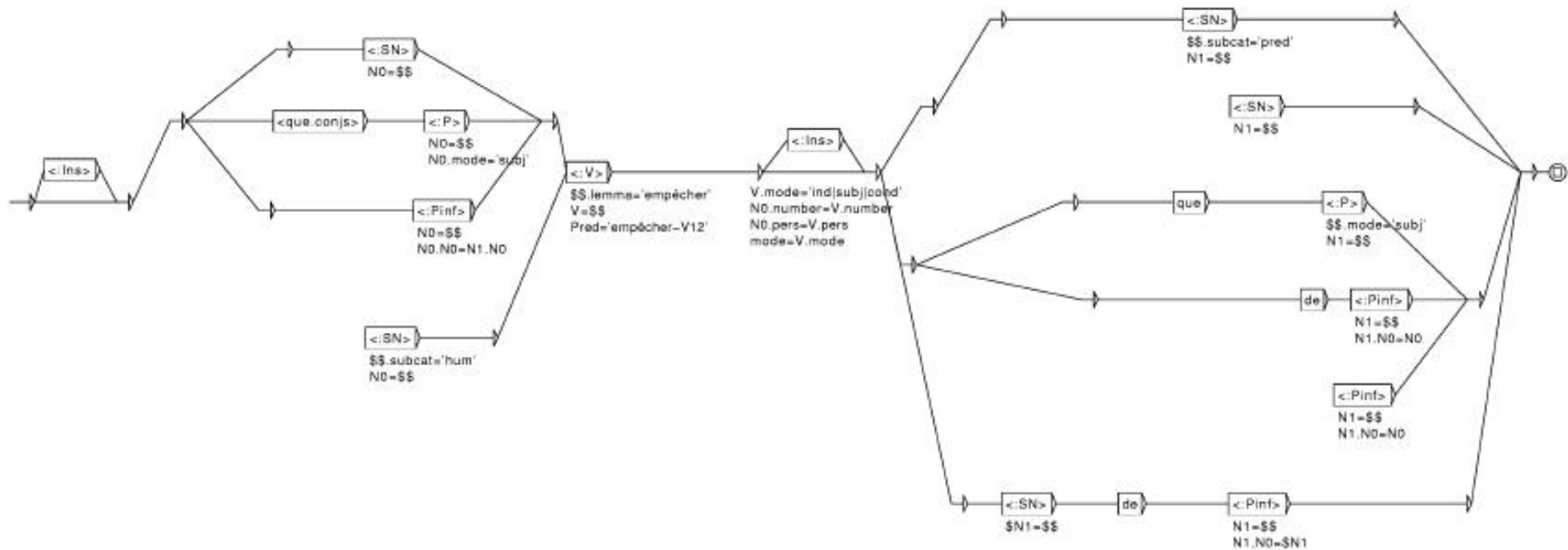


- Extraction d'informations nominatives (Martineau *et al.*, 2007)

# Outils de traitement par ressources

- Plates-formes de traitement des langues par ressources linguistiques et règles : Unitex et Outilex, une cinquantaine d'institutions utilisatrices, une quinzaine de langues
- Chunking
- Préparation de corpus de textes : corpus de SMS, corpus oral transcrit
- Analyse morphologique : granularité fine, rappel élevé
- Lexicalisation de grammaires syntaxiques

# Extrait de grammaire syntaxique lexicalisée



- Blanc & Constant (2005)

# Construction et gestion de ressources

- Une stratégie à long terme : construction et maintenance manuelles
- Une politique de qualité : granularité fine, contrôle sur les performances des systèmes
- Annotation de corpus : expressions multi-mots
- Évaluation de grammaires locales
- Normalisation de modèles formels de lexiques : ISO, language resource management
- Documentation et extension de lexiques du français : propriétés syntaxico-sémantiques
- Création de lexiques : français, coréen, grec moderne, roumain

# Extrait d'un lexique syntaxico-sémantique

	Complétive	IIT	IIP ce qui/ce que	IIP ≠ ce qui/ce que	IIP prépositionnel
	N0 V N1 sur ce Qu P N1 est Vpp sur ce Qu P N se V sur ce Qu P [pc z.]	N0 V N1 sur le fait Qu P N0 V N1 sur si P ou si P prép = ∅	N1 est Vpp sur ce qui/ce que P-N prép = ∅ N se V sur ce qui/ce que P-N prép = ∅	N0 V N1 sur Qu-P-Qu- prép = ∅ N1 est Vpp sur Qu-P-Qu- prép = ∅ N se V sur Qu-P-Qu- Qu- = qui Qu- = pourquoi, comment	N0 V N1 sur Prép Qu-P-Qu- prép = ∅ N0 V N1 sur sur Qu-P-Qu- N0 V N1 sur Qu-P-Qu- N1 est Vpp sur Prép Qu-P-Qu- prép = ∅ N1 est Vpp sur sur Qu-P-Qu- N1 est Vpp sur Qu-P-Qu- N se V sur Prép Qu-P-Qu- prép = ∅ N se V sur sur Qu-P-Qu- N se V sur Qu-P-Qu-
interroger	* * * *	* * * *	* * * *	* * * *	* * * *
questionner	* * * *	* * * *	* * * *	* * * *	* * * *
renseigner	* * * *	* * * *	* * * *	* * * *	* * * *
baser	+ + + *	+ + + *	+ + + *	+ + + *	+ + + *
fonder	+ + + *	+ + + *	+ + + *	+ + + *	+ + + *
asseoir	+ + + *	+ + + *	+ + + *	+ + + *	+ + + *

- Nakamura (2008)

# Contrats

- DoXa (2009-2011), Cap Digital, traitement des sentiments et opinions, 13 partenaires
- Infom@gic (2006-2009), Cap Digital, recherche et extraction d'informations, 20 partenaires
- Outilex (2002-2006), Ministère de l'Industrie, plate-forme de traitement de textes écrits, 10 partenaires
- LexSynt (2005-2007), Institut de Linguistique française, interface lexique/grammaire en analyse syntaxique, 10 partenaires



# Évaluation scientifique

- Comités de lecture de 4 revues dont 3 de linguistique
- Comités de programme de 12 colloques internationaux réguliers dont 10 d'informatique
- Expertises pour agences de recherche
- Direction d'une revue de linguistique : *Lingvisticæ Investigationes*
- Organisation d'un colloque international annuel sur le Lexique et la Grammaire

# Projets pour 2009-2012

- Réaliser des applications à l'aide de méthodes hybrides : statistiques et ressources linguistiques construites manuellement
- Comparer les performances obtenues avec celles des méthodes statistiques pures
- Évaluer l'utilisabilité applicative de ressources linguistiques construites manuellement
- Publier l'intégrale du lexique-grammaire du français dont nous disposons : verbes, noms, adjectifs, expressions, avec documentation

# Un exemple d'application

Le super-chunking



# Principe du chunking

- Un chunk (Abney, 1991) est un constituant non récursif simple
- En français, un chunk s'arrête à la tête du constituant
- Exemples:
  - GN -> Det N (*un lac*)
  - GN -> Det A N (*un grand lac*)
  - GN -> Det N A (*un lac admirable*)
  - GN -> Det N Prep N (*un lac d'Amérique*)

} limite=N

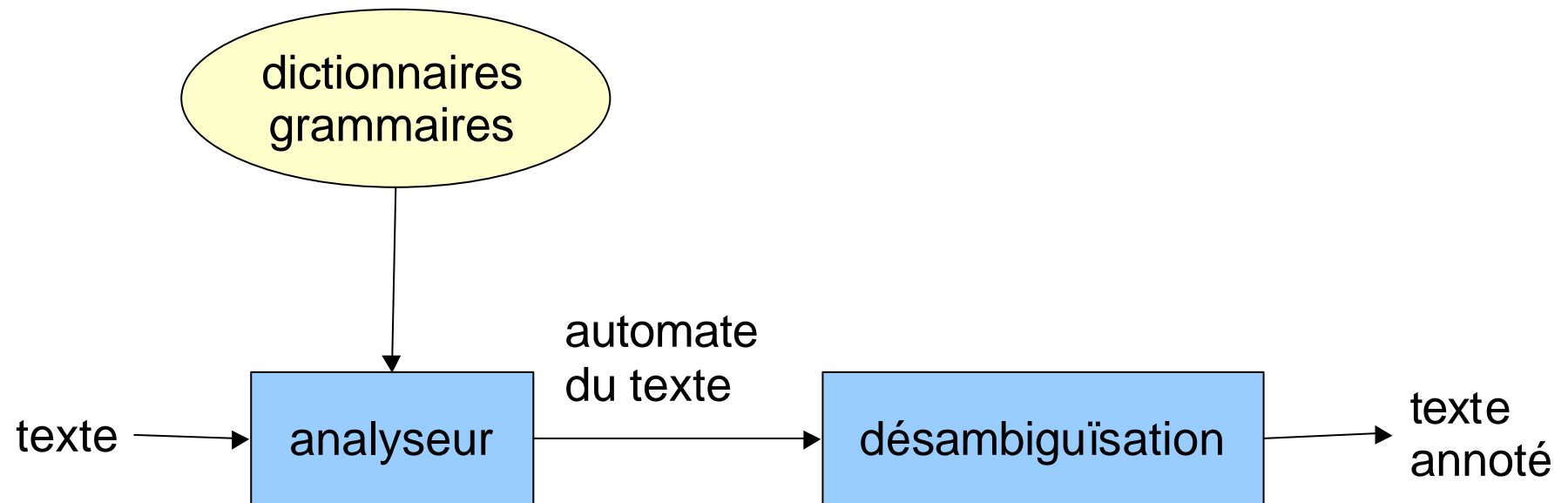
# Exemple d'analyse en chunks

- *La température a beaucoup augmenté à l'intérieur de quelques milliers de maisons*
- [*La température* XN] [*a* XV] [*beaucoup* XADV] [*augmenté* XV] [*à l'intérieur* XP] [*de quelques milliers* XP] [*de maisons* XP]

# Les super-chunks

- Un super-chunk (Blanc *et al.*, 2007) est un chunk intégrant les unités multi-mots appartenant à une catégorie lexicale
- [*La température XN*] [*a beaucoup augmenté XV*] [*à l'intérieur de quelques milliers de maisons XP*]
- Plus grande finesse d'analyse
- Réduction de la complexité de l'analyse syntaxique

# Architecture du super-chunker



# Exemple d'application

<http://fr.news.yahoo.com/2/20090119/tod-un-americain-parcourt-5-000-kilometr-7f81b96.html>

"Voir quelqu'un qui a de tels idéaux, et la question raciale, me donnent l'espoir que nous pouvons faire mieux dans ce pays", explique le jeune homme en souriant.

"Pour la première fois, j'ai le sentiment que ce ne serait pas honteux pour moi d'être patriote, d'être fier de mon pays", poursuit Ryan Bowen.

Son périple, "juste une idée folle qui a marché" selon lui, a été filmé par le documentariste Albert Velazquez, qui l'a suivi un moment dans la voiture balai.

Cerise sur le gâteau, Ryan Bowen a reçu l'un des 240.000 billets lui permettant d'assister de près à la prestation de serment de Barack Obama aux pieds du Capitole, des mains du parlementaire de l'Oregon Earl Blumenauer.

 Envoyer  Partager ▼  Imprimer

---

**PLUS D'INFOS INSOLITE**





# Exemple d'application

{S} {Son périple,périple.XN+3+m+s+head=périple} , " {juste une idée,idée.XN+3+f+s+head=idée} {folle,fou.XA+f+s+head=folle} {qui,qui.pro+rel} {a marché,marcher.XV+ind+s+3} " {selon lui,luire.XV+a+ppast+m+s} , {a été filmé,filmer.XV+ind+s+3+passive} {par le documentariste,documentariste.XP+3+f+s+prep=par+head=documentariste} {Albert Velazquez,Albert Velazquez.XN+3+det0} , {qui,qui.pro+rel} {l' a suivi,suivre.XV+ind+s+3+ppvacc} {un moment,moment.XN+3+m+s+head=moment} {dans la voiture balai,voiture balai.XP+3+f+s+prep=dans+head=voiture\_balai} . {S}

{S} {cerise sur le gâteau,cerise sur le gâteau.XN+3+f+s+det0+head=cerise\_sur\_le\_gâteau} , {Ryan Bowen,Ryan Bowen.XN+3+det0} {a reçu,recevoir.XV+ind+s+3} {l' un des 240 000 billets,billet.XN+3+m+p+head=billets} {lui permettant,permettre.XV+ppres+ppvdat} {d' assister,assister.XVP+inf+prep=de} {de,de.det+ind} {près,près.XADV} {à la prestation de serment,prestation de serment.XP+3+f+s+prep=à+head=prestation\_de\_serment} {de Barack Obama,Barack Obama.XP+3+prep=de} {aux pieds du Capitole,Capitole.XP+3+prep=aux\_pieds\_de} , {des mains du parlementaire,parlementaire.XP+3+f+s+prep=des\_mains\_de+head=parlementaire} {de l' Oregon Earl Blumenauer,l' Oregon Earl Blumenauer.XP+3+prep=de} .