

# Journée des anciens de Linguistiques & Informatique (96-99)

Djamé Seddah  
MCF Université Paris Sorbonne & Alpage

April 5, 2014

## ▶ Formation geeko-littéraire

- *IUT d'informatique d'Orsay*
- *DEUG de Lettres de Modernes*
- **Travail en** comme analyste sémio-linguistique dans un **bureau d'étude sociologique**

## ▶ Une formation initiale en TAL (Paris 7)

- *Licence & Maîtrise de Linguistique et Informatique*
- ⇒ **Inférence de grammaires stochastiques (stage 3 mois, Irisa, Lannion)**
- *DEA de Linguistique Théorique Formelle option TALN*
- ⇒ **Compilation et parsing des Grammaires d'Arbres Adjoints (Stage 6 mois, Inria)**

- ▶ Une Thèse en Informatique
  - *Doctorat d'Informatique, Loria, Langue et Dialogue*
  - ⇒ **Analyse Synt. & Interface syntaxe-sémantique**
- ▶ ATER
  - *Plein temps en Informatique (Univ. Nancy I et Epinal)*
  - *Plein temps en Linguistique (Univ. Paul Verlaine de Metz)*
- ▶ Une expérience de Recherche industrielle
  - *Position en Recherche et Développement (SSII GFI-Luxembourg, div. Intelligence Economique)*
  - ⇒ **Interface LN multilingue entre Ontologie de domaine et Base de Données**
- ▶ Un Post-doctorat à l'étranger
  - *Recherche et Encadrement, (DCU, NCLT, GramLab)*
  - ⇒ **Traduction automatique "syntax based" dans le cadre des LFG**
  - ⇒ **Parsing Deep Syntax out-of-domain**

- ▶ MCF en Informatique à Paris 4, rech. à Alpage
  - ⇒ **Interface Syntaxe Sémantique et TAG**
  - ⇒ **Parsing Robuste, à large couverture, Langues à morpho. riches**
  - ⇒ **Treebanking (hors domaine, bruité, de questions)**
  - ⇒ **Parsing Sémantique**

# Ce que la formation m'a apporté..

- ▶ **metaphysiquement**
  - *L'envie d'être totalement libre de m'intéresser à ce dont j'ai envie*
- ▶ **Plus prosaïquement**
  - *La possibilité de comprendre en temps réel l'évolution des outils qu'on considère comme acquis (correction orthographique, moteurs de recherche, traduction automatique, Siri)*
- ▶ **Soyons terre à terre**
  - *Les interactions entre les différents niveaux d'analyse*
  - *la possibilité de passer d'une suite d'octets à un graphe complexe interfacable avec n'importe quelle ontologie*
  - *le cours de langage formel et de parsing de M. Cori, interfacier ELisa et un parser CFG en Lisp avec Laurence, le Prolog impossible avec Pascal, les Tags avec Anne,..*

# Ce que j'avais délibérément choisi d'ignorer (ou d'oublier)

- ▶ la phonétique (sans pitié, ni remords)
- ▶ la logique (à mon grand regret quand j'ai du en faire plein en thèse)
- ▶ l'épistémologie (à ma décharge: locaux non chauffés et insalubres)
- ▶ tout ce qui relevait de l'énonciation à la Culioli

## Ce qui m'a le plus manqué..

- ▶ la dimension ontologique, graphes conceptuels à la Sowa (1984 pourtant..)
- ▶ le raisonnement type systèmes experts, franchement pas assez d'IA dans le cursus (réseau de neurone, arbre de décision,etc.)
- ▶ les PCFG, la dimension statistiques du parsing (Charniak 97, Collins 99)
- ▶ une méthodologie de validation expérimentale (on ne voyait pas ça du tout)
- ▶ Des maths (et de la vraie logique appliquée type Montague pour les nuls)
- ▶ de la sémantique lexicale

# Mes conseils...

- ▶ Vous trouver un mentor, quelqu'un qui vous donnera envie d'aller lire des papiers abscons à 2H du mat
- ▶ Trouver un problème, un phénomène qui vous obsède et que personne ne semble prêt à résoudre
- ▶ Travailler en binôme, en petit groupe, avec qui vous pouvez parler science, langue, algorithmes à la terrasse d'un café sans trouver ça incongru
- ▶ Être absolument "fluent" dans deux langages de programmation: un pour faire des tests, des prototypes rapidement et un autre pour développer sérieusement
- ▶ Maîtriser un des domaines de la linguistique: état de l'art, positions théoriques divergentes, application au TAL
- ▶ Commencer à écrire une grammaire, un treebank ou des données d'apprentissage à la main: rien de plus formateur
- ▶ Faire des stages intéressants, motivants. Pas simplement aller travailler dans une boîte à stagiaire qui se moquera de savoir si vous avez appris des choses (J'ai eu bcp de chance, merci