

Gérard DUPONT
Groupe 1
Dernière mise à jour : 26/12/10

La Linguistique Informatique

Public visé : Néophyte
Type de Document : Vulgarisation

Résumé

Cette étude présente le métier de linguiste informaticien. Après avoir décrit le domaine du Traitement Automatique des Langues à travers ces principaux objets d'études et les métiers qui y sont rattachés (partie I et II), nous entrerons dans le vif du sujet et évoquerons la profession ainsi que ses spécificités (partie III). Enfin nous nous intéresserons au marché de l'emploi et nous comparerons les exigences des employeurs et la qualité des formations (partie IV et V).

Mots Clefs

Traitement Automatique des Langues

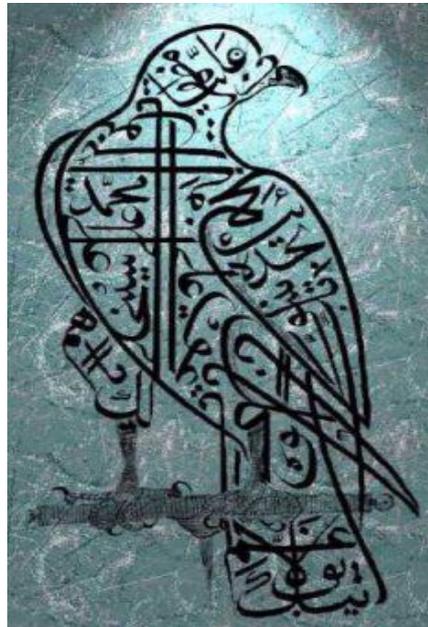
Langage Naturel

Formalisme Linguistique

Sommaire

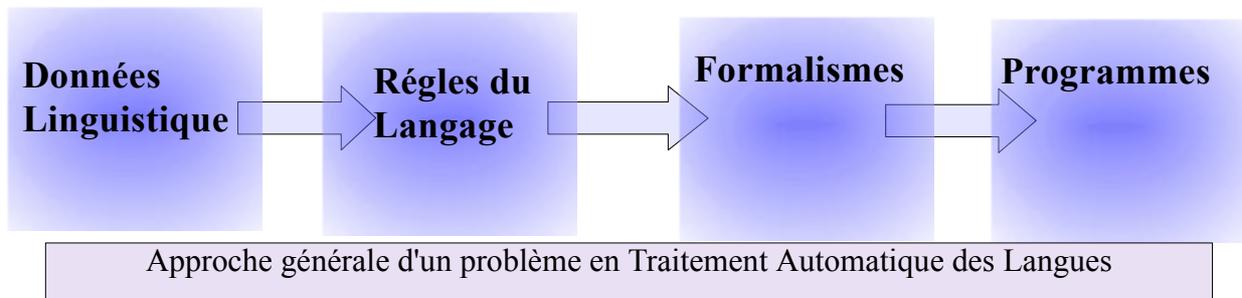
Partie I : Le traitement automatique du langage naturel	5
Partie II : Les métiers liés au traitement automatique des langues.....	12
Partie III : Linguiste informaticien.....	16
Partie IV : Le marché de l'emploi.....	20
Partie V : Les différentes filières et parcours de formation.....	34
Partie VI : Conclusion.....	40
Partie VII : Annexes.....	42
Glossaire.....	62
Index Lexical.....	64

Partie I : Le traitement automatique du langage naturel



I.1 Introduction

Que nous considérons les langues comme des entités singulières ou que nous nous intéressions au phénomène du langage dans sa globalité, nous constatons la présence de règles (conjugaison des verbes, place du sujet dans la phrase, accord des adjectifs etc...) réagissant la communication humaine. Ainsi rien d'étonnant à ce que informatique et linguistique se soient rapprochées pour donner naissance à une nouvelle discipline : le Traitement Automatique des Langues¹.



Aujourd'hui comme hier les défis ne manquent pas : du vieux rêve de la machine parlante aux logiciels de traduction automatisée, les idées d'applications ne font guère défaut. Cependant qu'en est-il des productions réelles en Traitement Automatique du Langage ?

Afin d'apporter une réponse à cette question nous explorerons cinq thématiques phares de cette discipline :

- x le travail sur synthèse et la reconnaissance vocale
- x la recherche de données

¹ Nommé aussi traitement automatique du langage naturel

Le traitement automatique du langage naturel

- x la génération automatique de texte
- x les applications de correction orthographique et grammaticale
- x la traduction de texte d'une langue vers l'autre.

I.2 Traduction automatique et traduction assistée [14]

Champ de recherche emblématique, s'il en est un, la traduction automatique affronte de nombreuses difficultés typiques du traitement automatique des langues.

Tout d'abord se pose le problème de retranscrire les règles grammaticales. Car s'il existe des régularités au sein d'une langue, des exceptions viennent souvent compliquer le problème. Mais aussi complexe que soit cette tâche, le véritable défi se situe au niveau du vocabulaire.

Ainsi, fournir un dictionnaire à l'ordinateur pour qu'il y cherche le mot adéquat ne saurait suffire. Lorsque nous nous exprimons, nous ne nous contentons pas de comprendre un mot grâce à sa définition. Le contexte de la phrase, comme celui de la discussion nous y aide. Et pour cause, le problème des homonymes constitue un des principaux écueils de la traduction automatique.

Ajoutons qu'il ne faudrait pas non plus espérer résoudre par l'informatique les problèmes spécifiques à la traduction - à savoir que chaque langue possède ses propres concepts, intraduisibles dans d'autres langages (par exemple, le « Heimat » allemand qui ne trouve pas d'équivalent en français) et que chaque texte a ses contraintes incompatibles (dans le Roi des Aulnes de Goëthe le traducteur ne peut conserver les rimes en plus de choisir des termes français proches de la signification

allemande et de garder le rythme général du poème qui rappelle le galop d'un cheval). Dès lors la traduction devient affaire de décision, complexifiant encore la réalisation d'une application de traduction automatique.

Malgré ces difficultés et les échecs rencontrés, la traduction automatique reste encore présente sur la scène informatique, puisque très appréciée du grand public à qui elle facilite la navigation internet sur des pages étrangères. En effet, des services de traductions automatiques comme ceux de Google ou de Yahoo, offrent une compréhension superficielle du texte alliée à la rapidité et la simplicité d'utilisation.

Enfin, la traduction assistée par ordinateur (T.A.O.), moins ambitieuse que la traduction automatique, s'avère elle aussi prometteuse. De la mémoire de traduction aux dictionnaires numériques (bien plus simple d'utilisation que leurs homologues en papier), l'informatique a transformé la traduction et apporte aujourd'hui une aide considérable aux traducteurs.

I.3 Correction orthographique et grammaticale [14]:

Au début cantonnée aux logiciels de traitements de texte, la correction orthographique et grammaticale s'invite désormais sur les forums, les gestionnaires de courriel et les navigateurs web, où elle s'effectue désormais en temps réel, c'est à dire au fur et à mesure que l'utilisateur saisit son texte, lui signalant d'éventuelles erreurs et lui proposant un certains nombres de corrections.

Contrairement à ce qui est communément admis, un logiciel de correction orthographique ne s'avère pas moins compliqué à mettre en œuvre que son homologue grammatical. En effet, l'anglais mis à part, la plupart des langues adaptent

Le traitement automatique du langage naturel

l'orthographe d'un mot à celle des termes voisins.

Exemple 1 : en français, « le » devient « l' » face à un nom commençant par un voyelle.

Exemple 2 : en allemand, des mots composés sont formés sans qu'un caractère tel que le « - » sépare distinctement les deux termes fusionnés.

Actuellement les recherches se focalisent sur le développement d'algorithmes capables de reconnaître un mot mal-écrit, même lorsque ce dernier figure dans le dictionnaire. Ainsi peut-on espérer qu'à l'avenir Cendrillon ne souffrira plus dans ces chaussures de verre et retrouvera ses confortables pantoufles en vairs. Actuellement l'algorithme le plus efficace a été mis au point par Andrew Golding et Dan Roth. Ce dernier est capable de reconnaître environ 96% des erreurs liées au contexte, en plus des détections de non-mots.

I.4 La génération automatique de texte (GAT) [14]

Autre fer de lance du Traitement Automatique des Langues, la génération automatique de textes se distingue des autres disciplines dans la mesure où elle n'a pas besoin de saisir le sens d'un texte source, mais au contraire à fournir une forme linguistique à partir de données brutes (dictionnaire, table de conjugaisons etc...).

Elle s'articule autour de quatre étapes :

x Quoi dire ?

Exemple : souhaiter la bienvenue à l'utilisateur

x Comment le dire ?

Exemple : registre de langue parlée

- ✗ Quand dire quoi ?

Exemple : lors du démarrage de l'application

- ✗ Le dire

Exemple : Bonjour maître

Au delà des contraintes lexicales ou grammaticales, il faut donc ajouter celle la prise en compte du contexte qui jouera un rôle déterminant dans la génération automatique d'énoncés.

I.5 Recherche de données [14]

La recherche de données, autre grande thématique du Traitement Automatique des langue, n'est rien d'autre qu'une spécialisation de la fouille de données et rejoint ainsi le domaine de l'intelligence artificielle.

Le processus se scinde généralement en deux étapes :

- ✗ D'abord une première analyse qui consiste à reconnaître les mots, les phrases, leurs rôles grammaticaux, leurs relations et leur sens.

- ✗ Ensuite une interprétation de cette analyse qui permet de sélectionner un texte (mot, phrase etc..) parmi d'autres.

Actuellement deux critères de sélection prédominent : celui de similarité (cibler les textes qui ressemblent le plus à l'énoncé source) et celui de la nouveauté (découvrir des relations non explicites ou indirectes entre deux éléments éloignés d'un texte).

Si les moteurs de recherche sont une application évidente des travaux de cette

Le traitement automatique du langage naturel

discipline, nous ne saurions oublier qu'elle prouve son utilité dès lors que l'utilisateur se trouve face à une masse conséquente de documents. En ce sens la recherche de données rejoint un des enjeux majeurs de l'informatique : celui du traitement de corpus d'informations toujours plus considérables.

I.6 Reconnaissance et synthèse vocale [14]

Au delà de l'attrait ludique que ces deux domaines comportent -faire parler une machine et lui répondre– il est bon de rappeler leur importance dans l'aide aux personnes handicapées. Ainsi la synthèse vocale permet-elle aux aveugles de se rapprocher d'une lecture « normale », en réglant le rythme de la voix de synthèse sur celui de leur écoute (souvent bien plus rapide qu'on ne le croirait) et en retrouvant un timbre de voix dénudé, qui ne joue et ne réinterprète pas le texte qu'il lit.

Rien d'étonnant donc à ce que ces disciplines fassent l'objet de nombreuses recherches. L'une comme l'autre reposent pour une part importante sur l'étude des phonèmes, mais tandis que l'enjeu de la synthèse vocale est plutôt de produire un texte audible à partir de caractères graphiques ; la reconnaissance vocale se heurte d'avantage à la difficulté de reconnaître n'importe quel type de voix.

I.7 Conclusion

Cette présentation du Traitement Automatique des Langues et des domaines de recherche qui lui sont associés ne saurait en aucun cas prétendre à l'exhaustivité. Nous espérons cependant avoir souligné quelques uns des enjeux majeurs de cette discipline et avoir permis au lecteur de mieux comprendre dans quel domaine et face à quels défis évolue le linguiste informaticien.

Partie II : Les métiers liés au traitement automatique des langues



II.1 Introduction

Ce bref descriptif du Traitement Automatique des Langues devrait vous en avoir convaincu : les domaines d'applications pour le traitement automatique des langues ne manquent pas. Aussi, si la recherche et l'innovation occupent une place importante dans cette discipline, il n'est pas surprenant que des ingénieurs spécialisés en traitement automatique du langage naturel soient également présents à de nombreux niveaux de la conception des programmes. Néanmoins, peu de métiers spécifiques au Traitement Automatique des Langues existent.

II.2 Des Ingénieurs présents à tous les niveaux de l'industrie de la langue [2]

Observons la liste des métiers disponible sur site internet de l'APIL (Association des Professionnels du Traitement de la Langue). Deux métiers attirent particulièrement notre attention.

Premièrement celui d'ingénieur support linguiste, chargé d'assurer « les différentes tâches nécessaires à l'accomplissement d'un projet client. »² Concrètement il s'agit d'exécuter du mieux possible chacune des étapes du projet, en respectant les délais, et d'en assurer ensuite la maintenance ainsi que la correction des erreurs éventuelles.

Deuxièmement la profession d'ingénieur qualité linguiste, à qui incombe le rôle de mettre en place des plans de test des logiciels de Traitement Automatique des Langues afin d'en évaluer les qualités linguistiques et de les comparer aux attentes

2 Site web de l'[APIL](#)

Les métiers liés au traitement automatique des langues

spécifiées par le client.

A priori ces deux métiers ne sont pas spécifiques au Traitement Automatique du Langage. Au contraire nous nous attendons à les rencontrer chez tous les éditeurs ou intégrateurs de logiciels, quel que soit le domaine que touchent ces programmes.

Ainsi constatons-nous une première catégorie de professions liées au Traitement Automatique des Langues : celles des métiers de l'informatique au sein de l'industrie des langues, principalement pour des postes de supervision et d'évaluation des applications. En effet dans le secteur des langues de simples connaissances en informatique ne sauraient suffire : les ingénieurs doivent avoir une bonne connaissance en linguistique s'ils veulent pouvoir répondre aux attentes de leurs employeurs.

II.3 Ergolinguiste [2]

Toutefois, sur le site de l'APIL, nous relevons au moins un métier qui n'existe que dans le domaine du T.A.L : celui d'Ergolinguiste. Ce professionnel est chargé d'adapter des applications de reconnaissance vocale au langage humain et à ses spécificités.

La profession n'est cependant pas très répandue, puisque seulement une dizaine de spécialistes œuvrent dans ce domaine.

II.4 Un important domaine de recherche

Enfin, si les applications en Traitement Automatique des Langues ne font pas défaut, nul n'est besoin de rappeler combien cette discipline appelle l'innovation. Les

La Linguistique Informatique

métiers de la recherche (privée comme publique) occupent donc une part importante des débouchés en Traitement Automatique des Langues.

Dans cette catégorie nous retrouvons le métier de chef de projet ou encore, celle qui va désormais nous intéresser : la profession de linguiste informaticien.

Partie III : Linguiste informaticien



L'APIL définit le linguiste informaticien comme une personne qui « assure la conception, le développement et la maintenance d'outils informatiques ayant pour objet des langues naturelles ». Vaste programme. Nous en retiendrons surtout la double spécialisation : en informatique et en linguistique.

III.1 Évolution du métier au cours des dernières années [4]

A l'origine de la profession de linguiste informaticien il y a un échec : celui des américains à mettre en place une application de traduction automatique du russe, durant la guerre froide. Alors que seuls des mathématiciens avaient été recrutés, l'idée s'impose qu'il faudra également s'attacher les services de spécialistes de la langue. La linguistique et l'informatique sellent ainsi leur première rencontre.

Durant les années soixante, c'est la révolution de Noam Chomsky et de sa grammaire généralisée. L'idée qu'il existe des principes communs à toutes les langues va s'imposer et transformer la linguistique comme l'informatique, permettant par exemple la conception de compilateur.

Actuellement cette manipulation de grammaire tend à disparaître au profit d'études statistiques. La discipline quitte sa sphère purement académique. Il est moins important de connaître le modèle théorique sur lequel reposent les outils que les industriels mettent en œuvre. Ainsi les linguistes informaticiens sont-ils davantage recrutés pour assurer la vérification des applications développées et en mesurer les performances.

III.2 Des compétences de linguistes [2]

Le linguiste informaticien doit d'abord se faire linguiste. Au programme une bonne connaissance des langues étrangères ou du moins de leurs fonctionnements puisque la maîtrise complète n'est, elle, pas nécessaire. L'anglais devra être parfaitement parlé avec si possible une, voire plusieurs, langues orientales.

Phonologie, formalismes grammaticaux, lexicologie, sémantique : toutes ces disciplines font également partie du bagage du linguiste informaticien. En effet il s'agit pour lui de savoir expliciter les règles de la langue afin de les représenter dans des formalismes opératoires.

III.3 ... et d'informaticiens [2]

C'est à ce niveau qu'entre en jeu l'informatique puisqu'à partir de cette description formelle le linguiste informaticien réalise une application concrète.

Le linguiste informaticien doit savoir comment développer et tester des programmes informatiques, comment mettre en place des algorithmes, implémenter le code, mettre en places des modules de tests, corriger des bugs, écrire un manuel etc... et à cette fin connaître plusieurs langages de programmation (le Java, le Perl et le Prolog principalement), ainsi que les différents systèmes d'exploitations.

Enfin, le linguiste informaticien doit être ouvert aux nouvelles technologies.

III.4 Formations et débouchés [2][4]

Pour ce faire plusieurs formations spécialisées existent, notamment en Traitement Automatique des Langues. Cependant il est aussi possible de se former d'abord dans l'un des deux domaines et de compléter par la suite ses connaissances

dans le second.

Quant aux marchés de l'emploi il concerne principalement le travail dans la recherche (par exemple dans de petites sociétés spécialisées à forte innovation) et débouche sur un salaire initial d'environ 1800 euros par mois³.

III.5 Conclusion

Le linguiste informaticien est donc avant tout un spécialiste qui travaille au croisement de deux domaines. Dès lors se pose la question des offres d'emplois possible pour qui brigue un pareil diplôme. N'y a t-il que quelques postes réservés à une élite ? Ou bien l'offre permet t-elle à un jeune diplômé d'espérer trouver rapidement un emploi stable, avec une rémunération descente ?

3 Chiffre de Ludovic Tanguy

Partie IV : Le marché de l'emploi



IV.1 Introduction

Dans cette présentation des offres d'emplois disponibles pour un diplômé en linguistique informatique, nous procéderons en deux temps. D'abord nous proposerons un rapide aperçu des entreprises et laboratoires traitant de problèmes du langage, et donc, en ce sens, susceptibles d'embaucher des linguistes informaticiens. Ensuite nous étudierons les offres de recrutement et en dégagerons quelques constats.

IV.2 Exemples d'entreprises dans l'industrie du langage

Selon le site internet de l'APIL il existe en France un peu plus d'une soixantaine d'entreprises dont le domaine d'activité concerne les problèmes langagiers. De la conception de dictionnaire électroniques, au text-mining en passant par la reconnaissance de caractères, nombreuses sont les disciplines qu'elles couvrent. Il peut même sembler difficile de s'y retrouver. Aussi, les exemples qui suivent n'ont-ils pas la prétention de présenter les entreprises les plus influentes, mais s'offrent plutôt comme un aperçu du milieu industriel orienté dans les langues.

IV.2.a Syllabs [11]

Syllabs, comme les consonances de son nom le laissent deviner, est une entreprise spécialisée en analyse sémantique et création automatique de textes. Localisée à Paris, elle propose divers services liés à l'analyse de données textuelles sur le Web. Par exemple : la récupération de pages pertinentes, l'extraction et la catégorisation des informations clef etc.. Ces compétences sont plutôt proposées à des sites marchands, soucieux de leur visibilité sur le net, ainsi que de l'environnement

Le marché de l'emploi

qu'ils offrent à l'utilisateur.

De plus, Syllabs est l'exemple type d'une société misant fortement sur l'innovation, puisque sur sa page internet elle met en avant sa participation à de nombreux projets de recherche, tant dans le cadre des programmes de l'ANR (l'Agence Nationale française de la Recherche) qu'à l'échelle européenne.

IV.2.b Synapse Développement [12]

Société toulousaine d'édition de logiciels, créée en 1994, Synapse Développement se donne pour vocation « le développement d'applications intégrant les techniques de la linguistique et de l'intelligence artificielle appliquées aux domaines de traitement de la langue »⁴ dont la correction orthographique, syntaxique, l'analyse de la langue, la traduction, le traitement automatique du langage naturel sont autant d'exemples.

Parmi les atouts que l'entreprise choisit de mettre en avant, il est intéressant de noter le plurilinguisme de ces employés, venant de pays francophones, germanophones, anglophones et hispanophones ainsi que les deux domaines de compétences que ces équipes disent couvrir : celui de l'informatique et celui de la linguistique. Nous sommes donc clairement devant une société susceptible d'embaucher des linguistes informaticiens.

IV.2.c Auralog [3]

Auralog se présente comme « le premier éditeur au monde à avoir appliqué la technologie de la reconnaissance vocale à l'apprentissage des langues étrangères »⁵. Cette société de renommée mondiale, notamment avec la collection TELL ME MORE, développe toute une gamme de produits pédagogiques et s'adresse tant aux

4 Site web de [Synapse Développement](#)

5 Site web d'[Auralog](#)

La Linguistique Informatique

professionnels qu'aux milieux éducatifs ou aux particuliers. A noter l'intégration de synthèse et reconnaissance vocale dans les logiciels proposés qui permet d'obtenir une évaluation automatique de la prononciation et ouvre ainsi la possibilité de travailler compréhension et expression orale en plus de la grammaire et du vocabulaire « écrit ».

Par ailleurs Auralog entretient une équipe de Recherche et de Développement, composée d'environ quatre vingt ingénieurs et linguistes, afin que les contenus des logiciels évoluent en permanence.

IV.2.d Linguattec [5]

Leader allemand des logiciels d'ingénierie linguistique pour la bureautique, ayant reçu trois fois le Prix Européen des Technologies de l'Information, Linguattec regroupe des experts en technologies linguistiques.

La société travaille autour de trois champs d'activité, qui structurent la conception de ses logiciels :



IV.3 Informatique et linguistique : quelques laboratoires

Il serait cependant dommage, pour un métier comme la linguistique

Le marché de l'emploi

informatique, de se contenter de présenter des entreprises privées. En effet la recherche publique occupe également une place importante quant aux places disponibles.

Une fois de plus, les laboratoires présentés si dessous n'y sont qu'à titre d'exemple et l'énumération n'a pas prétention à l'exhaustivité.

IV.3.a ALPAGE [1]

Le laboratoire ALPAGE (Analyse Linguistique Profonde A Grande Echelle), crée en 2007 et rattaché à l'Université Paris 7, effectue des travaux de recherche dans tout le domaine du Traitement Automatique des Langues.

Cette UMR (unité mixte de recherche) se recentre particulièrement autour de l'analyse syntaxique automatique, le développement de ressources lexicales et le traitement du discours pour la langue française. A la croisée de l'informatique, de la linguistique, des mathématiques probabilistes et de la cognition, ses projets actuels sont la compréhension automatique et la génération automatique en langue française principalement, mais aussi sur d'autres langues, dont l'anglais.

IV.3.b TALEP [13]

Acronyme de Traitement Automatique du Langage Ecrit et Parlé, TALEP désigne l'équipe créée le 1er janvier 2008. Elle résulte de la fusion de deux équipes : l'une orientée sur la compréhension automatique du langage naturel, l'autre sur la linguistique Informatisée.

Basée à Marseille, l'équipe TALEP réalise des travaux de linguistiques fondamentales et de traitements automatiques non triviaux de la langue. Selon son site web « elle constitue et diffuse des ressources linguistiques électroniques (corpus,

lexiques, grammaires, etc.) dont elle étend régulièrement le contenu »⁶.

Parmi les principaux domaines étudiés par l'équipe TALEP se trouvent :

- x la nature de la langue : écrit et oral
- x le style de la langue (normé et déviant - spontané et préparé/planifié - formel et informel)
- x la couverture linguistique (langue générale et sous-langages)
- x les modèles des connaissances sur la langue mis en œuvre (symbolique, numérique et mixte)
- x les méthodes employées pour l'acquisition des connaissances sur la langue (introspective, semi-automatique, apprentissage automatique)
- x les niveaux de compréhension : compréhension complexe et compréhension partielle
- x les processus, algorithmes et architectures : analyse et génération
- x l'évaluation des processus de traitement automatique de la langue
- x les types d'applications du traitement automatique : constitution automatisée de ressources linguistiques ; dialogue homme-machine en langage naturel
- x l'apprentissage de langues
- x aide à la traduction automatique, indexation

6 Site web de [T.A.L.E.P](#)

Le marché de l'emploi

- x extraction et recherche d'informations
- x l'aide aux handicaps

Et cela sur de nombreuses langues : français, anglais, japonais etc..

IV.3.c L'équipe de T.A.L de l'université du Mirail à Toulouse **[4]**

Les recherches de cette équipe se concentrent principalement sur l'utilisation, l'adaptation et le développement d'outils informatiques permettant un travail sur les données linguistiques. Ces travaux sont axés sur l'analyse de données complexes, présentes au sein de corpus de grande taille. Ces principales thématiques de recherche sont :

- x l'étude de la structure du lexique à travers des graphes issus de corpus et de dictionnaires
- x l'acquisition d'informations lexico-syntaxiques et de ressources lexicales à partir de corpus
- x l'étude de l'organisation discursive

IV.4 Deux exemples typiques d'offres d'emploi

Afin de terminer cette présentation, il est intéressant de consulter quelques offres d'emplois. En voici deux, à titre d'exemples :

IV.4.a CDD pour le Centre Commun de Recherche de la Commission Européenne.

Cette annonce étant en anglais, une traduction en français est proposée.

In conjunction with the European Commission Joint Research Center, ELDA offers a 6-month position to produce an updated version of the sentence aligned multilingual parallel corpus JRC-Acqui

En union avec le Centre Commun de Recherche de la Commission Européenne, l'ELDA offre un poste pour une durée de six mois afin de réaliser une application de mise à jour pour l'alignement sous phrastique multilingue pour le CCR-Acqui.

Purpose of the work / Tasks :	Travail/tâches à accomplir
<ul style="list-style-type: none">x Download multilingual EU documentation from a server via a dedicated	<ul style="list-style-type: none">x <i>Télécharger la documentation européenne multilingue via un serveur dédié.</i>
<ul style="list-style-type: none">x Java application	<ul style="list-style-type: none">x <i>Application Java</i>
<ul style="list-style-type: none">x Convert all documents to a standardised XML format	<ul style="list-style-type: none">x <i>Convertir tous les documents dans un format XML standardisé</i>
<ul style="list-style-type: none">x Clean and pre-process the data by identifying specific text parts	<ul style="list-style-type: none">x <i>Nettoyer et préparer les données en identifiant des parties spécifiques du texte.</i>
<ul style="list-style-type: none">x Such as document footers, lists of addresses and annexes	<ul style="list-style-type: none">x <i>Recherche de pieds de page, de liste d'adresses et d'annexes</i>
<ul style="list-style-type: none">x Possibly: run off-the-shelf tools to sentence align the documents	<ul style="list-style-type: none">x <i>Éventuellement : utiliser les outils dédiés d'alignement des phrases</i>

<ul style="list-style-type: none">✗ Carry out consistency checking of the data✗ Produce statistics on the data✗ Prepare the data for distribution✗ Various Perl scripts to produce the first version of the corpus exist and should be reused	<p><i>sur les documents</i></p> <ul style="list-style-type: none">✗ <i>Accomplir une vérification suivie des données</i>✗ <i>Préparer les données pour la distribution</i>✗ <i>Différents Scripts en Perl pour réaliser la première version pour le suivi et l'éventuelle réutilisation du corpus.</i>
<p>Profile and required skills:</p> <ul style="list-style-type: none">✗ Degree or MSc in computer science, computational linguistics, natural language processing or similar fields✗ Good knowledge of Perl to read and change existing data processing scripts.✗ Java and SQL, to use the application accessing the EU's document database.✗ XML and XSLT	<p><i>Profil et compétences requis :</i></p> <ul style="list-style-type: none">✗ <i>Licence ou Master en informatique et linguistique-informatique, traitement automatique du langage ou domaine similaire</i>✗ <i>Bonne connaissance de Perl pour écrire et changer des scripts de suivie de données.</i>✗ <i>Connaissance de Java et SQL pour utiliser l'application d'accès à la base de données des documents de l'union européenne.</i>✗ <i>XML et XSLT</i>

x Proficiency in English	x <i>maîtrise de l'anglais</i>
x At least passive knowledge of several of the 23 official EU languages (see the JRC-Acquis page for details)	x <i>avoir des connaissances de base de plusieurs des vingt trois langages officiels de l'Union Européenne (voir le site du JRC-Acquis pour plus de détails).</i>

Salary: Commensurate with qualifications and experience.

Salaire : Évalué vis à vis des qualifications et de l'expérience.

* * *

Suivent les informations pratiques (lieu de travail, adresse, numéros de téléphone, liens vers davantage d'informations etc...)

IV.4.b CDI pour le CNRS

Le CNRS recrute, pour le laboratoire C LLE-ERSS (UMR 5263, Toulouse), un ingénieur de recherche en traitement et analyse de bases de données. Il s'agit d'un emploi statutaire de la fonction publique, sur concours externe, ouvert aux titulaires d'un doctorat ou d'un diplôme d'ingénieur. Les détails de la procédure et du concours sont accessibles sur le site du C NRS : <http://www.dgdr.cnrs.fr /drhita/concoursita>.

Les dossiers doivent être déposés au CNRS avant le 9 Juillet 2010. La description du poste est reproduite ci-dessous à partir du profil officiel disponible ici.

Le poste est basé à Toulouse (université de Toulouse le Mirail). Les conditions et le traitement sont fixés par le CNRS, les détails sont disponibles ici. Mission : L'ingénieur de recherche soutiendra les axes du laboratoire dans l'analyse,

Le marché de l'emploi

l'exploitation et la valorisation de ses collections de données linguistiques.

Activités :

x Participer aux travaux de recherche de l'équipe sur les données linguistiques (analyse, exploitation, valorisation).

x Utiliser, adapter, voire développer des outils informatiques et statistiques permettant l'analyse de ces données.

x Définir en collaboration avec les chercheurs du laboratoire les protocoles de recueil des données, participer à la mise en forme et à l'analyse des résultats.

x Formaliser, mettre en valeur et diffuser les collections de données et les résultats des analyses.

x Participer à l'élaboration de projets nationaux et internationaux, en en assurant éventuellement la coordination.

x Impulser la participation de l'unité à des réseaux scientifiques internationaux.

Compétences :

x Connaissance approfondie des outils informatiques et statistiques pour l'analyse et la manipulation, des données linguistiques.

x Connaissance dans le domaine de l'informatique : XML et normes de codage, langage de script (Perl ou Python), utilisation d'analyseurs morphologiques et syntaxiques.

x Connaissance dans le domaine des statistiques : description et exploration,

tests d'hypothèses, corrélation-régression, analyses multivariées.

✕ Bonnes connaissances en linguistique, et notamment dans les domaines suivants : phonologie, morphologie, syntaxe, sémantique, discours, dialectologie, romanistique.

✕ Des connaissances en psycholinguistique et/ou en TAL seraient fortement appréciées.

✕ Connaissance de l'éthique et de la réglementation en matière de détention, usage, stockage, et diffusion des données.

Contexte :

L'agent exercera sa mission au sein du laboratoire C LLE (C ognition Langues Langage Ergonomie – UMR 5263), et plus spécialement dans sa composante linguistique, C LLE-ERSS (Equipe de Recherche en Syntaxe et Sémantique).

Le ou la candidat(e) retenu(e) contribuera aux recherches de plusieurs axes, et notamment à des projets transversaux.

IV.5 Synthèse des offres d'emploi

Au regard des différentes annonces, plusieurs points significatifs peuvent être relevés :

✕ D'abord le niveau requis en langues est assez élevé, avec une connaissance indispensable de l'anglais et souvent des compétences demandées dans des langues rares (arabe, chinois, japonais etc...). Le linguiste informaticien doit être un bon polyglotte et un anglophone confirmé.

Le marché de l'emploi

✕ Ensuite, parmi les connaissances en informatiques exigées, l'apparition récurrente des langages Java et Perl, ainsi qu'une certaine familiarité requise avec les bases de données, notamment SQL.

✕ Enfin le nombre d'années d'études minimal est assez élevé : dans le meilleur des cas (et rarement) un bac + 4 est nécessaire, très souvent les postulants sont recrutés après un master, voir un doctorat, toujours dans le domaine du Traitement Automatique des Langues. Pour les postes de haute responsabilité, une expérience professionnelle de deux à trois ans s'avère indispensable.

D'autres éléments, peuvent bien-sûr intervenir : certains postes nécessiteront des compétences particulièrement poussée en linguistique, le nombre de langages de programmation connu peut augmenter. Enfin, il est intéressant de relever que souvent, à une offre d'emploi correspond certaines langues à connaître. Ainsi tel poste demandera des connaissances en langues orientales, tel autre en arabe ou en espagnol.

Nous soulignerons donc une dernière fois ce point crucial : le linguiste informaticien doit connaître, au moins de manière succincte, un grand nombre de langues. Ce point s'avèrera déterminant quant à la quantité et la qualité des offres d'emplois auxquelles il peut espérer postuler.

Partie V : Les différentes filières et parcours de formation



Or, maintenant que les exigences des employeurs nous paraissent plus claires, il est judicieux de les mettre en regard des formations proposées. Rapidement une carence s'impose : celles des langues étudiées. Très souvent seul l'anglais figure au programme. Pour le reste d'une filière à l'autre les proportions de linguistique et d'informatique peuvent varier.

Trois types de formations existent. Tout d'abord nous nous pencherons sur les masters en Traitement Automatique des Langues. Puis nous étudierons un cas particulier : le parcours linguistique et informatique de l'université Paris 7. Enfin nous conclurons avec une filière davantage orientée vers les sciences cognitives proposée à l'université Stendhal de Grenoble.

V.1 Les Master informatique spécialité T.A.L. [4] [9][7][8]

A Nancy, Lille, Bordeaux, Marseille, Toulouse... : apparus dans les années quatre-vingt dix les masters d'informatique orientés Traitement Automatique des Langues ne manquent pas. La linguistique y est enseignée, ou du moins les domaines de la linguistique directement en rapport avec le T.A.L. L'informatique peut se limiter aux connaissances nécessaires pour une utilisation en Traitement Automatique des Langues (par exemple à Toulouse). Des stages en entreprises ou en laboratoire sont intégrés au cursus, mais il est regrettable que la pratique d'une seconde langue ne soit pas envisagée. Quant à l'anglais, il reste peu présent, n'étant parfois pratiqué qu'un semestre sur les quatre qui constituent le master.

Vision synthétique	
Description Générale	L'objectif de ce type de formation est

	<p>généralement d' amener des étudiants issus de la licence d'Informatique à poursuivre une formation d'informaticien tout en les orientant vers des compétences en Traitement Automatique de la Langue ainsi qu'en intelligence artificielle.</p> <p>Certaines formations sont davantage tournées vers le milieu professionnel, d'autres laissent la possibilité à l'étudiant de s'orienter vers la recherche (par exemple le master de Nancy).</p>
<p>Points intéressants des formations</p>	<p>L'étudiant s'ouvre au champs de la linguistique et de l'étude des langues (lexique, syntaxe, sémantique etc...) en même temps qu'il complète sa formation d'informaticien.</p>
<p>Conditions d'admissions</p>	<p>Entrée sur dossier</p>

V.2 Parcours informatique et linguistique Paris IV Diderot [10]

Pionnier de la linguistique informatique, ce parcours de l'université Paris IV s'impose comme l'une des rares formations dédiée à la profession de linguiste informaticien. Cependant, il est regrettable de constater que l'apprentissage de langues étrangères est très peu mis en relief.

Vision synthétique	
Description Générale	<p>Premier Coursus en linguistique informatique, créé en 1984 , ce parcours propose, dans un département de linguistique spécialisé en linguistique formelle, une formation conséquente informatique. L'étudiant choisit de s'orienter soit vers un parcours professionnel soit vers la recherche.</p>
Points intéressants de la formation	<p>Sur les trois années le parcours est assez bien équilibré entre les deux domaines. Cependant l'enseignement informatique ne devient sérieux qu'à partir du master, la troisième année de licence privilégiant la linguistique.</p> <p>Point négatif, le peu de place qu'occupent les langues, avec une seule langue disponible à l'apprentissage et, ce systématiquement à travers des options.</p>
Conditions d'admissions	<p>L'entrée se fait sur dossier à tous les niveaux. Le cursus est largement ouvert, tant aux étudiants issus de filières littéraires (linguistique, langues, sciences humaines...) qu'aux étudiants issus de</p>

	filières scientifiques (informatique, MASS...).
--	---

V.3 Master de recherche en Sciences Cognitives à Grenoble [6]

Contrairement aux autres formations décrites précédemment, ce master de science cognitive est moins tourné vers les problématiques linguistiques et leurs applications informatiques. Cependant, cette thématique n'en est pas absente et de bonnes connaissances des processus cognitifs peuvent être aussi un atout dans le traitement automatique des langues.

Vision synthétique	
Description Générale	La formation concerne l'étude des processus dans les systèmes intelligents naturels ou artificiels.
Points intéressants de la formation.	UE Linguistique et Philosophie du Langage Option de communication parlée possible
Conditions d'admissions	Admission sur dossier ou validation d'études ou validation d'acquis de l'expérience. La lettre de motivation est un élément important du dossier.

	<p>Un excellent niveau est exigé de la part des postulants qui peuvent venir de diverses filières (informatique, neurosciences, psychologie etc...).</p> <p>Les places disponibles sont limitées à une trentaine.</p>
--	---

Partie VI : Conclusion



S'il ne devait rester qu'un point autour duquel résumer le métier de linguiste informaticien, ce serait évidemment celui de la double compétence attendue. Cette dernière constitue à la fois le principal atout et la difficulté majeure de la profession.

En effet, c'est ce double champs d'activité qui rend le linguiste informaticien indispensable aux industries de la langue, puisqu'il peut superviser l'ensemble du projet sans être déstabilisé par l'un ou l'autre de ses aspects. Par contre obtenir un tel niveau de qualification ne paraît pas aisé, et le problème de l'apprentissage des langues et non de la linguistique pure se pose.

Cependant pour celui qui atteint ce sésame, la récompense est là : toujours à la pointe de l'innovation, reculant les limites des précédentes applications, le métier de linguiste informaticien s'annonce exempt d'ennui.

Partie VII : Annexes

VII.1 Syllabus détaillés des formations

VII.1.a Coursus de Linguistique Informatique à Paris 7

Année	Contenu
L3	<p>Premier Semestre</p> <ul style="list-style-type: none"> * UE1 Linguistique fondamentale 1 (6 ECTS) <ul style="list-style-type: none"> o 49LG1235 Phonétique 1 o 49LG2335 Phonologie 1 o 49LG3935 Morphologie 1 * UE2 Linguistique fondamentale 2 (6 ECTS) <ul style="list-style-type: none"> o 49LG2535 Théorie linguistique et grammaire du français (48h) o 49LG1635 Sémantique 1 * UE3 Option Linguistique informatique (9 ECTS) <ul style="list-style-type: none"> o 49LI0335 Programmation 1 (Langage Java) o 49LI0535 Utilisation d'Unix o 49LI0735 Introduction au TAL * UE4 Option générale (6 ECTS) <ul style="list-style-type: none"> o 49LG4435 Linguistique de corpus o 49LD0135 LANSAD Apprentissage d'une langue étrangère (cours géré par EILA) * UE5 Linguistique Informatique (3 ECTS) <p>Second Semestre</p> <ul style="list-style-type: none"> * UE1 Linguistique fondamentale 3 (6 ECTS) <ul style="list-style-type: none"> o 49LG1236 Phonétique 2 (contenu) o 49LG2336 Phonologie 2 (contenu) o 49LG3936 Morphologie 2 (contenu) * UE2 Linguistique fondamentale 4 (6 ECTS) <ul style="list-style-type: none"> o 49LG2436 Syntaxes génératives (48h, contenu) o 49LG1636 Sémantique 2 (contenu) * UE3 Option Linguistique informatique (9 ECTS) <ul style="list-style-type: none"> o 49LI0336 Programmation 2 (24h, contenu) o 49LI0436 Algorithmique (48h, contenu)

	<ul style="list-style-type: none"> o 49LI0636 Bases formelles du TAL (24h, contenu) o Projet en TAL * UE4 Option générale (3 ECTS) <ul style="list-style-type: none"> o Stage de 30 à 60 heures, à effectuer auprès d'une entreprise, d'une école de langue ou d'un laboratoire de recherche suivant le projet personnel et professionnel de l'étudiant * UE5 Option Linguistique informatique (6 ECTS) <ul style="list-style-type: none"> o 51MT2LO4 Introduction à la logique mathématique (contenu)
M1	<p>Premier Semestre :</p> <ul style="list-style-type: none"> * UE1 Langue (2 ECTS) <ul style="list-style-type: none"> o Anglais pour scientifiques Attention: à partir de 2011-12, il s'agira d'un cours "banalisé" du LANSAD, et non plus d'un cours spécifique * UE2 Algorithmique (5 ECTS) <ul style="list-style-type: none"> Un cours au choix (Le choix dépend de la formation initiale de l'étudiant. Il doit être validé par le responsable.) o Algorithmique Master LI o Algorithmique Master Info SI * UE3 Programmation (8 ECTS) <ul style="list-style-type: none"> Deux cours au choix (Le choix dépend de la formation initiale de l'étudiant. Il doit être validé par le responsable.) Tous ces cours commencent la semaine du 20 septembre o Programmation fonctionnelle (Cours de L3 Info avec TD spécifique) o Programmation orientée objet (Cours de L3 Info avec TD spécifique) o Prolog et programmation par contraintes (Cours de M1 Info) o Autre cours d'informatique du Master 1 Info (Compilation, Algorithmique, Intelligence Artificielle...) * UE4 Linguistique (15 ECTS) <ul style="list-style-type: none"> o Morphologie o Phonologie o Syntaxe o Sémantique <p>Deuxième Semestre :</p> <ul style="list-style-type: none"> * UE5 Linguistique (10 ECTS) <ul style="list-style-type: none"> o Nouvelles théories syntaxiques

	<ul style="list-style-type: none"> o Traitement du signal de parole * UE6 TAL (16 ECTS) <ul style="list-style-type: none"> o Langages formels o Traduction automatique o Méthodes probabilistes pour le TAL o Sémantique computationnelle * UE7 Expérience professionnelle (4 ECTS) <ul style="list-style-type: none"> o Projet en TAL (validé par un mémoire et une soutenance) <ul style="list-style-type: none"> o Stage en entreprise ou en laboratoire (le stage est obligatoire mais ne donne pas lieu à une note)
M2 R	<ul style="list-style-type: none"> * UE1 TAL (12 ECTS) <ul style="list-style-type: none"> o Analyse syntaxique automatique du LN o Analyse sémantique automatique du LN * UE2 Informatique (6 ECTS) <p>Un cours au choix (Le choix dépend de la formation initiale de l'étudiant. Il doit être validé par le responsable.)</p> <ul style="list-style-type: none"> o Prolog et programmation par contraintes (Cours M1 info) o Compilation (Cours M1 Info) o Intelligence artificielle (Cours M1 Info) o Autre cours de Master 2 Informatique ou Sciences Cognitives * UE3 Séminaires (18 ECTS) <ul style="list-style-type: none"> o Obligatoire Traitement automatique du discours (Séminaire de L. Danlos) <ul style="list-style-type: none"> o Au choix <ul style="list-style-type: none"> + Syntaxe formelle et comparée (Séminaire de A. Abeillé) + Modélisation des langues: mots et grammaire (Séminaire de S. Kahane) + Autre séminaire du CogMaster (avec composante syntaxe) o Obligatoire Autre séminaire de Master 2 Linguistique Paris Diderot (ou établissement en convention) * UE4 Langue (2 ECTS) <ul style="list-style-type: none"> o LanSAD * UE6 Mémoire (22 ECTS) <ul style="list-style-type: none"> o Travaux de recherche dans un laboratoire de l'Université Paris Diderot, validé par un mémoire et une soutenance publique.
M2 Pro	<ul style="list-style-type: none"> * UE1 TAL (8 ECTS) <ul style="list-style-type: none"> o Analyse syntaxique automatique du LN o Analyse sémantique automatique du LN

	<ul style="list-style-type: none"> * UE2 Informatique (4 ECTS) <ul style="list-style-type: none"> Un cours au choix (Le choix dépend de la formation initiale de l'étudiant. Il doit être validé par le responsable.) <ul style="list-style-type: none"> o Prolog et programmation par contraintes (Cours M1 info) o Compilation (Cours M1 Info) o Intelligence artificielle (Cours M1 Info) o Autre cours de Master 2 Informatique ou Sciences Cognitives * UE3 Séminaires (12 ECTS) <ul style="list-style-type: none"> o Obligatoire Traitement automatique du discours (Séminaire de L. Danlos) <ul style="list-style-type: none"> o Au choix <ul style="list-style-type: none"> + Syntaxe formelle et comparée (Séminaire de A. Abeillé) + Modélisation des langues: mots et grammaire (Séminaire de S. Kahane) o Obligatoire Traitement automatique de la parole (Séminaire de P. Martin) * UE4 Cours professionnels (16 ECTS) <ul style="list-style-type: none"> o Ontologie et réseaux sémantiques o Programmation en PERL o Terminologie o Dialogue homme-machine o Recherche d'information o XML o Bases de données (SQL) o Fouille de textes * UE5 Langue (2 ECTS) <ul style="list-style-type: none"> o LanSAD * UE6 Stage (18 ECTS) <ul style="list-style-type: none"> o Stage en entreprise d'au moins 6 mois, validé par un mémoire et une soutenance publique
--	--

VII.1.b Master en T.A.L. de Nancy : un des rare Master permettant une formation correcte en Anglais

<p>UEs d'initiation en M1</p>	<ul style="list-style-type: none"> * Initiation au Traitement Automatique des Langues <ul style="list-style-type: none"> Enseignant(s): Perrier, Guy 30 heures, 6 ECTS Ce cours sera donné en Français * Linguistique de Base
-------------------------------	---

	<p>Enseignant(s): Knittel, Marie-Laurence & Namer, Fiametta 30 heures, 5 ECTS Ce cours sera donné en Anglais * Lexiques</p> <p>Enseignant(s): Jacquy, Evelyne & Valette, Mathieu & Buchi, Eva 30 heures, 5 ECTS Ce cours sera donné en Anglais * Mathématiques pour l'Informatique</p> <p>Enseignant(s): Blackburn, Patrick 20 heures, 4 ECTS Ce cours sera donné en Anglais * Programmation pour le Traitement Automatique des Langues</p> <p>Enseignant(s): Perrier, Guy & Amblard, Maxime 30 heures, 6 ECTS Ce cours sera donné en Anglais * Projet Tutoré en Tal</p> <p>Enseignant(s): Perrier, Guy 0 heures, 5 ECTS Ce cours sera donné en Anglais * Technologie du Web</p> <p>Enseignant(s): Belaid, Abdel 40 heures, 6 ECTS Ce cours sera donné en Anglais</p>
<p>UEs d'ossature en M2</p>	<p>* Logique et statistiques pour la modélisation des langues Enseignant(s): Smaili, Kamel & Areces, Carlos 30 heures, 6 ECTS Ce cours sera donné en Anglais * Applications du TAL</p> <p>Enseignant(s): Gardent, Claire & Venant, Fabienne 30 heures, 6 ECTS Ce cours sera donné en Anglais * Outils et Algorithmes pour le TAL</p>

	<p>Enseignant(s): Perrier, Guy & Gaiffe, Bertrand 30 heures, 6 ECTS Ce cours sera donné en Anglais</p>
<p>UEs de différenciation en M2</p>	<p>* Sémantique Computationelle Enseignant(s): de Groote, Philippe 15 heures, 3 ECTS Ce cours sera donné en Anglais * Linguistique de Corpus, Ressources Linguistiques et Normalisation Enseignant(s): Sylvain Pogodalla 15 heures, 3 ECTS Ce cours sera donné en Anglais * Génération Enseignant(s): Gardent, Claire 15 heures, 3 ECTS Ce cours sera donné en Anglais * Formalismes Grammaticaux Enseignant(s): Guillaume, Bruno & Roussanaly, Azim 15 heures, 3 ECTS Ce cours sera donné en Anglais * Lexicologie Enseignant(s): Pierrel, Jean-Marie 15 heures, 3 ECTS Ce cours sera donné en Anglais * Représentation du Discours et du Dialogue Enseignant(s): Blackburn, Patrick 15 heures, 3 ECTS Ce cours sera donné en Anglais * Phonétique Enseignant(s): Bonneau, Anne 15 heures, 3 ECTS Ce cours sera donné en Anglais * Fouille de Données et Web Sémantique Enseignant(s): Napoli, Amedeo & Toussaint, Yannick</p>

	15 heures, 3 ECTS Ce cours sera donné en Anglais
--	---

VII.1.c Master en Science Cognitive à Grenoble

Semestre du M2	Contenu
1	<p>Tronc Commun (18 ECTS)</p> <p>Psychologie Cognitive (3 ECTS, 18h CM)</p> <p>Neurophysiologie, Imagerie cérébrale, Analyse et mise en œuvre expérimentale (6 ECTS, 33h CM 3h TD)</p> <p>Réseaux de Neurones Formels (3 ECTS, 12h CM 6h BE) Détails</p> <p>Intelligence Artificielle, Vie Artificielle et Cognition (3 ECTS, 18h CM)</p> <p>Linguistique et Philosophie du Langage (3 ECTS, 18h CM)</p> <p>MODULES D'OPTION : Choix entre les 2 profils sur 12 ECTS</p> <p>Profil "Cognition Naturelle et Artificielle" : Choix de 12 ECTS</p> <p>Profil CNA : Communication Parlée (3 ECTS, 18h CM) Détails</p> <p>Profil CNA : Signaux, Images et Modèles de Perception Visuelle (3 ECTS, 18h CM)</p> <p>Profil CNA : Modèles de Mémoire et Apprentissage dans les</p>

	<p>Systemes Naturels et Artificiels (3 ECTS, 18h CM)</p> <p>Profil CNA : Neurosciences Cognitives des Compétences Précoces : Développement de la mémoire, de la vision, de la parole, des interactions sociales et multimodales chez les bébés et nourrissons (3 ECTS, 18h CM)</p> <p>Profil CNA & ICH : Cognition Bayésienne : Modèles pour la Perception, l'Apprentissage et l'Action (3 ECTS, 18h CM)</p> <p>Profil "Informatique Centrée Humain" : Choix de 12 ECTS</p> <p>Profil ICH : Méthodes Numériques pour le Traitement de l'Oral et de l'Ecrit (3 ECTS, 15h CM)</p> <p>Profil ICH : Méthodes Expertes pour le Traitement Automatique du Langage Naturel et le Multilinguisme (3 ECTS, 15h CM)</p> <p>Profil ICH : Modèles Cognitifs pour les Environnements Informatiques d'Apprentissage Humain (3 ECTS, 15h CM)</p> <p>Mots clefs : EAIH, apprentissage, modèle d'apprenant, évolution des connaissances.</p> <p>Profil ICH : Ingénierie des Environnements Informatiques d'Apprentissage Humain (3 ECTS, 15h CM)</p> <p>Profil ICH & CNA : Cognition Bayésienne : Modèles pour la Perception, l'Apprentissage et l'Action (3 ECTS, 18h CM)</p>
2	Mémoire de recherche (27 ECTS)

La Linguistique Informatique

	Anglais (3 ECTS, 24h TD) ou Français Langue Etrangère (3 ECTS, formation Université de Grenoble)
--	--

VII.2 Offres d'emplois

VII.2.a Informaticien Linguiste Senior (CDI)

Dans le cadre de notre développement, nous recherchons un: Informaticien Linguiste Senior

Responsable de l'équipe linguistique (CDI) basé à Nantes (44) H/F.

Poste

Au sein de notre équipe développement, vous animez et encadrez l'équipe linguistique qui est en charge des activités suivantes :

Analyse des langues devant être supportées par notre technologie en combinant les approches théoriques avec des approches plus pragmatiques (constitution et analyse de corpus, analyse de l'alphabet et des styles d'écritures manuscrites communément utilisés, etc.).

Participation au processus de collecte d'échantillons d'écriture manuscrite.

Création et maintenance des ressources linguistiques et de leur intégration dans nos moteurs de reconnaissance d'écriture manuscrite.

Création et maintenance des outils internes et à destination des clients permettant de générer et de maintenir ces ressources linguistiques.

Recherche sur les modèles de langages, notamment statistiques, et leurs applications aux interfaces homme-machine.

Vous assurez le management de l'équipe linguistique et des projets qui lui sont confiés. Vous êtes notamment le garant de la qualité des ressources linguistiques mises à disposition des autres équipes :

Mise en place d'indicateurs permettant de mesurer la qualité des ressources par rapport à leur impact sur la reconnaissance d'écriture.

Propositions, en relation étroite avec le Directeur de la R&D, pour améliorer les processus de collecte et de création des ressources linguistiques.

Vous aurez également en charge la direction de l'équipe et des projets qui lui sont confiés.

Profil

De formation supérieure (Master, Ingénieur,...) en informatique complétée par une spécialisation sur le Traitement Automatique du Langage (TAL), possiblement une thèse en « Machine Learning », vous avez une connaissance approfondie des langues asiatiques (Chinois, Japonais, Coréen).

Vous intervenez depuis 5 à 10 ans sur des projets de développement logiciel mettant en œuvre vos compétences à la fois informatiques et linguistiques, ainsi que vos capacités à encadrer une équipe de 5 à 10 personnes.

Dans un contexte international, vous souhaitez acquérir des responsabilités importantes, de l'autonomie et rejoindre une structure en forte croissance qui saura répondre à vos attentes.

Anglais courant impératif, une bonne maîtrise d'au moins une langue asiatique serait un plus important.

VII.2.b CDI chez VISION OBJEC

Dans le cadre de notre développement, nous recherchons un Linguiste – Informaticien Basé Nantes (44) H/F

Au sein de notre équipe développement, vous reportez au Responsable des ressources linguistiques (4 perso)

Vous assurez la conception et le développement d'outils liés au traitement automatique de langage naturels

Votre double compétence vous permettra de :

comprendre et synthétiser les besoins, en liens avec l'équipe recherche

définir les spécificités des outils de traitement du langage

élaborer le cahier des charges et les formalismes opératoires

développer les outils : algorithmes, implémentation de codes, modules de tests...

ainsi que les applications de développement à la création de nouvelles langues

Profil

De formation supérieure (Master...), vous justifiez d'une compréhension des environnements linguistique et traitement automatique du langage (TAL) ainsi que d'une connaissance linguistique approfondie, idéalement la langue Arabe

Expertise C ; C ++ ; Perl...

Dans un contexte international, au sein d'une société en plein développement, vous souhaitez acquérir de l'autonomie et rejoindre une structure en très forte croissance qui saura répondre à vos attentes.

Anglais courant

VII.2.c CDD 7 mois : Linguiste informaticien / Terminologue

LexisNexis en France (600 collaborateurs, 139 M€ de CA), filiale du groupe anglo-néerlandais Reed Elsevier, est un acteur majeur dans les services d'information professionnelle. Ses activités couvrent trois domaines : l'information et l'édition juridiques, la diffusion de la presse et de l'information économique et financière sur Internet, les logiciels professionnels.

L'entreprise s'appuie sur une expertise éditoriale centenaire et sur une technologie de pointe pour apporter au monde du droit et aux professionnels de tous secteurs d'activités une vaste gamme de produits et services réputés : JurisClasseur, Litec, D.O, Bottin Administratif et les services en ligne LexisNexis.

LexisNexis vous propose de rejoindre le service Veille et management de l'information dans le cadre d'un CDD de 7 mois. Vous interviendrez principalement sur l'aide à la construction d'un thésaurus juridique monolingue informatisé qui servira à l'indexation et à la recherche documentaire de contenus juridiques.

COMPÉTENCES ET QUALITÉS :

✕ Maîtrise de l'outil informatique, très bonne orthographe et connaissances en linguistique.

✕ Des connaissances en droit seraient un plus.

✕ Rigoureux (se), organisé(e) et sérieux(se), vous faites preuve d'un bon relationnel.

VII.2.d Poste chez DGA/Ingénierie de Projet

En lien avec les acteurs scientifiques et industriels du traitement automatique du

langage oral et écrit et pour répondre aux besoins des opérationnels de la défense à court comme à long terme, vous serez en charge de concevoir, spécifier, suivre et évaluer des projets technologiques dans le domaine. De manière à mener à bien efficacement ces projets, vous effectuerez également une veille technologique active, des actions de coordination au niveau national et international, et des travaux d'étude et de développement informatique.

Profil :

Une expérience dans le domaine du traitement automatique du langage associée à des compétences en gestion de projet est recherchée.

La maîtrise de l'anglais et des relations internationales sont un plus.

Diplôme grande école ingénieur ou diplôme niveau bac+5 exigé.

VII.2.e Poste pour LIMSI – CNRS

Le LIMSI - CNRS recrute une personne ayant un doctorat ou un diplôme de Master Professionnel en informatique, avec des compétences en Traitement Automatique du Langage, et en espagnol parlé et écrit, pour une période de deux ans, ou un an renouvelable, à partir de novembre

2009.

La personne recrutée travaillera dans le cadre du projet CIDESAL (http://www.msh-m.fr/rubrique.php3?id_rubrique=315), financé par l'Union Européenne et réunissant des équipes de l'Argentine, de l'Uruguay et de la Colombie, et sera responsable de la partie Fouille de texte.

Description

Le but du projet est d'identifier et de localiser des réseaux de diasporas scientifiques (Diaspora Knowledge Networks), à partir de l'exploitation des données du Web. A l'instar de nombreux pays du sud, l'exode des compétences est un problème récurrent dans les politiques de développement de l'Amérique Latine depuis 40 ans. Cet exode soutenu et continu a débouché sur la constitution de diasporas de personnes hautement qualifiées, très présentes sur le Web. Des méthodes de fouille du Web devront être développées pour détecter et décrire ces réseaux.

Parmi les tâches impliquées par le projet figurent la constitution d'un corpus de documents pertinents du Web, pour caractériser la structure et l'organisation d'un réseau de diaspora scientifique, et le développement de méthodes de traitement du langage en espagnol pour identifier le domaine, la nature et les objectifs de l'activité collective, ainsi que le contexte de coopération. Ces tâches requièrent les compétences suivantes : en informatique : le langage C, ou perl (ou autre), Linux, MySQL, en traitement du langage : classification et extraction d'information, en langue : espagnol, anglais

VII.2.f Offre d'emploi en CDD chez Sinequa

Sinequa (www.sinequa.com) est le leader en France du moteur de recherche d'entreprise. Le produit phare de l'entreprise, Sinequa CS, est un moteur de recherche intelligent basé sur une technologie linguistique et sémantique brevetée.

Sinequa propose un CDD dans ses locaux à Paris pour une durée d'environ 5 mois (en raison d'un congé maternité) à partir du 15 janvier 2010. Le travail portera entre autres sur l'extraction d'entités nommées, la phonétisation et l'édition de grammaires locales et de dictionnaires sémantiques et morphosyntaxiques. La date limite pour poser sa candidature est

fixée au 11/11/2009. La rémunération sera en fonction de l'expérience du candidat.

Expérience et compétences demandées

x Profil : Bac +4 ou bac +5 en linguistique-informatique ou ayant une mention
TAL

x Langues : Excellente maîtrise du français et de l'anglais, la maîtrise d'autres langues est un plus

x Domaine d'expérience souhaité : Recherche d'information

VII.2.g a Emploi à mi-temps chez Syllabs

La société Syllabs (www.syllabs.com) est un jeune laboratoire de recherche privé spécialisé dans les domaines de la Gestion de l'Information et du Traitement Automatique des Langues. Syllabs est au coeur de trois activités complémentaires : La Recherche, les Développements Innovants et le Conseil.

Nous recherchons un(e) linguiste informaticien(ne) pour une mission dans le domaine du traitement de la parole. Il s'agit de corriger un alignement son/phonème des phrases en anglais US réalisé automatiquement.

Profil souhaité :

x bac+4 ou bac+5 en linguistique informatique

x bonnes connaissances en traitement de la parole

x maîtrise de la phonétique de l'anglais américain

x locuteur natif de l'anglais

x expérience avec un outil d'alignement type Praat

Durée : 6-8 semaines à partir du 15 juillet à mi-temps.

Lieu de travail : Possibilité de télétravailler, une formation à l'outil à utiliser pour la correction sera assurée dans nos locaux basés à Paris.

Salaire : à négocier selon expérience

VII.2.h CDI pour OWI Technologies

OWI Technologies développe des solutions de Compréhension Automatique du Langage Humain, reposant sur le procédé OWI. Nous recherchons un ingénieur ou docteur en informatique à très fort potentiel, spécialisé dans la reconnaissance vocale. Sur un sujet intellectuellement très motivant, mais particulièrement complexe, vous prendrez en charge la conception et la maintenance d'un ou plusieurs domaines (Compréhension, Mémoire Longue,

Énonciation, Explicitation, Traduction...), et serez tout particulièrement chargé d'explorer et de mettre en oeuvre la technologie OWI dans des environnements de traitement de la voix. Son activité comportera à la fois des tâches de recherche (veille, études, expérimentations) et de développement des solutions mises au point. Ses qualités et sa motivation lui permettront d'accompagner l'évolution très rapide de la société.

Profil :

x Docteur ou ingénieur en informatique

x Solides compétences de développement en C++

x Bonne connaissance de SQL et des bases de données relationnelles

x Bonne connaissance dans l'intercommunication de systèmes et les architectures informatiques distribuées (TCP/IP, Socket, SOAP, Web Services)

Annexes

✕ Idéalement, une première expérience dans le domaine des applications de reconnaissance vocale multi-locuteurs : Expérience pratique de l'intégration de solutions du marché (Nuance, Loquendo, Telisma) / Création et gestion de grammaires dynamiques / Interconnexion au moteur de reconnaissance vocale via protocole normalisé (type MRCP) ou API dédiées

✕ La connaissance du développement d'applications Web (HTML, Java, AJAX) serait un plus

✕ Anglais

✕ La connaissance d'une troisième langue est un plus

Rémunération et intéressement particulièrement motivants.

VII.3 Interview de Ludovic Tanguy et Nabil Hathout, enseignant chercheur et chercheur en linguistique informatique à l'université du Mirail, Toulouse 2

Question 1 : Pouvez vous me présenter les principales évolutions du traitement automatique des langues ?

Ludovic Tanguy : Le Traitement Automatique des Langues est une très vieille idée. La première apparition de la discipline eut lieu dans le cadre de la guerre froide : les américains voulaient traduire automatiquement du russe. A cette époque il n'y avait pas de linguistes informaticiens à proprement parler, juste des mathématiciens. Cette première tentative fut un échec total : les textes ne voulaient rien dire. D'où l'idée qu'il fallait une meilleure connaissance de la langue et l'entrée en jeu des linguistes.

Le deuxième grand apport fut celui de Noam Chomsky, notamment avec son concept de grammaire généralisée, lequel a beaucoup influencé l'informatique (par exemple en permettant la conception de compilateur). Les années soixante furent donc celle du formalisme, et la linguistique informatique suivit peu ou prou les évolutions de la linguistique.

Les master en T.A.L quant à eux sont apparus vers le début des années 90.

Nabil Hathout : Aujourd'hui, la linguistique informatique est beaucoup moins académique. Il y a eu une rupture avec l'apparition du web et des gros corpus (archives). On cherche davantage à faire des choses sur les données sans chercher à savoir ce à quoi ça correspond d'un point de vue théorique. Avant les linguistiques informaticiens manipulaient essentiellement des grammaires. Maintenant la discipline est très liée aux statistiques. Le linguiste a un rôle moindre, tant dans la recherche que dans l'industrie. Il sert surtout à évaluer un système, ainsi que ses performances, ou à travailler sur de petits outils dans de grandes industries.

Question 2 : Pouvez vous me parler de vos différents parcours ? Comment devient-on linguiste informaticien ?

Annexes

L.T. : J'ai d'abord fait une classe préparatoire, puis une école d'ingénieur (TELECOM Bretagne) qui proposait une option Intelligence Artificielle. En dernière année j'ai passé un DEA d'informatique à Rennes, puis j'ai enchaîné avec une thèse en informatique, très orientée vers la linguistique (en fait il y avait plus de linguistique que d'informatique).

N.H.: J'ai validé une maîtrise de mathématique puis refait un L3 et un master en informatique. Ensuite j'ai effectué une thèse d'informatique, le tout à l'Université Paul Sabatier.

Il y a des relations très étroites entre le laboratoire de Traitement Automatique des Langues du Mirail et l'IRIT. Beaucoup de personnes sont à cheval sur les deux laboratoires. Actuellement nous avons trois-quatre membres qui ont passé leurs thèses à l'IRIT .

L.T.: Il est plus courant pour un informaticien de se tourner vers la linguistique que le contraire. Ici nous formons des étudiants en sciences du langage, avec une spécialisation en T.A.L qui apparaît en L3. Nos étudiants commencent à faire de l'informatique en troisième année de licence. Dans les filières d'informatique c'est l'inverse : la linguistique apparaît dès la quatrième voir la cinquième année.

Question 3 : Pourrait-on revenir sur l'importance des langues dans le T.A.L. ? J'ai remarqué que les offres d'emplois demandaient des compétences assez poussées dans ce domaine et que les formations proposées n'étaient souvent pas à la hauteur.

L.T.: C'est effectivement un véritable problème. Pour ma part je maîtrise l'anglais et j'ai des notions d'allemand, d'arabe, et d'esperanto. Il est indispensable de maîtriser parfaitement l'anglais, et c'est un plus que d'avoir des notions dans d'autres langues, d'en connaître le fonctionnement. Après quand une entreprise a besoin d'un expert en langue arabe, par exemple, elle choisit quelqu'un dont c'est la langue natale. Actuellement les langues demandées correspondent généralement soit au pays émergent, soit aux nouveaux pays intégrés dans l'union européenne.

Les élèves en fac de lettres ne sont pas meilleurs en anglais que les élèves de filières scientifiques. Nous essayons de les former à l'expression et la compréhension écrite. Mais ici ils ont la possibilité de découvrir de nombreux langages, notamment des langues un peu rares ou exotiques. L'université propose des cours d'initiation à une quinzaine de langues.

Question 4 : Quels sont les langages informatiques les plus utilisés ?

L.T.: En premier lieu le Perl et le Python, ainsi que le Java lorsque nous avons besoin de réaliser une application. On pourrait éventuellement rajouter le Prolog, mais il tend à disparaître.

N.H.: J'ai fait ma thèse en Prolog avec contraintes.

L.T.: Il y a aussi des langages comme Matlab, le R ou le Ruby. Mais les deux principaux sont véritablement Perl et Python. Actuellement nous essayons de former nos élèves au Java.

Sur la façon de programmer des linguistes informaticiens il faut reconnaître qu'ils codent souvent mal, vite et de manière sale. Nous n'avons pas de formation sur les méthodes de programmation et ce n'est d'ailleurs pas important d'avoir un code propre. Si une entreprise a besoin d'améliorer le code elle fera appel à un informaticien pur.

N.H.: Le X.M.L et XSST sont également importants. Ainsi que la base de données et les statistiques bien sûr.

L.T.: Nous ne nous préoccupons pas d'interfaces graphiques et nos étudiants n'y sont pas formés. Nous faisons surtout de l'informatique d'utilisateur. Mais beaucoup d'étudiants s'orientent vers un vrai master d'informatique ensuite.

Question 5 : Pouvez vous me parlez de votre travail ? Par exemple combien d'heures passez-vous pour l'enseignement ? Quelles sont les évolutions de carrière possible ?

L.T.: Je suis maître de conférence, et donc enseignant-chercheur. J'ai un volume de 200 heures par ans à consacrer aux cours, c'est à dire une moyenne de huit heures par semaine. Mais il y a des semaines où je n'aurais que deux heures de cours et d'autres où j'en aurais une vingtaine. J'ai également en charge une partie de administration, en tant qu'enseignant et en tant que chercheur. Il s'agit par exemple d'organiser ou de participer à des réunions, d'ouvrir de nouveaux cours, de gérer les chargés de cours. Et puis il y a la recherche qui m'occupe le reste du temps. Je dois être à environ cinquante heures par semaine, avec la moitié du temps en recherche.

Nous avons des horaires très élastiques, marqués par de nombreuses contraintes : celles des cours bien-sûr, mais aussi toutes les échéances de nos divers projets avec des entreprises, publications etc... C'est assez difficile de se débloquer du temps pour travailler sur un domaine spécifique, par exemple pour produire une deuxième thèse au bout de 10-15 ans et ainsi de passer à l'échelon

Annexes

supérieur. Et il y a des pics de travail, qui nous occupent des weekends entiers.

N.H.: Je suis chercheur, plus précisément chargé de recherche, et je n'ai donc que quarante heures de cours par an. Et encore, je l'ai voulu. Il n'y a aucune contrainte de donner des cours. Par contre j'ai des obligations de publications beaucoup plus élevées. En moyenne deux fois plus que mon collègue.

L.T.: Ensuite pour l'évolution de carrière c'est essentiellement celle d'un fonctionnaire. Nous franchissons des échelons au fil du temps et nos salaires augmentent. Pour les enseignants chercheur il y a deux stades : d'abord celui du maître de conférence ensuite celui de professeurs des universités.

N.H.: Pour les chercheurs il a également deux niveaux : d'abord le chargé de recherche, ensuite le directeur de recherche.

L.T.: Généralement on reste dans sa voie. Sinon, il est quand même plus facile pour un chercheur de devenir enseignant chercheur que le contraire. Ce n'est pas qu'il y ait plus de contraintes, mais dans la pratique, c'est plus difficile. C'est important pour nous que les chercheurs aient du temps : ça nous permet d'avancer sur pas mal de projets. Il faut savoir que contrairement à la linguistique qui est un domaine où l'on travail seul, la linguistique informatique est avant tout un travail d'équipe nécessitant des compétences très diverses. De plus, le passage d'un échelon à l'autre se fait de manière beaucoup plus rapide en informatique pure, qu'en sciences du langage.

N.H.: Il faut dire aussi que nos chercheurs et enseignants chercheurs sont recrutés beaucoup plus tard : vers trente ans en fait. Cela tient au fait que les thèses sont souvent beaucoup moins financées dans les milieux littéraires.

L.T.: En T.A.L. nous allons mettre en place l'obligation d'avoir un financement pour débiter une thèse. C'est une décision saine car beaucoup de personnes traînent pendant sept voire huit ans une thèse qu'ils ne peuvent terminer, et qui les maintient dans des conditions de vie désastreuses. Il faut savoir qu'en France près d'une thèse sur deux ne se finit pas.

Question 6 : Quels sont les débouchés, autre que dans la recherche, après un master de T.A.L. ?

L.T.: Nous avons un master pro, qui offre de bons débouchés dans deux types de domaines :

- soit chez des fabricants d'outils T.A.L., pour évaluer les outils produits ou concevoir de petits

La Linguistique Informatique

outils

- soit dans de grands groupes (aéronautique, pétrole, voiture etc...) pour travailler sur des problèmes de documentations. C'est d'ailleurs une des orientations de notre master pro

Le fait de faire une thèse n'est toutefois pas handicapant pour travail dans le privé, au contraire. A mon époque, mon diplôme d'ingénieur m'a un peu handicapé pour rejoindre la recherche, mais aujourd'hui je ne crois pas que ce soit le cas.

Question 7 : Pouvez-vous me donner une idée de vos salaires ?

L.T.: Je gagne 2500 euros, avec 12,13 ans d'ancienneté.

N.H.: Je suis un peu plus ancien et gagne 3000 euros.

L.T.: Le salaire est généralement une question d'âge. Aujourd'hui un chercheur ou enseignant chercheur qui débute en linguistique informatique gagne environ 1800 euro

N.H.: Personnellement j'ai commencé avec 9000 Franc, soit 1500 E environ

L.T.: Il y a eu une revalorisation des salaires. A mon époque quand j'ai commencé à être enseignant chercheur je gagnais moins que durant mon post-doc. Contrairement à ce qu'on pourrait croire, nos collègues dans l'industrie ont à peu près les mêmes salaires que nous.

Glossaire

LINGUISTIQUE : Étude du langage humain

SCIENCES COGNITIVES : Sciences qui regroupent un ensemble de disciplines scientifiques dédiées à l'étude et la compréhension des mécanismes de la pensée humaine, animale ou artificielle

TEXT-MINING : ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes produits par des humains pour des humains

TRADUCTION ASSISTÉE : Traduction réalisée par un être humain, à l'aide d'un support informatique qui facilite l'opération sans en être l'intervenant principal

TRADUCTION AUTOMATIQUE: Traduction entièrement réalisée à partir d'un ou de plusieurs programmes informatiques

Index Lexical

Sources

- 1: ALPAGE ~ <http://www.inria.fr/domaines-epi/perception-cognition-interaction/langue-parole-et-audio/alpage>
- 2: APIL ~ <http://apil.asso.fr/>
- 3: Auralog ~ http://fr.tellmemore.com/a_propos_auralog
- 4: L. Tanguy & N. Hathout ~ Université du Mirail
- 5: Linguatec ~ <http://www.linguatec.net/about/profile>
- 6: Master de Sciences Cognitives Grenoble ~ <http://prevert.upmf-grenoble.fr/SpecialiteSC/index.htm>
- 7: Master T.A.L. Bordeaux ~ http://www.u-bordeaux3.fr/fr/formations/offre_de_formation/master/lettres_et_sciences_du_langage/master_recherche_linguistique_et_informatique.html
- 8: Master T.A.L. Marseille ~ http://www.univ-provence.fr/public_html/univ-provence/index.php?pages=article&id=161&numpage=644&reform=213
- 9: Master T.A.L. Nancy ~ <http://tal.loria.fr/talmaster/>
- 10: Parcours Linguistique Informatique Paris 7 ~ <http://www.univ-paris-diderot.fr/sc/site.php?bc=formations&np=PARCOURS?NP=124>
- 11: Syllabs ~ <http://www.syllabs.com>
- 12: Synapse Développement ~ http://www.synapse-fr.com/sub_soc.htm
- 13: TALEP ~ <http://www.lif.univ-mrs.fr/spip.php?article353>
- 14: Wikipedia ~ http://fr.wikipedia.org/wiki/Traitement_automatique_du_langage_naturel

Notions fondamentales

ALPAGE.....	24
Auralog.....	22
Correction orthographique et grammaticale.....	8
Ergolinguiste.....	14
génération automatique de texte.....	9
ingénieur qualité linguiste.....	13
ingénieur support linguiste.....	13
Linguatec.....	23
linguiste informaticien.....	17
Master de recherche en Sciences Cognitives à Grenoble.....	38
Master informatique spécialité T.A.L.....	35
Parcours informatique et linguistique Paris IV Diderot.....	36
Recherche de données.....	10
Reconnaissance et synthèse vocale.....	11
Syllabs.....	21
Synapse Développement.....	22
TALEP.....	24
Traduction automatique et traduction assistée [14].....	7

Table des Matières

Partie I : Le traitement automatique du langage naturel	5
I.1 Introduction.....	6
I.2 Traduction automatique et traduction assistée [14].....	7
I.3 Correction orthographique et grammaticale [14]:.....	8
I.4 La génération automatique de texte (GAT) [14].....	9
I.5 Recherche de données [14].....	10
I.6 Reconnaissance et synthèse vocale [14].....	11
I.7 Conclusion	11
Partie II : Les métiers liés au traitement automatique des langues.....	12
II.1 Introduction.....	13
II.2 Des Ingénieurs présents à tous les niveaux de l'industrie de la langue [2].....	13
II.3 Ergolinguiste [2].....	14
II.4 Un important domaine de recherche	14
Partie III : Linguiste informaticien.....	16
III.1 Évolution du métier au cours des dernières années [4].....	17
III.2 Des compétences de linguistes [2].....	17
III.3 ... et d'informaticiens [2].....	18
III.4 Formations et débouchés [2][4].....	18
III.5 Conclusion.....	19
Partie IV : Le marché de l'emploi.....	20
IV.1 Introduction.....	21
IV.2 Exemples d'entreprises dans l'industrie du langage.....	21
IV.2.a Syllabs [11].....	21
IV.2.b Synapse Développement [12].....	22
IV.2.c Auralog [3].....	22
IV.2.d Linguatec [5].....	23
IV.3 Informatique et linguistique : quelques laboratoires.....	23
IV.3.a ALPAGE [1].....	24
IV.3.b TALEP [13].....	24
IV.3.c L'équipe de T.A.L. de l'université du Mirail à Toulouse [4].....	26
IV.4 Deux exemples typiques d'offres d'emploi.....	26
IV.4.a CDD pour le Centre Commun de Recherche de la Commission Européenne.....	26
IV.4.b CDI pour le CNRS.....	29
IV.5 Synthèse des offres d'emploi.....	31
Partie V : Les différentes filières et parcours de formation.....	34
V.1 Les Master informatique spécialité T.A.L. [4][9][7][8].....	35
V.2 Parcours informatique et linguistique Paris IV Diderot [10].....	36
V.3 Master de recherche en Sciences Cognitives à Grenoble [6].....	38
Partie VI : Conclusion.....	40
Partie VII : Annexes.....	42
VII.1 Syllabus détaillés des formations.....	43
VII.1.a Cours de Linguistique Informatique à Paris 7.....	43
VII.1.b Master en T.A.L. de Nancy : un des rare Master permettant une formation correcte en	

La Linguistique Informatique

Anglais.....	46
VII.1.c Master en Science Cognitive à Grenoble.....	49
VII.2 Offres d'emplois.....	52
VII.2.a Informaticien Linguiste Senior (CDI).....	52
VII.2.b CDI chez VISION OBJEC.....	53
VII.2.c CDD 7 mois : Linguiste informaticien / Terminologue.....	53
VII.2.d Poste chez DGA/Ingénierie de Projet.....	53
VII.2.e Poste pour LIMSI – CNRS.....	54
VII.2.f Offre d'emploi en CDD chez Sinequa.....	54
VII.2.g a Emploi à mi-temps chez Syllabs.....	55
VII.2.h CDI pour OWI Technologies.....	55
VII.3 Interview de Ludovic Tanguy et Nabil Hathout, enseignant chercheur et chercheur en linguistique informatique à l'université du Mirail, Toulouse 2.....	57
Question 1 : Pouvez vous me présenter les principales évolutions du traitement automatique des langues ?.....	57
Question 2 : Pouvez vous me parler de vos différents parcours ? Comment devient-on linguiste informaticien ?	57
Question 3 : Pourrait-on revenir sur l'importance des langues dans le T.A.L. ? J'ai remarqué que les offres d'emplois demandaient des compétences assez poussées dans ce domaine et que les formations proposées n'étaient souvent pas à la hauteur.	58
Question 4 : Quels sont les langages informatiques les plus utilisés ?.....	59
Question 5 : Pouvez vous me parlez de votre travail ? Par exemple combien d'heures passez-vous pour l'enseignement ? Quelles sont les évolutions de carrière possible ?.....	59
Question 6 : Quels sont les débouchés, autre que dans la recherche, après un master de T.A.L. ?	60
Question 7 : Pouvez-vous me donner une idée de vos salaires ?.....	61
Glossaire.....	62
Index Lexical.....	64
Sources.....	65
Notions fondamentales.....	65