



HAL
open science

Centre de recherche en linguistique et traitement automatique des langues, Lucien Tesnière

Rapport Hcéres

► **To cite this version:**

Rapport d'évaluation d'une entité de recherche. Centre de recherche en linguistique et traitement automatique des langues, Lucien Tesnière. 2011, Université de Franche-Comté - UFC. hceres-02034428

HAL Id: hceres-02034428

<https://hal-hceres.archives-ouvertes.fr/hceres-02034428v1>

Submitted on 20 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



agence d'évaluation de la recherche
et de l'enseignement supérieur

Section des Unités de recherche

Rapport de l'AERES sur
l'unité :

Centre Tesnière - EA 2283

sous tutelle des
établissements et organismes :

Université de Franche-Comté

Mars 2011



agence d'évaluation de la recherche
et de l'enseignement supérieur

Section des Unités de recherche

Rapport de l'AERES sur l'unité :

Centre Tesnière - EA 2283

sous tutelle des
établissements et organismes :

Université de Franche-Comté

Le Président de l'AERES

Didier Houssin

Section des unités
de recherche

Le Directeur

Pierre Glorieux

Mars 2011



Unité

Nom de l'unité : Centre Tesnière

Label demandé :

N° si renouvellement : EA 2283

Nom du directeur : Mme Sylviane CARDEY-GREENFIELD

Membres du comité d'experts

Président :

M. Sylvain KAHANE, Université Paris Ouest Nanterre

Experts :

Mme Marie CANDITO, Université Paris Diderot

M. Nabil HATHOUT, Université Toulouse 2

Mme Annie BERTIN, Université Paris Ouest, représentant le CNU

Représentants présents lors de la visite

Délégué scientifique représentant de l'AERES :

Mme Mariapaola D'IMPERIO

Représentant(s) des établissements et organismes tutelles de l'unité :

M. Jacques BAHI, VP recherche



Rapport

1 • Introduction

- Date et déroulement de la visite :

La visite a commencé à 9h30. Elle s'est déroulée dans des bonnes conditions logistiques, dans une enfilade de salons de réception de l'université. L'unité a été présentée en une demi-heure par sa directrice. Une séance de questions d'une heure et demi a suivi où sont intervenus les partenaires industriels, mais très peu de membres titulaires à part la directrice. Le comité d'experts s'est ensuite retiré une demi-heure avec la tutelle. La séance s'est terminée à 14h après une discussion avec les membres de l'équipe autour d'un buffet, suivie d'une séance d'une demi-heure de démonstration des outils de l'équipe animée par un doctorant.

- Historique et localisation géographique de l'unité et description synthétique de son domaine et de ses activités :

Le Centre Tesnière a été créé en 1980 par M. Yves GENTILHOMME, mathématicien et linguiste franco-russe, qui a participé au développement du TAL en France.

Le Centre Tesnière développe un certain nombre d'outils en Traitement Automatique des Langues, notamment un système de génération automatique, un classifieur et un analyseur morphosyntaxique. Ces systèmes sont basés sur une architecture assez élémentaire, présentée comme une implémentation du système d'analyse systémique SyGULAC développé par le fondateur de l'unité, puis par sa directrice actuelle (voir sa thèse d'état en 1987).

- Equipe de Direction :

L'unité est dirigée par Mme Sylvaine CARDEY laquelle est secondée par un bureau de 8 membres (la directrice, 2 MCF, 1 Doctorant, 3 enseignants-chercheurs non-titulaires et partenaires d'entreprises, 1 Ingénieur de recherche chargé de cours).

- Effectifs de l'unité : (sur la base du dossier déposé à l'AERES) :

| | Dans le bilan | Dans le projet |
|---|---------------------------|-------------------------|
| N1 : Nombre d'enseignants-chercheurs (cf. Formulaire 2.1 du dossier de l'unité) | 7 | 7 |
| N2 : Nombre de chercheurs des EPST ou EPIC (cf. Formulaire 2.3 du dossier de l'unité) | 0 | 0 |
| N3 : Nombre d'autres enseignants-chercheurs et chercheurs (cf. Formulaire 2.2 et 2.4 du dossier de l'unité) | 8 | 7 |
| N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires (cf. Formulaire 2.5 du dossier de l'unité) | 1 | 1 |
| N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires (cf. Formulaire 2.6 du dossier de l'unité) | 2 (déjà comptés en N3) | 3 (différents de N3) |
| N6 : Nombre de doctorants (cf. Formulaire 2.7 du dossier de l'unité) | 16 | 18 |
| N7 : Nombre de personnes habilitées à diriger des recherches ou assimilées | 3 | 2 |



2 • Appréciation sur l'unité

- Avis global :

Malgré de forts liens avec l'industrie, et des contrats productifs, ainsi qu'un réseau de partenaires universitaires internationaux, notamment au travers d'un master Erasmus Mundus en Traitement Automatique des Langues et des conférences xTAL, le centre Lucien Tesnière apparaît extrêmement isolé sur le plan de la recherche nationale et internationale.

Le comité d'experts a été frappé par la proportion de publications endogènes, la quasi absence de publications dans des supports reconnus internationalement, ainsi que l'absence, au sein des publications de l'équipe, de confrontation aux travaux d'autres équipes.

D'après ce qui a été présenté au comité d'experts, les technologies utilisées pour la fouille de textes, la détection d'opinions, la génération multilingue, ou la désambiguïsation morpho-syntaxique, sont rudimentaires et consistent essentiellement en l'écriture manuelle de règles, sans prise en compte des avancées dans le domaine de ces 30 dernières années. Le comité d'experts a néanmoins noté un travail conséquent dans la mise en œuvre de ces techniques dans le cadre d'applications industrielles.

- Points forts et opportunités :

L'unité participe à l'adossement scientifique d'un master Erasmus Mundus en TAL qui fournit un flux important d'étudiants et de doctorants.

Elle a des collaborations nombreuses et durables avec des partenaires industriels (Nestlé, Airbus notamment) et est à l'origine de la création de cinq entreprises « spin-off » (dont une lauréate au concours national d'aide à la création d'entreprises 2006), ce qui est exceptionnel en SHS, tout particulièrement pour une équipe de cette taille. Cela lui permet d'assurer un financement des thèses ainsi que l'insertion professionnelle de ses diplômés.

Deux brevets déposés par l'équipe sont renouvelés et un est en cours de dépôt avec Airbus.

Deux projets ANR (LISE et un projet Emergence qui lui fait suite) et un projet européen (MESSAGE) témoignent du dynamisme de l'équipe. A noter cependant que pour les deux projets ANR, le centre Tesnière est le seul partenaire académique.

L'équipe est à l'initiative d'une série de conférences xTAL, déclinées en FracTAL en 97, VexTAL en 99, GoTAL etc., qui sont devenues au fil des années des conférences d'assez bon niveau.

- Points à améliorer et risques :

Le point central à améliorer concerne la méthodologie de recherche, qui ne prend pas en compte les travaux extérieurs à l'équipe.

Cette méthodologie a un impact sur les publications de l'équipe, qui ne se positionnent pas par rapport aux autres travaux et qui n'ont aucun impact dans la communauté (les travaux du centre Tesnière ne sont quasiment pas cités en dehors de l'équipe elle-même). En conséquence le taux de producteurs de l'unité est assez bas.

Le fonctionnement de l'équipe est trop centralisé, ce qui met en péril sa pérennisation.



- **Recommandations au directeur de l'unité :**

- Distribuer davantage les responsabilités.
- Inciter à la confrontation des résultats de l'équipe avec les travaux nationaux et internationaux, ce qui passe par le recours à des standards d'évaluation autres que des « benchmarks » d'entreprise.
- Limiter les contributions dans des publications éditées par des membres de l'équipe (revue Bulag, numéros de *Advances in Natural Language Processing*). L'existence de la revue Bulag nuit au rayonnement de l'équipe, par le fait qu'elle constitue une solution de facilité qui détourne les chercheurs du Centre Tesnière de publier dans des supports à plus large diffusion et scientifiquement plus sélectifs.
- Mettre en œuvre une politique d'amélioration du taux de producteurs.

- **Données de production :**

(cf. http://www.aeres-evaluation.fr/IMG/pdf/Criteres_Identification_Ensgts-Chercheurs.pdf)

| | |
|---|------|
| A1 : Nombre de producteurs parmi les chercheurs et enseignants chercheurs référencés en N1 et N2 dans la colonne projet | 1 |
| A2 : Nombre de producteurs parmi les autres personnels référencés en N3, N4 et N5 dans la colonne projet | 1 |
| A3 : Taux de producteurs de l'unité $[A1/(N1+N2)]$ | 0,42 |
| Nombre d'HDR soutenues | 0 |
| Nombre de thèses soutenues | 22 |
| Autre donnée pertinente pour le domaine (à préciser...) | |

Le taux de producteurs indique que 3 membres sur 7 ont une production non négligeable. Néanmoins l'un des 3 membres publie dans une thématique qui n'est aucunement relayée dans le rapport de l'unité et aucun des 3 membres n'a de publications sur des supports figurant dans la liste de l'AERES.



3 • Appréciations détaillées :

- **Appréciation sur la qualité scientifique et la production :**

L'équipe ne prend pas en compte une des dimensions de la recherche, qui est la production de connaissances, et est uniquement centrée dans la réutilisation de techniques existantes selon une théorie élaborée en son sein et qui lui est propre.

- **Pertinence et originalité des recherches, qualité et impact des résultats :**

L'impact des résultats est quasiment inexistant dans la communauté scientifique du domaine. Les travaux n'ont pas d'originalité démontrée : ni les supports choisis pour la publication, ni la nature des articles sans positionnement par rapport à l'état de l'art ne permettent d'attester la moindre originalité des travaux de l'équipe.

Les questions du comité d'experts portant sur l'originalité des outils développés par l'unité se sont heurtées à des réponses du genre « Cet outil a fait l'objet d'un dépôt de brevet et nous ne pouvons donc pas vous expliquer ce qu'il a d'original, mais le fait qu'il ait été breveté prouve qu'il l'est ». De même les questions sur l'évaluation des outils ont amené des réponses telles que « Cet outil a fait l'objet d'une étude comparative de la part d'un de nos partenaires industriels et s'ils l'ont retenu c'est qu'il était meilleur ». A noter également que le partenaire industriel en question est membre associé de l'unité, chargé de cours dans le master et ancien étudiant de l'université de Besançon.

Le comité d'experts s'étonne également d'une confusion entre l'activité déclarée de « traduction automatique à partir de langue contrôlée » (activité supposée phare) et l'activité effective, qui relève plutôt selon lui du domaine de la « génération multilingue », étant donné que l'entrée traitée par le système doit être préalablement structurée par l'utilisateur. C'est un symptôme de l'isolement scientifique de l'équipe, et un facteur de renforcement de cet isolement.

- **Quantité et qualité des publications, communications, thèses et autres productions :**

Les publications sont nombreuses mais leur qualité est extrêmement problématique, pour ce qui concerne

1) le type de support : sur les 53 publications présentées comme relevant de la rubrique ACL (supposée rassembler les articles dans des revues nationales et internationales avec comité de lecture répertoriées par l'AERES ou dans les bases de données internationales), 26 publications sont dans la revue Bulag dont la rédactrice en chef est la directrice du centre Tesnière, qui n'est pas répertoriée par l'AERES et qui, selon le comité d'experts, ne remplit aucun des critères d'une revue internationale ou même nationale. Cette revue, en effet, n'a pas sollicité de comité de lecture international, elle n'a aucun impact dans la communauté scientifique, elle n'a pas de taux de sélection connu. 3 publications sont dans un volume *Advances in Natural Language Processing*, également édité par la directrice du centre. Le comité d'experts remarque un manque de publications dans les conférences ou revues majeures en traitement automatique des langues, à l'exception d'une communication à LREC 2010, et deux communications dans des workshops de COLING en 2006. Par exemple aucune publication à la revue TAL, ni à TALN, revue et conférence de référence de la communauté TAL francophone.

2) Le contenu des articles que les membres du comité d'experts ont pu lire : le comité relève en effet l'absence totale de positionnement par rapport à l'état de l'art, des citations exclusivement endogènes, l'absence de contenu innovant des travaux (comme l'absence de tentative pour démontrer ce caractère innovant).

- **Qualité et pérennité des relations contractuelles :**

L'université n'apporte pas d'aide particulière à l'équipe ni en termes financiers, ni en termes de postes. L'équipe possède un réseau de partenaires industriels qui suffit à assurer son fonctionnement. Elle n'a aucune collaboration avec d'autres équipes de l'université (il existe dans une autre équipe de linguistes, le LASELDI, des chercheurs en informatique). Il est difficile au comité d'experts de recommander un soutien plus important des tutelles sans un changement majeur dans la politique scientifique de l'unité.



- **Appréciation sur le rayonnement, l'attractivité, et l'intégration de l'unité de recherche dans son environnement :**

Le rayonnement et l'intégration de l'unité est relativement faible du fait de son isolement, tant sur le plan international, national que local. Les travaux du Centre Tesnière semblent inconnus du reste de la communauté tant locale, nationale qu'internationale (si l'on excepte les quelques contacts du consortium qui compose le Master Erasmus Mundus en Traitement Automatique des Langues dont l'unité constitue l'adossement scientifique). Le modèle SyGULAC n'est utilisé dans aucun autre laboratoire national ou international. Les travaux sur la génération multilingue ne sont pas repris ni discutés dans cette communauté. Il est probable que la politique consistant à privilégier les dépôts de brevets et les relations avec des entreprises industrielles est sur ce point un handicap.

- Nombre et renommée des prix et distinctions octroyés aux membres de l'unité, y compris les invitations à des manifestations internationales :

La principale distinction du Centre Tesnière est la nomination de sa directrice comme membre senior de l'IUF. Il n'y a pas d'autre distinction au plan académique. Par ailleurs, l'une des spin off issue du Centre Tesnière a reçu une distinction nationale. Le comité d'experts relève deux conférences invitées dans le bilan de l'unité (colloque de traductologie à Oran, ambassade de France en Thaïlande).

- Capacité à recruter des chercheurs, post-doctorants ou étudiants de haut niveau, en particulier étrangers :

Il n'y a pas eu de recrutement au Centre Tesnière au cours du quadriennal. Il n'y a pas eu non plus de recrutement de post-doc de haut niveau. On note cependant qu'une grande partie des doctorants du Centre Tesnière sont recrutés à l'étranger pour leur maîtrise de leurs langues nationales.

- Capacité à obtenir des financements externes, à répondre ou susciter des appels d'offres, et à participer à l'activité des pôles de compétitivité :

C'est là l'un des points forts de l'unité qui dispose d'un réseau de spin off et d'interlocuteurs dans quelques entreprises importantes qui lui permettent de valoriser ses outils et d'obtenir des financements externes conséquents (de l'ordre de 500 000 euros au cours du quadriennal). Le Centre Tesnière ne participe pas à des pôles de compétitivité.

- Participation à des programmes internationaux ou nationaux, existence de collaborations lourdes avec des laboratoires étrangers :

Le Centre Tesnière est l'un des laboratoires auquel est adossé un Master Erasmus Mundus en TAL. Les collaborations au sein du consortium du Master (Algarve, Barcelone, Besançon, Wolverhampton) ont donné lieu au projet MESSAGE financé par l'UE, sur la thématique de recherche principale de Centre Tesnière : la génération multilingue de messages. On note que ce financement ne s'inscrit pas dans le cadre des appels des 6^e/7^e PCRD.

- Valorisation des recherches, et relations socio-économiques ou culturelles :

La valorisation semble être la priorité absolue du laboratoire dont l'activité est tournée pour l'essentiel vers la collaboration avec des partenaires industriels. Les succès sur ce plan sont remarquables pour une unité de SHS aussi petite. On signalera notamment 2 brevets déposés et 1 en cours de dépôt avec Airbus.



- **Appréciation sur la stratégie, la gouvernance et la vie de l'unité:**

- Pertinence de l'organisation de l'unité, qualité de la gouvernance et de la communication interne et externe :

Il est à noter que le dossier scientifique présenté aux experts est exclusivement centré sur les travaux de la directrice de l'unité. Le MCF HDR qui dirige de nombreuses thèses (7 thèses soutenues dans le quadriennal, 10 en cours) n'est pas membre du bureau et n'a pas participé à la présentation.

L'élaboration des projets semble impulsée par la directrice d'unité en relation avec ses contacts réguliers dans les entreprises, sans que l'on perçoive une élaboration en commun au sein de l'équipe. La périodicité des réunions du bureau n'est pas régulière ; le financement des missions n'est pas régi par des règles précises, selon ce qui apparait au cours de l'entretien.

- Pertinence des initiatives visant à l'animation scientifique, à l'émergence, et à la prise de risques :

Le comité d'experts note le caractère très positif des collaborations avec les industriels et le CHU de Besançon. La directrice de l'équipe a également impulsé des rencontres internationales régulières et a maintenu le réseau de collaborations internationales. Le comité a cependant l'impression qu'il s'agit davantage de la gestion d'un acquis que d'actions permettant l'émergence d'avancées dans la recherche et la montée en responsabilités de chercheurs susceptibles d'apporter un renouvellement voire une pérennité aux travaux de l'équipe.

- Implication des membres de l'unité dans les activités d'enseignement et dans la structuration de la recherche en région :

Les membres de l'unité participent à des enseignements dans le domaine du TAL et en langues vivantes.

L'unité est fortement impliquée dans un master Erasmus Mundus en Traitement Automatique des Langues.

L'unité n'a aucune activité structurante au niveau de la recherche en dehors de sa collaboration avec des industriels. Le comité d'experts a été étonné de l'absence de contacts avec les autres équipes et chercheurs travaillant dans le domaine du TAL et, plus généralement, de la linguistique ou de l'informatique, au sein de l'Université de Franche-Comté.

- **Appréciation sur le projet :**

- Existence, pertinence et faisabilité d'un projet scientifique à moyen ou long terme :

L'unité n'a pas présenté de projet scientifique autre que la continuation de ses activités. Le dossier présenté aux experts contient très peu d'informations scientifiques précises. Le même texte, très court, est utilisé pour renseigner différentes rubriques.

- Existence et pertinence d'une politique d'affectation des moyens :

Aucun élément porté à la connaissance du comité d'experts ne permet d'accréditer l'existence d'une politique d'affectation des moyens.

- Originalité et prise de risques :

L'unité ne fait preuve d'aucune originalité ni d'aucune prise de risques : le projet est dans la continuité totale du précédent ; le comité d'experts n'a pu relever aucune nouveauté dans le projet pour le prochain contrat.



Notation :

| Intitulé UR / équipe | C1 | C2 | C3 | C4 | Note globale |
|--|----|----|----|----|--------------|
| Centre de Recherche en Linguistique et Traitement automatique des Langues, Lucien Tesnière | C | B | C | C | C |

C1 Qualité scientifique et production

C2 Rayonnement et attractivité, intégration dans l'environnement

C3 Gouvernance et vie du laboratoire

C4 Stratégie et projet scientifique

Statistiques de notes globales par domaines scientifiques (État au 06/05/2011)

Sciences Humaines et Sociales

| Note globale | SHS1 | SHS2 | SHS3 | SHS4 | SHS5 | SHS6 | Total |
|--------------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| A+ | 2 | 8 | 2 | 11 | 5 | 6 | 34 |
| A | 12 | 33 | 12 | 13 | 32 | 18 | 120 |
| B | 11 | 37 | 6 | 22 | 19 | 5 | 100 |
| C | 8 | 4 | 2 | 6 | 1 | | 21 |
| Non noté | 1 | | | | | | 1 |
| Total | 34 | 82 | 22 | 52 | 57 | 29 | 276 |
| A+ | 5,9% | 9,8% | 9,1% | 21,2% | 8,8% | 20,7% | 12,3% |
| A | 35,3% | 40,2% | 54,5% | 25,0% | 56,1% | 62,1% | 43,5% |
| B | 32,4% | 45,1% | 27,3% | 42,3% | 33,3% | 17,2% | 36,2% |
| C | 23,5% | 4,9% | 9,1% | 11,5% | 1,8% | | 7,6% |
| Non noté | 2,9% | | | | | | 0,4% |
| Total | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

Intitulés des domaines scientifiques

Sciences Humaines et Sociales

- SHS1 Marchés et organisations
- SHS2 Normes, institutions et comportements sociaux
- SHS3 Espace, environnement et sociétés
- SHS4 Esprit humain, langage, éducation
- SHS5 Langues, textes, arts et cultures
- SHS6 Mondes anciens et contemporains



Observations générales non parvenues à l'AERES sous forme publiable.