

## Un outil exploratoire de corpus d'apprenants

Issu du projet IDILL (<http://www.idill.org>) mené au sein du réseau européen d'excellence Kaleidoscope (<http://www.noie-kaleidoscope.org>), ce travail vise la définition et la réalisation d'un « exemplier ». Destiné à la fois aux enseignants de langue, aux chercheurs en didactique des langues et aux chercheurs en TAL (Traitement Automatique des Langues), cet outil permet l'interrogation du corpus d'apprenants FRIDA afin d'en extraire des exemples de phrases comportant certaines caractéristiques notamment au niveau des erreurs. Le corpus, composé de 764 textes soit 9466 phrases, 20 474 erreurs et 179 642 mots, possède un double système d'annotation :

- Balisage (manuel) des erreurs d'apprenants avec correction
- Balisage morphosyntaxique (automatique) de l'ensemble des formes (texte, erreurs et correction)

L'exemplier offre ainsi une interrogation de ce corpus selon différents critères issus de ces annotations : langue maternelle des apprenants, niveau, densité d'erreurs, type d'erreurs, mots (forme, lemme, catégorie syntaxique, traits morphologiques), corrections, contextes...

Les utilisations d'un tel système sont variées tant sur un plan de didactique appliquée que sur un plan de recherche en didactique ou en TAL. Les pédagogues et didacticiens peuvent, entre autres, cerner et quantifier les erreurs typiques des allophones, visualiser les fréquences d'emploi de mots et de structures, comparer des productions entre différents groupes d'apprenants (langue maternelle, niveau), etc. Pour les chercheurs en TAL, l'exploration de tels corpus devrait permettre une amélioration des modèles de détection et d'évaluation des erreurs pour la production de rétroactions adaptées.

(242 mots avec le titre)

### Bibliographie

- Sylviane Granger (2000). Didactique des langues étrangères, linguistique de corpus et traitement automatiques des langues. dans M. Marquillo Larruy (ed.) *Questions d'épistémologie en didactique du français (langue maternelle, langue seconde, langue étrangère)*. Cahiers Forell. Université de Poitiers, pp. 105-109.
- Sylviane Granger, Anne Vandeventer & Marie-Josée Hamel (2001). Analyse de corpus d'apprenants pour l'ELAO basé sur le TAL. *TAL* 42(2), 609-621.
- Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T. & Ponton, C. (2005). "Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 8, n° 1. pp. 65-79. [http://alsic.u-strasbg.fr/v08/antoniadis/alsic\\_v08\\_04-rec4.htm](http://alsic.u-strasbg.fr/v08/antoniadis/alsic_v08_04-rec4.htm), mis en ligne le 15/11/2005.