

**VIRGINIE ANDRÉ**  
**Analyse et Traitement Informatique de la Langue Française (ATILF)**  
**Université de Lorraine et CNRS**  
**Nancy, France**  
**Virginie.Andre@univ-lorraine.fr**

### ***Apprendre une langue étrangère grâce à un dispositif numérique d'apprentissage. L'exemple de FLEURON***

Deux évolutions majeures des sociétés actuelles font avancer, orientent et modifient les travaux concernant l'enseignement et l'apprentissage des langues étrangères. Tout d'abord, la mobilité s'accroît. De plus en plus de personnes choisissent de voyager ou de partir travailler à l'étranger, quel que soit leur statut : étudiants, enseignants, chercheurs, médecins, commerciaux, curieux, etc. Ensuite, avec l'apparition de nouvelles technologies, la didactique des langues a bénéficié de nouveaux outils qui s'accompagnent de nouvelles approches. Ces deux dynamiques se rejoignent dans la mesure où la maîtrise de langue du pays choisi garantit une intégration plus facile et de meilleure qualité. C'est pourquoi, nombreux sont ceux qui apprennent la langue du pays avant d'entreprendre leur voyage. Cet apprentissage bénéficie aujourd'hui de la combinaison de plusieurs avancées scientifiques, notamment en linguistique, en didactique et en informatique. Ces interactions disciplinaires permettent de proposer de nouvelles méthodologies d'appropriation des langues.

Nous examinerons ici un dispositif numérique d'apprentissage du français dédié aux étudiants étrangers qui veulent faire un séjour universitaire en France. Ce dispositif s'intègre dans le cadre d'un projet de recherche intitulé FLEURON - Français Langue Etrangère Universitaire – Ressources et Outils Numériques - et se concrétise par un site Internet du même nom : <https://apps.atilf.fr/fleuron/>. Ce site met à disposition librement et gratuitement une plateforme d'apprentissage du français permettant aux futurs étudiants en France de se préparer aux situations de communication qu'ils vont rencontrer. Nous verrons que FLEURON s'inscrit dans les recherches actuelles en sociolinguistique des interactions verbales, en didactique de l'oral et en apprentissage sur corpus (*data-driven learning*). Notre objectif est de montrer de quelle façon un corpus multimédia de situations de communication authentiques peut aider les étudiants étrangers à acquérir des compétences sociolangagières facilitant leur intégration.

#### **1. Mobilité et besoins sociolangagiers**

L'origine du projet FLEURON est liée à l'accroissement de la mobilité étudiante depuis les vingt dernières années (André 2016). De plus en plus d'étudiants étrangers viennent faire un séjour universitaire ou la totalité de leurs études en France. Certains viennent pour obtenir un diplôme particulier qui enrichira ou complètera leur cursus initial dans leur pays d'origine. Les étudiants étrangers viennent très majoritairement rejoindre les universités françaises mais certains intègrent également les écoles d'ingénieurs et les grandes écoles (de management, de commerce, d'arts, par exemple). Un nombre important d'étudiants en mobilité s'inscrit en doctorat, 41 % des doctorants en France sont étrangers. De plus, la France cherche à attirer de plus en plus d'étudiants étrangers. En quinze ans, le nombre d'étudiants a doublé. Cependant, à Nancy (France), lieu de naissance du projet FLEURON dans les années 2000, des problèmes de communication ont été constatés lors d'interactions entre le personnel universitaire (enseignant et administratif) et les étudiants étrangers. C'est pourquoi, nous avons souhaiter mettre en place un dispositif d'apprentissage de la langue afin d'améliorer l'intégration de ces étudiants. En proposant l'appropriation de besoins langagiers nécessaires à mener une vie d'étudiant, FLEURON permet aux apprenants de se préparer aux interactions verbales auxquelles ils seront susceptibles de participer en France.

Selon Campus France, l'agence gouvernementale qui gère la mobilité étudiante française, 90 % des étudiants étrangers en France sont satisfaits de leur séjour universitaire. Néanmoins, il reste certains problèmes d'intégration liés à la maîtrise du français. En effet, ces problèmes sont fréquents bien que les étudiants étrangers préparent leur séjour universitaire. Les pays d'origine dispensent des cours de français, souvent spécialisés et sur objectifs universitaires. Les alliances françaises proposent également une offre de formations en français langue étrangère (FLE) dans tous les pays du monde. Certains étudiants suivent des cours intensifs de français dans des centres de langues en France, dès leur arrivée et avant de commencer leur année universitaire. Cependant, certaines compétences sociolangagières manquent aux étudiants et ces manques nuisent à une bonne intégration. Les échecs universitaires des étudiants migrants sont souvent dus à des problèmes d'intégration.

Les formations en français sur objectif universitaire (FOU) cherchent à faire acquérir aux étudiants étrangers des compétences langagières disciplinaires. Celles-ci leur permettent de comprendre un cours magistral ou plus généralement un discours scientifique, de présenter un exposé à l'oral, de construire une argumentation, de rédiger un écrit scientifique, de comprendre une consigne à un examen encore de prendre la parole dans un cours. Mangiante et Parpette (2011) ont très bien montré les enjeux du FOU ainsi que les compétences langagières visées. Notre objectif, avec FLEURON, n'est pas d'intervenir en FOU. Nous proposons de former les étudiants à participer à toutes les situations de communication universitaires, en dehors des cours. Au-delà des compétences langagières nécessaires à la réalisation d'un cursus universitaire, les étudiants ont d'autres besoins langagiers qui leur permettent de s'intégrer en tant qu'étudiant au sein d'un système universitaire ainsi qu'en tant que citoyen migrant dans une société d'accueil. Les compétences sociolangagières acquises grâce à ce dispositif

permettent aux étudiants d'interagir de façon efficace, des points de vue linguistique, pragmatique, interactionnel et culturel.

## 2. Interface linguistique et didactique

Notre recherche se situe au croisement de plusieurs disciplines. Le projet FLEURON est né des avancées scientifiques en linguistique et en didactique, ainsi qu'en informatique. Plus précisément, ce projet se situe à l'interface entre linguistique et didactique des langues. Il met à profit les résultats de l'analyse sociolinguistique des interactions verbales pour l'enseignement et l'apprentissage du français oral. Ces résultats sont issus, plus généralement, de la linguistique de corpus et de ses outils d'exploitation des données.

Les recherches actuelles en didactique des langues s'orientent vers l'utilisation de corpus à des fins d'enseignement et d'apprentissage (Boulton, Tyne 2014). Les travaux anglo-saxons dans le domaine du *data-driven learning* (Johns 1991, Aston 2001) sont désormais bien ancrés en didactique, notamment de l'anglais écrit. Cette nouvelle méthodologie trouve son origine dans l'utilisation des documents authentiques, la linguistique de corpus et les nouvelles technologies. Tout d'abord, l'utilisation des documents authentiques, prônée depuis les années 1970, est indispensable à l'acquisition de compétences sociolangagières. Pour rappel, un document authentique est une ressource qui n'est pas constituée à des fins didactiques. En ce qui concerne le document authentique oral, il présente une interaction qui aurait eu lieu même si elle n'avait pas été recueillie pour servir de document support à l'apprentissage de la langue (Duda et al. 1972, Abé et al. 1979). Les documents authentiques permettent de confronter les apprenants à la réalité des pratiques langagières. Ces documents authentiques peuvent être constitués en corpus, à l'instar des corpus de données langagières élaborés par les linguistes. Nous définissons un corpus, selon Sinclair (1996 : 4) comme « a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language ». Il existe plusieurs corpus de français parlé qui sont des échantillons plus ou moins grands de la langue française, sélectionnés en fonction d'un certain nombre de critères extralinguistiques. Il existe peu de corpus constitué à des fins didactiques, FLEURON en est un. Il rassemble différents documents dans le domaine universitaire. Nous décrirons ce corpus dans le point suivant. L'analyse du français parlé et du fonctionnement de la langue parlée en interaction est permise grâce aux différents corpus oraux (voir par exemple les travaux pionniers en France de Blanche-Benveniste, Jeanjean, 1987; Gadet, 1989; Kerbrat-Orecchioni, 1990-1994). A partir du corpus multimédia FLEURON, les enseignants et les apprenants de français peuvent procéder à des analyses du fonctionnement de la langue à des fins d'appropriation de celles-ci. Ils utilisent les outils d'analyse des interactions verbales et des corpus de français parlé. Nous verrons des exemples dans les points suivants.

## 3. Le dispositif numérique d'apprentissage FLEURON

Le projet de recherche FLEURON se concrétise en un site internet libre et gratuit (voir page d'accueil, image 1 ci dessous). Il est constitué principalement d'un corpus multimédia, d'un concordancier et d'autres outils d'aide à l'apprentissage de la langue et à l'appropriation de savoirs permettant d'améliorer l'intégration des étudiants étrangers en France. Par exemple, sur le site, les apprenants peuvent trouver des liens utiles pour gérer leur arrivée (logement, aide financière, aide médicale, etc.), un glossaire, des conseils méthodologiques d'utilisation du site, des commentaires d'autres utilisateurs, des possibilités de créer des parcours de navigation. Nous allons présenter plus précisément le corpus et ses possibles exploitations.

### 3.1. Contenu au cœur de FLEURON

Le cœur de FLEURON est composé du corpus et de son concordancier. Le concordancier est un outil emprunté à la linguistique de corpus permettant d'interroger les données, nous montrerons en 3.3. comment les apprenants peuvent l'utiliser. Le corpus multimédia de FLEURON est composé de documents audio (pour les plus anciens et les situations impossibles à filmer) et vidéo. Ces documents authentiques sont de différentes natures.

- Des interactions naturelles
  - entre étudiants et personnels administratifs de l'université : scolarité, secrétariat, bibliothèque, santé, maison de l'étudiant, appareil, service des relations internationales, centre de langues, service d'aide aux logements, service d'assistance sociale et médicale, service informatique, associations étudiantes, etc. ;
  - entre étudiants et enseignants : choix de cours, constitution d'emplois du temps, soutenance, etc. ;
  - des conversations entre étudiants français ou étrangers ;
  - des interactions de la vie quotidienne en dehors du campus universitaire : à la banque, à la gare, avec des voisins, etc.
- Des témoignages
  - d'étudiants français et étrangers ;
  - des situations impossibles à filmer (le déroulement d'une visite médicale, d'un entretien à la préfecture, etc.) ;
  - des narrations (modalités et qualité de l'intégration, chocs culturels, etc.).
- Des explications sollicitées

- du fonctionnement de la bibliothèque universitaire, de la carte du restaurant universitaire, des horaires d'ouverture, de la sécurité sociale, des cours informatiques, des aides au logement, de location des vélos de la ville, etc.

Toutes les ressources sont transcrites et sous-titrées (voir image 2). En fonction de leurs objectifs d'apprentissage et de leur niveau, les apprenants peuvent faire apparaître ou masquer les sous-titres et les transcriptions.

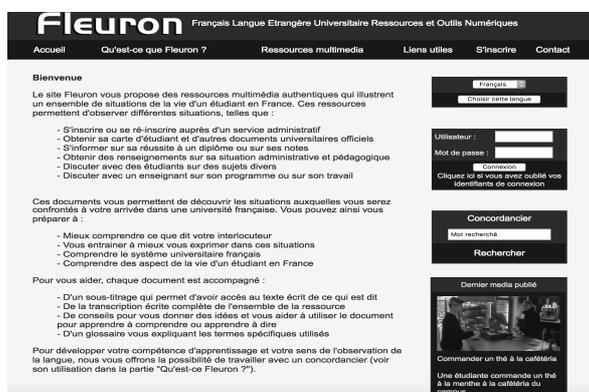


Image 1 : Page d'accueil du site FLEURON

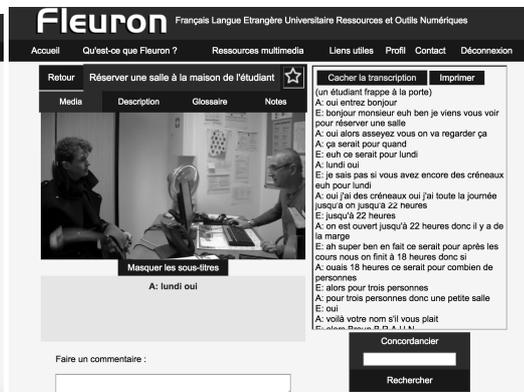


Image 2 : Ressource vidéo avec sous-titres et transcription

Les apprenants peuvent ainsi observer des situations de communications auxquelles ils participeront quand ils seront en France. Ils peuvent également comprendre le fonctionnement universitaire et les aspects de la vie étudiante ainsi que s'informer sur les pratiques et les différences culturelles, tout en apprenant la langue. De plus, deux façons différentes notamment permettent d'apprendre le français en exploitant le corpus FLEURON.

### 3.2. Analyse de situations de communications naturelles et authentiques

La première façon d'exploiter le corpus FLEURON à des fins d'apprentissage de la langue est de faire un travail d'analyse en sélectionnant une ressource. En s'inspirant et en investissant les résultats de l'analyse sociolinguistique des interactions verbales (André 2014), les apprenants peuvent eux-mêmes, éventuellement aidés de leur enseignant, observer et comprendre le fonctionnement des interactions en français. Ils observent des phénomènes rarement voire jamais écrits dans les grammaires traditionnelles ou les manuels de langue. Par exemple, en analysant la séquentialité des interactions, les apprenants peuvent découvrir les mécanismes de prise de paroles, les répétitions, les ruptures de construction, les énoncés inachevés, les hésitations, les reformulations, etc. Tous ces phénomènes sont non seulement absents des manuels mais sont également considérés comme des « dysfluences » alors que tout discours oral se construit avec l'ensemble de ces phénomènes.

Par exemple, en visionnant une interaction, dont la transcription est ci-dessous, entre un étudiant (E) et un agent administratif (A) au Service Inter-Universitaire d'Activités Physiques et Sportives (SIUAPS), un apprenant peut repérer différents éléments (ceux que nous commentons sont en gras).

- |   |  |
|---|--|
| <ol style="list-style-type: none"> <li>1. E: bonjour</li> <li>2. A: bonjour</li> <li>3. E: <b>euuh j'aurais voulu</b> me renseigner <b>euuh</b> pour <b>euuh</b> savoir <b>comment on s'inscrit</b> au SIUAPS <b>s'il vous plait</b></li> <li>4. A: <b>alors vous êtes étudiant sur Nancy</b></li> <li>5. E: <b>euuh</b> oui</li> <li>6. A: oui</li> <li>7. A: vous devez posséder obligatoirement votre carte d'étudiant</li> <li>8. E: <b>euuh</b> oui</li> <li>9. A: de l'année universitaire en cours</li> <li>10. E: je l'ai là</li> <li>11. A: vous l'avez très bien</li> <li>12. E: voilà</li> <li>13. A: et avec votre numéro INE vous inscrire via notre site internet</li> <li>14. E: d'accord</li> </ol> | <ol style="list-style-type: none"> <li>15. A: voilà l'inscription est obligatoire avant de venir retirer votre carte SIUAPS</li> <li>16. E: d'accord très et l'<b>adresse du site pour s'inscrire c'est</b></li> <li>17. A: <b>alors</b> le site internet vous le trouverez sur notre plaquette en format papier</li> <li>18. E: d'accord</li> <li>19. A: voilà</li> <li>20. E: oui</li> <li>21. A: et s- oui</li> <li>22. E: oui</li> <li>23. A: sur le devant et c'est ce programme contient toutes les activités qu'on propose avec les horaires et les différents lieux</li> <li>24. E: d'accord ok très bien</li> <li>25. E: <b>alors</b> euuh ensuite pour les activités <b>ça se passe comment je m'inscris où</b></li> </ol> |
|---|--|

Cette interaction montrant un étudiant qui demande des renseignements au service des sports est une situation à laquelle les étudiants étrangers sont susceptibles de participer. L'utilisateur de FLEURON peut analyser l'interaction dans sa séquentialité, c'est-à-dire au fur et à mesure de son déroulement. Tout d'abord, il peut observer les salutations suivies d'une demande formulée par l'étudiant (tour 3). Cette demande est influencée par la politesse, très présente dans les interactions françaises. Le tour 3 est composé de trois parties : une préparation de la question, avec un conditionnel pour adoucir davantage la requête, la question, et une formule de politesse. Il est intéressant de constater que cet énoncé comporte plusieurs hésitations (*euuh*), notamment une en début de tour de parole qui sert justement à prendre le tour et à introduire l'énoncé. De plus, la question en

elle-même (*comment on s'inscrit au SIUAPS*) ne présente pas une forme canonique, telle que les questions décrites dans les grammaires ou les manuels. Nous pouvons constater le même phénomène pour les deux questions du tour 25 : *ça se passe comment* et *je m'inscrit où*. Dans ces interrogations, le mot interrogatif se situe en fin d'unité. Ce type d'interrogation n'est pas présenté dans les manuels. Ces derniers préconisent la structure *mot interrogatif + verbe + sujet*. Une autre question, tour 16, présente des dislocations syntaxiques, constructions spécifiques à la langue parlée. Ensuite, cette interaction montre que le discours des interlocuteurs est co-construit, chacun s'appuie sur l'autre pour formuler ses paroles. L'apprenant peut également relever, puis apprendre à comprendre, l'enchaînement de paires adjacentes, de questions de l'agent et de réponses de l'étudiant, ce qui est typique de ce type d'échange (du tour 4 au tour 12). On peut également mentionner une répétition de la part de l'agent (tour 6) de la réponse de l'étudiant (tour 5) qui confirme la compréhension mutuelle. Une autre séquence, du tour 18 au tour 22, montre les traces d'attention et d'intercompréhension des interlocuteurs. Enfin, parmi les nombreuses pratiques langagières spécifiques au discours en interaction à observer, l'apprenant qui travaille avec cette ressource peut constater que deux tours de parole commencent par « alors » et trois autres par « euh » dont un début de tour qui combine les deux « alors euh » (tour 25). En plus de travailler la compréhension orale, les apprenants peuvent aussi travailler l'expression orale en s'appuyant sur ce qu'ils observent.

Les apprenants observent également différentes pratiques langagières situées, réalisant des activités langagières au sein de genres de discours et de situations de communication particulières. Par exemple, ils constatent qu'une question peut être posée différemment en fonction des éléments de la situation de communication. S'ils s'adressent à un service administratif, comme nous l'avons vu précédemment, les précautions de politesse sont recommandées alors que s'ils posent une question à un autre étudiant ou à un ami, ces précautions ne sont pas nécessairement requises. L'utilisation de corpus permet ainsi une sensibilisation à la variation langagière.

### 3.3. Exploitation outillée de corpus

La seconde façon d'exploiter le corpus FLEURON est d'utiliser le concordancier intégré au site internet (voir image 2, en bas de la transcription). Ce concordancier est adapté des outils de la linguistique de corpus mais il est simplifié afin d'être utilisé facilement par des apprenants de langue. Ce concordancier permet aux apprenants de rechercher des chaînes de caractères, autrement dit des mots ou plusieurs mots. La recherche se fait dans les transcriptions des ressources et les résultats s'affichent en liste avec le cotexte droit et gauche (voir image 3). La particularité de ce concordancier est qu'il est « aligné texte-son-vidéo », cela signifie qu'il permet de cliquer sur une occurrence et de visualiser la ressource au moment où l'occurrence recherchée est prononcée (voir image 4).

bon en ce qui concerne les	cours	par exemple du sport
parce que quand vous faites les	cours	de sport
je pense que les	cours	vont pas tarder euh la rentrée c'est le 28
donc les	cours	en général commencent euh
hein ensuite euh vous discuterez avec lui hein pour le choix des	cours	
hein pour vous rencontrer et discuter des	cours	
où les	cours	sont notés où vous devez
euh respect- euh où vous avez les	cours	euh
normalement hein les	cours	c'est fin fin juin on met fin juin
donc le 1er semestre le 1er jour de de	cours	c'est le 28
pour euh pour vos	cours	
euh si vous souhaitez prendre des	cours	de français
hein ces	cours	sont gratuits mais ils ne donnent pas de points ECTS
Lise XX c'est elle qui va vous aider pour le choix de vos	cours	hein toutes les questions sur les
et pour les	cours	maintenant les
alors au niveau des	cours	euh normalement hein vous devez voir votre professeur coordinateur

Image 3 : Extrait de la liste des résultats pour la recherche de « cours » dans le concordancier



Image 4 : Vidéo correspondant à la 11<sup>ème</sup> occurrence de « cours » dans l'extrait de l'image 3

Cette méthodologie est inspirée des travaux d'analyse de grands corpus de données, écrites ou orales. Une fois les résultats affichés, il s'agit de faire une analyse distributionnelle du phénomène étudié. Les apprenants se prêtent parfaitement à cet exercice, même s'ils ont souvent besoin de l'aide de leur enseignant, au moins lors des premières recherches. Lors des expérimentations de FLEURON, nous avons suggéré aux apprenants des listes de mots à rechercher (*sans, comme, quoi, pour, chaque, en, qui, bien, juste, voulais, etc.*). Ils ont d'abord suivi nos suggestions, parfois liées à des contenus de séquences pédagogiques précédentes (c'est le cas de « voulais » lors d'un travail sur le passé) puis ils ont rapidement choisi leurs propres requêtes. Nous avons constaté que presque tous les apprenants recherchent d'abord des petits mots de l'oral ou autres marqueurs interactionnels, tels que : *tu sais, du coup, bon, ben, genre, en fait, alors, etc.* Ensuite, les apprenants recherchent des connecteurs tels que : *parce que, car, bien que, donc, même (si), aussi, ainsi, etc.* Par exemple, un apprenant a souhaité comprendre le fonctionnement de *même*, il l'a recherché avec le concordancier et a analysé la totalité des occurrences (59) dans le corpus. Une à une, il a cherché le sens, observé la place dans le tour de parole, regardé ce qui précède et ce qui suit, a eu recours à la vidéo en cas d'incompréhension, a rédigé des paraphrases et trouvé des synonymes, etc. En résumé, il a réalisé une analyse fine des différents sens, fonctions, valeurs de *même*. A la fin de son analyse, bien qu'il n'ait pas saisi le sens de toutes les occurrences de *même* dans le corpus, il a déclaré avoir compris ce qu'il n'avait pas réussi à comprendre jusqu'à présent avec des dictionnaires et des grammaires.

Les apprenants saisissent vite que l'analyse des résultats du concordancier facilite leur compréhension du fonctionnement de la langue, ses usages ainsi que ses variations. En effet, toutes les ressources du corpus FLEURON sont décrites d'un point de vue sociolinguistique (Hymes 1972) pour que les apprenants observent différentes façons d'interagir, pour qu'ils fassent le lien entre les pratiques langagières repérées et les éléments contextuels. En outre, une des premières choses que les apprenants remarquent, c'est la différence entre l'oral et l'écrit. En effet, parfois ils recherchent des mots qui n'existent pas dans FLEURON, c'est le cas de *cependant* que les apprenants souhaitent souvent rechercher. Pour qu'ils comprennent les raisons de cette absence, nous demandons aux apprenants de faire la même recherche dans un corpus écrit (Frantext<sup>1</sup>, par exemple). Les utilisateurs apprennent également que toutes les pratiques langagières ne sont pas actualisables dans toutes les situations de communication. Ainsi, ils s'aperçoivent que les situations entre étudiants autorisent le tutoiement et les salutations avec *salut* contrairement aux situations administratives, plus formelles. Comme nous l'avons mentionné dans le point précédant, les apprenants saisissent la variation.

## Conclusion

Les nouvelles technologies permettent désormais de créer des dispositifs d'apprentissage numériques capables de stocker, gérer et interroger un grand nombre de données multimédia. Néanmoins, la mise à disposition de telles données ne suffit pas à l'appropriation d'une langue étrangère. De nouvelles méthodologies accompagnent ces innovations et permettent de confronter les apprenants à des corpus multimodaux, comme c'est le cas avec FLEURON. Le travail réalisé avec ces corpus permet aux utilisateurs de comprendre le fonctionnement de la langue cible et notamment de saisir les spécificités de la langue parlée en interaction. Celles-ci ne sont que rarement évoquées dans les salles de classes. L'apprentissage sur corpus apporte un complément aux dictionnaires, grammaires et manuels de langue.

Le numérique apporte de nouvelles pratiques pédagogiques et de nouveaux outils auxquels les enseignants de langue doivent être formés. L'utilisation des corpus, tels que FLEURON pensé pour des apprenants de français ou tels que TCOF<sup>2</sup> ou CLAPI<sup>3</sup> qui sont des corpus de linguistes, reste un défi pour la didactique des langues. Les premières expérimentations sont très prometteuses. Les pratiques des enseignants évoluent. Les explorations qualitative, par ressource, et quantitative, par concordancier, permettent une exposition directe à la langue cible et à ses variations. Les pratiques des apprenants évoluent également et s'orientent vers plus d'autonomie. Au delà du fait que le dispositif numérique soit librement accessible, les apprenants peuvent choisir eux-mêmes leurs objectifs, les situations de communication qu'ils souhaitent observer, les pratiques langagières qu'ils souhaitent s'approprier.

## Bibliographie

- Abe, D., Carton, F., Cembalo, S. M., Régent, O. (1979). Didactique et authentique: du document à la pédagogie. *Mélanges pédagogiques*, 10, p.1-14.<http://www.atilf.fr/spip.php?article3580>
- André, V. (2016). FLEURON: Français Langue Étrangère Universitaire–Ressources et Outils Numériques. Origine, démarches et perspectives. *Mélanges Crapel*, 37, p.69-92.<http://www.atilf.fr/spip.php?article4164>
- André, V. (2014). L'énonciation conjointe : trace et ressource de la construction collaborative du discours. Congrès Mondial de Linguistique Française – CMLF 2014, p.1891-1904,[http://www.linguistiquefrancaise.org/articles/shsconf/pdf/2014/05/shsconf\\_cmlf14\\_01323.pdf](http://www.linguistiquefrancaise.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01323.pdf)
- Aston, G. (ed) (2001). *Learning with corpora*. Houston, Athlestan.
- Blanche-Benvéniste, C. Jeanjean, (1987). *Le français parlé. Transcription et édition*. Paris : Didier érudition.
- Boulton, A., Tyne, H. (2014). *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris : Didier.
- Duda, R., Esch E., Laurens, J.P., (1972), Documents non didactiques et formation en langues. *Mélanges pédagogiques*, p. 1-47.<http://www.atilf.fr/spip.php?article3528>
- Gadet, F. (1989). *Le français ordinaire*. Paris: Armand Colin.
- Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz J., Hymes D. (Eds), *Directions in Sociolinguistics. The Ethnography of Communication*. Oxford : Basil Blackwell, p.35-71.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In Johns T., King P. (éds.), *Classroom Concordancing, English Language Research Journal*, 4, p.1-16.
- Kerbrat-Orecchioni, C. (1990-1994). *Les interactions verbales*. Vol. 1, 2, 3. Paris: Armand Colin.
- Parpette, C. Mangiante, J. M. (2011). *Le français sur objectif universitaire*. Grenoble, Presses Universitaires de Grenoble.
- Sinclair J. (1996), « Preliminary recommendations on corpus typology », EAGLES Document TCWG-CTYP/P, <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>

<sup>1</sup><https://www.frantext.fr/>

<sup>2</sup><http://www.cnrtl.fr/corpus/tcof/>

<sup>3</sup><http://clapi.ish-lyon.cnrs.fr/>