



Lidil

Revue de linguistique et de didactique des langues

58 | 2018

**L'enseignement et l'apprentissage de l'écrit
académique à l'aide de corpus numériques**

Questionner des corpus pour questionner la langue. L'exemple des virgules en allemand

About Corpus Inquiries and Language Inquiries. The Case of Commas in German

Eva Schaeffer-Lacroix



Édition électronique

URL : <http://journals.openedition.org/lidil/5515>

DOI : 10.4000/lidil.5515

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-37747-064-8

ISSN : 1146-6480

Référence électronique

Eva Schaeffer-Lacroix, « Questionner des corpus pour questionner la langue. L'exemple des virgules en allemand », *Lidil* [En ligne], 58 | 2018, mis en ligne le 02 novembre 2018, consulté le 22 avril 2019.

URL : <http://journals.openedition.org/lidil/5515> ; DOI : 10.4000/lidil.5515

Ce document a été généré automatiquement le 22 avril 2019.

© Lidil

Questionner des corpus pour questionner la langue. L'exemple des virgules en allemand

About Corpus Inquiries and Language Inquiries. The Case of Commas in German

Eva Schaeffer-Lacroix

1. Introduction

- 1 Dans le contexte de l'enseignement-apprentissage d'une langue étrangère (LE), l'exploration d'un corpus numérique est souvent associée à l'action d'un apprenant chercheur (Johns, 1991 ; Boulton & Tyne, 2014, p. 75 ; Schaeffer-Lacroix, 2015). Il s'agit, entre autres, d'arriver à formuler des questions permettant d'étudier certains aspects du fonctionnement de la langue. Si l'on opte, comme Kennedy et Miceli (2016), pour une utilisation directe des outils de corpus par les apprenants, ces derniers doivent arriver à transformer leurs questions linguistiques en requêtes techniques permettant d'explorer le corpus. La mise en mots de questions par rapport à certains faits linguistiques intéresse également les chercheurs du domaine du *language awareness* (Gombert, 1996 ; Stegu, 2008 ; Landure, 2013), terme que l'on peut traduire par « prise de conscience linguistique ». Swain (2006, p. 96), Pinto et El Euch (2015) ainsi que Grobet et Vuksanović (2017), qui adhèrent également à ce courant, travaillent plus particulièrement sur des situations d'interaction verbale, plus précisément des discussions métalinguistiques.
- 2 Il a semblé pertinent de combiner ces différentes entrées centrées soit sur les requêtes sur corpus, soit sur le questionnement de la langue et d'en faire l'élément central d'un module de soutien à l'écrit académique proposé à de futurs enseignants de l'allemand. Formuler des requêtes sur corpus portant sur l'élément à réviser — en l'occurrence la virgule — peut permettre aux apprenants scripteurs (Leblay, 2014) de tester des références linguistiques par rapport à cet élément, de réviser et stabiliser leurs connaissances et de les mettre en pratique lors de la phase de réécriture. Le dispositif élaboré à partir de ce positionnement a été le suivant : deux groupes d'apprenants se sont

vus proposer un atelier dédié à la prise en main d'outils de corpus et à la conceptualisation de la virgule. L'un des groupes a été invité à aborder le corpus d'experts *Aufsätze* (rédactions) (Schaeffer-Lacroix, 2014) à partir de ses propres questions et hypothèses (approche inductive) et l'autre a été informé sur des références existantes par rapport à la virgule allemande avant qu'il ne consulte ce corpus (approche déductive). Les deux traitements ont mené à des modes différents d'exploration des données.

2. Recherche-action

- 3 Le plan de cet article reflète le va-et-vient entre théorie et terrain qui, selon Narcy-Combes (2005, p. 112), caractérise une recherche-action. Lors de ce type d'intervention, les données sont recueillies selon un procédé d'observation participante (Montagne-Macaire, 2007). Impliquée à plusieurs niveaux dans la recherche, j'y joue à la fois le rôle de conceptrice du module d'enseignement, d'enseignante et de chercheuse.

2.1. Contexte de la recherche

- 4 À l'ÉSPÉ de Paris, le suivi de la production écrite en allemand langue étrangère d'étudiants inscrits en première année de master MEEF (Métiers de l'enseignement, de l'éducation et de la formation) est intégré dans un module d'une durée de 18 heures, dédié aux courants majeurs en didactique. Avant chaque rencontre en présentiel, les futurs enseignants d'allemand doivent lire un article en langue allemande portant sur le thème de la séance à venir. L'un des membres du groupe en fait un compte rendu, et il le dépose sur l'espace numérique de travail du module afin que son texte puisse être lu et commenté par les pairs. L'enseignante procède ensuite à une annotation des erreurs dans le compte rendu et les commentaires afin de signaler aux apprenants scripteurs les endroits méritant une révision. Les dernières six heures du module sont consacrées à un atelier de prise en main du système de gestion de corpus *Sketch Engine* (Kilgarriff, Rychly & Pomikalek, n. d.) et à la conceptualisation de la virgule. Cet atelier se termine par une phase individuelle de révision de texte lors de laquelle les scripteurs sont encouragés à chercher des solutions alternatives en se servant de *Sketch Engine*.
- 5 L'annotation des erreurs dans les textes a mis à jour que dans pratiquement toutes les productions, une proportion assez importante de virgules est absente ou apparaît à des endroits non pertinents. Cette constatation trouve un écho dans le rapport de jury du CAPES d'allemand de la session 2016. Ce rapport contient des « Indications aux candidat (e)s quant à l'orthographe allemande » censées les aider à améliorer leur production écrite lors de la phase d'admissibilité du concours (Goullier, 2016, p. 72). Certaines de ces indications concernent l'emploi de la virgule ; cependant, elles n'aident pas forcément à faire des choix éclairés : « l'emploi ou non de la virgule avant *und* (et), *oder* (ou), etc. coordonnant des propositions [...] est libre et obéit à des considérations de sens et de lisibilité » (*ibid.*).
- 6 Dans la présente recherche, le thème de la ponctuation émerge de l'observation du terrain. On peut constater que ces derniers temps, aussi bien en France qu'en Allemagne, ce thème est également bien représenté dans le secteur des recherches en didactique et linguistique (Boch & Frier, 2015 ; Boettcher, 2016 ; Hidden, Alday & Portine, 2016 ; Eluerd, 2017).

2.2. Méthode de récolte et d'analyse de données

- 7 La recherche a permis de récolter des traces tangibles des actions et productions écrites des apprenants. Elles sont complétées par des données de type interaction verbale, à savoir les enregistrements audio des échanges en groupe entier, les discussions des binômes lors de l'exploration du corpus documentées à l'aide de films d'écran avec option audio et les entretiens semi-dirigés filmés qui ont clos le projet. Les écrans ont été filmés à l'aide de *ScreenCast-o-Matic* (Big Nerd Software, n. d.). Cet outil numérique permet de capter en même temps les actions à l'écran et les conversations qui les accompagnent. Les analyses faites à partir des données récoltées sont de type qualitatif (analyse fine des films d'écran, des échanges verbaux en groupe entier et des entretiens) et quantitatif ; des calculs ont été apportés aux données issues des productions écrites et des transcriptions des films d'écran à l'aide de l'outil de statistique R (R Development Core Team, 2008).

3. Définitions et positionnement scientifique

3.1. Types d'écrit académique

- 8 Tout comme Gaskell et Cobb (2004), Chambers (2005), Ädel (2010) et O'Sullivan (2010), je m'intéresse au potentiel des corpus pour soutenir la rédaction d'écrits académiques. Selon Rinck (2011), le terme d'*écrit académique* recouvre des genres textuels qui varient d'une discipline universitaire à une autre. À partir de cette définition large, il semble légitime de dire qu'un écrit académique en LE est une production réalisée au sein d'un module de formation universitaire qui s'adresse aux spécialistes de la langue ou aux non-spécialistes et qui peut aussi bien viser une formation portant sur le fond que sur la forme.
- 9 Le contexte précisé en section 2.1 concerne des étudiants francophones, spécialistes de l'allemand, ayant produit deux types d'écrit académique, à savoir le compte rendu d'un article scientifique et le commentaire portant sur l'un des comptes rendus rédigés par des pairs (cf. la démarche d'implication des pairs pour la relecture d'écrits académiques proposée par Lejot, 2017). Le cadrage formel pour ces deux genres textuels a été souple : les étudiants ont été encouragés à rédiger un compte rendu sélectif d'une demi-page environ dans lequel ils esquissent les idées principales de l'article ainsi que les conséquences qu'ils pensent pouvoir en tirer pour leurs futures actions d'enseignant. Le commentaire a été présenté aux participants comme un écrit assez informel dans lequel ils entrent en discussion avec l'auteur du compte rendu.

3.2. Corpus d'experts *Aufsaetze*

- 10 Un corpus d'experts contient exclusivement des productions rédigées par des personnes considérées comme compétentes par rapport à la langue et au contenu couverts par le corpus (Boch & Frier, 2015, p. 91). On peut élargir cette définition et considérer que des étudiants germanophones sont des experts du point de vue de la langue allemande davantage que les étudiants francophones. Ce positionnement a motivé mon choix de regrouper les comptes rendus produits par des germanophones d'une session antérieure en un corpus d'experts nommé *Aufsaetze* (rédactions) (Auteur). Ce dernier a été publié sur

Sketch Engine et étiqueté de façon automatique au niveau des parties du discours à l'aide de l'outil *RFTagger* (Schmid & Laws, 2008) qui détecte le genre, le nombre et le cas des unités qui en sont pourvues.

3.3. Questionnement linguistique et requête sur corpus

- 11 Le questionnement est une composante essentielle du protocole expérimental d'un chercheur (Charaudeau, 2009). Le terme de *requête* utilisé dans le domaine de l'exploration des corpus pour désigner une ligne de commande¹ en est proche. On peut distinguer deux façons différentes de questionner les données d'un corpus, à savoir l'approche inductive et l'approche déductive.

3.3.1. Approche inductive

- 12 L'approche inductive prévoit la formulation de références par les apprenants qui abordent la langue en partant de leurs propres questions et hypothèses. Elle se matérialise en l'appellation *apprenant chercheur* qui confère à l'apprenant le rôle d'explorateur et de créateur de références linguistiques (Johns, 1991 ; Kennedy & Miceli, 2001, 2016 ; St John, 2001 ; Cheng, Warren & Xun-Feng, 2003, p. 178). Bernardini (2004, p. 16) évoque cette expression dans le contexte de l'apprentissage par la découverte appliqué aux corpus. Selon Johns (1991, p. 2), la consultation de lignes de concordance donne aux apprenants l'occasion de trouver des réponses à leurs propres questions. Une approche inductive est également adoptée par Kennedy et Miceli (2016) qui laissent une grande liberté aux utilisateurs de corpus dans des situations de production écrite. Dans leur scénario, il n'est pas prévu que les apprenants formulent des références grammaticales après avoir observé des données ; ils doivent pouvoir aborder le corpus de façon ouverte et personnelle et choisir, à partir des occurrences obtenues, les formulations qu'ils estiment utiles pour améliorer leur production écrite. L'exploration du corpus ne doit pas être bridée par un guidage trop serré ; la curiosité des apprenants pour les phénomènes langagiers attestés doit pouvoir se déployer librement.

3.3.2. Approche déductive

- 13 Aussi concluant que l'on puisse trouver l'approche inductive, il est un fait qu'elle est souvent chronophage et que son bénéfice dépend grandement d'un accompagnement pédagogique intense et pertinent (Schaeffer-Lacroix, 2015). Ces gestes ont un certain coût et ne peuvent pas être offerts de façon systématique aux apprenants. L'approche déductive incite les formés à observer la langue en se servant de références trouvées ailleurs, par exemple, dans des ouvrages ou sites de référence. Placée dans un contexte d'ouverture d'esprit et de questionnement des références existantes, cette approche peut tout aussi bien mener à des résultats intéressants. Tsui (2004, p. 39) décrit une expérience lors de laquelle des enseignants d'anglais LE ont été encouragés à formuler des questions linguistiques dont ils pensaient devoir connaître la réponse afin de pouvoir fournir aux apprenants des repères grammaticaux pertinents (Tsui, 2004, p. 57). Le formateur a ensuite sélectionné des lignes de concordance susceptibles de donner des éléments de réponse à ces questions. Il résulte de cette recherche que l'observation de données de corpus a aidé les enseignants à mettre à l'épreuve leurs représentations et connaissances dans le domaine de la grammaire pédagogique (Tsui 2004, p. 57).

3.3.3. Qui pose les questions ?

- 14 Les chercheurs interrogent les corpus afin d'y trouver des réponses à leurs questions scientifiques. Qui est censé poser les questions dans le contexte du DDL (*data-driven learning*)² ? Les questions correspondent en principe à des besoins des apprenants et devraient donc être formulées par eux-mêmes. Dans la littérature, on trouve deux cas de figure extrêmes : Leńko-Szymańska (2017) décrit une expérience lors de laquelle c'est l'enseignant qui prend en charge la formulation des questions qui ne sont pas explicitement motivées par des besoins recensés sur le terrain : « T³ asks a question: *What is the common feature of these sentences?* » (E pose une question : Quel est le trait commun entre ces phrases ?). Kennedy et Miceli (2016) souhaitent rendre les apprenants capables de formuler, de façon autonome, des « *open-ended searches* », donc des requêtes sur corpus non préformatées correspondant à des questions que les formés se posent. Les chercheuses s'attendent à ce que le développement de cette capacité mène à l'observation et à la conceptualisation de phénomènes linguistiques.

3.4. Recherches sur la virgule

- 15 Les données que j'exploite dans cet article attestent d'une maîtrise perfectible de la virgule dans pratiquement toutes les productions en allemand. Selon Favriaud (2011), ce signe de ponctuation contribue à la structuration du texte : il informe sur le lien et les frontières entre ses différents éléments ; la ponctuation aurait « une vertu réflexive et réfléchissante en regard de l'écrit » (*ibid.*). On peut donc conclure qu'en situation d'apprentissage de la production écrite, l'insertion (ou non) des virgules est un domaine qui renseigne sur certaines compétences du scripteur, en l'occurrence son degré de compréhension du fonctionnement de la langue.
- 16 Les recherches sur l'enseignement-apprentissage de la virgule en allemand sont majoritairement pensées pour des contextes d'un apprentissage institutionnel de la langue maternelle (Bredel & Primus, 2007 ; Sappok, 2011) et plus rarement de la langue étrangère (Kirchhoff & Primus, 2016 ; Diehl, Pistorius & Fayolle Dietl, 2002). Kirchhoff et Primus (2016, p. 78) proposent une analyse multilingue du phénomène : elles comparent l'emploi de virgules dans cinq situations différentes en allemand, anglais, espagnol et russe. Dans deux des situations, des différences peuvent être observées. La première concerne ce que Boettcher (2016, p. 336) appelle le « marquage de territoire⁴ » entre une structure verbale d'accueil et un groupe subordonné, pouvant correspondre à une complétive ou un groupe conjonctionnel. Comme l'illustre l'exemple suivant de Kirchhoff et Primus (2016, p. 87), contrairement aux autres langues observées, en allemand et en russe, ce marquage est effectué à l'aide d'une virgule : « *Ich glaube, dass niemand zur Party gekommen ist* » ; « *Я полагаю, что никто не прибыл на вечеринку* » (Je crois que personne n'est venu à la fête).
- 17 La deuxième différence concerne la présence ou non-présence d'une virgule devant des groupes infinitifs en allemand (Kirchhoff & Primus, 2016, p. 88). La dernière réforme de l'orthographe allemande a mené à un assouplissement des références concernant ce point ce qui oblige les scripteurs à décider en fonction du contexte (*Regeln und Wörterverzeichnis*, 2006, § 75(2) et § 76) s'il faut mettre une virgule ou pas. Un autre domaine qui oblige à faire des choix est la coordination d'éléments de même nature grammaticale par *und* (et),

oder (ou), sowie (ainsi que), wie (comme), beziehungsweise (respectivement), entweder... oder (ou... ou), weder... noch (ni... ni). Ces éléments ne sont en principe pas séparés par une virgule (*Regeln und Wörterverzeichnis*, 2006, § 72) ; toutefois, il peut arriver qu'une virgule séparatrice soit nécessaire afin de désambiguïser le sens.

- 18 J'ai choisi comme cadre descriptif et d'analyse les catégories suivantes proposées par Boettcher (2016, p. 330) qui paraissent, davantage que les 132 règles fournies pour l'entrée *Komma* (virgule) dans le dictionnaire *Duden* (Bibliographisches Institut GmbH, 2017), convenir à un public d'apprenants scripteurs ayant besoin de faire des choix linguistiques raisonnés dans des délais acceptables.
- *Schlaglöcher* (nids de poule), donc les incisives : « *Dies sind, meine Damen und Herren, die schönsten Bilder unserer Ausstellung* » (Voici, Mesdames et Messieurs, les tableaux les plus beaux de notre exposition).
 - *Drillinge* (triplets), donc les énumérations de plus de deux éléments de même nature : « *Meine Lieblingsmaler sind Picasso, Dix und Klimt* » (Mes peintres préférés sont Picasso, Dix et Klimt).
 - *Territoriumsmarkierung* (marquage du territoire), donc le marquage de séparations entre groupes syntaxiques dont l'un est subordonné à l'autre : « *Picasso, der aus Spanien kommt, hat lange in Frankreich gelebt* » (Picasso qui est originaire d'Espagne a vécu pendant longtemps en France).
- 19 Les catégories de Boettcher, à teneur métaphorique, sont susceptibles d'offrir à des apprenants scripteurs des pistes pour tester la pertinence de références existantes lors de l'exploration de la virgule dans des corpus contenant des textes de scripteurs experts traitant un sujet comparable. Cette exploration peut ensuite être mise à profit pour la révision de texte.

3.5. Questions de recherche

- 20 Les données récoltées lors de l'expérimentation présentée en section 4 sont analysées à l'aide des questions suivantes : Qu'est-ce que les enregistrements de l'écran et de la conversation des binômes lors de la prise en main de l'outil peuvent nous apprendre sur la démarche de recherche sur corpus des apprenants, en particulier dans le domaine du questionnement et de la formulation des requêtes (Q1) ? Dans quelle mesure les questions correspondent-elles à une reprise de ce qui a été proposé pendant la formation (Q2) ? Peut-on observer la formulation personnelle de questions ouvertes (Q3) ? Le questionnement soutient-il la conceptualisation du fait linguistique enquêté, à savoir la présence ou absence de la virgule (Q4) ?

4. Dispositif

- 21 J'ai proposé à deux cohortes d'apprenants (groupe 1 : n = 9 ; groupe 2 : n = 5) un atelier d'une durée de six heures dédié à la virgule et à la prise en main de *Sketch Engine*. Les deux groupes ont été, à des moments différents de l'atelier, confrontés aux catégories de Boettcher. Le groupe 1, soumis à une approche déductive, a eu connaissance de ces références avant l'exploration de la virgule dans le corpus *Aufsätze* et le groupe 2 y a été exposé après la phase de travail en binômes. J'ai présenté l'approche déductive au groupe 1 sans la nommer explicitement.

Après la pause, je vais parler de la virgule. Ce sera un apport linguistique. [...] on peut [...] se dire, hein : [...] il y a des règles qui existent, des références, par exemple

dans le *Duden*, par rapport à la virgule, ou dans d'autres dans des grammaires, ou parce que je pense connaître une règle [...] et on peut [...] venir avec ces idées-là et ensuite tester leur adéquation dans le corpus. [...] C'est ce qu'on va faire tout à l'heure.

- 22 Le groupe 2 a obtenu deux informations par rapport à la démarche à suivre, la première pendant la session de formation technique et la deuxième, rappelée ci-dessous, directement avant le travail en binômes⁵.

C'est donc le moment où on va faire l'exploration du corpus hein *Aufsätze* [...] l'idée, c'est que [...] vous partiez de vos de vos:: représentations par rapport à la virgule, dans quelle situation on en a besoin/ ou pas besoin. Vous pouvez carrément l'écrire [...] pour vous mettre d'accord sur ce que vous avez envie de tester. à l'aide de ce corpus-là. Après, vous essayez [...] d'explorer. [...] si ça peut être utile quand vous explorez: le corpus. Sur *Edmodo*, je vous ai mis à disposition des requêtes-type en langage mathématique. [...] Donc, on veut savoir dans quelle situation il faut: la virgule en allemand, voilà, hein, bon. [...] Vous essayez de formuler parce qu'on sait que avec 132 règles, on ne peut pas fonctionner/. Vous essayez avec ce que vous observez là de trouver, hein, admettons, trois, quatre, cinq choses qui vous aideraient, VOUS, quand vous devrez hein écrire des choses à faire moins d'erreurs parce que vous en avez presque presque tous fait.

- 23 On peut trouver que cette approche n'est que partiellement inductive vu que les deux groupes avaient accès à trois modèles de requête CQL (*Corpus Query Language*), fournis sur *Edmodo* (Borg & O'Hara, 2008), l'espace numérique de travail du module. Cette aide technique met en quelque sorte à disposition des hypothèses possibles concernant la co-apparition de virgules et de conjonctions ou d'infinitives.

```
[word=","][tag="CONJ.*"]
```

```
[tag="CONJ.*"] : Cliquer sur "node forms" pour voir les formes les plus fréquentes.
```

```
[word=","][0,4][word="zu"][tag="VINF.Full"]
```

- 24 La première de ces formules permet de trouver des virgules suivies d'une conjonction et la dernière peut être utilisée quand on cherche une virgule qui est suivie, à une distance entre 0 et 4 mots, de *zu* et d'un verbe à l'infinitif, par exemple, *zu machen* (pour faire). La deuxième ligne d'aide incite à utiliser la fonctionnalité *Node forms* (formes du pivot⁶) qui permet d'obtenir une liste de fréquence de l'élément recherché, en l'occurrence des conjonctions.
- 25 Après la phase d'exploration du corpus en binômes, nous avons procédé à un échange oral en groupe entier afin de classer ensemble dans les catégories de Boettcher les occurrences observées dans le corpus. Nous y avons également classé une sélection d'énoncés erronés produits par les étudiants. Les scripteurs avaient ensuite l'occasion d'explorer individuellement le corpus dans le but de réviser leurs textes.

5. Analyse des données

- 26 La recherche a permis de créer un corpus d'apprenants⁷ d'écrits académiques de 7 634 mots en langue allemande, dont 12 comptes rendus et 22 commentaires. S'y ajoutent les transcriptions des deux entretiens finaux (2 259 et 4 353 mots), les annotations sélectives des films d'écran avec enregistrement audio des explorations de quatre binômes⁸ et les transcriptions des enregistrements audio des échanges en groupe entier. Ces échanges illustrent la façon dont les binômes s'emparent des outils de corpus tout en réfléchissant ensemble sur les caractéristiques des virgules. Les éléments pouvant soutenir des

observations quantitatives ont été rassemblés dans un fichier Excel afin de permettre des calculs avec R (R Development Core Team, 2008).

5.1. Types d'erreurs

- 27 Afin de pouvoir analyser les productions écrites du projet à l'aide de calculs faits avec R, j'ai choisi des catégories renseignant, entre autres, sur le genre textuel, la longueur des textes en termes de nombre de mots, le nombre de conjonctions de subordination, la nature d'erreur des virgules (absentes, à remplacer par un autre signe de ponctuation ou superflues) et les catégories d'erreur inspirées par Boettcher. « NA's » veut dire ici « absence de données ». Le terme de « territoire » désigne des groupes syntaxiques pouvant être bornés par une virgule ; « violation de territoire » signale la présence non pertinente d'une virgule au sein d'un tel groupe, et « territoire non marqué » veut dire qu'il y a absence d'une virgule à un endroit où elle est nécessaire pour marquer la limite entre deux groupes syntaxiques. La catégorie « double partiellement marqué » renseigne sur l'absence d'une des deux virgules d'une virgule double (*paariges Komma*), p. ex., pour encadrer une incise.

Tableau 1. – Aperçu des données des deux cohortes (n = 14).

```

genre_textuel  nombre_mots  conjonctions.de.subordination
commentaire :34  Min.   : 49.0  Min.   : 1.000
compte-rendu:81  1st Qu.:155.5  1st Qu.: 3.000
                Median :356.0  Median : 8.000
                Mean   :339.7  Mean   : 6.613
                3rd Qu.:519.0  3rd Qu.:11.000
                Max.   :572.0  Max.   :12.000
                NA's   :9

erreur_virgule  interprétation
absente       :71  double partiellement marqué: 2
signe erroné:10  incise non marquée       : 3
superflue    :34  jumeau marqué           : 8
                territoire non marqué   :76
                violation de territoire  :26

```

- 28 L'analyse des comptes rendus et des commentaires a permis de constater que les catégories « virgule absente » et « territoire non marqué » contiennent les scores d'erreur les plus élevés (respectivement 71 et 76 erreurs). Elles sont suivies des catégories « virgules superflues » (32 erreurs) et « violation de territoire » (26 erreurs). Les résultats concernant les virgules absentes et la lecture cursive des productions soutiennent la supposition que plus les apprentis scripteurs utilisent de conjonctions de subordination, plus ils ont tendance à faire des erreurs de virgule.

5.2. Types de requête

- 29 Comme précisé *supra*, explorer un corpus peut être associé à une quête ayant une valeur opérationnelle pour des apprenants d'une LE tout en étant scientifiquement défendable. Cela motive le repérage d'indices d'une démarche de recherche effectuée par les binômes, à savoir des questions explicites ou plus généralement des formulations signalant l'intention ou la réalisation d'une démarche d'exploration (Q1).

Tableau 2. – Aperçu des requêtes.

nombre_requêtes_	type_requête	nombre_occurrences.obtenues	formation.préalable
bertille_agnès:17	CQL	:24 0	:11 non:11
chantal_fanny : 7	CQL_context	: 1 100<	: 8 oui:27
germine_sarah :10	phrase	: 1 entre 1 et 5	: 2
habiba_laura : 4	simple	: 4 entre 10 et 49:	1
	simple_context:	8 entre 50 et 99:	3
		entre 6 et 49 :	5
		NA's	: 8

- 30 Pendant les 15 minutes d'exploration du corpus qui leur étaient allouées, les binômes ont réalisé entre 4 et 17 requêtes. 27 requêtes ont été effectuées par les membres du groupe 1 (Bertille et Agnès ; Germiné et Sarah⁹) soumis à une démarche déductive à travers une formation préalable sur les catégories de Boettcher. 11 requêtes ont été faites par les deux binômes du groupe 2 à qui on avait proposé une démarche inductive : Chantal et Fanny ainsi que Habiba et Laura avaient été invitées à formuler leurs propres questions et hypothèses.
- 31 On peut recenser en tout 24 requêtes de type CQL (*Corpus Query Language*) et 13 requêtes de type simple, c'est-à-dire, en langage naturel, dont 8 avec indication d'un contexte. 11 requêtes sur 37 ont mené à un résultat zéro, presque toujours en raison d'une erreur de formulation de la requête, et 8 (signalés dans le tableau comme « NA's ») n'ont pas été exécutées jusqu'au bout.

5.3. Types de questionnement des données

- 32 Dans les données, on peut distinguer trois types de démarche de questionnement des données, à savoir l'exploration ouverte, l'exploration fermée et l'exploration aléatoire.

5.3.1. Exploration ouverte

- 33 L'exploration ouverte correspond à une démarche de recherche très générale, comme la recherche de toutes les virgules dans le corpus :

Laura : Mais: on peut tout d'abord tout simplement . faire une recherche de la virgule et: voir ce que ça donne.

- 34 Ce mode peut être utile s'il mène par la suite à la précision de la question de travail et à un examen ciblé des données. Dans une certaine mesure, une démarche d'exploration ouverte est nécessaire pour obtenir des résultats inattendus : chercher toutes les conjonctions et voir lesquelles sont précédées d'une virgule peut renseigner sur le fait qu'en allemand, toutes les conjonctions introduisant un groupe subordonné sont précédées d'une virgule et que les conjonctions de coordination en sont dépourvues, sauf si elles ont un sens d'opposition comme *aber* (mais) et *sondern* (mais, au contraire), sens illustré dans cet exemple tiré du corpus *Aufsaetze*.

*So kann ein jeder Zwang ausüben, wenn er nur die wirtschaftlichen oder physischen Druckmittel dazu hat, **aber** Autorität ist nie garantiert, **sondern** muss gebildet werden!*

(Ainsi tout un chacun peut exercer une pression à condition d'avoir les moyens économiques ou physiques de coercition pour le faire **mais** l'autorité n'est jamais garantie — **au contraire**, elle doit être constituée !)

5.3.2. Exploration fermée

- 35 Le terme d'*exploration fermée* désigne des requêtes qui restreignent la recherche à des objets précis. Si une exploration cible de façon trop étroite son objet, elle risque de mener à un raisonnement circulaire : chercher des virgules directement suivis d'un pronom relatif permet simplement de constater qu'il y a un certain nombre d'occurrences de ce type dans le corpus.

Bertille : Maintenant, on peut regarder s'il y a une virgule avec les justement avec les groupes verbaux relatifs. [...] Effectivement, la virgule apparaît avant tous les pronoms relatifs.

- 36 Toutefois, des requêtes fermées ne sont pas forcément sans effet heuristique : dans l'exemple suivant, la longue liste d'occurrences obtenues confirme les représentations déjà là de Sarah et Germinie par rapport à la nécessité absolue de mettre une virgule avant les conjonctions de subordination en allemand.

Sarah : Donc, on va essayer: avec .. les conjonctions de SUBordination.
 <84 lignes de concordance s'affichent à l'écran.>
 (Sarah) : Alors=
 Germinie : =Ouh là !
 Sarah : Eh ben, là aussi, xxx
 Germinie : Et là, on voit que c'est o- [obligaTOire.]

5.3.3. Exploration aléatoire

- 37 Une démarche d'exploration aléatoire se traduit par des gestes qui ne semblent pas motivés par des questions pertinentes ou par des tâches reconnaissables. On observe des requêtes qui ne paraissent pas logiques, comme celle-ci effectuée par Fanny : <[word=","] [tag="CONJ.Sub.*"] & conjonction & déterminer>. Ici, l'outil est censé trouver une virgule directement suivie par une conjonction de subordination. Ces deux éléments sont combinés avec une conjonction et un déterminant apparaissant à une distance de 5 mots à leur gauche ou leur droite. Le film d'écran qui documente le travail de Chantal et de Fanny renseigne sur le fait que c'est cette dernière qui saisit, à un rythme effréné, toutes les requêtes. Fanny a tendance à cocher plusieurs options à la fois dont certaines sont incompatibles entre elles. Chantal reste en retrait ; elle se contente de faire des commentaires et des suggestions. Parfois, elle essaie de freiner l'élan de Fanny.

Fanny : Si on change(ait). On n'a qu'à mettre celui-là.
 <Fanny clique sur une commande de CQL préenregistrée qui permet de chercher les conjonctions de subordination précédées d'une virgule. Elle coche en même temps le contexte *conjunction* [conjonction] et *déterminer* [déterminant] ce qui ne permet pas de trouver d'occurrences.>
 Fanny : Cela avait marché tout à l'heure.
 Chantal : Là, on ne bouge plus.

5.3.4. Aperçu général des démarches adoptées

- 38 Les résultats présentés dans le tableau 3 permettent de constater que l'exploration fermée est la démarche la plus souvent adoptée, avec un déséquilibre important entre les deux groupes (25 fois par le groupe 1, soumis à l'approche déductive, et 4 fois par le groupe 2, soumis à l'approche inductive). L'exploration aléatoire ne concerne qu'un seul des quatre binômes (groupe 2), et les actions d'exploration ouverte, en principe les plus propices à faire découvrir de nouvelles choses, ont été entreprises par un des binômes du

groupe 1 et par les deux du groupe 2. La démarche de Germinie et Sarah (groupe 1) est la plus équilibrée : deux explorations ouvertes sont secondées de huit explorations fermées. Dans le script de leur interaction verbale, on constate davantage de moments de discussion métalinguistique que dans ceux des autres binômes.

Tableau 3. – Types d'exploration.

Binôme	Approche	Exploration ouverte	Exploration fermée	Exploration aléatoire	Total
Bertille, Agnès	déductive	0	17	0	17
Germinie, Sarah	déductive	2	8	0	10
Chantal, Fanny	inductive	2	1	4	7
Habiba, Laura	inductive	1	3	0	4
Total		5	29	4	38

- 39 Le test *Fisher's exact* fournit les scores de probabilité suivants pour un seuil de $p < 0,05$: 0,047 (exploration ouverte/fermée), 0,001 (exploration fermée/aléatoire), 0,444 (exploration ouverte/aléatoire). Les deux approches peuvent donc servir de variables indépendantes, sauf pour la corrélation « exploration ouverte/aléatoire ».

5.4. Interprétation des occurrences par les apprenants

5.4.1. Absence d'interprétation

- 40 Les observations des films d'écran avec son intégré montrent qu'une bonne partie des requêtes s'enchaînent sans qu'elles ne soient suivies d'observations précises des occurrences obtenues. En particulier Fanny paraît plusieurs fois pressée de formuler des références sans prendre le temps de réfléchir : « Alors là, on a des *und, doch, sondern, aber* (et, ou, mais). C'est quoi, la règle ? » Dans l'exemple ci-dessous, Germinie et Sarah recherchent tous les pronoms du corpus sans faire figurer la virgule dans la requête. Leur interprétation des occurrences obtenues ne se concentre toutefois pas sur les pronoms, sauf pour constater (à tort) une erreur d'étiquetage pour la forme *mich* (me). À la place, les deux observent d'autres phénomènes liés à la virgule, en l'occurrence l'incise qu'elles appellent ici « apposition ».

Sarah : [Si on] met pronom tout simple ? . Là . au lieu de mettre Rel, on met juste Pronom.

Germinie : °°On peut essayer.°°

Sarah : Ssupprimer/ .. Ah voilà, on a une sacré liste, donc tous les: pronoms/

Germinie : Oui. .. Enfin, pas TOUS. (Il y a) des exemples un peu bizarres avec *mich*.

Sarah : Parfois, c'est . des virgules qui sont ENTre crochets, tu vois, là, c'est *MEIner MEInung NACH/* (selon mon avis)

Germinie : Oui, c'est il y a (de nouveau une) virgule.

Sarah : C'est des appositions.

Germinie : Oui.

5.4.2. Formulation de références

- 41 D'autres situations illustrent un effort de formulation d'une référence.

Chantal : Donc, si on résume : on a déjà TROIS . raisons de mettre une virgule : devant une conjonction de subordination &=

Fanny : =Mhm.

Chantal : & devant une conjonction de coordination¹⁰, devant un pronom relatif=&

Fanny : Ouais.

Chantal : & qui introduit enfin .

Fanny : Okay.

- 42 Les éléments évoqués dans l'exemple ci-dessus ne correspondent toutefois pas à une découverte d'éléments auparavant inconnus. Germine et Sarah vont davantage dans cette direction en formulant une hypothèse pertinente par rapport à l'infinitive introduite par la forme *um* qui, effectivement, est toujours précédée d'une virgule (« *um etwas zu machen* » veut dire « pour faire quelque chose »). Il aurait été possible de rapprocher ce phénomène du concept « marquage du territoire » de Boettcher.

Germine : Il y a le : c'est quoi, *um zu* ? Une proposition infinitive ?

Sarah : En fait, *um zu*, c'est devant le *um*, enfin c'est devant le *um* qu'il y a peut-être une virgule.

5.4.3. Références non fondées sur les données du corpus

- 43 Laura (tout comme Agnès) sort du cadre de l'observation empirique des données du corpus pour choisir la prosodie comme cadre de référence : selon elle, la virgule correspondrait à une pause repérable à l'oral.

Laura : Je pense que la virgule, en fait, elle est assez claire : il suffit de de de se dire la phrase, en fait, à l'oral [...] parce que je pense qu'on fait automatiquement une pause quand quand il y a une virgule.

- 44 L'exemple précédant soutient l'idée que la démarche inductive peut amener les apprenants à avoir recours à des références antérieures dont certaines sont moins pertinentes que celles offertes lors d'une démarche déductive.

- 45 Une autre façon d'éviter l'emploi des outils de corpus peut être observée par Fanny qui cherche plusieurs fois sur Internet des réponses à ses questions : elle vérifie la définition de conjonction de coordination et de subordination, et elle cherche le terme métalinguistique correspondant à *um zu* (pour).

Fanny : Comment on appelle ça, *um zu* ? De toute façon, il faut que je révise ça . pour le CAPES¹¹.

- 46 Fanny va jusqu'à formuler dans son moteur de recherche Internet la requête « Utilisation de la virgule en allemand » sans toutefois la mener jusqu'au bout.

6. Discussion des résultats

- 47 Les types de questionnement qui peuvent être recensés dans les données de cette expérience varient entre questions de recherche motivées et démarches tâtonnantes ou aléatoires. Dans les transcriptions des discussions faites pendant l'exploration du corpus, on trouve quelques traces de questionnements susceptibles de mener à la découverte de nouveaux savoirs ou à la remise en question de représentations erronées. Bertille émet l'hypothèse suivante : « Il n'y a peut-être jamais de verbe après (une virgule) », hypothèse qu'elle n'arrive toutefois pas à transformer en une requête sur corpus. On peut observer quelques tentatives potentiellement fructueuses de partir d'un concept, en l'occurrence l'énumération et l'incise.

Chantal : On fait quoi maintenant ?

Fanny : Ben, il faut qu'on trouve l'énumération mais: je ne sais pas comment faire.

- 48 Sarah et Germinie visent à enquêter sur le phénomène d'incise qu'elles appellent « apposition ». Elles ne trouvent pas l'étiquette *RFTagger* qui y correspond, ce qui est normal vu qu'une incise se compose de plusieurs mots ; le *RFTagger* s'applique uniquement aux lexèmes simples.

Sarah : Ah oui. Comment on va trouver apposition ?
 <Parcourt la page d'aide de *Sketch Engine* contenant les étiquettes pour *RFTagger*>
 Sarah : Appr ah oui. Euh: ..
 Germinie : Appo c'est quoi ?
 Sarah : Tu vois Appo ?
 Germinie : Oui, là. Je ne sais pas ce que c'est.
 Sarah : On va essayer. On va bien voir.
 Germinie : ça ne marche pas [...] Parce qu'il y a zu là-haut.
 Sarah : Oui.
 <Efface zu dans le champ de recherche non pertinent.>
 Sarah : Non, ça ne marche pas. Apposition, on ne peut pas trouver.

- 49 La plupart du temps, les questionnements ressemblent davantage à la formulation d'une tâche assez peu motivée par une quête susceptible de créer des connaissances : il n'est pas plausible de voir un lien entre l'apparition d'une virgule et d'une préposition, éléments que tentent de rechercher Agnès et Bertille, ou un lien entre la coprésence d'une virgule et d'un déterminant, combinaison ayant fait l'objet d'une requête de Chantal et Fanny. Plusieurs fois, les étudiantes combinent la virgule avec une classe de mot qui semble avoir été choisie au hasard, apparemment sans rapport avec les difficultés que l'on peut avoir quand on doit placer les virgules en allemand. Les moments les plus propices à la conceptualisation ont été les discussions métalinguistiques entamées par deux des binômes. L'une des discussions entre Chantal et Fanny part d'un désaccord par rapport au terme de coordonnant que Fanny utilise pour désigner les conjonctions de subordination. Elle ne semble pas avoir une conscience précise de la différence entre les conjonctions de subordination, les conjonctions de coordination et les prépositions. Pendant l'interaction verbale, Fanny passe d'un terme à l'autre pour se référer aux mêmes phénomènes.

Chantal : Alors, *wenn, dass, ganz, ob, weil*= (quand, que, tout à fait, si, parce que)
 <Fanny coupe Chantal qui avait commencé à regarder les occurrences pour stipuler une règle.>
 Fanny : =La première règle, c'est que dès qu'il y a une subordonnée, la virgule se trouve avant le subordonnant/ [...]
 Chantal : Comment est-ce que ça se fait ? Je n'appelle pas ça subordonnant, moi.
 Fanny : Le mot ? Le mot, on appelle ça un subordonnant. Le *dass*, c'est un subordonnant.
 Chantal : Ah bon ?
 Fanny : [...] C'est pas des conjonctions de coordination, mais où est donc or ni car ?
 Chantal : J'crois hein.
 <Fanny commence alors à chercher sur Internet des informations sur les conjonctions de coordination et de subordination en langue française.>
 Fanny : Donc, là, nous sommes dans la coordination. [...]
 <Fanny se trompe ici.>
 Chantal : [Il y a les deux] puisqu'en allemand, c'est: les deux.

- 50 Chantal clôt la discussion en mettant le doigt sur le fait que certains des éléments observés sont pluricatégoriels (Boutet & Deprez, 2002). Ce fait est également l'objet d'une discussion métalinguistique menée entre Germinie et Sarah. Germinie s'étonne de découvrir *ohne* (sans) dans la liste des occurrences des conjonctions de subordination.

Germinie : *ohne* c'est pas une conjonction de subordination, non ? . Ah mais c'est parce qu'il y a zu ? Ah non.

Sarah : Mais SI. à, dans, par, pour, envers, de, sans, sous . ça, c'est les conjonctions de subordination.

Germine : =[Ah non,]

Sarah : c'est des conjonctions [de coordination]

Germine : *ohne*, °°c'est (pas)°° une conjonction de (léger rire)) subordination.

Sarah : xxx

Germine : On va le: garder en tête.

Sarah : Oui.

Germine : On va (le noter.)

Sarah : Oui, tu peux noter que: que: hein, on se posé la question du *ohne*.

Germine : Oui. Hein .

- 51 Dans cette interaction, un flottement catégoriel peut être constaté pour Sarah. En effet, la préposition *ohne* (sans) peut apparaître au sein de la locution *ohne dass* (sans que) introduisant un groupe conjonctionnel et dans *ohne zu* (sans) introduisant un groupe infinitif. Un questionnement comparable se trouve dans les interactions verbales de Germine et Sarah et de Chantal et Fanny quand elles cherchent le métaterme correspondant à *um zu* (pour).

7. Conclusion

- 52 Cette recherche a examiné la façon dont de futurs enseignants d'allemand se sont emparés d'outils de corpus qui leur ont été présentés comme des instruments pouvant soutenir la conceptualisation de la virgule allemande. L'exploration de ce fait linguistique dans le corpus a eu comme conséquence que les apprenants réfléchissent, entre autres, sur la structuration du discours et sur le rôle que jouent les conjonctions pour cette structuration. Dans les lignes qui suivent, je présenterai les résultats que la recherche a permis d'obtenir par rapport aux quatre questions de travail formulées en section 3.5.

- 53 L'analyse des données permettant de répondre à la première des questions (« Qu'est-ce que les enregistrements de l'écran et de la conversation des binômes lors de la prise en main de l'outil peuvent nous apprendre sur la démarche de recherche sur corpus ? ») a mené à l'identification de trois niveaux d'exploration du corpus, à savoir l'exploration ouverte, fermée et aléatoire. Il s'est avéré que la combinaison des deux premiers types d'exploration peut soutenir une démarche de réflexion sur la langue. Je me suis attendue à ce que l'aspect technique de la formulation des requêtes pose problème aux apprenants ; ce n'était pas vraiment le cas même si certaines difficultés subsistent.

Enquêtrice : (Cela) ne vous a pas perturbé de devoir taper comme ça un langage pas naturel ?

Laura : Non, non. Mais c'est très simple, une fois [...] qu'on trouve les codes, on sait les employer.

- 54 Par contre, la formulation des questions de travail a parfois été laborieuse. Lors de l'entretien du groupe 1, Sarah et Agnès thématisent cette difficulté en mentionnant deux conditions différentes pour interroger un corpus de façon pertinente, la nécessité d'avoir un doute par rapport à un fait linguistique et d'être au clair par rapport à sa question de travail.

Sarah : [Il faut] avoir un doute ou savoir qu'on a fait une faute=

Agnès : Il faut avoir une question précise, sinon . si c'est vague, on ne trouve pas.

- 55 Pour la deuxième question de recherche (« Dans quelle mesure les questions correspondent-elles à une reprise de ce qui a été proposé pendant la formation ? »), les

résultats suivants sont disponibles : les catégories de Boettcher n'ont pas été nommément citées dans les explorations du corpus par le groupe 1 soumis à une approche déductive. Les requêtes préformatées, publiées sur *Edmodo*, ont été reprises par tous. Les entretiens contiennent des observations sur les méthodes d'investigation proposées aux étudiantes ; les approches deductives et inductives sont discutées dans l'entretien du groupe 2 soumis à une approche inductive.

56 La troisième question s'est intéressée aux traces de formulations personnelles de questions ouvertes permettant d'enquêter sur les virgules allemandes. J'ai identifié peu de questions ouvertes explicitement dédiées aux virgules mais les discussions métalinguistiques des binômes contiennent des questions spontanées concernant d'autres domaines, comme la catégorie syntaxique à laquelle appartient un terme donné.

57 J'ai également cherché à savoir dans quelle mesure le questionnement soutient la conceptualisation du fait linguistique enquêté. Sur la base des données explorées pour cet article, il n'est pas possible de donner une réponse détaillée à cette question ; l'analyse des transcriptions des séances plénières et les révisions individuelles pourraient apporter des éléments de réponse plus précis. Les entretiens attestent toutefois d'un début de prise de conscience par rapport à l'usage de la virgule en allemand.

Habiba : <Parler des virgules,> ça nous a éclairées sur la façon dont on utilise la la virgule . en allemand puisque ce n'est pas forcément un thème qu'on aborde . pendant notre: cursus hein scolaire ou universitaire [...] et puis on: maintenant (on) fait plus attention.

58 En section 3.3.3, la formulation d'une question de travail a été présentée comme élément central d'une activité de recherche. On peut noter qu'avant le début du travail à deux, l'enseignante donne comme envoi sa propre question, même au groupe 2 soumis à une approche inductive : « Donc, on veut savoir dans quelle situation il faut: la virgule en allemand ». On peut se demander si cette façon de présenter le travail a trop poussé les participantes à faire des descriptions linguistiques de phénomènes qui leur étaient familiers. Il aurait peut-être été judicieux d'encourager les apprenantes — déjà pendant cette phase d'interrogation et non seulement pendant la phase de révision — à partir de leurs propres erreurs et de chercher des solutions dans le corpus. Comme cette recherche l'a illustré, traduire un besoin d'apprentissage en une question et ensuite en une requête sur corpus n'est toutefois pas un geste banal. Selon Amadieu et Tricot (2014, p. 29), plus on a de connaissances dans un domaine, plus on est en mesure de les augmenter de façon autonome à l'aide d'un outil numérique. On peut en conclure que l'approche déductive offerte à des apprenants, explorateurs de corpus, peut être un instrument heuristique pertinent s'il est proposé en parallèle avec des démarches de découverte plus libres. Il reste à démontrer, à l'aide d'analyses complémentaires, dans quelle mesure ces approches mènent à la révision pertinente de productions d'écrits académiques.

BIBLIOGRAPHIE

- ÄDEL, Annelie. (2010). Using Corpora to Teach Academic Writing: Challenges for the Direct Approach. Dans M. C. Campoy-Cubillo, B. Belles-Fortuño & M. L. Gea-Valor (dir.), *Corpus-Based Approaches to English Language Teaching* (p. 39-55). Londres, New York : Continuum.
- AMADIEU, Frank & TRICOT, André. (2014). *Apprendre avec le numérique. Mythes et réalités*. Paris : Retz.
- BERNARDINI, Silvia. (2004). Corpora in the Classroom. An Overview and Some Reflections on Future Developments. Dans J. Sinclair (dir.), *How to Use Corpora in Language Teaching* (p. 15-36). Amsterdam, Philadelphie : John Benjamins Publishing Company.
- BOCH, Françoise & FRIER, Catherine. (2015). Travailler le texte : ponctuation, anaphores et collocations. Dans F. Boch & C. Frier (dir.), *Écrire dans l'enseignement supérieur : des apports de la recherche aux outils pédagogiques* (p. 53-109). Grenoble : UGA Éditions.
- BOETTCHER, Wolfgang. (2016). Komma & Co unter dem Kooperationsprinzip: Interpunktionslernen im Kompetenzbereich "Schreiben und Sprachreflexion". Dans R. Olsen, C. Hochstadt & S. Colombo-Scheffold (dir.), *Ohne Punkt und Komma... Beiträge zu Theorie, Empirie und Didaktik der Interpunktion* (p. 326-361). Berlin : RabenStück Verlag.
- BOULTON, Alex & TYNE, Henry. (2014). *Des documents authentiques aux corpus*. Paris : Didier.
- BOUTET, Josiane & DEPREZ, Christine. (2002). Ici et là-bas, public et privé. Des catégories à interroger. *Cahiers de l'Institut de linguistique de Louvain*, 28(3/4), 37-46.
- BREDEL, Ursula & PRIMUS, Beatrice. (2007). Komma & Co: Zwiegespräch zwischen Grammatik und Performanz. *Zeitschrift für Sprachwissenschaft*, 26, 81-131. <<https://doi.org/10.1515/ZFS.2007.006>>.
- CHAMBERS, Angela. (2005). Integrating Corpus Consultation in Language Studies. *Language Learning & Technology*, 9(2), 111-125. Disponible en ligne sur <www.lltjournal.org/item/2509>.
- CHARAUDEAU, Patrick. (2009). Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. *Corpus*, 8, 37-66. Disponible en ligne sur <<https://journals.openedition.org/corpus/1674>>.
- CHENG, Winnie, WARREN, Martin & XUN-FENG, Xu. (2003). The Language Learner As Language Researcher: Putting Corpus Linguistics on the Timetable. *System*, 31(2), 173-186.
- DE COCK, Sylvie & TYNE, Henry. (2014). Corpus d'apprenants et acquisition des langues. *Recherches en didactique des langues et des cultures - Les Cahiers de l'Acedle*, 11(1). <<https://doi.org/10.4000/rdlc.1716>>.
- DIEHL, Erika, PISTORIUS, Hannelore & FAYOLLE DIETL, Annie. (2002). Grammatikerwerb im Fremdsprachenunterricht - Ein Widerspruch in sich? Dans W. Börner & K. Vogel (dir.), *Grammatik und Fremdspracherwerb. Kognitive, psycholinguistische und erwerbstheoretische Perspektiven* (p. 143-163). Tübingen : Narr.
- ELUERD, Roland. (2017). *La Ponctuation française. Règles, usages et plaisir du texte*. Paris : Garnier.
- FAVRIAUD, Michel. (2011). Approches nouvelles de la ponctuation, diachroniques et synchroniques. *Langue française*, 4(172), 3-18. <<https://doi.org/10.3917/lf.172.0003>>.
- GASKELL, Delian & COBB, Thomas. (2004). Can Learners Use Concordance Feedback for Writing Errors? *System*, 32(3), 301-319.

- GOMBERT, Jean-Émile. (1996). Activités métalinguistiques et acquisition d'une langue. *Acquisition et interaction en langue étrangère (Aile)*, 8, 41-55. Disponible en ligne sur <<https://journals.openedition.org/aile/1224>>.
- GOULLIER, François. (2016). *Rapport de jury CAPES Externe, section Allemand, session 2016*. Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Disponible en ligne sur <http://media.devenirenseignant.gouv.fr/file/externe/53/7/rj-2016-capex-externe-allemand_628537.pdf>.
- GROBET, Anne & VUKSANOVIĆ, Ivana. (2017). Variations de l'élaboration conceptuelle dans différentes modalités d'enseignement bilingue. *Bulletin VALS-ASLA (Bulletin suisse de linguistique appliquée)*, vol. spécial (2), 89-100.
- HIDDEN, Marie-Odile, ALDAY, Maria Victoria, PORTINE, Henri & SHAO, Baoqing. (2016). La ponctuation en langue étrangère peut-elle devenir un objet d'apprentissage ? Dans S. Pétilion, F. Rinck & A. Gautier (dir.), *La ponctuation à l'aube du XXI^e siècle. Perspectives historiques et usages contemporains* (p. 231-250). Limoges : Lambert Lucas.
- JOHNS, Tim. (1991). Should You Be Persuaded: Two Samples of Data-Driven Learning Materials. Dans T. Johns & P. King (dir.), *Classroom Concordancing, English Language Research Journal*, 4, 1-13.
- KENNEDY, Claire & MICELI, Tiziana. (2001). An Evaluation of Intermediate Student's Approaches to Corpus Investigation. *Language Learning & Technology*, 5(3), 77-90. Disponible en ligne sur <www.lltjournal.org/item/3045>.
- KENNEDY, Claire & MICELI, Tiziana. (2016). Cultivating Effective Corpus Use by Language Learners. *Computer Assisted Language Learning*, 30(1-2), 91-114. <<http://dx.doi.org/10.1080/09588221.2016.1264427>>.
- KIRCHHOFF, Frank & PRIMUS, Beatrice. (2016). Das Komma im mehrsprachigen Kontext [La virgule en contexte multilingue]. Dans R. Olsen, C. Hochstadt & S. Colombo-Scheffold (dir.), *Ohne Punkt und Komma... Beiträge zu Theorie, Empirie und Didaktik der Interpunktion* (p. 78-97). Berlin : RabenStück Verlag.
- LANDURE, Corinne. (2013). Corpus et dictionnaires : prise de conscience linguistique chez des apprenants Lansad. *Bulletin VALS-ASLA (Bulletin suisse de linguistique appliquée)*, 97, 59-76. Disponible en ligne sur <http://doc.rero.ch/record/11876/files/bulletin_vals_asla_2013_097.pdf>.
- LEBLAY, Christophe. (2014). Les écritures intermédiaires réflexives en littérature avancée. *Le français aujourd'hui*, 1(184), 103-115.
- LEJOT, Ève. (2017). La relecture entre pairs en formation doctorale : de l'analyse des commentaires à l'élaboration d'une grille d'accompagnement. *Lidil*, 55. Disponible en ligne sur <<https://journals.openedition.org/lidil/4255>>.
- LEŃKO-SZYMAŃSKA, Agnieszka. (2017). Training Teachers in Data-Driven Learning: Tackling the Challenge. *Language Learning & Technology*, 21(3), 217-241. <www.lltjournal.org/item/3017>.
- MONDADA, Lorenza. (2005). *Chercheurs en interaction. Comment émergent les savoirs*. Lausanne : Presses polytechniques et universitaires romandes.
- MONTAGNE-MACAIRE, Dominique. (2007). Didactique des langues et recherche-action. *Les Cahiers de l'Acedle*, 4, 93-120.
- NARCY-COMBES, Jean-Paul. (2005). *Didactique des langues et TIC : vers une recherche-action responsable*. Paris : Éditions Ophrys.

- O'SULLIVAN, Íde. (2010). Using Corpora to Enhance Learners' Academic Writing Skills in French. *Revue française de linguistique appliquée*, 2(15), 21-35.
- PINCEMIN, Bénédicte. (2006). Concordances et concordanciers – De l'art du bon KWAC. Dans C. Duteil-Mougel & B. Foulquié (dir.), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation : actes du 17^e colloque d'Albi Langages et Signification* (p. 33-42). Texto! Disponible en ligne sur <www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes_ALBI-06.pdf>.
- PINTO, Marie-Antonietta & EL EUCH, Sonia. (2015). *La conscience métalinguistique. Théorie, développement et instruments de mesure*. Laval : Presses de l'Université Laval.
- Regeln und Wörterverzeichnis*. (2006). Munich et Mannheim : IDS Mannheim. Disponible en ligne sur <www1.ids-mannheim.de/fileadmin/service/reform/regeln2006.pdf>.
- RINCK, Fanny. (2011). Former à (et par) l'écrit de recherche. Quels enjeux, quelles exigences ? *Le français aujourd'hui*, 3(174), 79-89. <<https://doi.org/10.3917/lfa.174.0079>>.
- SAPPOK, Christopher. (2011). *Das deutsche Komma im Spiegel von Sprachdidaktik und Prosodieforschung - Forschungslage - "Parsing vs. Phrasing" - Experimente*. Berlin, Münster : LIT.
- SCHAEFFER-LACROIX, Eva. (2015). Impact de discussions métalinguistiques sur l'apprentissage de la production écrite en allemand, langue étrangère. Dans S. Garnier, F. Rinck, F. Sitri & S. De Vogüe (dir.), *Former à l'écrit universitaire, un terrain pour la linguistique ?*, *Linx (Revue des linguistes de l'université Paris Ouest Nanterre La Défense)*, 72, 193-211.
- STEGU, Martin. (2008). Linguistique populaire, language awareness, linguistique appliquée : interrelations et transitions. *Pratiques*, 139-140, 81-92. Disponible en ligne sur <<http://pratiques.revues.org/1193>> ; <<https://doi.org/10.4000/pratiques.1193>>.
- ST JOHN, Elke. (2001). A Case for Using Parallel Corpus and Concordancer for Beginners of a Foreign Language. *Language Learning & Technology*, 5(3), 185-203. Disponible en ligne sur <www.lltjournal.org/item/3054>.
- SWAIN, Merrill. (2006). Languaging, Agency and Collaboration in Advanced Second Language Learning. In H. Byrnes (dir.), *Advanced Language Learning: The Contributions of Halliday and Vygotsky* (p. 95-108). Londres : Continuum.
- TSUI, Amy Bik May. (2004). What Teachers Have Always Wanted to Know—and How Corpora Can Help. Dans J. McH. Sinclair (dir.), *How to Use Corpora in Language Teaching* (p. 39-61). Amsterdam, Philadelphie : John Benjamins Publishing.
- TYNE, Henry. (2013). Corpus et apprentissage-enseignement des langues. *Bulletin VALS-ASLA (Bulletin suisse de linguistique appliquée)*, 97, 7-15. Disponible en ligne sur <http://doc.rero.ch/record/11876/files/bulletin_vals_asla_2013_097.pdf>.

Sites, outils et corpus

- BIBLIOGRAPHISCHES INSTITUT GMBH. (2017). *Duden*. Dictionnaire électronique. Disponible en ligne sur <www.duden.de>.
- BIG NERD SOFTWARE. (n. d.). *ScreenCast-O-Matic*. Enregistreur audiovisuel d'écran.
- BORG, Nic & O'HARA, Jeff. (2008). *Edmodo*. Plateforme d'apprentissage. Disponible en ligne sur <www.edmodo.com>.
- KILGARRIFF, Adam, RYCHLY, Pavel & POMIKALEK, Jan. (n. d.). *Sketch Engine*. Système de gestion de corpus. Disponible en ligne sur <www.sketchengine.co.uk>.

SCHAEFFER-LACROIX, Eva. (2014). *Aufsätze*. Corpus d'experts.

SCHMID, Helmut & LAWS, Florian. (2008). Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-Grained POS Tagging. Dans *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)* (p. 777-784). Manchester, Angleterre. Disponible en ligne sur <www.cis.uni-muenchen.de/~schmid/papers/Schmid-Laws.pdf>.

R DEVELOPMENT CORE TEAM. (2008). *R: A Language and Environment for Statistical Computing*. Vienne : R Foundation for Statistical Computing. Disponible en ligne sur <www.R-project.org>.

ANNEXES

Conventions de transcription

Extrait de *Mondada* (2005, p. 31-34).

Les locuteurs non identifiés sont notés par un x.

[note le début du chevauchement entre les locuteurs.

] note la fin du chevauchement, lorsque cela a été jugé nécessaire.

= note un enchaînement rapide entre deux tours de parole.

& note la continuation du tour par le même locuteur, au-delà de l'interruption de la ligne de la transcription pour l'introduction d'un chevauchement par un autre locuteur.

. et .. et ... notent des pauses petites, moyennes, longues.

: et :: notent des allongements syllabiques (de manière iconique par rapport à la durée, on pourra avoir : ou ::).

- note la troncation d'un mot esquissé.

/ et \ notent les montées et descentes intonatives.

Les CAPITALES notent une accentuation.

Les degrés ° ° et ° ° ° notent un volume bas, murmuré, de la voix (° ° ° pour un volume très bas).

< > délimitent le segment sur lequel porte un commentaire.

xxx notent un segment incompréhensible.

(ravère) note un segment incertain.

NOTES

1. Exemple d'une requête sur corpus : pour trouver des virgules suivies d'une conjonction dans un corpus déposé sur *Sketch Engine*, on formulera la ligne de commande suivante : [word=",",] [tag="CONJ.*"].

2. Tyne (2013) traduit ce sigle par « apprentissage conduit ou porté par les données ».

3. T (*the teacher*) ; E (l'enseignant).

4. Dans ses termes : « *die Territorien der beteiligten Prädikate [...] markieren* » (Boettcher, 2016, p. 337).

5. Les conventions de transcription sont celles proposées dans Mondada (2005, p. 31-34) ; voir annexe.
 6. Le pivot est « le mot ou motif linguistique dont on veut étudier les occurrences en contexte » (Pincemin, 2006).
 7. Pour plus de précisions, voir De Cock et Tyne (2014).
 8. Il n'y a que quatre binômes car cinq personnes sur quatorze n'ont pas pu être présentes pendant la phase d'exploration. La neuvième personne présente a perdu son film d'écran.
 9. Prénoms anonymisés.
 10. Chantal se trompe ici : la plupart des conjonctions de coordination s'utilisent sans virgule.
 11. CAPES : Certificat d'aptitude au professorat de l'enseignement secondaire.
-

RÉSUMÉS

Le concept d'apprenant chercheur attribue à l'apprenant les traits d'un linguiste cherchant à comprendre le fonctionnement de la langue. Souvent mentionné dans le contexte de l'utilisation des corpus pour l'enseignement-apprentissage d'une langue étrangère, ce concept nécessite de la part des apprenants la capacité de formuler une question portant sur un fait linguistique et de transformer cette question en une requête sur corpus. Afin d'observer la façon dont de telles compétences peuvent se mettre en place, une expérience a été proposée à deux groupes de futurs enseignants d'allemand ayant produit des écrits académiques pendant un module de didactique. L'analyse d'erreurs appliquée à leurs textes a motivé le choix de proposer à ce public un travail de conceptualisation de la virgule. Lors d'un atelier de formation linguistique et technique d'une durée de 6 heures, les étudiants avaient l'occasion d'observer la virgule dans un corpus d'experts. Ils étaient regroupés en binômes afin d'être incités à une interaction verbale pendant leur travail d'exploration. Pour l'un des groupes, une approche déductive a été choisie et pour l'autre une approche inductive. L'analyse des données a permis de déterminer trois types d'exploration du corpus par les apprenants et d'établir un lien entre leurs modes d'investigation et le traitement spécifique auquel ils ont été soumis. Les interactions verbales et les films d'écrans contiennent des indices selon lesquels l'approche déductive a, davantage que l'approche inductive, mené à des démarches s'apparentant à une enquête scientifique.

The concept of the learner as a researcher assigns the learner features of a linguist striving to understand the functioning of language. This concept, often mentioned within the context of corpus use for foreign language learning and teaching, implies the learner's capacity to formulate questions about a given language feature and to transform these questions into corpus queries. In order to observe how related competencies can be set up, a research project was designed for future German teachers. Two groups were invited to write academic texts during a Master's degree course in applied linguistics. The error analysis of their texts revealed that the students needed assistance in the use of commas in German. A one-day workshop was dedicated to linguistic and technical training focusing on the comma and on corpus tools. Grouped into pairs in order to stimulate verbal interaction, the students explored the comma in a specialised corpus. A deductive approach was instigated for half of the group and an inductive approach for the other half. The research data fell into three different exploration modes. The interactions and the screen casts of the deductive approach group contain stronger evidence of scientific investigation methods and results than those of the inductive approach group.

INDEX

Mots-clés : apprenant chercheur, écrit académique, corpus, question, virgule, allemand

Keywords : learner as researcher, academic writing, corpora, inquiries, comma, German

AUTEUR

EVA SCHAEFFER-LACROIX

Sorbonne Université – ESPÉ de Paris

schaeffer-lacroix@orange.fr

Eva Schaeffer-Lacroix est maître de conférences en études germaniques à Sorbonne Université – ESPÉ de Paris. Ses recherches portent sur l'usage des technologies pour l'enseignement-apprentissage des langues étrangères. Elle s'intéresse en particulier aux corpus numériques, à la production écrite et à la réflexion sur la langue.