SE SERVIR DE CORPUS D'APPRENANTS ET DE CORPUS D'EXPERTS POUR FORMER EN LANGUE ÉTRANGÈRE

Eva Schaeffer-Lacroix
Sorbonne Université – Espé de l'académie de Paris

Journée d'étude "Lier recherche et formation – Des corpus à l'interface". Université Clermont Auvergne.

Plan

- Méthodologie
- Cycle de recherches incluant les corpus
- Obstacles d'enseignement-apprentissage
- Apports des corpus d'experts
- Apports des corpus d'apprenants
- Scénario *Audiodeskription*
- Scénario *Buettenwarder*
- À faire par la suite

Recherche orientée par la conception (RoC)

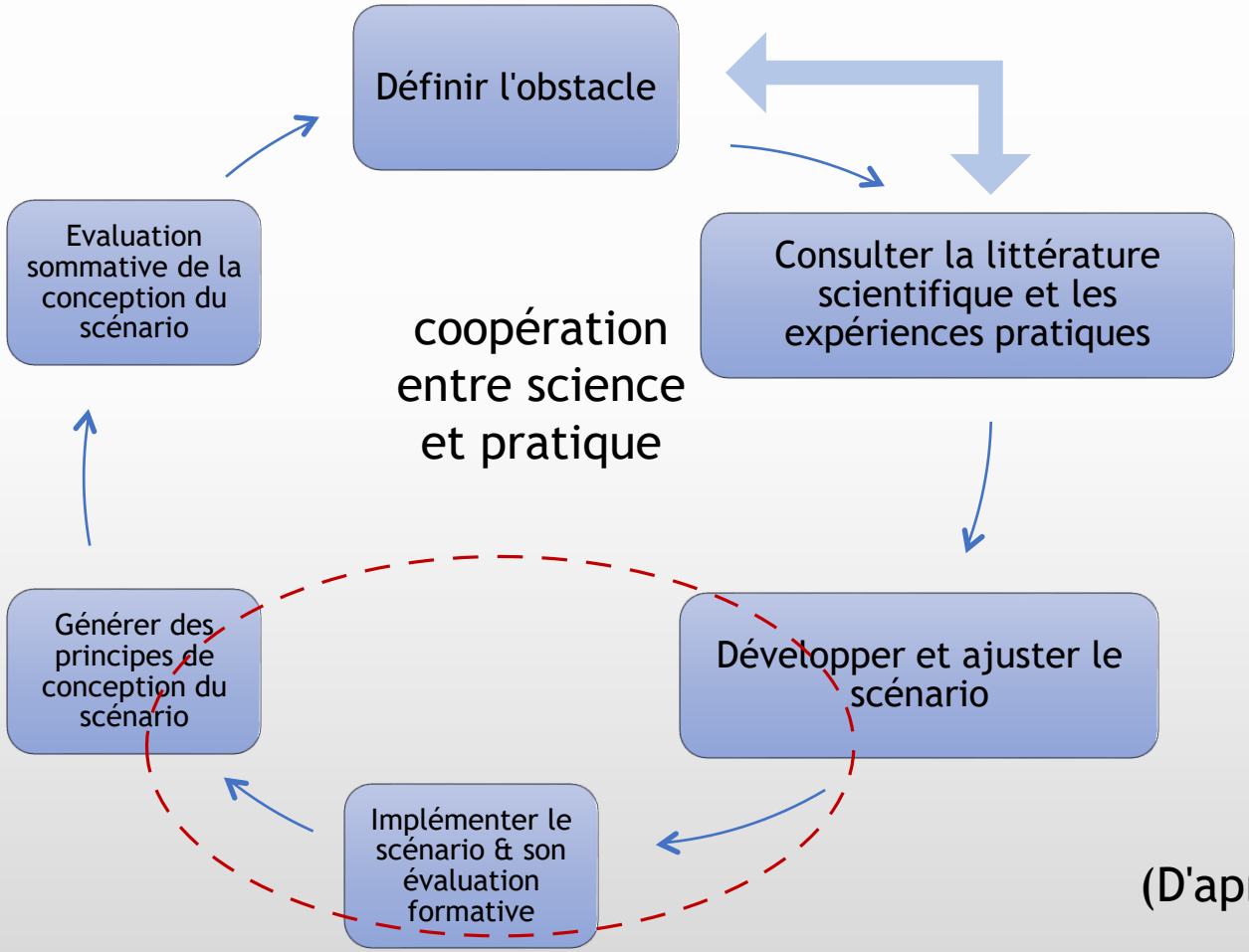
- Définition :

- "[La RoC est] l'étude systématique de concevoir, développer et évaluer des interventions pédagogiques (telles que […] des matériels, produits et systèmes) pouvant contribuer à résoudre des problèmes complexes qui se manifestent dans la pratique pédagogique" (Plomp, 2007 : 13).

- Objectifs (Sanchez & Monod-Ansaldi, 2015) :

- "produire, dans le champ de l’éducation, des connaissances ou de lever des incertitudes théoriques (recherche)".

Phases d'un processus de type RoC

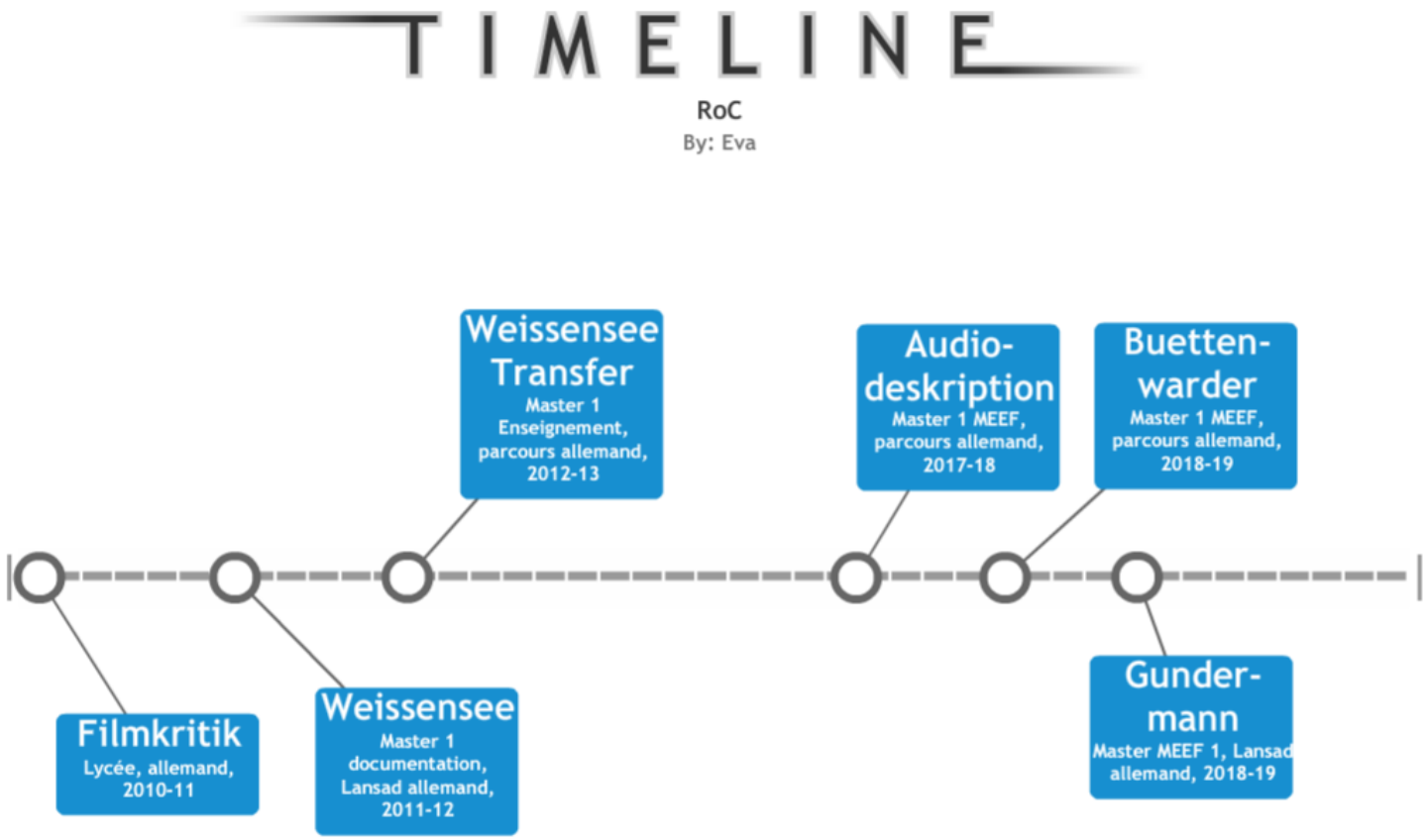


(D'après Euler, 2014)

Projets incluant des corpus

Items:

- **Filmkritik**
  Ecrire un mini-script de film ; réviser à l'aide du corpus Filmlokal
- **Weissensee**
  Ecrire une scène pouvant remplacer ou continuer une scène du script de la série Weissensee (Hess, 2010) ; corpus Weissensee (monolingue ; 1er épisode : bilingue allemand) - français
- **Weissensee Transfer**
  Ecrire un scénario incluant au moins une activité faisant appel au corpus Weissensee
- **Audiodeskription**
  Ecrire un scénario incluant une activité permettant de réaliser une audiodescription
- **Buettenwarder**
  Ecrire un scénario incluant des activités sur le corpus Buettenwarder (manuscrits d'audiodescription)
- **Gundermann**
  Ecrire l'audiodescription d'une des scènes du film Gundermann (Dresen, 2018), en se servant du corpus Buettenwarder (manuscrits d'audiodescription)



Conditions pour un usage des corpus par les futurs enseignants de langue

- Faire comprendre que les corpus aident à surmonter des obstacles d'apprentissage
- Aider à se servir d'outils de corpus suffisamment conviviaux
- Aider à construire des scénarios et des activités faisant appel aux corpus

Obstacles identifiés lors d'un stage d'observation dans le secondaire

Journal de bord de Mira

- Matériaux d'apprentissage peu pertinents
- Activités inutiles
- Savoirs insuffisants sur la langue (son fonctionnement et la métalangue)
- Surcharge cognitive des apprenants

Définir l'obstacle

Conditions d'apprentissage d'une langue étrangère

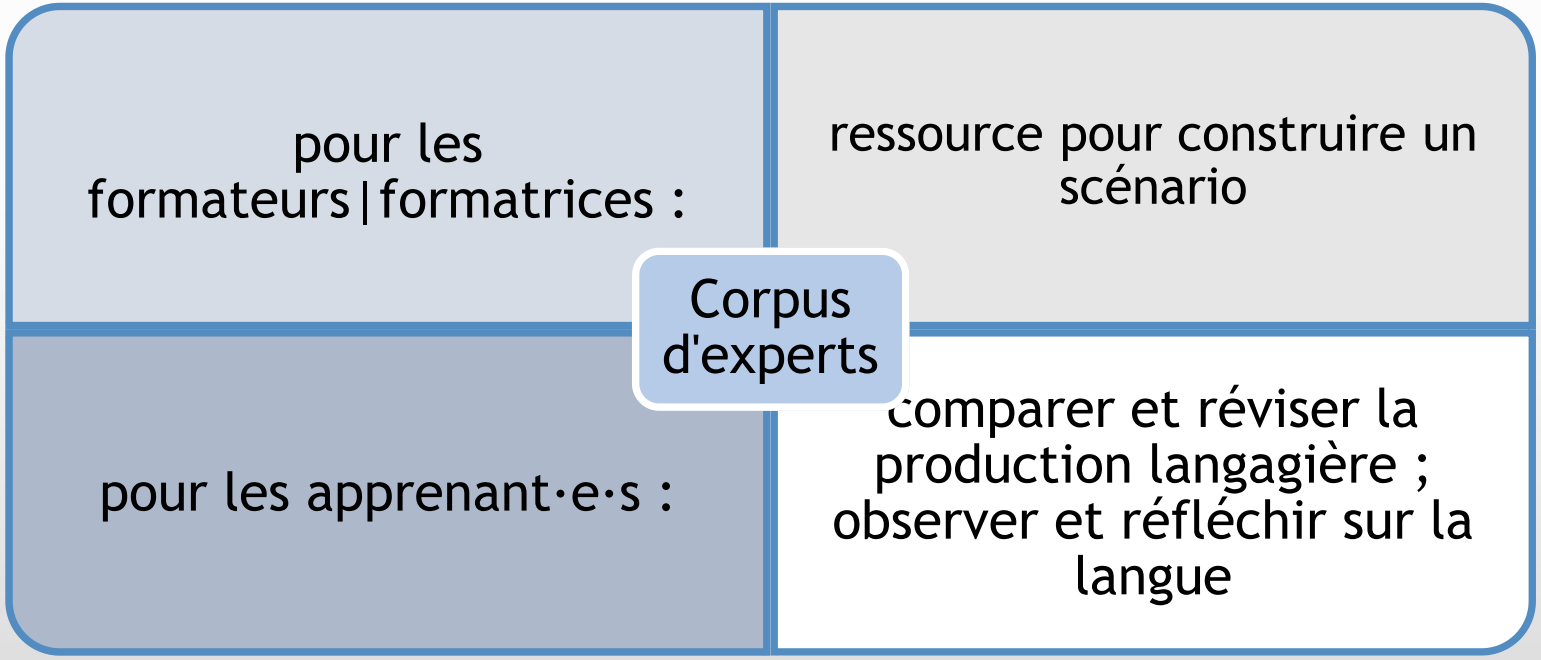
- Hulstijn (2007)

- La nature et la quantité de l'input
- l'environnement social
- les facteurs qui dépendent de l'apprenant (motivation pour apprendre la LE, son attitude par rapport à la LE et par rapport aux locuteurs natifs de la LE et du pays de la LE)
- l'aptitude de l'apprenant :

➔ sa capacité à traiter l'information
➔ ses connaissances métalinguistiques
➔ ses stratégies de résolution de problème

Consulter la littérature scientifique

Les corpus d'experts pour enseigner et apprendre une langue



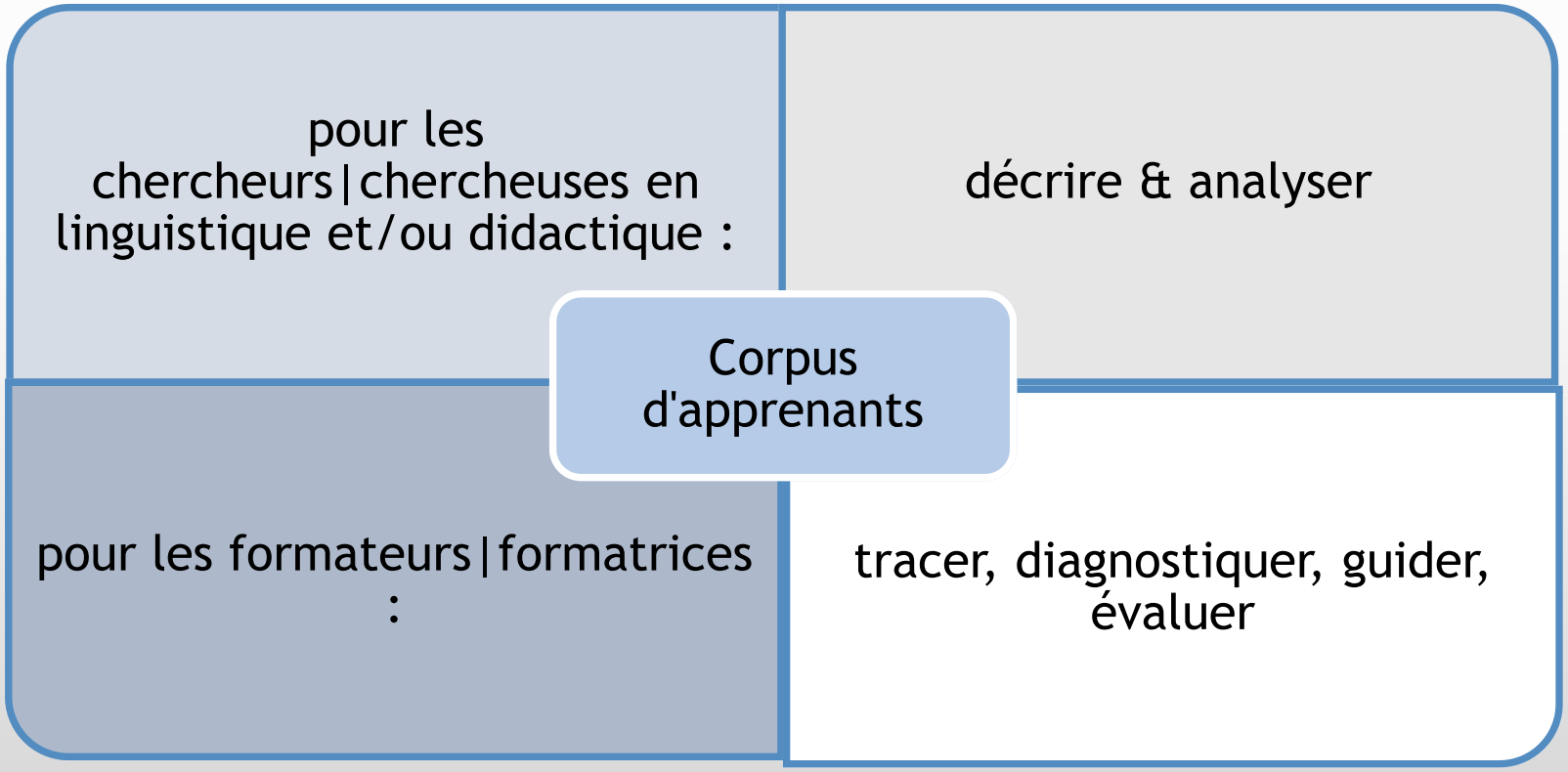
Apports de corpus d'experts

- **Input ayant des caractéristiques favorables**
  - Quantité et qualité suffisante de données langagières
  - Répositoire de termes et formules spécifiques à un genre textuel (liste de ngrams ou de mots-clés spécifiques)

- **Interface susceptible de réduire la charge cognitive des apprenants**
  - Possibilité d'aborder les données de façon flexible

- **Ressources pour créer des activités d'apprentissage**
  - Énigmes linguistiques ou culturelles, discussions métalinguistiques, révision de texte

Les corpus d'apprenants pour la recherche et pour la planification de l'enseignement



Apports de corpus d'apprenants

- Répositoire d'erreurs récurrentes

- Caractéristiques lexicales ou autres des textes des apprenants versus celles rassemblées dans un corpus d'experts

➔ Différences que j'ai constatées sur le terrain :

- Fréquence de signes de ponctuation
- Variété lexicale
- Formules spécifiques à un genre textuel

Gilquin, Granger & Paquot (2007) :

- connaissances lacunaires des caractéristiques des registres langagiers,
- usage inapproprié de formules (semi-)figées,
- erreurs sémantiques.

Consulter la littérature scientifique et les expériences pratiques

RoC adossée à un module de création d'un scénario incluant une activité d'audiodescription

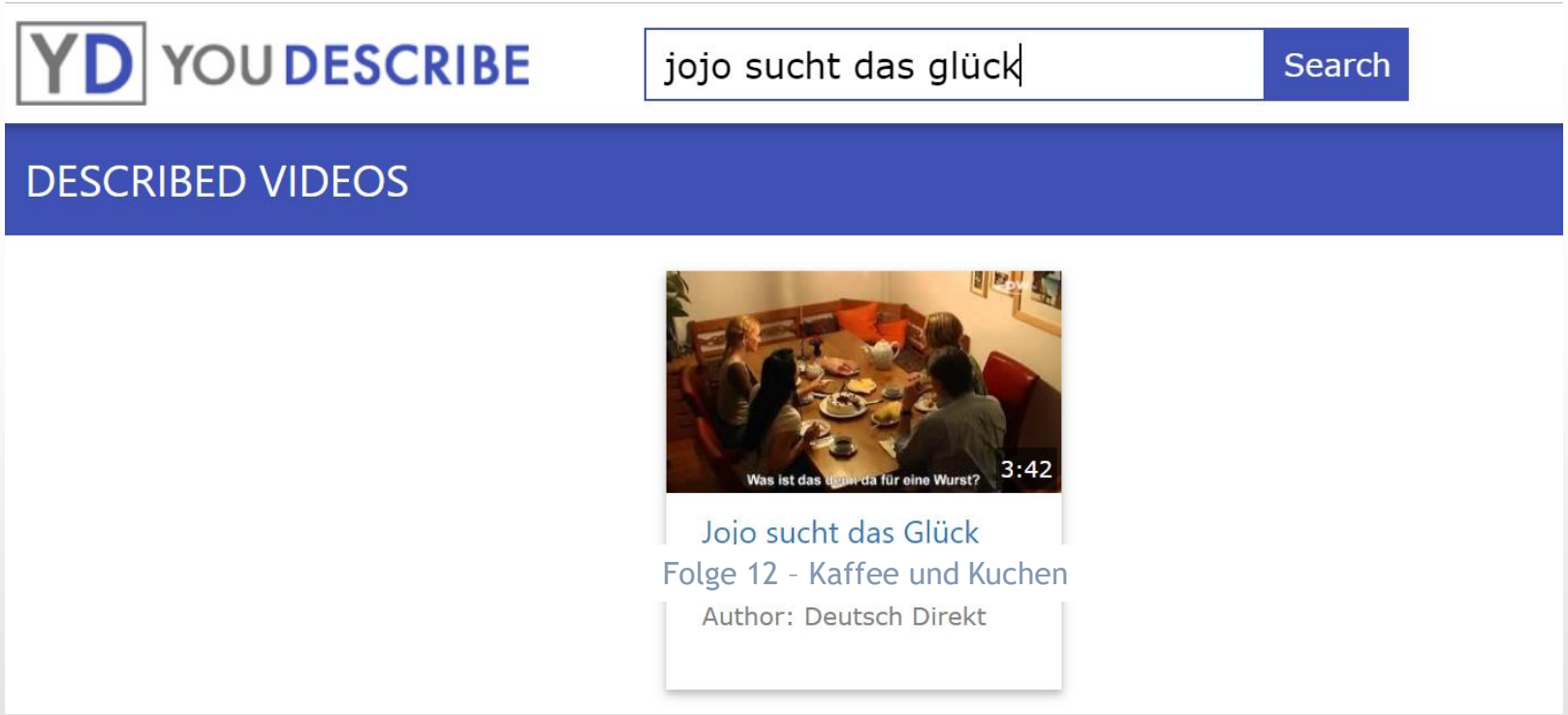
- 14 étudiants Master 1 MEEF parcours allemand (Espé de Paris)
- En petits groupes, création d'un scénario pour le secondaire incluant une activité pouvant permettre de réaliser l'audiodescription (AD) d'un court extrait de film
- Production complète d'une audiodescription par un des groupes

➔ Cette production = point de départ pour l'analyse des besoins pour créer une audiodescription

➔ Création d'activités d'apprentissage faisant appel aux corpus

Développer et implémenter un module

Audiodescription des apprenants de la RoC



Série télévisée *Jojo sucht das Glück* [Jojo à la recherche du bonheur]

Épisode 12 : *Kaffee und Kuchen [À l'heure du goûter]*

Outil de publication de l'audiodescription : Youdescribe.org

Caractéristiques du genre textuel AD

- AD = description d'événements visuels
- Manuscrits AD & indications pour les narrateurs et les narratrices

Recommandations aux auteurs des manuscrits (NDR Fernsehen, nd) : adopter une écriture…

- concise
- précise
- objective
- empathique

On peut voir…

Non, je ne peux pas !

Moyens linguistiques pour écrire un manuscrit AD en langue allemande

être concis et bref	mots composes
	participes présents
être précis	prépositions
	préverbes
	éventail large de verbes
être empathique	choix terminologique adapté aux personnes malvoyantes ou aveugles

Corpus d'experts : Buettenwarder

Manuscrits AD pour *Neues aus Buettenwarder* (Eberlein, 1997-2017), une série télévisée racontant des histoires d'un milieu rural en Allemagne du nord.

Corpus déposé sur Sketch *Engine*

336.723 tokens, 69 fichiers texte + fichier métadonnées

Annoté au niveau des classes de mots à l'aide du Rftagger (Schmid & Laws, 2008)

Photo : *Neues aus Büttenwarder*, episode *Goethe-Jahr*.

Manuscrit AD d'experts, extrait de Buettenwarder

10:01:46
"Ich bin beim Essen!" ["Je suis en train de manger !"]
ss Adsche nimmt Griem den Teller weg. [Adsche enlève à Griem son assiette.]

Repères visuels pour enregistrer les pistes AD :

- indications temporelles ;
- fins des dialogues du script de film signalisées par des guillements ;
- indications concernant la manière de parler mises en gras (p.e., "**ss**" : parler très vite).

Manuscrit AD d'apprenants, extrait de Jojo

*<…> = erreur

1.19 Aufnahme *<auf dem Essen> auf dem Tisch: (…)
1.36 Jojo ist verlegen, Lena senkt den Blick.
1.50 Jojo zeigt *<xxx> einen Teller.

1.19 *<Prise de vue *sur le repas> sur la table : (…)
1.36 Jojo est embarassée, Lena baisse les yeux.
1.50 Jojo *<fait voir> une assiette.

Utiliser des mots composés pour être concis

Aufnahme auf dem Essen auf dem Tisch	[Prise de vue *sur le repas sur la table] ➔ Trop long ; erreur de préposition|de cas
> Großaufnahme des Essens	[> Gros plan du repas] = mot composé suivi d'un nom au génitif

Mots composés dans Buettenwarder

Traduction	Forme	Nombre de caractères
ayant la forme d'un bateau	schiffchenförmige	17
noir(s) et luisant(s)	schwarzglänzenden	17
ressemblant à un tracteur	treckerähnlichen	16
planté(s) d'arbres	baumbewachsenen	15
à rayures bleues	blaugestreiften	15

Utiliser des prépositions pour être précis

Jojo zeigt xxx einen Teller

>

Jojo zeigt auf einen Teller

["montrer une assiette" versus "désigner une assiette, la pointer du doigt"]

Photo : *Jojo sucht das Glück* (nd), épisode 12.

Zeigen dans le corpus Buettenwarder

SKETCH ENGINE

accusative objects of "zeigen"		35.25
bierdeckel	3	10.62
vogel	2	10.30
daumen	3	10.22
foto	2	10.00
adsche	2	6.79

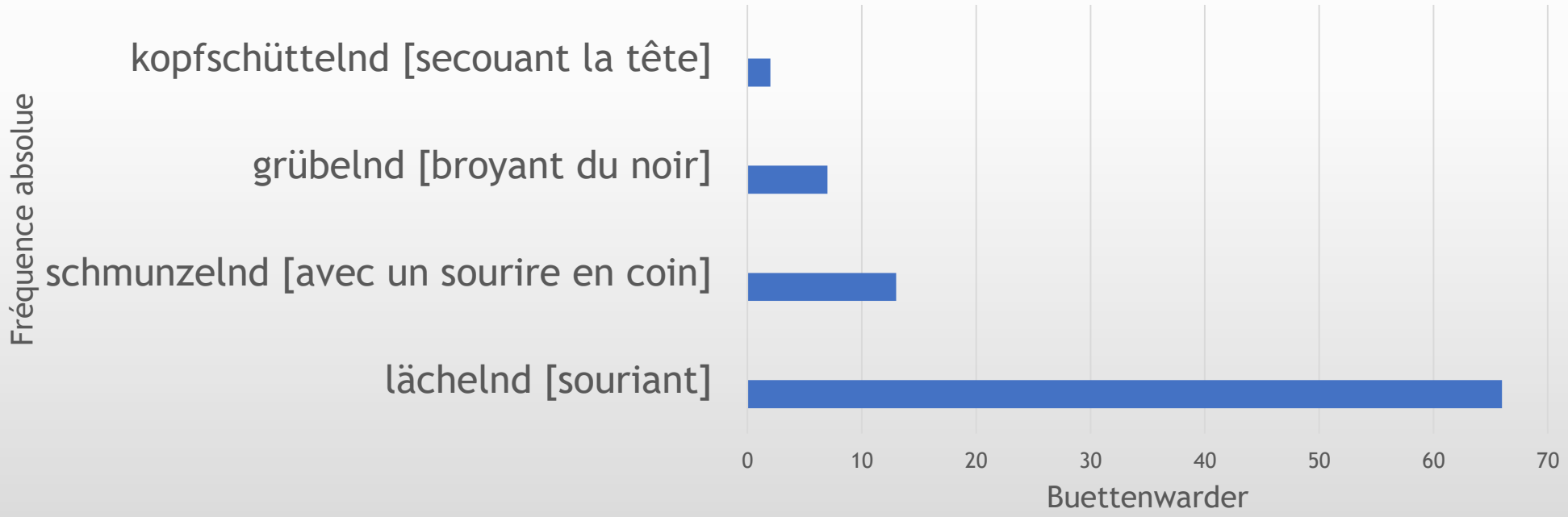
On peut montrer à quelqu'un…
un dessous de verre, un
"oiseau", son pouce, une
photo, …

prepositions of "zeigen"		38.85
auf	39	10.84
Er zeigt auf		
hinter	3	9.72
nach	3	9.25
mit	5	7.94
zu	3	7.04

On peut pointer…
à/derrière/en direction de quelque
chose/quelqu'un
On peut pointer…
avec quelque chose

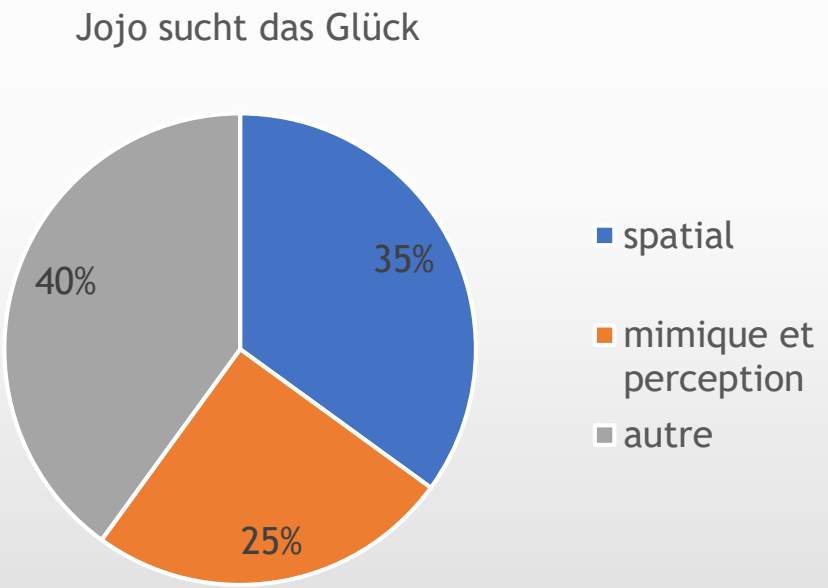
Utiliser des participes présents pour être concis ✓

"Vier lachende Jugendliche"	"Quatre jeunes souriants"



Varier les verbes et éviter des verbes faisant référence à la capacité visuelle du spectateur ✓

- haben
- erscheinen [apparaître]
- bringen [apporter]
- servieren [servir]
- trinken [boire]
- 3x essen [manger]
- 2x sein [être]
- (den Blick) ab·wenden [détourner (le regard)]
- antworten [répondre]
- (den Blick) senken [baisser les yeux]
- (auf etwas) zeigen [pointer (vers quelque chose)]
- nehmen [prendre]
- etw auf etwas legen [poser quelque chose sur quelque chose]
- sich zu jm wenden [se tourner vers quelqu'un]
- lachen [rire]
- zu·blinzeln [faire un clin d'œil à quelqu'un]
- lächeln [sourire]



RoC adossée à un module de création d'un scénario d'audiodescription & corpus

Tâche du groupe recherche	Tâche du groupe témoin	Variable indépendante	Variable dépendante
Créer un scénario AD faisant, entre autres, appel à des outils de corpus	Créer un scénario AD faisant appel des outils d'aide (p.e., dictionnaires et traducteurs en ligne) mais pas à des outils de corpus	Type d'outils	- Types de descriptions de la langue ; - types d'activités d'apprentissage créées par les groupes

Implémenter le module réajusté

Erreurs constatées dans les manuscrits AD de futurs enseignants

- Sémantique lexicale et formules (semi-)figées

"Schwimmbad > Schwimmbecken" [piscine : établissement > bassin] ; "er versetzt im Schrecken" > "er erschrickt" [il prend peur]

- Syntaxe de groupes de sens

"*Junge hält ein Referat über seine Mutter *im Klassenzimmer." [*Garçon fait un exposé sur sa mère *dans la salle de classe.]

- Prépositions et préverbes

*an Nacht [la nuit] ; "Sie zeigt ❑ den Möbelhaufen" [Elle montre le tas de meubles]

- Techniques pour faire court

Commencer une audiodescription par un nom commun non précédé d'un déterminant

- Empathie

Décrire des parties malgré le fait que l'on puisse les comprendre grâce à la piste son existante

Définir l'obstacle

Comment dire "[Il traverse la ville] la nuit" ?
(v. *an Nacht)

Contexte de gauche	KWIC
. # Brakelmanns Hof **bei**	**Nacht**
. # Brakelmanns Hof **bei**	**Nacht**
08 Brakelmanns Hof **bei**	**Nacht**

**an Nacht* >
- *bei Nacht* (6 fois)
- *Nachts* (18 fois)

Les corpus pour surmonter l'obstacle

Peut-on renoncer au déterminant d'un nom commun pour faire court ?

"*Junge hält ein Referat (…)" [*Garçon fait un exposé (…)]

CONCORDANCE manuskriptebuettenwarder

CQL [tag="CARD"][word=".*"][tag="N.*"] 2 768 (8 431 par M)

Fréquence MODIFIER LES CRITÈRES RETOUR AU CONCORDANCIER

Montrer la fréquence par million

	Word	
1	# Die Häsin	
2	# Der Pritschenwagen	
3	# Der Peterwagen	
4	# Der Hof	1
5	# Der Hase	1
6	# Der Bürgermeister	1
7	# Der Bärtige	1
8	# Der Blick	1
9	# Der Bauer	1
10	# Das Publikum	1
11	# Das Cabrio	1

# Das Cabrio	Le cabrio
# Das Publikum	Le public
# Der Bärtige	Le barbu
# Der Bauer	Le paysan
# Der Blick	Le regard
# Der Bürgermeister	Le maire
# Der Hase	Le lapin
# Der Hof	La cour

Les corpus pour surmonter l'obstacle

Comment dire "pousser dans une direction" ?

"Sie schieben ein Sofa in *einem *Schwimmbad"
[Ils poussent un canapé (en étant) dans la piscine]

➔ piscine = bassin

➔ déplacement & accusatif

"Sie schieben ein Sofa in ein Schwimmbecken"
[Ils poussent un canapé dans (le bassin de) la piscine]

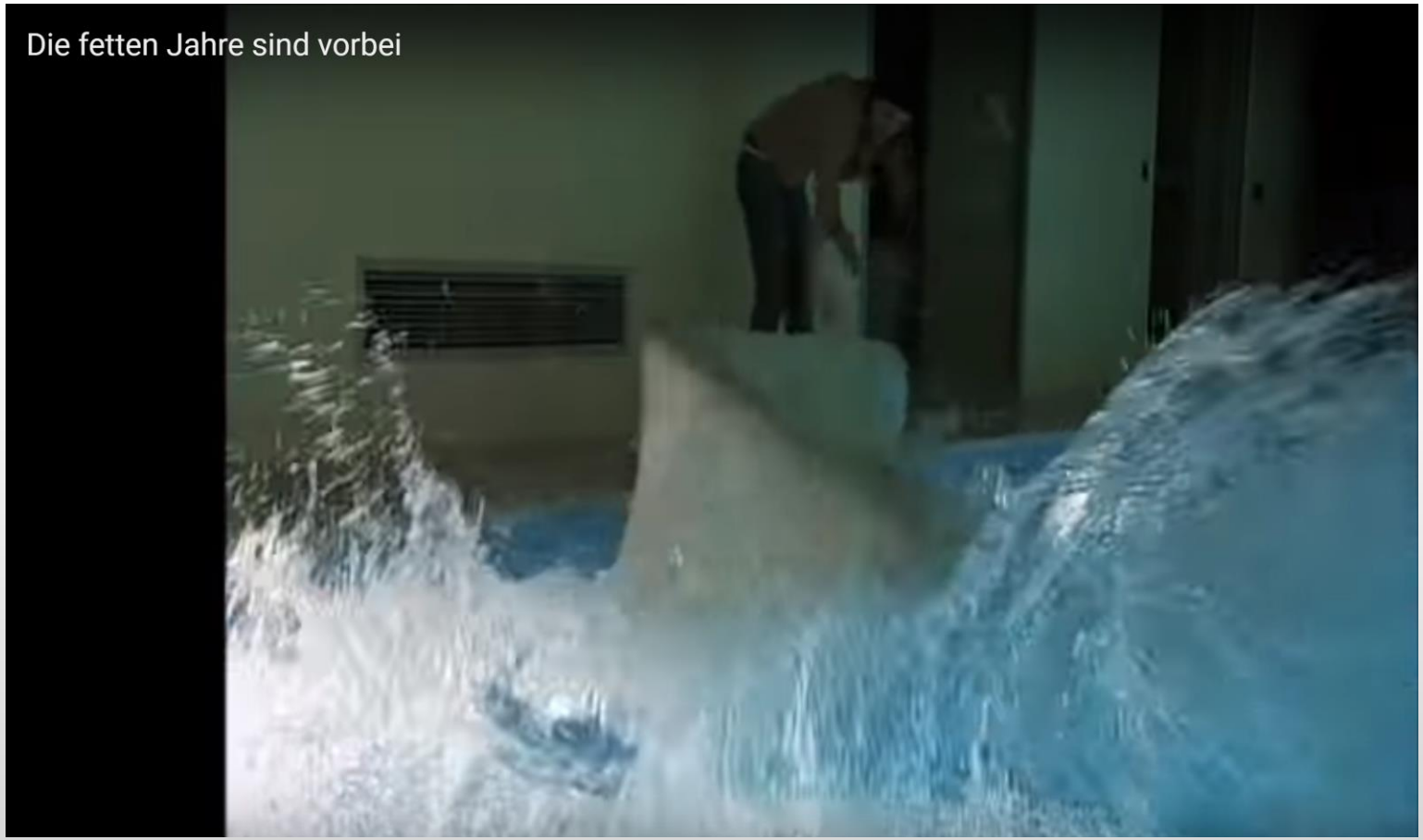
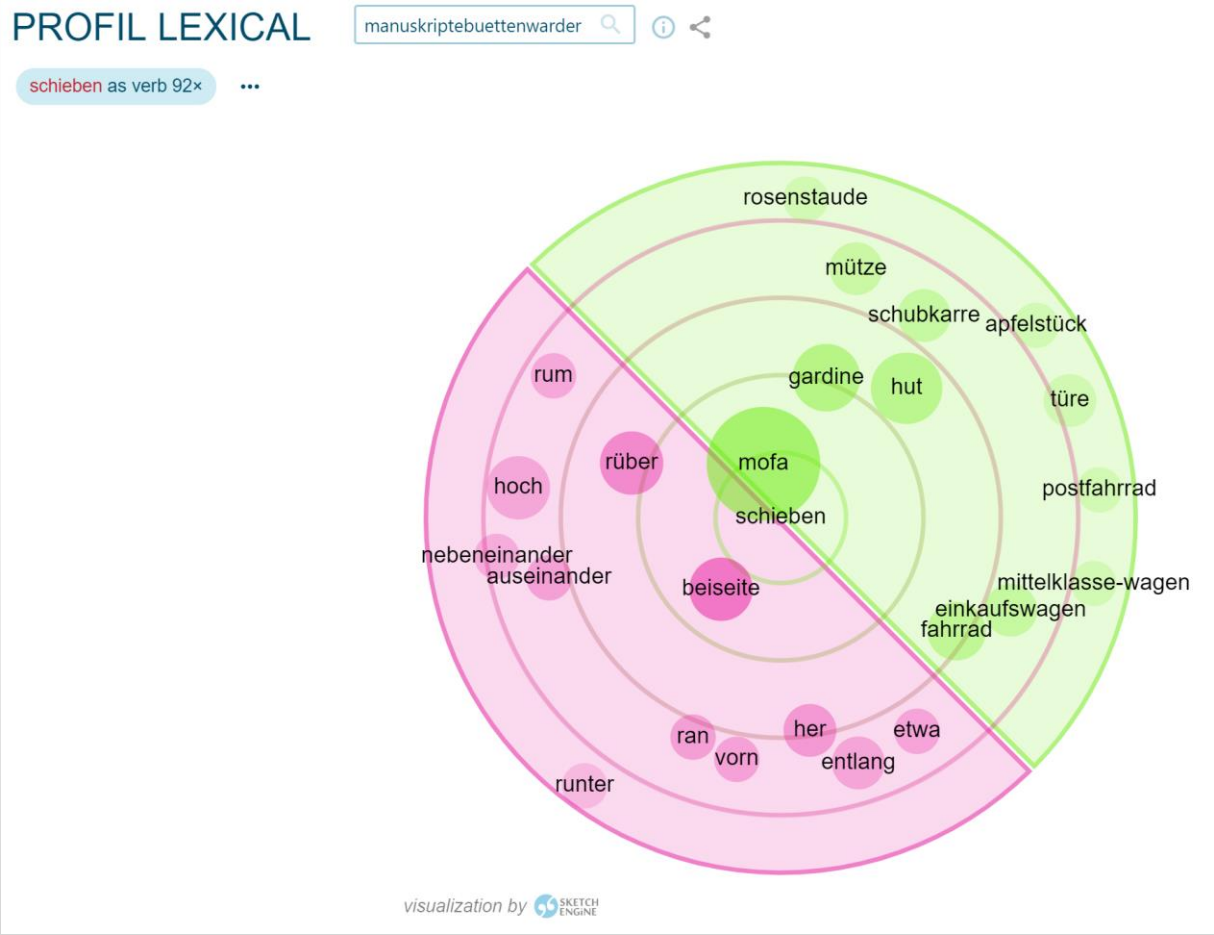


Photo : Générique de *Die fetten Jahre sind vorbei* (Weingartner, 2004).

Le verbe "schieben" [pousser] dans le corpus Buettenwarder



92 occurrences du verbe "schieben" [pousser]

On pousse une mobylette, un rideau, un chapeau, etc., pour les écarter, changer de côté, rapprocher, etc.

Toutes les occurrences désignent un déplacement (& accusatif), aucune un séjour (& datif).

À faire par la suite

- Annoter, à l'aide XML, les manuscrits d'audiodescriptions
- Partager le corpus de manuscrits d'audiodescriptions
- Pour des projets à venir, partir de corpus d'apprenants annotés existants (Falko, Merlin) pour identifier des besoins d'apprentissage à une échelle plus large

Annoter les manuscrits AD

- Annotations XML du corpus d'experts

- Indications temporelles
- Indications pour la narration
- Mots composés
- Annotations sémantiques ?
- …

Partager le corpus de manuscrits AD

- Dépôt du corpus Buettenwarder sur Ortolang – CLARIN (droits accordés par le propriétaire des données)

Toutes les ressources déposées sur ORTOLANG sont désormais moissonnées par le **Virtual Language Observatory** de CLARIN et accessibles via la recherche par facettes.

CLARIN
Common Language Resources and Technology Infrastructure

Ortolang

Partir de corpus d'apprenants

German	A1	206
	A2	209
	B1	210
	B2	204
	C1	204

MERLIN-Korpus

Échantillons de tests de placement TELC (allemand et italien) et d'examens de l'université de Prague (tchèque). Les tâches sont standardisées et alignées au CECR (Cadre européen de référence pour les langues).

Weber (2018) : analyse de l'usage des prépositions *an* [à, près de, en lien étroit avec…] et *auf* [sur] dans le Merlin-Korpus. Résultat : les sens dits "abstraits" des prépositions sont appris à des niveaux plus avancés que les sens dits "concrets".

Cf. résultats projet Weissensee (Schaeffer-Lacroix, 2015) : *an* [à, près de, en lien étroit avec…] est un indicateur des progrès faits par les étudiantes ayant participé à ce projet.

Merci pour votre attention !

- elacroix@espe-paris.fr
- http://didaktik.hautetfort.com/

Références

Cadre européen commun de réference pour les langues. Apprendre, enseigner, évaluer, 2001, Division des politiques linguistiques, Strasbourg. Paris : Didier.

Eberlein, Norbert, 1997-2017), Neues aus Büttenwarder. Television series. Audio descriptions by the team of the broadcast company NDR Fernsehen (Norddeutscher Rundfunk).

Ellis, Nick & Wulff, Stefanie, 2018, Usage-based approaches to second language acquisition, in D. Miller, F. Bayram, J. Rothman & L. Serratrice (dir.), *Bilingual Cognition and Language: The state of the science across its subfields*, 37–56.

Euler, Dieter, 2014, Design-Research – a paradigm under development, in D. Euler & P.F.E. Sloane (dir.). *Design-Based Research*. Stuttgart : Franz Steiner Verlag, 15-44.

Gilquin, Gaëtanelle, Granger, Sylviane & Paquot, Magali, 2007, Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 319-335. Hess, A. 2010. *Weissensee*. Script d'une série télévisée. http://www.stichwortdrehbuch.de/drehbuch

Hulstijn, Jan, 2007, Psycholinguistic perspectives on second language acquisition. In Cummins, J. & Davison, C . (dir). *The international handbook on English language teaching*. Norwell, MA : Springer, 701-713.

Jojo sucht das Glück, nd, soap opera, Deutsche Welle.

Kilgarriff, Adam, Rychly, Pavel and Pomikalek, Jan (nd). *Sketch Engine*. Corpus management system. http://www.sketchengine.co.uk/

NDR Fernsehen, nd, Audiodeskription: Mit denn Ohren sehen.

ORTOLANG (Open Resources and TOols for LANGuage). Répositoire de corpus. www.ortolang.fr

Plomp, Tjeerd, 2007, Educational Design Research: An Introduction. In T. Plomp & N. Nieveen (dir.), An Introduction to Educational Design Research. Enschede : SLO – Netherlands institute for curriculum development, 9–36.

Samdereli, Yasemin, Almanya, 2011, Générique du cinéfilm.

Sanchez, Éric, & Monod-Ansaldi, Réjane, 2015, Recherche collaborative orientée par la conception. *Education & didactique*, *9*(2), 73–94.

Schaeffer-Lacroix, Eva, 2015, Impact de discussions métalinguistiques sur l'apprentissage de la production écrite en allemand, langue étrangère. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (72), 233–237. https://doi.org/10.4000/linx.1676

Schmid, Helmut & Laws, Florian, 2008, Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-grained POS Tagging. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1* (pp. 777–784). Stroudsburg, PA, USA : Association for Computational Linguistics.

Weber, Tassja, 2018. "Grammatik und Lernerkorpora: Eine korpusorientierte Untersuchung von Präpositionalphrasen im deutschen MERLIN-Korpus", in E. Fuß, M. Konopka, B. Trawiński & U. H. Waßner (dir.). *Grammar and Corpora 2016*. Heidelberg : Heidelberg University Publishing.

Weingartner, Hans, 2004, Die Fetten Jahre sind vorbei [The Educators]. Générique du cinéfilm

Wisniewski. Katrin, 2014, MERLIN - a multifunctional trilingual learner corpus related to the CEFR. 11th EALTA Conference in Warwick, UK, May 29 -