

Exploiter les corpus allemands pour soutenir le développement d'habiletés cognitives

Eva Schaeffer-Lacroix
Université Paris-Sorbonne / ESPE de Paris

Journée d'études du GERALS - « Les corpus et l'allemand de spécialité »
Université Le Havre - 7 novembre 2014

- Quels types corpus, et quels types d'apports ?
- Quelles habiletés cognitives ?
- Quelles activités d'apprentissage ?
- Quelles perspectives pour l'exploitation des corpus en allemand de spécialité ?

Langue de spécialité / LANSAD
versus
langue généraliste

Corpus spécialisés
versus
corpus généraux

Quelle langue de spécialité ?
Quels corpus ?

Types de corpus	Caractéristiques	Taille
Corpus général	Les textes qu'il contient doivent "représenter" une langue dans sa globalité.	Plus il est grand, mieux c'est.
Corpus parallèle ou comparable (général ou spécialisé)	Les textes qu'il contient proviennent de deux ou plusieurs langues. Ils contiennent des données traduites ou traitant du même sujet.	La taille n'est pas un critère primordial.
Corpus spécialisé (aussi dit « corpus réflexif » ; cf. Mayaffre, 2002)	Les textes qu'il contient sont rassemblés en raison de leurs caractéristiques comparables (appartenance à un même genre textuel, un même type de discours, etc.).	La taille n'est pas un critère primordial.

Types de corpus

- Bain de langue (Zanettin, 2001) ;
- apprentissage incident (Bernardini, 2004) ;
- attitude d'"apprenant-chercheur" (Johns, 1997) ;
- ... et quoi d'autre ?

Que peut-on attendre des corpus
pour apprendre une langue de
spécialité ?

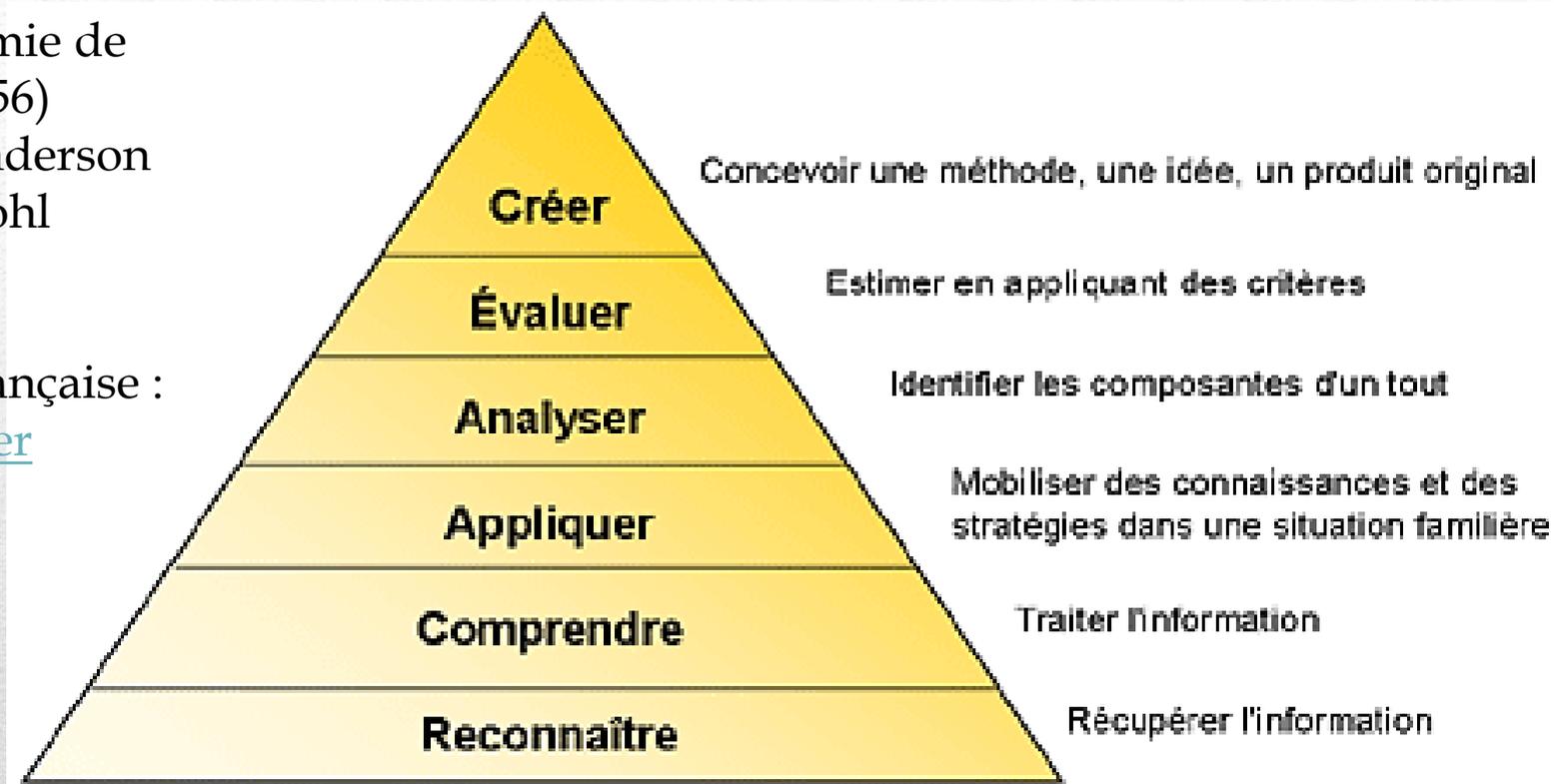
Une habileté cognitive est une compétence...

- qui se développe au niveau de la cognition individuelle ;
- qui peut être illustrée par une action observable (sélectionner, trier, etc. ; cf. Anderson & Krathwohl, 2001) ;
- qui peut être favorisée par une interaction avec des experts/des pairs (Matthey, 2010).

Qu'est-ce qu'une habileté cognitive ? 6

La taxonomie de Bloom (1956)
d'après Anderson & Krathwohl (2001).

Version française :
[Anna Vetter](#)
(2009)



Les habiletés cognitives selon
Bloom (1956) / Anderson (2001)

Capacités démontrées	PREMIER DEGRE (A2)	DEUXIEME DEGRE (B1-1)	TROISIEME DEGRE (B1-2)
Caractéristiques générales	Repérage de quelques éléments. Transmission intelligible dans des situations de communication parfaitement prévisibles.	Repérage et transmission d'un ensemble d'éléments d'informations en relation avec des situations de communication prévisibles.	Sélection et classement d'éléments d'informations en adéquation avec la tâche délimitée. Autonomie partielle dans des situations de communication qui restent prévisibles.
Compréhension de l'écrit et de l'oral	Repérage de quelques éléments factuels simples	Compréhension de la plupart des éléments explicites marquants.	Recueil d'informations multiples provenant de sources diversifiées.
Production écrite et orale	Reproduction intelligible des éléments repérés.	Restitution des éléments compris dans un format simple.	Réorganisation et réutilisation personnelles des éléments retenus.

Extrait du descriptif des capacités démontrées du DCL (Diplôme de compétence en langues, 2002)

Les habiletés cognitives dans la grille d'évaluation du DCL

Série télévisée *Weissensee* (Hess, 2010)



Intégration d'activités basées sur des corpus dans un scénario pédagogique 9

Lieu	IUFM de Paris
Apprenants	Étudiantes en Master 1 documentation, 22-28 ans
Niveau	A2/B1/B2
Corpus	Corpus <i>Weissensee</i> (1 ^{ère} épisode en bilingue, le reste en monolingue)
Tâche de production écrite	Écrire une scène qui remplace ou prolonge une scène d'un script de film existant
Tâches intermédiaires linguistiques	- Comparer sa production non pertinente à celle d'un expert (corpus) - Déterminer le sens de particules verbales
Travail de réflexion et de révision	- Cinq discussions métalinguistiques - Une séance de révision collective
Outils de communication	<i>Edmodo, Skype, TitanPad</i>
Type d'intervention pédagogique	pro-active et réactive

Projet *Weissensee*

Language	Corpus name	Tokens	Words	
Chinese, Simplified	zhTenTen11	2,106,661,021	1,729,867,455	 
English	British National Corpus	112,181,015	96,048,950	 
English	enTenTen12	12,968,375,937	11,191,860,036	 
French	frTenTen12	12,369,868,562	10,666,617,369	 
German	deTenTen10	2,844,839,761	2,338,036,362	 
Japanese	jpTenTen11 [SUW]	10,321,875,664	8,432,256,386	 
Russian	ruTenTen11	20,162,118,568	15,763,181,803	 
Spanish	esTenTen11 (European, Freeling)	2,341,159,406	1,991,879,282	 

[Show all corpora](#) | [Parallel corpora](#)

My corpora

Corpus ID	Corpus name	Language	Tokens	
WeissenseeFrenchAligned	weissensee.franzoesisch.teil1	French	15,832	    
WeissenseeGermanAligned	weissensee.deutsch.teil1	German	13,640	    
WeissenseeKomplett	weissensee.komplettes.drehbuch	German	82,030	    

Sketch Engine pour explorer les corpus du projet

L'enseignant(e)...

- relève des passages non pertinents ;
- propose des requêtes préformatées ;
- commente les versions révisées à l'aide du corpus ;
- propose de nouvelles actions de révision.

Révision de texte basée sur les corpus –
Les actions de l'enseignant(e)

Projet *Weissensee* - Cornélia (V0)

Requêtes à faire dans le corpus *deTenTen*, sauf autre mention.

DUNJA: Du kannst angstfrei sprechen, Julinka. Ich bin da, ich **höre dir**.

ich höre dir (requête Sketch Engine)

Dunja erkennt die Unruhe **bei/in** Julia.

Unruhe & ihr (right) (2)

JULIA: Ich **weiss** nicht mehr weiter. Einerseits gibt es Robert, mit dem ich mich in **der** Sicherheit fühle. Andererseits gibt es auch Martin, der mir wirklich gefällt...

Sicherheit fühlen

Julia **setzt zu sprechen fort** aber Dunja hört sie nicht. Dunja verzieht keine Miene.

Pons, "continuer". Quelle particule apparaît fréquemment ?

(Sie ist **in ihren Gedanken versinkt**).

Gedanken & versinken

Révision individuelle guidée

L'apprenant(e)...

- exécute les requêtes préformatées ;
- repère dans le corpus des segments pertinents ;
- choisit un des segments repérés et l'insère dans son texte ;
- adapte le segment du corpus à son texte ;
- ...et, avec le temps, trouve lui-/elle-même des idées de requête qui lui permettent de tester l'acceptabilité de sa production.

Révision de texte basée sur les corpus –
Les actions de l'apprenant(e)

Passage à réviser (Cornélia, V0)
*Sie ist in *ihren Gedanken *versinkt.*

Requête conseillée
Gedanken & versinken

1 tokens: 111,502 / 1,000,000 days left: 13 [subscribe](#)

Corpora

Corpus name	Language	Tokens	Words	
British National Corpus	English	112,181,015	96,048,950	 
enTenTen	English	3,268,798,627	2,759,340,513	 
deTenTen	German	2,844,839,761	2,338,036,362	 
JpWaC	Japanese	409,384,405	333,246,192	 
Russian web corpus	Russian	187,965,822	147,930,261	 

[Show 61 more corpora](#)

My corpora

Corpus ID	Corpus name	Language	Tokens	
WeissenseeFrenchAligned	weissensee.französisch.teil1	French	15,832	    
WeissenseeGermanAligned	weissensee.deutsch.teil1	German	13,640	    
WeissenseeKomplett	weissensee.komplettes.drehbuch	German	82,030	    

 [Create corpus](#) /  [WebBootCaT](#)

Repérer des occurrences pertinentes 15

Cornélia (V1)

Passage sélectionné	<i>aber Ruth ist sehr in Gedanken versunken</i>
Passage révisé	<i>Sie ist in Gedanken versunken.</i>

Repérer un élément et le restituer
dans un format simple

The screenshot shows the Sketch Engine web interface. At the top, there is a navigation bar with links for Home, Settings, Change password, and Log out. A search bar is present with the text 'Search' and 'in Help'. Below this, the user information is displayed: 'user: Eva Lacroix corpus: deTenTen [2010]'. A secondary search bar is also visible with 'Search' and 'in deTenTen [2010]'. On the left side, there is a vertical menu with links: Concordance, Word List, Word Sketch, Thesaurus, Find X, Sketch-Diff, and Corpus Info. The main content area is titled 'Context' and contains two filter sections: 'Lemma filter' and 'PoS filter'. The 'Lemma filter' section has a 'Window' dropdown set to 'right' and a value of '5' tokens, and a 'Lemma(s)' input field with an 'all' dropdown. The 'PoS filter' section has a 'Window' dropdown set to 'both' and a value of '5' tokens, and a 'PoS' dropdown menu with options: adjective, adposition, adverb, conjunction, and determiner. Both filter sections have an 'all' dropdown for 'of these items'. At the bottom of the main area, there are 'Make Concordance' and 'Clear All' buttons. In the bottom right corner, there is a logo for 'Lexical Computing' and version information: 'Sketch Engine (ver:2.29-SkE-2.109.10-3.48.17)'. At the bottom center, there is a language selection bar: 'Interface language: English | český | 简体中文 | 繁體中文 | Gaeilge | slovenščina | hrvatski'. In the bottom left corner, there is a watermark for 'Screencast-O-Matic.com'.

Analyser et réutiliser des éléments repérés

Julia dreht das Autoradio auf.

Julia allume la radio.

Martin hört auf zu tippen und sieht Julia an.

Martin arrête de taper et regarde Julia.

Görlitz und Martin treten auf den Flur.

Görlitz et Martin vont dans le couloir.

Julia nimmt ihre Tasche, sie steht auf.

Julia prend son sac, elle se lève.

- **Nature d'un mot** – « Particule ou préposition ? »
 - Propriétés syntaxiques ;
 - propriétés morphologiques.
- **Distinctions sémantiques** – « Quel est le sens de *auf* ? »
 - *auf/drehen* vs. *auf/hören*, etc.

Les particules verbales comme
objet d'apprentissage

Drehbuch *Weissensee*, Folge 1, 21. Bild

Martin betritt seine Einliegerwohnung im Kellergeschoss. Er hat hier ein eigenes Bad, eine kleine Küche. Es stehen noch ein paar Umzugskisten **herum**. Der Raum ist noch nicht richtig **eing**erichtet. Augenscheinlich ist Martin vor kurzem erst **ein**gezogen.

Martin stellt seine Tasche **ab**, öffnet sie und nimmt etwas **heraus**: es ist Julias Kasette. Er betrachtet die Kasette: die Aufkleber sind **ab**gelöst. Martin legt die Kasette in ein Abspielgerät **ein**. Er legt sich **an**gezogen auf sein Bett und hört interessiert **zu**:

MÄNNLICHE STIMME

...My name ist Peter. My name is Peter.
I am from Munich... I am from Munich...

78 mots, 9 particules verbales.

Query PTKVZ 1,610 (19,623.6 per million)

Sprachrätsel « *Los, weg* oder *xxx*? » - Révision collective des solutions de Solène.

- | | |
|---|--|
| <p>11 Robert fährt weg > LOS.</p> <p>12 Julia reicht Martin die Hand, sie schütteln sie, beide lassen sich nicht mehr los.</p> <p>14 Martin packt Falk beim Handgelenk und drückt dessen Hand los > WEG.</p> <p>15 Julia streift eilig im Gehen ihren Kittel los > xxx.</p> <p>16 Martin schließt den Spind weg > xxx.</p> <p>19 Sie biegt los > AB, da ertönt das Martinshorn direkt hinter ihr.</p> <p>20 JULIA Ich muss weg > LOS, wir haben heute Inventur...</p> <p>21 Martin springt von der Schaukel los > xxx und geht auf Julia zu.</p> <p>22 Die Ampel springt auf Grün. Martin biegt los > xxx und gibt Gas.</p> <p>23 LEUTNANT BÖSNER Na los, aufmachen! [den Kofferraum]</p> | <p>Robert démarre...</p> <p>Julia donne la main à Martin, ils se serrent la main, les deux ne se lâchent plus...</p> <p>Martin saisit le poignet de Falk et écarte la main de celui-ci.</p> <p>En partant, Julia enlève à la hâte sa blouse.</p> <p>Martin ferme son casier.</p> <p>Au moment où elle tourne, la sirène retentit directement derrière elle.</p> <p>JULIA Je dois partir, aujourd'hui, nous faisons l'inventaire...</p> <p>Martin saute de la balançoire, et il se dirige vers Julia.</p> <p>Le feu passe au vert. Martin tourne et accélère.</p> <p>LEUTNANT BÖSNER Allez, ouvrez-le ! [le coffre de la voiture]</p> |
|---|--|

(...) je mène les étudiantes à dire que *los* & un lien qui va se défaire ; il faut une certaine énergie pour le défaire – très bien expliqué par Cornélia & l'association "brutal".

Los & idée de départ (Marjorie), & lâcher (Marjorie), détacher. *Weg* & une idée d'absence (Marjorie). Je précise que c'est l'idée qui compte (peu importe où va qn ; ce qui compte c'est qu'il n'est plus là).

(Extrait du journal de bord de l'enseignante)

La différence entre *los* et *weg*

Solène : Cornélia dit qu'il y a un changement dans la continuité. (...)

EL : (...) Attendez, je regarde vite fait. "Un changement dans la continuité." Je ne pense pas. Non.

EL : Alors, dites-moi, pour le 14 par exemple, "er bremst ab", alors ça, c'est un petit peu bizarre, non ? ++++ Je n'ai pas d'explication. (...)

Marjorie : Ça a peut-être avoir avec la rupture justement, il était en train de rouler, et hop, il s'arrête.

EL : Voilà. (...) Là, l'idée de continuité de Cornélia me convainc un peu plus, ou en fait, c'est presque le contraire, n'est-ce pas ? On arrête une activité. C'est la rupture. On ne fait plus ce qu'on faisait avant.

[4/2/2012 7:05:07 PM]

Cornélia: Je dirais "interrompre"

[4/2/2012 7:05:47 PM]

Cornélia: Un changement dans la continuité peut-être?

[4/2/2012 7:06:07 PM]

Cornélia: pour "sauter", "démarrer" [4/2/2012 7:07:12 PM] Cornélia: C'est parce qu'elle saute de la balançoire après s'être balancé*

*"Jetzt springe ich ab!"

Analyser l'information recueillie –

La fonction aspectuelle de *ab*

aufhören zu sprechen } ab/brechen

aufhören, geradeaus zu fahren } ab/biegen

ab/fahren

ab/hauen

wegfahren (nach Fahrplan)

weggehen (wobei eine
Beziehung zerbricht)

Analyser l'information recueillie –
Ab ayant la fonction de borne

LISA

Wieder! *Macht es wieder! Ich will ein Flugzeug sein! Trag mich wieder in deinen Armen!

Martin ist völlig außer Atem. Er drängt seine Tochter nett **weg**.

MARTIN

Lisa, bitte! Ich bin so müde! Du bist kein Baby mehr. Du fängst **an**, schwer zu sein!

Das kleine Mädchen lächelt. Sie hat jetzt eine andere Idee im Kopf. Sie sieht ihren Vater mit großen Augen **an**.

LISA

Papaaa... Ich war ein nettes Mädchen heute.

Martin weiß wohl, was seine Tochter will. Er antwortet, unschuldig. (...)

- Des scénarios intégrant le travail basé sur des corpus spécialisés ;
- des activités d'exploration de corpus ciblant, dès le début de l'apprentissage, le développement des compétences linguistiques dont les étudiants ont le plus besoin.

Perspectives pour l'allemand
de spécialité

Public : futur(e)s chanteurs et chanteuses

Contenus des corpus : lieder, libretti, cantates, etc.

Compétence linguistique	Types de corpus	Outils
Compréhension de l'écrit	Corpus écrits monolingues ; corpus écrits bilingues.	<i>Sketch Engine, AntConc, etc.</i>
Production orale (aspects prononciation, prosodie, rythme)	Corpus oraux monolingues avec sous-titres (<i>ties</i>) écrits en alphabet phonétique international et /ou codés pour indiquer la longueur d'une syllabe, etc.	<i>Folker, EXMARalDA</i>
Compréhension de l'oral	Corpus oraux monolingues avec sous-titres écrits en langue étrangère et/ou en langue maternelle.	<i>Folker, EXMARalDA</i>

Domaine de spécialité « musique et musicologie »

Isolation du groupe Knulp ([Jamendo](#))

The screenshot shows an audio waveform with a cyan highlighted segment. Below the waveform is a control bar with buttons for 'Add event...', 'Append interval', and various playback controls. At the bottom is a table with columns for time intervals and linguistic annotations.

	0 [00:00]	1 [00:27.9]	2 [00:30.0]	3 [00:34.8]	4 [00:38.4]	5 [00:42.6]
X [v]		Geh' ich durch die Stadt,	dann nur noch mit Mähndrescher, Drescher,	der schlägt nach links aus und nach rechts.	Also, kommt bitte nicht näher.	Ich weiss, das ist nicht fair.
api		ge: ʔɪç duʁç di: ʃtat,	dan nuʁ nox mit mædʁəʃe, dʁəʃe,	dɛʁ ʃlɛkt naʃ liŋz aus ʔund naʃ rɛçtʃ.	ʔalzo: komt bite niçt næhe.	ʔɪç wɛis, daz ɪʃt niçt fæɪ.
voy		* * * % 0 *	* % * * 0 * * * *	% 0 * * * * * * *	* 0 * * * * 0 *	* * 0 * * *

Communication	Speaker	Left Context	Match	Right Context
	X		nicht	. Also, kommt bitte näher. Ich weiss, d
	X		nicht	Ich weiss, das ist fair. Schliesslich

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives (Abridged edition)*. New York : Longman.
- Bange, P., Carol, R. & Griggs, P. (2005). *L'apprentissage d'une langue étrangère : cognition et interaction*. Paris, Budapest, Torino : l'Harmattan.
- Bereiter, C. & Scardamalia, M. (1983). "Does learning to write have to be so difficult?". In Freedman, A., Pringle, I. & Yalden, J. (dir.). *Learning to write: First language, second language*. New York : Longman.
- Bernardini, S. (2000). "Systematising serendipity: Proposals for concordancing large corpora with language learners". In Burnard, L. & McEnery, T. (dir.). *Rethinking language pedagogy from a corpus perspective*. Frankfurt am Main : Peter Lang. pp. 225-234.
- Gautier, L. (2014). "[Des langues de spécialité à la communication spécialisée : un nouveau paradigme de recherche à l'intersection entre sciences du langage, info-com et sciences cognitives?](#)". *Études Interdisciplinaires en Francophonie. Sciences humaines (EISFH)*, vol. 2/2014, n° 1. pp. 225-245.
- Hess, A. (2010). [Weissensee](#). Script de film.
- Johns, T. (1991). "Should you be persuaded - two samples of data-driven learning materials". In Johns, T. & King, P. (dir.) (1991). pp.1-13.
- Kilgarriff, A., Rychly, P. & Pomikalek J. (nd). [Sketch Engine](#). Système de gestion de corpus.
- Krause, M. 2004. Konkurrenz, Komplementarität und Kooperation im Bereich der Präpositionen und Verbalpartikeln oder Wie lange noch müssen Präpositionen und Verbalpartikeln in Grammatiken ein Schattendasein führen?. *Linguistik online* 18-I/04, 2004, 35-69.
- Matthey, M. (2010). Matthey, M. 2010. Interaction : lieu, moyen ou objet d'acquisition ?. In Vargas, C., Calvet, L.-J., Gasquet-Cyrus, M., Véronique, D., Vion, R. (dir.) *Langues et sociétés : Approches sociolinguistiques et didactiques*. Paris, L'Harmattan, 31-42.
- Mayaffre, D. (2002). "[Les corpus réflexifs : entre architextualité et hypertextualité](#)". *Corpus* 1 | 2002.
- Schmidt, T. & Wörner, K. (nd). *EXMARALDA*. Outil de transcription de données audiovisuelles et d'exploration des données du texte transcrit. <http://www.exmaralda.org/>
- Zanettin, F. (2001). "Swimming in Words: Corpora, Translation, and Language Learning". In Aston, G. (dir.). *Learning with Corpora*. Houston TX : Athelstan.