

Pourquoi les préfabriqués de la langue posent un problème à l'IA dans l'enseignement des langues

Préfabriqués de la langue et apprentissage d'une langue étrangère : le miracle de l'IA?

Natalie Kübler
ALTAE, Université Paris Cité

De l'anecdote au scientifique

Sky my husband!



Jean-loup Chiflet, 1999: *Sky my husband! Ciel mon mari*. Editions du Seuil.

Plan

1. Linguistique de corpus
2. Préfabriqués
3. Difficultés pour les apprenants de langues étrangère
4. Comment l'IA peut aider...ou pas
5. Conclusion

Linguistique de corpus

Étudie le langage à partir de vastes collections structurées de textes (corpus) qui sont analysées à l'aide d'un outil d'interrogations de corpus afin d'identifier des modèles, des structures et des tendances d'utilisation.

⇒ Données empiriques

⇒ Dépasse notre intuition (surprise)

⇒ Informations systématiques

Qu'est-ce qu'un « préfabriqué » ?

Much of our language is not newly invented each time we speak or write, but is largely composed of preconstructed phrases that are stored and retrieved as wholes. (Sinclair 1991)

Un segment de langue est une séquence de mots qui exprime une manière privilégiée de dire les choses par les locuteurs

=> Vision qui s'éloigne de l'approche générativiste

Deux principes

Sinclair 1991

Principe du libre choix : le choix des mots dans les positions possibles syntaxiquement est libre, pourvu qu'ils entrent dans une classe acceptable :

Ex: *j'ai bu de l'eau/du thé/du jus*

Deux principes

Principe idiomatique : « Les mots sont choisis par séquences sur la base de leur attraction lexicale mutuelle, ce qui détermine leurs schémas collocationnel » (Sinclair 1991)

=> recours préférentiel à des séquences « préfabriquées »

Prendre une décision,

faire table rase,

prendre une douche,

commettre un crime

Deux principes

Principe idiomatique
sur la base de leur
détermine leurs sc

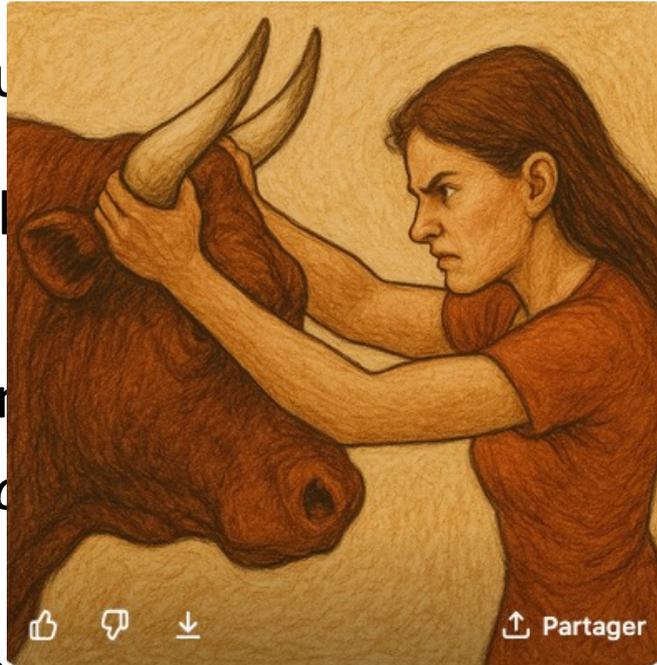
=> recours préfér

Prendre une décision

faire table rase,

prendre une douche,

commettre un crime



par séquences
le, ce qui
(Inclair 1991)

« éfabriquées »

Différents types de (semi-)préfabriqués

- Collocations : *prendre une douche, prendre une décision, steak bleu, café serré, mais aussi en fin de journée, à terme, autrement dit*
- Expr. formulaïques : *je vois ce que tu veux dire, comment ça va?, je vais y réfléchir*
- Schémas lexico-grammaticaux : *il est ADJ(important, crucial) de Vinfinitif (comprendre, décrire, exposer, rappeler), En cas de N négatif (problème, perte, dommage, etc.), Vinfinitif (appeler, contacter, s'adresser, etc.)*

Différents types de (semi-)préfabriqués

- Collocations spécialisées : *obtenir une grossesse*
- Schémas lexico-grammaticaux spécialisés : *Dans cet article, nous V (abordons, explorons, cherchons à , décrivons, etc.)*
- Prosodie sémantique : N (connotation négative) comme N(connotation négative) = faute, vol, crime, assassinat, erreur, etc...

Identification par le corpus

- Corpus: ensemble de textes authentiques assemblés à l'aide de critères explicites en vue d'un objectif précis, sur support électronique afin d'être interrogé par des outils d'interrogation de corpus
 - Mesures d'association: MI score, Log-likelihood, log dice etc.
⇒ Indiquent la force de l'association, plus ou moins rares ou fréquents
- exemple de *douche* dans Sketchengine

Exemple de *douche* dans Sketchengine

douche as noun 819,569x ...

verbs with "douche" as object			
prendre	73,296	7.3	...
encastrier	676	6.9	...
séparer	2,639	6.7	...
carreler	508	6.7	...
glacer	1,199	6.6	...
parer	680	6.0	...
boucher	474	5.9	...
parfumer	548	5.8	...
vitrer	434	5.6	...
installer	2,391	5.5	...
déboucher	210	5.3	...
deboucher	177	5.3	...

verbs with "douche" as subject			
gommer	74	6.6	...
nettoyer	139	6.2	...
fuir	158	5.9	...
habiller	85	5.9	...
consommer	183	5.7	...
spa	52	5.6	...
laver	66	5.5	...
relaxer	51	5.4	...
exfolier	27	5.2	...
évacuer	35	5.1	...
hydrater	35	5.0	...
couler	141	4.8	...

douche comme sujet de verbe: *douche* + *gommer*

CQL *douche* + *gommer* • 74

less than 0.01 • 2.7e-7%



KWIC ▾



Details

Left context

KWIC

Right context

- ⓘ topsante.com : nombreux gels douche.</s><s>Noix de coco : un gel **douche gommant** à faire soi-même</s><s>- 2 cuillères à soup
- ⓘ skyrock.com æ crème hydratante pour le corps à la fraise et une gel **douche gommant** à la fraise également : Body Shop.</s><s>L
- ⓘ blogspot.com 1 épouvantail.</s><s>Bisous ♥</s><s>HmMMM le gel **douche gommant** YR ;) Le jus de tomate je ne peux pas... c'e
- ⓘ beaute.fr >Pour une peau lissée et une silhouette amincie : Gel **douche gommant** à l'huile d'amande de L'occitane (15 euros €
- ⓘ beaute.fr 50 les 200 ml).</s><s>Pour une peau assouplie : Gel **douche gommant** ultra-sensorielle aux grains de fruits d'Yves
- ⓘ promoalert.com :</s><s>Douche gommage douceur caramel</s><s>Le **douche gomme** Cottage au caramel exfolie la peau en douce
- ⓘ jadefromparis.f... 15€.</s><s>Pour les curieux, on y trouve aussi un gel **douche gommant** à la fraise, des gels douche format voyage,
- ⓘ wordpress.com 'et, dedans il y avait tout un tas de produit:</s><s>Gel **douche gommant** énergisant, Yves Rocher (Mangue Coriandr
- ⓘ blogspot.com ouche à la mangue de Séphora sent tellement bon Le **douche gommant** de Yves rochers sent tellement bon La crèr

Tendances

- Ce que montrent les préfabriqués, c'est que ce sont des tendances préférentielles,

=>contrairement aux règles de grammaire, elles peuvent être transgressées

- Le principe de libre choix intervient à nouveau lorsque ces préférences sont transgressées: humour, création littéraire, mettre de l'emphase sur qqch

Cf. commettre + crime, vol, erreur, attentat, etc...

Commettre un roman, un ouvrage, etc...

Tendances

- Canard enchaîné: mars 2010

« Bonjour, je m'appelle Alain. J'ai commis un aller-retour Paris-Martinique dans un avion privé »

D'une langue à l'autre

Caractéristique principale des pré-fabriqués: pas **prédictibles**

Divers degrés de figement :

collocation => transparentes, peu figées

nom composé lexicalisé et expressions figées (idiomatiques)

=> non transparents, sens non compositionnel

=> Différents d'une langue à l'autre

 Allemand	 Français	 Anglais	 Italien	 Español
im Gedächtnis behalten	garder en mémoire	keep in mind	tenere a mente	guardar en la memoria
Stellung nehmen	prendre position	take a stand	prendere posizione	tomar posición
Gepäck aufgeben	enregistrer les bagages	check in luggage	registrare i bagagli	facturar el equipaje
weiches Herz	cœur tendre	soft heart	cuore tenero	corazón blando
zu Ende gehen	toucher à sa fin	come to an end	giungere alla fine	llegar a su fin
Schlagzeilen machen	faire la une	make headlines	fare notizia	ser noticia / salir en portada
Vortrag halten	donner une conférence	give a lecture	tenere una conferenza	dar una conferencia
Radio hören	écouter la radio	listen to the radio	ascoltare la radio	escuchar la radio
werdende Mutter	future mère	expectant mother	futura madre	futura madre
enger Zusammenhang	lien étroit	close connection/link	stretta relazione	estrecha relación
Koffer packen	faire sa valise	pack a suitcase	fare la valigia	hacer la maleta
tiefe Stimme	voix grave	deep voice	voce profonda	voz grave
Sorgen machen	se faire du souci	worry / be worried	preoccuparsi	preocuparse
in Kraft treten	entrer en vigueur	come into force	entrare in vigore	entrar en vigor

Quid pour les apprenants?

Non prédictibilité des préfabriqués

Variation d'une langue à l'autre : *prendre une douche, fare la docha*

=> Faux amis collocationnels : *faire une erreur *do a mistake, *commit a novel, I* have hunger, *je suis 8 ans*

Différences dans les langues de spécialité

⇒ Exposition

⇒ Apprentissage

Variation dans les préfabriqués

Corpus bilingue Open Parallel Corpus (OPUS) FR-DE

Prendre une décision : eine Entscheidung treffen, einen Beschluss fassen, entscheiden, make ou ?take, issue, reach a decision, came to a decision, decide, made up my mind

Tout ce qui menace la viabilité
des écosystèmes menace
également à **terme les**
sociétés humaines

Whatever threatens the
viability of ecosystems also
ultimately threatens human
societies

Insertion dans une autre préfabriqué en LSP

*les investissements en titres (actions, obligations, warrants, options, **contrats à terme** ou tous autres titres au sens large du terme ainsi que des contrats de souscription, d'acquisition ou de cession de tels titres),*

*Investment in securities (shares, bonds, warrants, options, **futures** or any other securities in the broadest sense of the term, as well as contracts to subscribe to, acquire or dispose of such securities),*

Open Parallel Corpus FR-EN

L'IA comme aide?

- Modèles entraînés sur corpus géants
- Traduction automatique neuronale : corpus bilingues alignés => parallèles
- Grands Modèles de Langue : corpus monolingues dans plusieurs langues => comparables
- Prédiction sensible aux co-textes: IA générative

L'IA ne comprend rien...

L'IA manipule des symboles, un peu comme si ces systèmes manipulaient les signifiants, sans comprendre le signifié

Les enfants courent dans les prés bleus en tirant derrière eux une théière qui vole dans le ciel (Wisniewski, conférence novembre 24, Rabat, Maroc)

Les IA ne sont pas ancrées! Dans la réalité...

...mais génère des textes corrects

La capacité à prédire un mot caché a permis une application révolutionnaire: générer du texte. => IA générative

Il suffit de donner un début de texte à l'IA et elle complète.

Le début de texte est appelé amorce ou prompt et permet de générer du texte, une réponse, une traduction dans une autre langue, un script python, etc.

IA et apprentissage des langues

L'IA comme dictionnaire: ça marche bien!

 Allemand	 Français	 Anglais	 Italien	 Espagnol
im Gedächtnis behalten	garder en mémoire	keep in mind	tenere a mente	guardar en la memoria
Stellung nehmen	prendre position	take a stand	prendere posizione	tomar posición
Gepäck aufgeben	enregistrer les bagages	check in luggage	registrare i bagagli	facturar el equipaje
weiches Herz	cœur tendre	soft heart	cuore tenero	corazón blando
zu Ende gehen	toucher à sa fin	come to an end	giungere alla fine	llegar a su fin
Schlagzeilen machen	faire la une	make headlines	fare notizia	ser noticia / salir en portada
Vortrag halten	donner une conférence	give a lecture	tenere una conferenza	dar una conferencia
Radio hören	écouter la radio	listen to the radio	ascoltare la radio	escuchar la radio
werdende Mutter	future mère	expectant mother	futura madre	futura madre
enger Zusammenhang	lien étroit	close connection	stretta relazione	estrecha relación
Koffer packen	faire sa valise	pack a suitcase	fare la valigia	hacer la maleta
tiefe Stimme	voix grave	deep voice	voce profonda	voz grave
Sorgen machen	se faire du souci	worry / be worried	preoccuparsi	preocuparse
in Kraft treten	entrer en vigueur	come into force	entrare in vigore	entrar en vigor

Améliorer la compréhension

L'IA comme dictionnaire: ça marche bien

lien étroit

Prompt: Donne-moi un exemple d'emploi naturel de l'expression « lien étroit »

ChatGPT: *Il existe un lien étroit entre la qualité de l'environnement de travail et la productivité des employés.*

Améliorer le passage à l'autre langue

ChatGPT:

*There is a **close connection** between the quality of the work environment and employee productivity.*

Ou selon le contexte : "a strong correlation" ou "a close relationship" peuvent aussi convenir.

Narrow link , traduction littérale du français *lien étroit* ne convient pas.

Enrichir la connaissance des préfabriqués

ChatGPT: Oui; il existe plusieurs expressions synonymes ou proches de « **lien étroit** », selon le **registre** (formel, courant), le **contexte** (scientifique, affectif, logique, professionnel), et le **champ sémantique** (relation humaine, lien de cause, relation institutionnelle, etc.).

Close relationship/strong bond/tight connection/strong link/strong ties

Génération du mot manquant

close link

Prompt: Insère l'expression qui convient dans l'espace entre guillemets

Pro-Israel activists with « » to the administration are also busy in the information arena inside and outside government. <Sketchengine

*Pro-Israel activists with "**close ties**" to the administration are also busy in the information arena inside and outside government. < ChatGPT*

Risque de nivellement

Prendre le taureau par les cornes => ChatGPT : *take the bull by the horns*

Variante légère : *grab the bull by the horns* < Sketchengine

Variantes : *I've decided to bite the bullet, grab the bull by the horns, grasp the nettles, gild the lily, polish the turd, come to grips* < Sketchengine

Difficultés en LSP => « Créativité »

*a limiting distance beyond which a **moonquake** and a fault are considered unrelated must be established.*

DeepL : *il faut établir une distance limite au-delà de laquelle un **séisme lunaire** et une faille sont considérés comme n'étant pas liés.*

Hésitations

ChatGPT (about DeepL) : "Moonquake" Translation: The term "**moonquake**" is translated as "**séisme lunaire**", which is correct **but loses the specific connotation of 'moonquake'**. A more direct translation like "**tremblement de lune**" could have been considered, though "**séisme lunaire**" is not incorrect

Limites

Machine:

Peu de variation proposée

Exemples et analyses très générales, explications d'emplois selon les registres, superficielles

Pas toujours fiable dans les domaines spécialisés

Créativité parfois débridée, par exemple dans l'extraction terminologique....ou silence

Risque de nivellement de la langue

Limites

Humains:

Risque d'accorder une entière confiance aux réponses générées

=> Les étudiants tendent à accorder une trop grande confiance aux systèmes, qu'il s'agisse de TAO ou d'IA générative

Conclusion

Intégrer l'IA dans l'enseignement => nécessité
=> nécessité de développer un esprit critique
=> nécessité d'un encadrement didactique

Inventer de nouveaux moyens de développer l'esprit critique

=> L'exploitation de corpus reste une approche fiable et utile
pour développer l'approche critique et vérifier des faits
authentiques

Merci de votre attention !

