ALF: Un jeu de données d'analogies françaises à grain fin pour l'évaluation de la connaissance lexicale des grands modèles de langue

Alexander Petrov[1], Antoine Venant[2], François Lareau[2], Yves Lepage[3], Philippe Langlais[1]

[1] RALI/Département d'informatique et de recherche opérationnelle, Université de Montréal
alexander.petrov@umontreal.ca, felipe@iro.umontreal.ca

[2] OLST/Département de linguistique et de traduction, Université de Montréal
antoine.venant@umontreal.ca, francois.lareau@umontreal.ca

[3] Waseda University
yves.lepage@waseda.jp

RÉSUMÉ

La révolution apportée par les grands modèles de langue (LLM) provient de l'étonnante fluidité des textes qu'ils génèrent. Cette fluidité soulève une question scientifique essentielle : quelle quantité de connaissance lexicale les LLM capturent-ils réellement afin de produire un langage aussi fluide ? Pour y répondre, nous présentons ALF, un jeu de données analogiqes librement accessible et doté de riches informations lexicographiques fondées sur la théorie Sens-Texte. Il comprend 2600 analogies lexicales à grain fin avec lesquelles nous évaluons la capacité lexicale de quatre LLM standards : `ChatGPT-4o mini`, `Llama3.0-8B`, `Llama3.1-8B` et `Qwen2.5-14B`. En moyenne, `ChatGPT` et la série `Llama` obtiennent une précision aux environs de 55%, tandis que `Qwen` est juste en dessous du seuil des 60%, ce qui montre qu'ALF pose un défi considérable. Nous identifions en outre certains types d'analogies et de méthodes d'invite qui révèlent des disparités de performance.

ABSTRACT

ALF : A Fine-Grained French Analogy Dataset for Evaluating Lexical Knowledge of Large Language Models

The revolution brought forth by Large Language Models (LLMs) stems from the amazing fluency of the texts they generate. This fluency raises a key scientific question : how much lexical knowledge do LLMs actually capture in order to produce such fluent language ? To address this, we present ALF, a freely-available, analogy dataset endowed with rich lexicographic information grounded in Meaning-Text Theory. It comprises 2600 fine-grained lexical analogies with which we evaluate the lexical ability of four off-the-shelf LLMs, namely `ChatGPT-4o mini`, `Llama3.0-8B`, `Llama3.1-8B`, and `Qwen2.5-14B`. On average, `ChatGPT` and the `Llama` series perform at around 55% accuracy, whereas `Qwen` reaches just below the 60% threshold, thus qualifying ALF as a challenging dataset. We further identify certain types of analogies and prompting methods that reveal performance disparities.

MOTS-CLÉS : Grands modèles de langue, sémantique du langage naturel, ressources et évaluation, théorie Sens-Texte, analogies.

KEYWORDS: Large language models, natural language semantics, resources and evaluation, Meaning-Text theory, analogies.

1 Introduction

La résolution automatique d'équations analogiques telles que *charpentier* : *bois* :: *maçon* : x (dont la solution est *pierre*) va au-delà des analogies formelles que l'on trouve couramment dans les manuels de linguistique, comme cet exemple en swahili [1] *atanipenda (il m'aimera)* : utanipenda *(tu m'aimeras)* :: atanipiga *(il me battra)* : x, avec la solution $x =$ *utanipiga* (tu me battras). Cette tâche va également au-delà des simples oppositions binaires, telles que *chaise* : *fauteuil* :: *tabouret* : x, où x est *siège curule*. Dans ces cas, les mots sont analysés sur la base de caractéristiques binaires, telles que la présence ou l'absence d'un dossier ou d'accoudoirs.

Des résultats convaincants dans la résolution automatique des analogies de sens des mots, avec une précision proche de la performance humaine moyenne, ont d'abord été obtenus par Turney dans une série d'articles (Turney, 2006,

1. Pris dans le répertoire des analogies morphologiques https://github.com/EMarquer/morpho-analogy-amair

2008; Turney *et al.*, 2011) utilisant des questions du Scholastic Assessment Test (SAT). (Mikolov *et al.*, 2013a) ont par la suite considérablement renouvelé l'intérêt pour la résolution analogique en développant le jeu de test d'analogies de Google. Sur la base d'observations empiriques (Mikolov *et al.*, 2013b) et de réflexions théoriques (Levy & Goldberg, 2014), les analogies de mots ont été considérées comme une référence raisonnable pour évaluer la qualité des modèles de plongement lexical statique.

Le jeu de test d'analogies Google comprend à la fois des relations sémantiques (ou encyclopédiques) spécifiques (e.g. `capital-world`) et des relations morphologiques (par exemple, `gram-plural`). Cependant, (Drozd *et al.*, 2016) suggère que ces analogies pourraient être trop spécifiques voire trop faciles pour les modèles entraînés sur des corpus encyclopédiques. Pour y remédier, des jeux d'analogies plus difficiles ont été développés, comme le Bigger Analogy Test Set (BATS) (Gladkova *et al.*, 2016) et les jeux de données U2 et U4 (Kumar & Schockaert, 2023), ou encore ANALOGYKB (Yuan *et al.*, 2023), dérivé des graphes de connaissances. Un autre travail pertinent est le réseau lexical JeuxDeMots (Roux *et al.*, 2024), qui regroupe des relations sémantiques variées entre lexies. Toutefois, ces analogies ne ciblent pas directement les distinctions lexicales fines que nous cherchons à évaluer.

La facilité de construction de modèles de plongement lexical dans différentes langues a conduit à la traduction de jeux de données, e.g., en islandais (Friðriksdóttir *et al.*, 2022), en japonais (Karpinska *et al.*, 2018) ou en bangla (Akter *et al.*, 2023), ainsi qu'à la création de jeux de tests pour des langues spécifiques, comme pour le portugais (Rodrigues *et al.*, 2016; Hartmann *et al.*, 2017). Afin de minimiser les biais centrés sur l'anglais (Ulčar *et al.*, 2019), des jeux de tests multilingues ont également été construits, soit par des méthodes semi-automatiques (Abdou *et al.*, 2018), soit de manière presque entièrement automatique (Ushio *et al.*, 2021), y compris l'extraction à partir d'arbres syntaxiques (Qiu *et al.*, 2015), une approche explorée avant même l'essor des modèles de plongement lexical (Chiu *et al.*, 2007).

Bien que ces jeux de données analogiques soient utiles, ils regroupent des ensembles de relations plutôt arbitraires. De plus, aujourd'hui, un plus grand intérêt est porté à l'évaluation des LLM qu'à celle des modèles de plongement lexical (Petersen & van der Plas, 2023).

Dans ce travail, nous tirons parti de la richesse linguistique d'une ressource lexicale appelée *Réseau lexical du français* (RL-fr) (Ollinger & Polguère, 2023; ATILF, 2024) pour construire ALF (*Analogies lexicales du français*), un jeu de données analogiques qui rassemble un ensemble de relations correspondant aux fonctions lexicales définies dans la théorie linguistique Sens-Texte (Mel'čuk, 1998; Mel'čuk & Polguère, 2021). Cela fait d'ALF un jeu de données analogiques unique qui peut être utilisé pour mesurer la capacité des LLM à capturer la connaissance lexicale fine.

Nous décrivons au §2 la ressource linguistique que nous avons utilisée pour construire ALF, qui est à son tour présenté au §3. Nous décrivons au §4 la méthodologie avec laquelle nous évaluons plusieurs LLM et rapportons les résultats au §5. Nous analysons ensuite nos résultats au §6 et concluons au §7.

2 La ressource RL-fr

Le *Réseau Lexical Français* [2] offre une description fine du lexique français, construite sur la notion de **fonction lexicale** (FL) (Mel'čuk, 1996). Les FL représentent des schémas sémantiques et syntaxiques réguliers, tels que (entre autres) la synonymie, l'antonymie, l'intensification ou les verbes supports, exemplifiées par différentes paires d'unités lexicales. Par exemple, les paires *présence* : *absence*, *en santé* : *malade* et *réussire* : *échouer* sont toutes des exemples d'antonymie, une relation qui implique à la fois une opposition sémantique et une similarité de distribution dans la syntaxe (les éléments appariés doivent avoir la même partie du discours (PdD)). Cette relation est représentée formellement par la FL `Anti`. Les FL sont appelées ainsi parce qu'elles sont définies comme des fonctions d'un élément lexical vers l'ensemble des éléments qui lui sont liés : par exemple `Anti`(*en santé*) = {*malade*, *souffrant*}.

RL-fr représente le lexique français sous la forme d'un graphe orienté dont les noeuds sont étiquetés par des unités lexicales et les arêtes par des FL. Les unités lexicales sont des mots pris dans un sens spécifique ; par exemple, le mot *aimer* correspond à plusieurs unités lexicales : $aimer_1$ (avoir une profonde affection pour quelqu'un), $aimer_2$ (apprécier beaucoup quelque chose), etc. Une arête étiquetée f entre deux unités lexicales L (également appelée *mot-clé*) et L' indique que $L' \in f(L)$, c'est-à-dire que $L : L'$ satisfait la relation lexicale sous-jacente à la FL f. Le graphe est très dense, avec environ 30 000 nœuds et plus de 65 000 arêtes. Comme il est par conséquent difficile à

2. L'ensemble des données et sa documentation technique officielle sont disponibles sur `https://www.ortolang.fr/market/lexicons/lexical-system-fr/v3,1`.

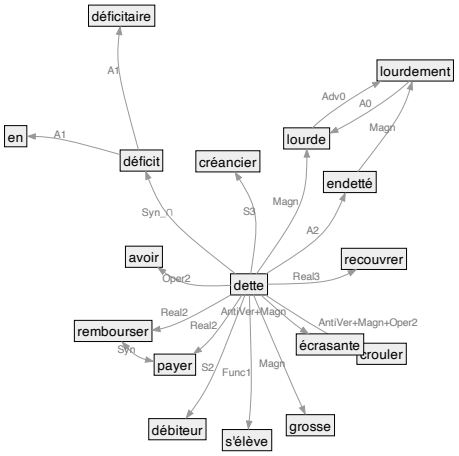

FIGURE 1 – *dette* et certaines de ses unités lexicales apparentées.

visualiser, nous proposons un exemple construit dans la Fig. 1, centré autour de l'unité lexicale *dette*.

Il existe deux types principaux de FL. **Les FL paradigmatiques**, comme `Anti` évoquée plus haut ou `Syn` pour les synonymes, relient une unité lexicale à des alternatives qui pourraient être utilisées à sa place (avec ou sans changement de sens). Parmi les FL paradigmatiques figurent aussi celles qui modifient la PdD tout en préservant le sens, à savoir $\mathtt{S}_0$, $\mathtt{V}_0$, $\mathtt{A}_0$ et $\mathtt{Adv}_0$. Celles-ci renvoient, respectivement, l'équivalent nominal, verbal, adjectival et adverbial d'une unité lexicale, e.g., $\mathtt{A}_0$(*lourdement*) = *lourde* [3] dans la Fig. 1. **Les FL syntagmatiques** relient une unité lexicale à des cooccurrents qui se combinent avec elle dans une phrase. En fonction de la FL, la phrase résultante hérite du sens du mot-clé ou lui ajoute une nuance, tirée d'un ensemble relativement restreint de sens communément exprimés par collocation dans les langues. Par exemple, l'intensification est généralement exprimée par collocation, et la FL `Magn` relie une unité lexicale à des modificateurs idiomatiques d'intensification, par exemple, `Magn`(*dette*)={*grosse*, *lourde*}, `Magn`(*endetté*)=*lourdement*.

Une classe importante de FL syntagmatiques renvoie des collocations verbales. Cette classe comprend, entre autres, des FL représentant formellement différents types de constructions verbales dans lesquelles un verbe sémantiquement faible ou vide est utilisé pour relier un prédicat à ses arguments. Les FL $\mathtt{Oper}_i$ sont des exemples de cette classe. Considérons à nouveau l'unité *dette* et prenons-la comme un prédicat ternaire : 'montant x dû par y à z', où x, y et z sont respectivement les 1er, 2e et 3e arguments du prédicat. Selon cette convention, $\mathtt{Oper}_2$(*dette*)=*avoir* capture formellement la connaissance lexicale selon laquelle on peut spécifier le débiteur (le deuxième argument) en disant *y a une dette*. Les FL établissent une distinction importante entre les verbes de support en fonction de leurs propriétés à l'interface syntaxe/sémantique. Les valeurs de $\mathtt{Oper}_i$ doivent toujours prendre le mot-clé comme objet syntaxique le plus direct. Les collocats verbaux qui ne possèdent pas cette propriété relèvent par conséquent d'autres FL, comme $\mathtt{Func}_i$. On peut par exemple spécifier le montant dû (le premier argument) avec *la dette s'élève à $20 000*, mais *dette* est ici un sujet et non un objet syntaxique, et donc ce type de construction relève de $\mathtt{Func}_1$ et pas d'$\mathtt{Oper}_1$: $\mathtt{Func}_1$(dette)=s'élever. Bien que $\mathtt{Func}_i$ ne fasse pas partie des FL exploitées dans le présent travail, c'est le cas d'$\mathtt{Oper}_1$ et $\mathtt{Real}_1$. La distinction est donc importante, car elle fait partie des connaissances lexicales testées : les paires analogiques correspondant au modèle $\mathtt{Func}_i$ seront considérées comme des réponses incorrectes pour les requêtes portant sur $\mathtt{Oper}_1$ ou $\mathtt{Real}_1$. Une dernière distinction importante concerne l'ajout de sens : du point de vue du lexicographe, $\mathtt{Oper}_i$ n'ajoute aucun sens à L, contrairement à $\mathtt{Real}_i$, qui ajoute une nuance d'« usage typique ou d'accomplissement télique » tout en se comportant de la même manière à l'interface syntaxe/sémantique. Dans notre exemple, le débiteur (argument 2) peut *payer* sa dette ($\mathtt{Real}_2(\textit{dette}) = \textit{payer}$), ce qui ajoute une dimension télique à *avoir* la dette.

Certaines FL fonctionnent de manière paradigmatique ou syntagmatique selon les cas. Par exemple, $\mathtt{A}_1$ dans la Fig. 1 renvoie des expressions adjectivales typiquement utilisées pour qualifier le premier argument du mot-clé. Cette FL peut être utilisée de manière paradigmatique, par exemple $\mathtt{A}_1$(*déficit*)=*déficitaire* (comme dans *le budget est déficitaire*), ou de manière syntagmatique, par exemple $\mathtt{A}_1$(*déficit*)=*en* (comme dans *le budget est en déficit*). La ressource RL-fr est toujours explicite quant à la nature paradigmatique ou syntagmatique d'une paire donnée.

Dans le présent travail, cela n'a d'incidence que sur les fonctions `Magn` et $\mathtt{Real}_1$. Par souci de simplicité, nous traitons chacune d'entre elles comme divisée en une version syntagmatique, pour laquelle nous conservons le nom

3. Par convention dans la littérature des FL, nous abusons des notations et utilisons $F(x) = y$ comme raccourci pour $y \in F(x)$.

standard, et une version paradigmatique que nous désignons par un astérisque. Par exemple, nous avons Magn(*dette*) = *lourde* et $\text{Magn}^*(\textit{lumineux}) = \textit{radiant}$.

Il existe environ 60 FL élémentaires, chacune correspondant à un patron de relation lexicale fréquemment observé dans les langues. Les plus fréquentes sont indiquées dans le tableau 4 au §9.2 de l'annexe. Ces FL élémentaires peuvent être combinées par composition fonctionnelle pour former des **FL composées**. Par exemple, les FL élémentaires Anti et Magn (qui dénotent respectivement l'antonymie et l'intensification) se composent en AntiMagn pour dénoter l'atténuation, où $\text{AntiMagn}(L)=\text{Anti}(\text{Magn}(L))$. En outre, les FL élémentaires et les FL composées peuvent se combiner pour former ce que l'on appelle des **configurations de FL** (notées par un « + »), où deux ou plusieurs FL contribuent indépendamment à la signification de l'expression. Par exemple, $\text{AntiVer+Magn+Oper}_2(\textit{dette})=\textit{crouler}$ dans la Fig. 1 renvoie un verbe qui combine les propriétés de Magn et Oper_2 discutées ci-dessus avec un sens '(déontiquement, éthiquement ou moralement) incorrect' associé à AntiVer, c'est-à-dire, 'Paul croule sous les dettes'≈'Paul a une dette trop importante'.

Les FL représentent formellement des connaissances essentielles à la compétence linguistique. Les locuteurs habiles d'une langue sont capables de reformuler leurs pensées de multiples façons. Par exemple, dans un scénario où Anne doit trop d'argent, on pourrait dire qu'elle *croule sous les dettes* ou parler de sa *dette écrasante*. Cela se fait en tirant parti du type de connaissances encodées dans RL-fr. Ainsi, la mesure de la capacité des LLM à résoudre de telles analogies est un moyen d'évaluer leur capacité à capturer des connaissances lexicales fines, au-delà de la prédiction du mot suivant ou des tâches de complétion de phrases.

3 ALF

À partir de la ressource RL-fr existante, nous avons créé ALF[4], un jeu de données de 2 600 analogies impliquant 25 fonctions lexicales de 2 catégories (paradigmatique et syntagmatique). Nous avons sélectionné des FL comportant suffisamment de paires pour construire 110 analogies, dont 100 ont été utilisées pour tester les modèles, tandis que les 10 restantes ont été réservées aux invites k-*shot*. Nous faisons une exception à cette règle pour deux FL paradigmatiques (qui ont une contrepartie syntagmatique) Magn^* et $\text{Real}_1{}^*$. En raison de leur faible occurrence dans RL-fr, nous n'avons recueilli que 35 analogies pour chacune d'elles, tout en conservant 10 analogies pour les invites k-*shot*.

Pour chaque FL sélectionnée, nous avons rassemblé toutes les instances dans RL-fr, et nous avons échantillonné aléatoirement et sans remplacement 110 (ou 35 pour les deux exceptions ci-dessus) paires d'instances, désormais appelées analogies. ALF rassemble principalement des fonctions lexicales paradigmatiques (21), plus 4 fonctions syntagmatiques. Certaines analogies, ainsi que les FL associées, sont présentées dans le tableau 1. Il convient de noter que, comme les fonctions lexicales renvoient généralement des ensembles de valeurs pour un mot-clé donné, certaines analogies dans ALF se retrouvent avec plusieurs solutions. En effet, nous acceptons comme solutions l'union des ensembles des images de la fonction pour tous les mots-clés s'écrivant de la même manière. Par exemple, *bandit* : *gang* :: *guêpe* : x admet *colonie*, *essaim*, *horde*, *nuage* et *nuée* comme solutions. En moyenne, nous obtenons 2,7 réponses possibles pour les analogies paradigmatiques et 3,8 pour les analogies syntagmatiques.

Paradigmatique	**Syntagmatique**
Mero *boulevard* : *voie* :: *avion* : *carlingue*	Oper_1 *crime* : *commettre* :: *bilan* : *effectuer*
Anti *impardonnable* : *pardonnable* :: *sale* : *propre*	Magn *conte* : *à dormir debout* :: *ressembler* : *pas mal*

TABLE 1 – Exemples d'analogies dans ALF pour une sélection de fonctions lexicales des deux catégories.

4 Méthodologie

4.1 Méthodes d'invite

Des travaux récents (Sahoo *et al.*, 2024; Chang *et al.*, 2024; Yasunaga *et al.*, 2024) ont montré l'importance croissante des invites lorsqu'il s'agit d'optimiser les résultats d'un LLM. Nous concevons donc un vaste espace

4. ALF est disponible sur notre page GitHub : https://github.com/rali-udem/alf.

de recherche de stratégies d'invites. En effet, nous explorons la combinaison de cinq éléments d'invite, chacun représenté ci-dessous par un symbole et une valeur séparée par deux points.

Historique de la conversation [`Hist:0|1`] Avec `Hist:1`, nous conservons un historique des interactions précédentes. Cela est particulièrement utile lorsque le modèle reçoit des informations en retour au cours d'une session. En revanche, avec `Hist:0`, le modèle n'a aucun souvenir d'une interaction précédente.

Style de l'invite [`Pr:d|e`] Les invites peuvent être directes ou plus élaborées. Dans le premier cas (`Pr:d`), l'invite est *Résolvez l'équation analogique suivante :* $a : b :: c : x$. Dans le second cas (`Pr:e`), nous utilisons une invite plus explicite :

Considérez ce premier terme : a
Considérez ce deuxième terme : b
Considérez ce troisième terme : c
Sachant cela, résolvez l'équation analogique suivante : $a : b :: c : x$

Langue [`L:fr|en`] Nous avons testé l'invite en français (`L:fr`) et en anglais (`L:en`), même si le matériel analogique est uniquement en français. Le fait d'inviter un modèle dans différentes langues peut entraîner des différences significatives, comme cela a été observé par exemple dans le *jail-breaking* des modèles (Shen *et al.*, 2024).

Exemples *k-shot* [`k:0|1|3|5|7|10`] Nous cherchons à savoir si le fait de demander k exemples analogiques aide le modèle à résoudre des équations. Nous avons testé l'invite à l'aide de k exemples présélectionnés, où $k \in \{0, 1, 3, 5, 7, 10\}$.

Rétroaction [`Fb:0|1`] Cette configuration n'est possible que si le mode historique est actif (`Hist:1`). Lorsque la rétroaction est activée (`Fb:1`), nous fournissons au modèle cette rétroaction pour les analogies auxquelles il a bien répondu : *Bon travail ! Ceci est une des réponses que l'on cherchait*, tandis que lorsque le modèle se trompe, nous lui fournissons les réponses attendues : *Mauvaise réponse. Des exemples de bonnes réponses que l'on cherchait étaient :* .... Pour décider si une analogie a reçu une réponse correcte, nous nous appuyons sur la métrique CM décrite au §4.2.

Pour chaque FL dans ALF, nous organisons une session *chat*. Tout au long de la session, nous demandons au modèle une analogie et attendons sa réponse. Nous répétons l'opération jusqu'à ce que toutes les analogies aient été soumises au modèle. Chaque session commence par cette invite du système (en français ou en anglais, selon le paramètre `L`) : *Vous êtes un expert en analogies françaises. Vous répondez toujours par une lexie française.* Des exemples de sessions sont présentés au §9.1 de l'annexe.

4.2 Métriques

Les réponses des modèles aux questions d'analogie sont évaluées en fonction des trois métriques suivantes.

Exact Match (EM) attribue un point au modèle si sa réponse correspond exactement à l'une des solutions de référence. Cette métrique, souvent utilisée dans les études sur les analogies, est stricte, car même s'ils ont reçu l'instruction de ne pas le faire, les modèles génératifs ont tendance à ajouter du texte pour présenter leur solution. Par exemple, pour l'analogie *profondément* : *profond* :: *proprement* : x, le système pourrait répondre *L'analogie porte sur la relation entre l'adverbe et l'adjectif [...] Ainsi, la réponse correcte est : propre*, qui ne correspond pas exactement à la référence *propre*, mais qui devrait tout de même être considérée comme valide.

Contain Match (CM) attribue un point à un modèle si l'une des solutions de référence est une sous-chaîne de sa sortie. Dans certains cas ($< 1\%$ des analogies), cette métrique peut produire des faux positifs, par exemple, si le modèle répond *abusement* (alors que la solution de référence est *abus*) pour l'analogie S_0 *enlaidir* : *enlaidissement* :: *abuser* : *abus*.

Semantic Match (SM) tient compte du fait qu'une réponse peut différer d'une référence tout en étant proche dans un espace de plongement sémantique. En nous inspirant de l'hypothèse du parallélogramme (Peterson *et al.*, 2020), nous considérons qu'une prédiction d' est correcte vis-à-vis d'une équation $a : b :: c : x$ dont la solution de référence est d, si sa représentation vectorielle $v(d')$ est suffisamment proche, en termes de distance euclidienne, de $v(d)$. Plus précisément, d' est une solution correcte si $\|v(d') - v(d)\|_2 \leq \frac{\epsilon}{2}$, où l'on définit $\epsilon := \min\{\|v(b) - v(d)\|_2, \|v(c) - v(d)\|_2\}$. À titre d'illustration, `ChatGPT` produit la solution *se battre* à l'équation *broiement* : *broyer* :: *bagarre* : x, dont la solution est *se bagarrer*. Cette solution est considérée comme correcte par SM, mais pas par EM ou CM. Pour l'implémentation, nous utilisons les plongements lexicaux de `fastText` (Bojanowski *et al.*, 2017). [5]

Parce que les modèles sont parfois plus verbeux que nécessaire, ils ont tendance à produire la bonne solution, mais avec des éléments supplémentaires. Nous avons observé quatre cas notables dont nous tenons compte dans notre protocole : **a)** la réponse à une analogie est formulée comme une analogie complète (par exemple, `Llama3.0` répond *crawl* : *nage* :: *astre* : *corps céleste* à l'équation correspondante), **b)** la réponse répète le troisième terme de l'équation (comme dans *souci* : *soucieux* produit par `Llama3.1` pour l'équation *singularité* : *singulier* :: *souci* : x), **c)** la réponse contient une ponctuation supplémentaire (comme dans *dentition.* proposé par `ChatGPT` pour l'équation *enfant* : *enfance* :: *dent* : x), et **d)** la réponse contient des majuscules supplémentaires (comme dans *explicITEMENT*, produit par `Qwen` au lieu de *explicitement* pour l'équation *sévère* : *sévèrement* :: *explicite* : x).

5 Expériences

Nous avons testé quatre LLM dans nos expériences : `ChatGPT-4o mini` (Achiam *et al.*, 2023), en tant que représentant d'un modèle de langue performant mais privé, deux modèles `Llama` (Dubey *et al.*, 2024) en tant que modèles populaires dont l'architecture et les poids sont librement accessibles : `Llama3-8B-Instruct` et `Llama3.1-8B-Instruct`, ainsi que le modèle `Qwen2.5-14B-Instruct`, qui a montré un potentiel prometteur (Bai *et al.*, 2023).

À titre de vérification, nous avons également testé un modèle de plongement lexical `fastText` (Bojanowski *et al.*, 2017) suivant l'approche inaugurée par (Mikolov *et al.*, 2013a). Nous avons utilisé les plongements `fastText` à la fois en anglais et en français.

5.1 Détails de la mise en œuvre

`ChatGPT` Nous avons utilisé l'API OpenAI [6] via Python pour utiliser `ChatGPT-4o mini` avec ses paramètres par défaut. L'API est sans état, ce qui signifie qu'elle ne simule pas un environnement de *chat*. Pour cela, nous avons utilisé la fonctionnalité 'role' de l'API où une invite est définie par son contenu (l'instruction littérale envoyée au modèle), mais aussi par son rôle qui indique comment le modèle doit traiter le contenu. Nous avons utilisé trois rôles pour nuancer nos invites : `system`, `user` et `assistant`.

Le contenu du rôle `system` est interprété comme une instruction de haut niveau (par exemple, être un assistant utile qui donne des réponses pertinentes) ; le contenu du rôle `user` peut être interprété comme une requête habituelle du client, alors que le contenu du rôle `assistant` correspond à la réponse du modèle. En établissant soigneusement une invite `system` et en alternant les invites `user` et `assistant`, nous parvenons à émuler un historique de conversation. Des sessions à plusieurs tours sont illustrées au §9.1 de l'annexe.

`Llama` et `Qwen` En raison des limitations matérielles, nous nous sommes concentrés sur les versions à 8 milliards de paramètres pour les modèles `Llama` et à 14 milliards de paramètres pour `Qwen`, et nous avons utilisé une quantification à 4 bits pour économiser de la mémoire. Nous avons sauvegardé localement les modèles pour faciliter l'expérimentation, mais même ainsi, il n'y avait pas de moyen simple de simuler une session préservant l'historique. Nous avons donc simulé l'historique du *chat* de la même manière que pour `ChatGPT`. Nous avons utilisé les valeurs par défaut pour les paramètres avancés, à l'exception du paramètre `max_new_tokens`, que nous avons réduit à 64 pour éviter que des réponses trop longues ne causent des problèmes de mémoire au niveau du GPU.

5. Nous avons utilisé le modèle `cc.fr.300` de (Grave *et al.*, 2018).
6. https://platform.openai.com/docs/overview

GPU et temps Nous avons exécuté les modèles `Llama` et `Qwen` sur un GPU NVIDIA GeForce RTX 4090 avec 26 GB de mémoire. L'exécution de 100 analogies prend généralement de 30 à 300 secondes, en fonction de la configuration. Pour la configuration `Hist:1-Pr:d-L:en-k:10-Fb:1`, le calcul nécessite 2,0, 2,1 et 2,2 heures pour `ChatGPT`, `Qwen` et `Llama3.0`/`Llama3.1` respectivement, et un peu plus lorsque les modèles sont invités à s'exprimer en français.

`fastText` Nous avons simplement utilisé la méthode `get_analogies` de l'API `fastText`.

5.2 Résultats

5.2.1 BATS

Nous comparons d'abord nos modèles sur le jeu de données populaire BATS (Gladkova *et al.*, 2016) qui différencie quatre grandes catégories de relations linguistiques : flexionnelle, dérivationnelle, encyclopédique et lexicographique, chacune ayant 10 sous-catégories avec 50 paires de mots anglais. En combinant ces paires pour construire des analogies, nous obtenons un ensemble sur lequel nous avons testé un certain nombre de modèles. Les résultats sont présentés dans le tableau 2 qui appelle plusieurs commentaires.

modèle	EM	CM	SM
`ChatGPT`	84,00	86,00	84,70
+ feedback	86,60	88,20	86,70
`Llama3.0`	80,40	81,80	80,60
+ feedback	82,60	83,70	83,00
`Llama3.1`	79,30	80,80	79,60
+ feedback	82,80	83,80	83,20
`Qwen`	86,60	87,40	86,80
+ feedback	86,40	87,90	86,80
`fastText(en)`	32,20	35,90	33,40
`fastText(fr)`	13,10	16,40	13,90
(Yuan *et al.*, 2023)	84,00	(`ChatGPT`)	
(Gladkova *et al.*, 2016)	28,50	(`GloVe`)	

TABLE 2 – Performances sur le jeu de données BATS. Les variantes sans rétroaction (feedback) sont invitées avec `Hist:1-Pr:d-L:en-k:0-Fb:0`, tandis que celles avec rétroaction utilisent `Pr:d-k:0-Hist1-Fb1`. Les modèles `fastText(en)` et `fastText(fr)` utilisent des plongements de dimension 300, tout comme `GloVe` (Pennington *et al.*, 2014) dans (Gladkova *et al.*, 2016).

Nous observons que les LLM offrent des performances beaucoup plus élevées que les approches `fastText`. Ceci corrobore les résultats de (Gladkova *et al.*, 2016) pour les plongements statiques et ceux de (Yuan *et al.*, 2023) pour les LLM. La forte performance des LLM est également cohérente avec les résultats de (Webb *et al.*, 2023), qui indiquent une capacité émergente de GPT-3 dans la résolution d'analogies, souvent meilleure que la performance humaine.

Les modèles `Qwen` et `ChatGPT` sont plus performants que les modèles `Llama`, mais la marge qui les sépare est relativement petite. Toutefois, il convient de noter qu'avec des installations GPU plus importantes, nous aurions pu utiliser des modèles `Llama` plus grands, ce qui aurait probablement permis de réduire l'écart. Nous observons également que le modèle `Qwen`, avec « seulement » 14 milliards de paramètres, offre des performances comparables, voire supérieures, à celles de `ChatGPT`, sans nécessiter de bande passante Internet. Un autre point à noter est que fournir de la rétroaction aux LLM conduit généralement à de légères améliorations, ce qui appelle à la nécessité de documenter de manière adéquate la façon dont les modèles sont interrogés.

Enfin, nos résultats avec `ChatGPT` sont très proches de ceux rapportés par (Yuan *et al.*, 2023), tout comme le sont nos résultats avec `fastText` de ceux de (Gladkova *et al.*, 2016). En effectuant cette comparaison, nous soulignons

que les deux études ne fournissent pas suffisamment de détails pour que nous puissions reproduire exactement leurs résultats.

5.2.2 ALF

Nous discutons ici les performances de nos modèles sur le jeu de données ALF. Le tableau 3 présente les performances de la meilleure variante de chaque modèle, sélectionnée en fonction de la métrique CM. Les résultats complets sont disponibles au §9.5 de l'annexe.

modèle	EM	CM	SM
`ChatGPT`	51,11	55,15	51,49
`Llama3.0`	50,17	53,74	51,19
`Llama3.1`	49,49	55,49	51,02
`Qwen`	**54,00**	**59,66**	**55,11**
`fastText(fr)`	11,92	16,00	12,13

TABLE 3 – Meilleures performances (selon le score CM) sur ALF. Les scores sont calculés sur l'ensemble des analogies, regroupant toutes les fonctions lexicales et catégories. `ChatGPT` utilise la configuration `Hist:1-Pr:d-L:en-k:7-Fb:1`, `Llama3.0` utilise `Hist:1-Pr:d-L:en-k:10-Fb:0`, `Llama3.1` et `Qwen` utilisent `Hist:1-Pr:d-L:fr-k:10-Fb:1`.

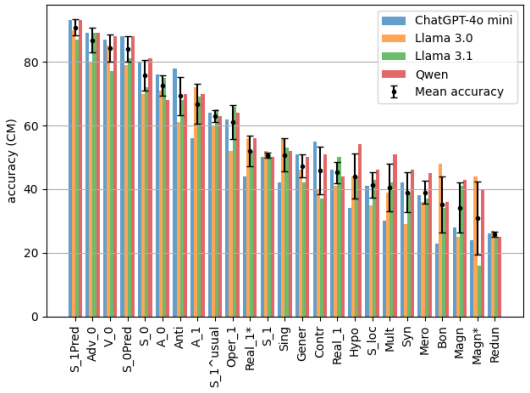

FIGURE 2 – Métrique CM par FL pour les LLM les plus performants et les 25 FL de ALF. Les FL sont classées par ordre décroissant de la performance moyenne des 4 modèles.

Nous observons que la performance sur ALF est significativement plus faible que sur BATS, ce qui indique que ALF n'est pas un jeu de données facile. Deuxièmement, `Qwen` obtient les meilleures performances indépendamment de la métrique et les LLM surpassent de loin les performances de `fastText`.

Sur la Fig. 2, nous observons en outre (voir §6) que les analogies impliquant des fonctions lexicales paradigmatiques sont les plus faciles à résoudre, avec quelques FL, comme $\texttt{S}_1\texttt{Pred}$ atteignant jusqu'à 91% pour CM. $\texttt{S}_1\texttt{Pred}$ est une FL composée qui, étant donné un adjectif A, renvoie un substantif dénotant *quelque chose qui est* A, comme $coupable_{\text{ADJ}}$: $coupable_{\text{N}}$. Dans ce cas, les paires apparentées sont, à quelques exceptions près, morphologiquement très similaires ou identiques en français. Les performances obtenues sur cette FL ne sont donc pas trop surprenantes.

La tendance générale à un meilleur classement des fonctions paradigmatiques n'est pas non plus inattendue, puisque les fonctions paradigmatiques produisent des paires généralement plus proches de celles qu'on peut trouver dans BATS. Cette tendance est corroborée par le fait que $\texttt{Real}_1{}^*$ a environ 7% d'avance sur son homologue syntagmatique $\texttt{Real}_1$. En revanche, la performance de $\texttt{Magn}^*$, inférieure d'environ 3% à celle de `Magn`, indique que les modèles peuvent présenter une préférence pour l'organisation syntagmatique par rapport à l'organisation paradigmatique dans des FL spécifiques.

La Fig. 3 montre les performances de `ChatGPT` agrégées sur les configurations des invites partageant une propriété spécifique (comme l'utilisation de la rétroaction). Par exemple, parmi toutes les variantes de `ChatGPT` que nous avons testées, les 24 variantes incorporants un retour d'information (`Fb:1`) conduisent au score CM moyen le plus élevé de 53,53. En fait, comme le montre le graphique, le score moyen de cette configuration est supérieur d'un écart type à ceux des autres configurations. Cela suggère que même en tenant compte de la variabilité, cette méthode offre un avantage qui témoigne de son efficacité. D'autre part, nous observons que la langue de l'invite (français ou anglais) n'a pas beaucoup d'impact sur les performances, même si l'invite en français entraîne une légère amélioration en moyenne. De même, l'utilisation de l'invite directe est globalement meilleure que l'utilisation de la variante élaborée. Nous fournissons le graphique correspondant pour les trois autres modèles au §9.2 de l'annexe.

Pour voir l'effet de l'augmentation du nombre d'exemples dans les invites, nous représentons les performances de tous nos modèles en fonction de k dans la Fig. 4 dans la variante qui conduit aux meilleurs résultats pour chacun d'entre eux. Nous remarquons que le modèle augmente généralement de façon monotone à mesure que k augmente. Une observation intéressante est que tous les modèles atteignent leur maximum à $k = 10$, sauf `ChatGPT` qui atteint son apogée à $k = 7$.

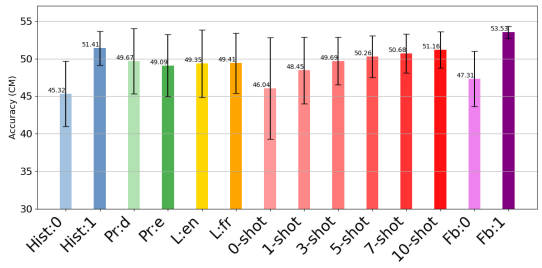

FIGURE 3 – Résultats moyens de CM et écart-type pour toutes les variantes partageant une propriété d'invite pour ChatGPT.

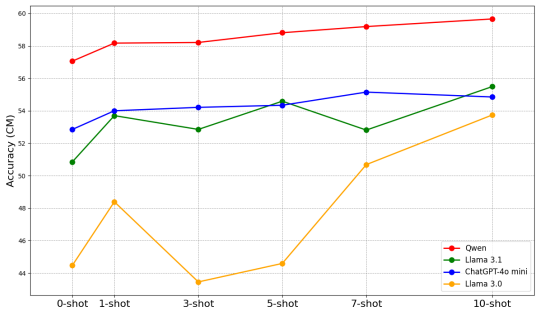

FIGURE 4 – CM des modèles dans la configuration conduisant à la meilleure performance, en fonction du nombre k d'exemples k-shot.

6 Analyse

Un examen plus détaillé des performances et des sorties des LLM permet de faire différentes observations sur la difficulté du jeu de données et la façon dont les modèles parviennent, ou échouent, à mobiliser l'information lexicale pour résoudre les différents types d'analogies. La Fig. 2 offre un apperçu global des difficultés posées par différentes fonctions lexicales, à travers la performance de la meilleure configuration de chacun des quatre modèles sur les 25 FL représentées dans ALF.

Difficulté des FL Les profils de performance des 4 modèles sont remarquablement alignés, dans le sens où les écarts significatifs de performance entre différentes FL sont le plus souvent partagés par tous les modèles (par exemple, tous traitent mieux $\mathtt{S}_0$ que $\mathtt{Oper}_1$). Cet alignement permet d'ordonner partiellement les fonctions lexicales par niveau de difficulté : $\mathtt{S}_1\mathtt{Pred}$, $\mathtt{Adv}_0$ et $\mathtt{V}_0$ sont les plus faciles à résoudre puisque tous les modèles obtiennent de bonnes performances sur ces FL, ChatGPT et Qwen dépassant les 80%. $\mathtt{S}_0$ et $\mathtt{A}_0$ sont légérement plus difficiles, induisant des performances resserrées autour de 70% CM. Anti, $\mathtt{A}_1$, $\mathtt{Real}_1{}^*$, $\mathtt{S}_1\mathtt{usual}$ (paradigmatiques) et $\mathtt{Oper}_1$ (syntagmatique) posent un peu plus de problèmes que les précédentes : sur ces FL, la performance des modèles varie entre 60% et 70%, à quelques exceptions près. ChatGPT se démarque positivement par ses performances sur Anti et négativement par ses performances sur $\mathtt{A}_1$, et Llama3.0 traite $\mathtt{Oper}_1$ moins bien que ses concurrents. Enfin, les FL restantes semblent globalement plus difficiles, avec des performances systématiquement inférieures à 60% pour tous les modèles, et particulièrement mauvaises pour les fonctions syntagmatiques Bon (à l'exception de Llama 3.0) et Magn, et des fonction paradigmatiques Redun et $\mathtt{Magn}^*$, pour lesquelles les performances s'étalent entre 20 et 40%, selon les modèles. Ces observations sont, dans une large mesure, indépendantes de la métrique considérée (avec bien sûr quelques différences mineures, par exemple, la métrique CM surestime les performances de ChatGPT sur $\mathtt{A}_0$ alors que les deux autres métriques sous-estiment largement les capacités de Qwen sur $\mathtt{Adv}_0$, en raison de sa verbosité).

Morphologie Les modèles semblent particulièrement bien maîtriser la morphologie dérivationnelle, puisqu'ils obtiennent leurs meilleures performances sur les FL faisant intervenir des motifs morphologiques réguliers. C'est en tout cas ce que suggère les résultats des FL $\mathtt{V}_0$, $\mathtt{Adv}_0$ $\mathtt{S}_0$ et $\mathtt{A}_0$, pour lesquelles les paires analogiques (*e.g.* nom-verbe, verbe-adverbe ou adjectif-adverbe) sont le plus souvent liées par suffixation. Comme mentionné précédemment, c'est également le cas pour $\mathtt{S}_1\mathtt{Pred}$ et $\mathtt{S}_0\mathtt{Pred}$ qui sont elles aussi très bien traitées. Cependant, les erreurs commises sur ces FL montrent également des difficultés de la part des modèles à reconnaître les cas particuliers, ou a négliger d'autres caractéristiques des paires analogiques (notamment les liens de sens, la grammaticalité du résultat) au profit de la proximité morphologique. Par exemple, Qwen et ChatGPT proposent tous deux *consciencieusement* au lieu de *consciemment* comme solution à l'équation $\mathtt{Adv}_0$ *lexical* : *lexicalement* :: *conscient* : x, et ils inventent régulièrement des néologismes en appliquant des schémas morphosyntaxiques de manière inappropriée : ChatGPT répond par exemple *trébuchage* à l'équation $\mathtt{S}_0$ *tailler* : *taillage* :: *trébucher* : x et Qwen propose *jugulaire* comme solution à l'équation *orageux* : *orage* :: *jugal* : x, ainsi que *sororité* (au lieu de *soeur*) à l'équation *pendulaire* : *pendule* :: *sororal* : x, ou encore *giclure* comme solution à *bâiller* : *baîllement* :: *gicler* : x (toutes trois correspondant à la FL $\mathtt{S}_0$). Les deux modèles répondent respectivement *chauvrerie* et *chauvité* (au lieu de *calvitie*) pour l'entrée *vengeur* : *vengeance* :: *chauve* : x ($\mathtt{S}_0\mathtt{Pred}$).

Distinction événement/participant Même si elles font également intervenir des connaissances morphologiques, les fonctions $\mathtt{A}_0$ et $\mathtt{S}_0$ semblent plus difficiles à résoudre que $\mathtt{V}_0$ et $\mathtt{Adv}_0$. Une explication possible de ce phénomène est que $\mathtt{A}_0$ et $\mathtt{S}_0$, contrairement à $\mathtt{V}_0$ et $\mathtt{Adv}_0$, sont en compétition avec $\mathtt{A}_1$ et $\mathtt{S}_1$/$\mathtt{S}_1$usual. Les modèles peuvent avoir de la difficulté à distinguer entre ces FL, car cela requiert, en plus de dériver un mot de la bonne catégorie, de distinguer les paires dont les éléments expriment un même événement ou état, comme *détruire* : *destruction*, des paires qui relient un événement à ses participants, comme *détruire* : *destructeur*. De fait, les modèles n'y parviennent pas toujours, et proposent à plusieurs reprises des sorties ne respectant pas les contraintes de sens de la FL cible. On observe par exemple *tranche* proposée au lieu de *tranchage* en tant qu'analogue $\mathtt{S}_0$ de *trancher*, ou *situationnel* (une $\mathtt{A}_0$) proposée au lieu de *situé* en tant qu'analogue $\mathtt{A}_1$ de *situation*.

Types d'erreur À partir d'un examen approfondi des réponses incorrectes de Qwen et ChatGPT sur les 25 FL représentées dans ALF dans lequel nous avons cherché à regrouper les erreurs de nature similaire, nous identifions les cas suivants :

a) **Violation de contraintes syntaxiques ou combinatoires**, survenant lorsque les modèles ne parviennent pas à respecter les contraintes imposées sur la PdD de la valeur des FL paradigmatiques, ou lorsque la combinaison mot-clef/valeur proposée pour une fonction syntagmatique n'est pas grammaticale. Par exemple, cela se produit lorsque Qwen propose *vers* comme solution de l'équation Anti *desserer* : *serrer* :: *envers* : x, ou alors, propose l'adverbe dangeureusement (plutôt qu'un adjectif comme *long* ou *interminable*) pour l'équation Magn *bien* : *formidablement* :: *route* : x. De même ChatGPT propose *mnémotechnique* (au lieu de *mémoire*) comme solution à l'équation $\mathtt{S}_0$ *empiler* : *empilage* :: *mnémonique* : x ou encore *violemment* (un adverbe plutôt qu'un adjectif comme *violent* ou *terrible*) pour l'équation Magn *ressembler* : *fortement* :: *tempête* : x. On peut d'ailleurs souligner à nouveau l'influence de la proximité morphologique dans certains de ces exemples.

b) **Violation de contraintes sémantiques**, survenant quand la réponse d'un modèle ne véhicule pas le sens requis par la FL sous-jacente. Régulièrement, les modèles ne parviennent pas à cerner précisément l'ensemble des caractéristiques de la relation analogique à partir des exemples k-shot et de la paire analogique à résoudre (et, pour les configurations avec feedback, des paires précédentes). En conséquence, ils proposent des réponses reliées au mot-clef de manière plus ou moins arbitraires. Par exemple, ChatGPT répond *éveillé*, au lieu de *e.g. merveilleux* comme solution à l'équation Bon *yaourt* : *crémeux* :: *rêve* : x. Bon est une fonction lexicale qui relie un mot-clef à un modificateur **mélioratif** typique du mot-clef, et ChatGPT semble donc n'avoir pas pu inférer la dimension méliorative de la relation analogique. S'il n'est pas surprenant que cette caractéristique soit difficile à cerner dans une configuration 0-shot, il est important de noter que les exemples discutés ici ont été obtenus dans un contexte 7-shot avec feedback, dans laquelle le modèle a été exposé à au moins 7 analogies (donc 14 paires) comme *jambe* : *joli* :: *air* : *frais*. Dans d'autres cas, ce sont les contraintes de sélections des verbes sont violées, par exemple lorsque ChatGPT répond *prendre* comme solution de *autocar* : *conduire* :: *relâche* : x ($\mathtt{Real}_1$).

c) **Production de néologismes ou de locution non-attestées**. C'est typiquement ce qui se produit lorsque les modèles appliquent à mauvais escient des transformations morphologiques, comme dans les examples donnés plus haut (*trébuchage*, *chauvité*, ...).

d) **Préférence d'expressions inusuelles à des expressions idiomatiques**. Dans ces cas, les modèles respectent les contraintes sémantiques et syntaxiques de la FL sous-jacente, mais en privilégiant une construction peu fréquente à un équivalent idiomatique. Ce type d'erreur est le moins fréquent et concerne surtout les FL syntagmatiques. Qwen propose ainsi *habilement* (plutôt que *bien* ou *comme un rossignol*) comme Bon de *chanter*, et ChatGPT préfère *circuler* à *rouler* comme verbe support $\mathtt{Real}_1$ de *mobylette*.

Nous avons manuellement annoté 497 exemples de réponses de ChatGPT et Qwen jugées incorrectes par CM [7] en leur assignant soit l'étiquette faux négatif, soit l'une des catégories de la typologie ci-dessus. Sur la base de ces annotations nous estimons que les faux-négatifs représentent environ 8% des erreurs. Parmi les vrais négatifs, nous estimons que les erreurs de type a représentent environ 8% des cas, les erreurs de type b environ 69%, les erreurs de type c environ 20% et les erreurs de type d environ 3%.

Influence de l'anglais Dans certains cas, l'influence de l'anglais pourrait expliquer les choix des modèles. Par exemple, dans l'exemple de *relâche* discuté plus haut, *prendre* n'est pas un $\mathtt{Oper}_1$ valide en français, tandis que *take* est un $\mathtt{Oper}_1$ de *break* en anglais.

7. 10 par FL et par modèle, sauf pour $\mathtt{S}_1$Pred pour laquelle les modèles ont fait moins de 10 erreurs.

7 Discussion

Dans cet article, nous avons présenté un ensemble unique de données analogiques pour évaluer la profondeur de la connaissance lexicale dans les LLM. Allant au-delà des relations morphologiques ou encyclopédiques typiques, nous adoptons le cadre linguistique des fonctions lexicales de la théorie sens-texte, en tirant parti des riches descriptions de la ressource RL-fr.

En comparant avec le jeu de données analogiques standard BATS, nous trouvons que ALF est significativement plus difficile. Malgré quelques variations, les LLM testés atteignent une précision comparable à travers ALF, avec une performance plus élevée (>70%) pour les transformations de PdD, mais une précision plus faible (40–60%) pour des fonctions comme $\texttt{Oper}_1$ ou $\texttt{Real}_1$, cruciales pour la sélection de verbes. Pour les fonctions relatives aux aspects plus idiomatiques du langage, comme Bon ou Magn, les performances sont moins bonnes.

Nos expériences confirment que les LLMs encodent, et peuvent restituer abstraitement, beaucoup de connaissances lexicales fondamentales. Dans l'ensemble, les performances des modèles sur ALF révèlent une très bonne aptitude à manipuler la morphologie dérivationnelle. Les performances, bien que plus faibles, obtenues sur des FL comme Anti, $\texttt{Oper}_1$ et $\texttt{Real}_1{}^*$ témoignent également une bonne aptitude à reconnaître des motifs sémantiques, combinatoires ou collocationnels. L'analyse d'erreur révèle qu'identifier les distinctions fines des relations analogiques présente un défi considérable pour les modèles, y compris dans des configurations avec exemples et feedback. Par ailleurs, même les fonctions les mieux traitées présentent des motifs d'erreur réguliers (tendance à sur-généraliser les dérivations morphologiques, difficultés à distinguer éventualité et actant) et donc une possible marge de progression.

Nos résultats ouvrent des pistes pour explorer la connaissance lexicale des LLM, en questionnant la manière dont cette connaissance est explicitement représentée. Des travaux en cours indiquent que, même sans analogies, les modèles rencontrent encore des difficultés avec des questions aussi simples que « Quel est le mot auxiliaire utilisé pour intensifier le mot *dette* ? » Nous soulignons la nécessité de méthodes alternatives pour évaluer la connaissance lexicale, ce qui constitue une direction importante pour les recherches futures.

8 Limitations

Bien que nous nous sommes limités à des modèles de taille moyenne en raison des contraintes de coût et de matériel, mesurer les performances de très grands modèles et plus généralement l'impact de la taille des modèles serait sans doute très informatif.

Dans certaines conditions, notamment dans des configurations 0-shot, sans feedback ou sans historique, les analogies obtenues à partir de certaines FL (comme Bon, ou $\texttt{A}_1$) sont probablement difficiles à désambiguiser. Par ailleurs, certaines analogies du RL-fr peuvent paraître contre-intuitives à première vue, notamment pour un observateur non-linguiste. Par exemple, sur le plan **sémantique**, il est commun de considérer que des adjectifs, des verbes et des noms peuvent dénoter des prédicats intensifiables. Une même opération sémantique comme l'intensification peut donc s'appliquer également aux trois catégories, et il en résulte des analogies pour Magn faisant intervenir des analogues de PdD différentes, comme dans la paire *félicitation* : *chaleureux* :: *malade* : *complétement*. Nous avons pris le parti d'adhérer aux prescriptions de la théorie sens-texte en conservant de tels cas, mais évaluer les effets d'une sélection plus fine pour un jeu de données plus 'intuitif' constitue une piste de recherche intéressante.

Notre ressource évalue les modèles sur leur maîtrise du lexique français, mais nous soulignons que la théorie sous-jacente est en soi indépendante de toute langue et qu'elle vise à capturer des relations lexicales présentes dans diverses langues. De plus, le français est bien représenté dans les données d'entraînement de tous les modèles testés, ce qui en fait un point de départ pertinent pour sonder leurs capacités. Néanmoins, disposer d'un tel jeu de données pour plusieurs langues ouvrirait des perspectives considérables, permettant notamment d'explorer les analogies interlinguistiques et d'autres pistes multilingues.

Nous ne disposons pas encore d'une mesure fiable des performances humaines, expertes ou non, sur ALF (ce qui nécessiterait un protocole d'annotation soigneusement conçu et un volume important d'annotations), mais nous y travaillons. Des expériences préliminaires suggèrent que des experts humains obtiendraient également de meilleures performances sur les fonctions paradigmatiques, avec une précision parfaite sur les fonctions dérivationnelles $\texttt{Adv}_0$, $\texttt{V}_0$ et $\texttt{A}_0$, et qu'ils maîtriseraient nettement mieux des fonctions telles que Magn que les modèles testés. Toutefois, ces résultats reposent sur un échantillon très restreint de 10 analogies par FL, induisant une variance élevée entre annotateurs pour certaines fonctions plus complexes.

Références

ABDOU M., KULMIZEV A. & RAVISHANKAR V. (2018). Mgad : Multilingual generation of analogy datasets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

ACHIAM J., ADLER S., AGARWAL S., AHMAD L., AKKAYA I., ALEMAN F. L., ALMEIDA D., ALTENSCHMIDT J., ALTMAN S., ANADKAT S. *et al.* (2023). Gpt-4 technical report. *arXiv preprint arXiv :2303.08774*.

AKTER M., SARKAR S. & SANTU S. K. K. (2023). On evaluation of bangla word analogies. *arXiv preprint arXiv :2304.04613*.

ATILF (2024). Réseau lexical du français (rl-fr). ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

BAI J., BAI S., CHU Y., CUI Z., DANG K., DENG X., FAN Y., GE W., HAN Y., HUANG F. *et al.* (2023). Qwen technical report. *arXiv preprint arXiv :2309.16609*.

BOJANOWSKI P., GRAVE E., JOULIN A. & MIKOLOV T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, **5**, 135–146. DOI : 10.1162/tacl_a_00051.

CHANG K., XU S., WANG C., LUO Y., XIAO T. & ZHU J. (2024). Efficient prompting methods for large language models : A survey. *ArXiv*, **abs/2404.01077**.

CHIU A., POUPART P. & DIMARCO C. (2007). Generating lexical analogies using dependency relations. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, p. 561–570.

DROZD A., GLADKOVA A. & MATSUOKA S. (2016). Word embeddings, analogies, and machine learning : Beyond king - man + woman = queen. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics : Technical Papers*, p. 3519–3530 : The COLING 2016 Organizing Committee.

DUBEY A., JAUHRI A., PANDEY A., KADIAN A., AL-DAHLE A., LETMAN A., MATHUR A., SCHELTEN A., YANG A., FAN A. *et al.* (2024). The llama 3 herd of models. *arXiv preprint arXiv :2407.21783*.

FRIÐRIKSDÓTTIR S. R., DANÍELSSON H., STEINGRÍMSSON S. & SIGURÐSSON E. (2022). Icebats : An icelandic adaptation of the bigger analogy test set. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, p. 4227–4234.

GLADKOVA A., DROZD A. & MATSUOKA S. (2016). Analogy-based detection of morphological and semantic relations with word embeddings : what works and what doesn't. In J. ANDREAS, E. CHOI & A. LAZARIDOU, Éds., *Proceedings of the NAACL Student Research Workshop*, p. 8–15, San Diego, California : Association for Computational Linguistics. DOI : 10.18653/v1/N16-2002.

GRAVE E., BOJANOWSKI P., GUPTA P., JOULIN A. & MIKOLOV T. (2018). Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

HARTMANN N., FONSECA E., SHULBY C., TREVISO M., RODRIGUES J. & ALUISIO S. (2017). Portuguese word embeddings : Evaluating on word analogies and natural language tasks. *arXiv preprint arXiv :1708.06025*.

KARPINSKA M., LI B., ROGERS A. & DROZD A. (2018). Subcharacter information in japanese embeddings : When is it worth it ? In *Proceedings of the Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP*, p. 28–37.

KILGARRIFF A., BAISA V., BUŠTA J., JAKUBÍČEK M., KOVÁŘ V., MICHELFEIT J., RYCHLÝ P. & SUCHOMEL V. (2014). The sketch engine : ten years on. *Lexicography*, **1**(1), 7–36.

KUMAR N. & SCHOCKAERT S. (2023). Solving hard analogy questions with relation embedding chains. In H. BOUAMOR, J. PINO & K. BALI, Éds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, p. 6224–6236, Singapore : Association for Computational Linguistics. DOI : 10.18653/v1/2023.emnlp-main.382.

LEVY O. & GOLDBERG Y. (2014). Linguistic regularities in sparse and explicit word representations. In *Proceedings of the eighteenth conference on computational natural language learning*, p. 171–180.

MEL'ČUK I. (1998). Collocations and Lexical Functions. In *Phraseology : Theory, Analysis, and Applications*, p. 23–53. Oxford University Press. DOI : 10.1093/oso/9780198294252.003.0002.

MEL'ČUK I. & POLGUÈRE A. (2021). Les fonctions lexicales dernier cri. In S. MARENGO, Éd., *La Théorie Sens-Texte. Concepts-clés et applications*, Dixit Grammatica, p. 75–155. L'Harmattan.

MEL'ČUK I. A. (1996). Lexical functions : A tool for the description of lexical relations in a lexicon. In L. WANNER, Éd., *Lexical Functions in Lexicography and Natural Language Processing*, volume 31 de *Studies in language companion series*, p. 37–102. Amsterdam/Philadelphia : John Benjamins.

MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. & DEAN J. (2013a). Efficient estimation of word representations in vector space. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

MIKOLOV T., YIH W.-T. & ZWEIG G. (2013b). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics : Human language technologies*, p. 746–751.

OLLINGER S. & POLGUÈRE A. (2023). Distribution des systèmes lexicaux. ANALYSE ET TRAITEMENT DE LA LANGUE FRANÇAISE INFORMATIQUE. Ressource distribuée sous licence : Creative Commons – Attribution 4.0 International (CC BY 4.0).

PENNINGTON J., SOCHER R. & MANNING C. D. (2014). Glove : Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, p. 1532–1543.

PETERSEN M. R. & VAN DER PLAS L. (2023). Can language models learn analogical reasoning ? investigating training objectives and comparisons to human performance. *arXiv preprint arXiv :2310.05597*.

PETERSON J. C., CHEN D. & GRIFFITHS T. L. (2020). Parallelograms revisited : Exploring the limitations of vector space models for simple analogies. *Cognition*, **205**, 104440.

QIU L., ZHANG Y. & LU Y. (2015). Syntactic dependencies and distributed word representations for analogy detection and mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, p. 2441–2450.

RODRIGUES J., BRANCO A., NEALE S. & SILVA J. (2016). LX-DSemVectors : Distributional semantics models for Portuguese. In J. SILVA, R. RIBEIRO, P. QUARESMA, A. ADAMI & A. BRANCO, Éds., *Computational Processing of the Portuguese Language*, p. 259–270, Cham : Springer International Publishing.

ROUX J., GUENOUNE H., LAFOURCADE M. & MOOT R. (2024). Explaining metaphors in the french language by solving analogies using a knowledge graph. In *International Conference on Text, Speech, and Dialogue*, p. 43–54 : Springer.

SAHOO P., SINGH A. K., SAHA S., JAIN V., MONDAL S. & CHADHA A. (2024). A systematic survey of prompt engineering in large language models : Techniques and applications. *arXiv preprint arXiv :2402.07927*.

SHAPIRO S. S. & WILK M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3-4), 591–611. DOI : 10.1093/biomet/52.3-4.591.

SHEN L., TAN W., CHEN S., CHEN Y., ZHANG J., XU H., ZHENG B., KOEHN P. & KHASHABI D. (2024). The language barrier : Dissecting safety challenges of llms in multilingual contexts.

TURNEY P., NEUMAN Y., ASSAF D. & COHEN Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, p. 680–690, Edinburgh, Scotland, UK. : Association for Computational Linguistics.

TURNEY P. D. (2006). Similarity of semantic relations. *Computational Linguistics*, **32**(2), 379–416.

TURNEY P. D. (2008). The latent relation mapping engine : Algorithm and experiments. *J. Artif. Int. Res.*, **33**(1), 615–655.

ULČAR M., VAIK K., LINDSTRÖM J., DAILIDĖNAITĖ M. & ROBNIK-ŠIKONJA M. (2019). Multilingual culture-independent word analogy datasets. *arXiv preprint arXiv :1911.10038*.

USHIO A., CAMACHO-COLLADOS J. & SCHOCKAERT S. (2021). Distilling relation embeddings from pre-trained language models. *arXiv preprint arXiv :2110.15705*.

WEBB T., HOLYOAK K. J. & LU H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, **7**(9), 1526–1541. DOI : 10.1038/s41562-023-01659-w.

WILCOXON F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**(6), 80–83. DOI : 10.2307/3001968.

YASUNAGA M., CHEN X., LI Y., PASUPAT P., LESKOVEC J., LIANG P., CHI E. H. & ZHOU D. (2024). Large language models as analogical reasoners. In *The Twelfth International Conference on Learning Representations*.

YUAN S., CHEN J., SUN C., LIANG J., XIAO Y. & YANG D. (2023). Analogykb : Unlocking analogical reasoning of language models with a million-scale knowledge base. *arXiv preprint arXiv :2305.05994*.

9 Annexe

9.1 Illustration de sessions de *chat*

system
You are an expert on French analogies. You always respond in one French lexeme.
user
Consider the following examples.
goût : *boulimie* :: *pluie* : *cordes*
ressemblance : *portrait* :: *bien* : *merveilleusement*
rouge : *cramoisi* :: *similaire* : *identique*
Now, use this pattern to solve the following analogical equation :
user
Solve the following analogical equation : lumineux : *resplendissant* :: *flot* : x
assistant
ondoyant
user
Wrong answer. Examples of correct answers we were looking for were : 'geyser', and 'torrent'
user
Solve the following analogical equation : averse : *déluge* :: *barbiche* : x
assistant
boucle
user
Wrong answer. Examples of correct answers we were looking for were : 'barbe'
⋮

FIGURE 5 – Illustration d'une interaction de *chat* pour la fonction lexicale `Magn*(Para)` en mode *3-shot* avec feedback `Hist:1-Pr:d-L:en-k:3-Fb:1`. Les rôles des invites sont indiqués en gras.

system
Vous êtes un expert des analogies françaises. Vous répondez toujours en une lexie française
user
Considérez les exemples suivants.
goût : *boulimie* :: *pluie* : *cordes*
ressemblance : *portrait* :: *bien* : *merveilleusement*
rouge : *cramoisi* :: *similaire* : *identique*
Maintenant, utilisez ce patron pour résoudre l'équation analogique suivante :
user
Résolvez l'équation analogique suivante : lumineux : *resplendissant* :: *flot* : x
assistant
ondoyant
==
Réinitialisation du *chat*
==
system
Vous êtes un expert des analogies françaises. Vous répondez toujours en une lexie française
user
Considérez les exemples suivants.
goût : *boulimie* :: *pluie* : *cordes*
ressemblance : *portrait* :: *bien* : *merveilleusement*
rouge : *cramoisi* :: *similaire* : *identique*
Maintenant, utilisez ce patron pour résoudre l'équation analogique suivante :
user
Résolvez l'équation analogique suivante : averse : *déluge* :: *barbiche* : x
assistant
moustache
⋮

FIGURE 6 – Illustration d'une interaction de *chat* pour la fonction lexicale `Magn*(Para)` en mode *3-shot* avec feedback `Hist:0-Pr:d-L:fr-k:3-Fb:0`. Les rôles des invites sont indiqués en gras.

9.2 Tableaux et graphiques supplémentaires

FL	Sens	Compte
Syn	Synonyme	29275
S_1	Nom de l'argument 1	3821
Gener	Terme générique	3119
Anti	Antonyme	2690
S_0	Équivalent nominal	2377
S_2	Nom de l'argument 2	1676
V_0	Équivalent verbal	1608
A_0	Équivalent adjectival	1224
Magn	Intensificateur	1164
A_1	Modificateur de l' argument 1	1036
$Oper_1$	Verbe léger	890
$Real_1$	Verbe de réalisation	826

TABLE 4 – Les FL les plus fréquentes dans RL-fr.

Métrique	P	S	MG
EM	13,80	2,75	11,92
CM	18,51	3,75	16,00
SM	14,00	3,00	12,13

TABLE 5 – Performances de `fastTextfr` sur ALF. P=paradigmatique, S=syntagmatique, MG=moyenne générale

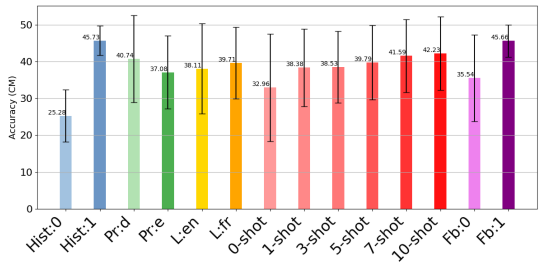

FIGURE 7 – Résultats moyens de CM et écart-type pour toutes les variantes partageant une propriété d'invite pour `Llama3.0`.

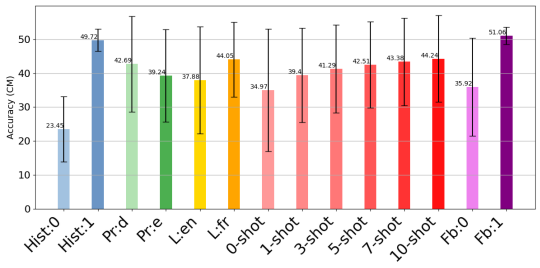

FIGURE 8 – Résultats moyens de CM et écart-type pour toutes les variantes partageant une propriété d'invite pour `Llama3.1`.

9.3 Remarque sur la fréquence des mots dans ALF

Afin de vérifier que notre jeu d'analogies ne repose pas sur des mots trop rares ou trop fréquents, nous avons extrait l'ensemble des mots présents, réduit à leur forme lemmatisée à l'aide de spaCy [8], puis récupéré leur rang de fréquence dans un corpus de référence [9]. La médiane des rangs ainsi obtenus est de 3548, ce qui correspond au mot lampe. Cela suggère que les mots dans notre jeux de données sont d'un usage courant, sans être excessivement fréquents ni trop rares.

9.4 Robustesse des résultats expérimentaux

Pour évaluer la robustesse de nos résultats, nous avons mené une seconde série d'expériences. Nous désignons les moyennes obtenues lors du premier et du deuxième essai par *essai1* et *essai2*, respectivement. Notre objectif est

8. https://spacy.io
9. Nous utilisons le corpus *frTenTen17*, accessible via *Sketch Engine* (Kilgarriff *et al.*, 2014)

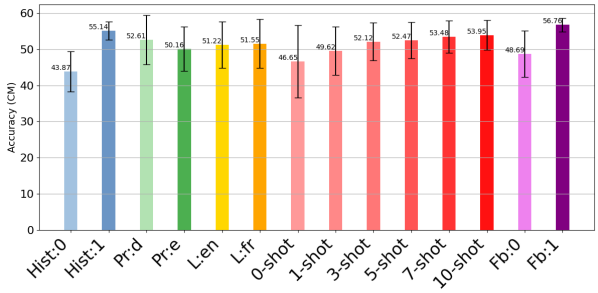

FIGURE 9 – Résultats moyens de CM et écart-type pour toutes les variantes partageant une propriété d'invite pour `Qwen`.

de comparer les performances entre *essai1* et *essai2*, en posant comme hypothèse nulle l'absence de différence significative entre les deux. Pour vérifier la normalité des différences, nous avons réalisé un test de Shapiro-Wilk (Shapiro & Wilk, 1965). Le test a montré que les différences ne suivent pas une distribution normale ($p < 0{,}001$). En conséquence, nous avons utilisé le test de Wilcoxon Signed-Rank (Wilcoxon, 1945) pour évaluer les différences entre les deux essais. La valeur p obtenue est de 0,20. Cette valeur étant supérieure au seuil de 0,05, nous ne rejetons pas l'hypothèse nulle, ce qui suggère que les résultats sont robustes vis-à-vis de la répétition des expériences.

Pour évaluer la robustesse de nos résultats, nous avons mené une seconde série d'expériences. Nous désignons les moyennes obtenues lors du premier et du deuxième essai par *essai1* et *essai2*, respectivement. Notre objectif est de comparer les performances entre *essai1* et *essai2*, en posant comme hypothèse nulle l'absence de différence significative entre les deux. Pour vérifier la normalité des différences, nous avons réalisé un test de Shapiro-Wilk (Shapiro & Wilk, 1965). Le test a montré que les différences ne suivent pas une distribution normale ($p < 0{,}001$). En conséquence, nous avons utilisé le test de Wilcoxon Signed-Rank (Wilcoxon, 1945) pour évaluer les différences entre les deux essais. La valeur p obtenue est de 0,20. Cette valeur étant supérieure au seuil de 0,05, nous ne rejetons pas l'hypothèse nulle, ce qui suggère que les résultats sont robustes vis-à-vis de la répétition des expériences.

9.5 Les performances des LLM sur ALF dans toutes les configurations

Les tableaux ci-dessous présentent toutes les moyennes sur ALF pour toutes les configurations.

TABLE 6 – EM des variantes ChatGPT (Hist:0), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	37,49	18,50	34,26
d	en	0	1	—	—	—
d	en	1	0	43,80	20,25	39,79
d	en	1	1	—	—	—
d	en	3	0	46,97	22,50	42,81
d	en	3	1	—	—	—
d	en	5	0	48,57	25,00	44,56
d	en	5	1	—	—	—
d	en	7	0	48,82	29,25	45,49
d	en	7	1	—	—	—
d	en	10	0	50,41	30,00	46,94
d	en	10	1	—	—	—
d	fr	0	0	37,85	17,00	34,30
d	fr	0	1	—	—	—
d	fr	1	0	44,71	18,50	40,25
d	fr	1	1	—	—	—
d	fr	3	0	48,11	24,25	44,05
d	fr	3	1	—	—	—
d	fr	5	0	49,64	25,50	45,53
d	fr	5	1	—	—	—
d	fr	7	0	50,77	27,25	46,77
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**52,41**	**28,75**	**48,38**
d	fr	10	1	—	—	—
e	en	0	0	38,10	15,75	34,29
e	en	0	1	—	—	—
e	en	1	0	43,18	23,00	39,75
e	en	1	1	—	—	—
e	en	3	0	45,54	25,00	42,05
e	en	3	1	—	—	—
e	en	5	0	46,41	25,75	42,90
e	en	5	1	—	—	—
e	en	7	0	47,43	26,50	43,87
e	en	7	1	—	—	—
e	en	10	0	47,65	28,00	44,30
e	en	10	1	—	—	—
e	fr	0	0	38,82	13,75	34,55
e	fr	0	1	—	—	—
e	fr	1	0	44,82	19,25	40,47
e	fr	1	1	—	—	—
e	fr	3	0	47,18	22,00	42,89
e	fr	3	1	—	—	—
e	fr	5	0	47,85	23,50	43,71
e	fr	5	1	—	—	—
e	fr	7	0	49,12	26,75	45,32
e	fr	7	1	—	—	—
e	fr	10	0	48,46	27,50	44,89
e	fr	10	1	—	—	—

TABLE 7 – EM des variantes ChatGPT (Hist:1), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	48,51	25,25	44,55
d	en	0	1	52,77	32,50	49,32
d	en	1	0	49,89	24,00	45,49
d	en	1	1	53,23	35,00	50,13
d	en	3	0	50,36	24,00	45,87
d	en	3	1	54,15	33,75	50,68
d	en	5	0	52,10	28,00	48,00
d	en	5	1	54,62	33,00	50,94
d	en	7	0	50,82	24,75	46,38
d	en	7	1	54,62	34,00	51,11
d	en	10	0	49,85	25,50	45,70
d	**en**	**10**	**1**	**54,36**	**36,75**	**51,36**
d	fr	0	0	48,92	23,00	44,51
d	fr	0	1	52,82	27,75	48,56
d	fr	1	0	49,85	24,25	45,49
d	fr	1	1	53,39	29,50	49,32
d	fr	3	0	50,56	24,00	46,04
d	fr	3	1	54,15	30,50	50,13
d	fr	5	0	50,15	21,75	45,32
d	fr	5	1	54,21	31,50	50,34
d	fr	7	0	50,87	23,75	46,25
d	fr	7	1	53,54	31,75	49,83
d	fr	10	0	50,26	26,00	46,13
d	fr	10	1	53,70	33,75	50,30
e	en	0	0	48,87	23,50	44,55
e	en	0	1	51,75	31,75	48,34
e	en	1	0	49,85	23,25	45,32
e	en	1	1	52,25	32,00	48,81
e	en	3	0	49,89	22,25	45,19
e	en	3	1	53,90	31,50	50,09
e	en	5	0	51,18	22,00	46,21
e	en	5	1	53,49	29,50	49,41
e	en	7	0	50,92	22,50	46,08
e	en	7	1	53,69	33,75	50,30
e	en	10	0	51,13	23,75	46,47
e	en	10	1	54,72	32,25	50,90
e	fr	0	0	49,79	21,75	45,02
e	fr	0	1	52,36	32,00	48,89
e	fr	1	0	49,65	21,00	44,77
e	fr	1	1	52,20	31,75	48,72
e	fr	3	0	50,30	21,50	45,40
e	fr	3	1	52,57	29,50	48,64
e	fr	5	0	49,90	22,50	45,23
e	fr	5	1	53,44	28,50	49,19
e	fr	7	0	51,95	24,75	47,32
e	fr	7	1	53,08	29,50	49,07
e	fr	10	0	50,88	24,25	46,34
e	fr	10	1	53,75	29,75	49,66

TABLE 8 – EM des variantes `Llama3.0` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	13,85	10,25	13,24
d	en	0	1	—	—	—
d	en	1	0	20,06	14,00	19,03
d	en	1	1	—	—	—
d	en	3	0	22,21	11,75	20,43
d	en	3	1	—	—	—
d	en	5	0	22,82	11,50	20,90
d	en	5	1	—	—	—
d	en	7	0	23,95	14,00	22,26
d	en	7	1	—	—	—
d	en	10	0	24,20	14,50	22,55
d	en	10	1	—	—	—
d	fr	0	0	12,06	7,00	11,20
d	fr	0	1	—	—	—
d	fr	1	0	27,89	15,25	25,74
d	fr	1	1	—	—	—
d	fr	3	0	31,39	16,75	28,90
d	fr	3	1	—	—	—
d	fr	5	0	32,21	17,25	29,66
d	fr	5	1	—	—	—
d	fr	7	0	35,54	21,25	33,11
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**36,21**	**25,50**	**34,38**
d	fr	10	1	—	—	—
e	en	0	0	12,16	8,50	11,54
e	en	0	1	—	—	—
e	en	1	0	22,72	13,25	21,11
e	en	1	1	—	—	—
e	en	3	0	21,85	9,75	19,79
e	en	3	1	—	—	—
e	en	5	0	22,57	10,50	20,51
e	en	5	1	—	—	—
e	en	7	0	25,03	11,25	22,68
e	en	7	1	—	—	—
e	en	10	0	25,29	13,25	23,24
e	en	10	1	—	—	—
e	fr	0	0	10,62	10,75	10,64
e	fr	0	1	—	—	—
e	fr	1	0	26,41	15,25	24,51
e	fr	1	1	—	—	—
e	fr	3	0	29,89	18,25	27,91
e	fr	3	1	—	—	—
e	fr	5	0	29,95	21,00	28,42
e	fr	5	1	—	—	—
e	fr	7	0	32,41	22,00	30,64
e	fr	7	1	—	—	—
e	fr	10	0	32,71	26,00	31,57
e	fr	10	1	—	—	—

TABLE 9 – EM des variantes `Llama3.0` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	43,07	17,25	38,68
d	en	0	1	46,56	22,50	42,47
d	en	1	0	45,28	33,25	43,23
d	en	1	1	47,53	34,75	45,36
d	en	3	0	45,69	31,25	43,24
d	en	3	1	47,07	36,25	45,23
d	en	5	0	47,07	23,75	43,10
d	en	5	1	47,54	32,25	44,94
d	en	7	0	48,15	32,50	45,49
d	en	7	1	47,90	32,75	45,32
d	en	10	0	48,46	25,00	44,46
d	en	10	1	48,56	40,25	47,15
d	fr	0	0	43,02	16,75	38,55
d	fr	0	1	40,41	20,25	36,98
d	fr	1	0	48,57	30,00	45,41
d	fr	1	1	48,46	31,50	45,58
d	fr	3	0	43,79	25,75	40,72
d	fr	3	1	46,87	32,25	44,38
d	fr	5	0	43,69	31,25	41,57
d	fr	5	1	51,28	33,50	48,25
d	fr	7	0	49,90	35,00	47,37
d	fr	7	1	52,83	33,25	49,49
d	**fr**	**10**	**0**	**52,77**	**37,50**	**50,17**
d	fr	10	1	47,08	35,25	45,07
e	en	0	0	40,51	24,25	37,74
e	en	0	1	44,00	14,50	38,98
e	en	1	0	40,05	23,00	37,15
e	en	1	1	43,90	28,25	41,23
e	en	3	0	43,18	12,00	37,87
e	en	3	1	44,93	16,25	40,04
e	en	5	0	46,00	13,50	40,47
e	en	5	1	45,18	28,00	42,25
e	en	7	0	46,20	12,75	40,51
e	en	7	1	43,08	29,25	40,72
e	en	10	0	47,59	26,75	44,04
e	en	10	1	46,87	15,50	41,53
e	fr	0	0	36,41	24,75	34,43
e	fr	0	1	37,03	25,00	34,98
e	fr	1	0	44,35	27,00	41,40
e	fr	1	1	38,05	23,75	35,62
e	fr	3	0	45,90	27,75	42,81
e	fr	3	1	37,69	25,50	35,62
e	fr	5	0	46,05	28,25	43,02
e	fr	5	1	39,85	27,50	37,75
e	fr	7	0	47,59	28,75	44,39
e	fr	7	1	40,97	25,50	38,34
e	fr	10	0	48,62	29,75	45,40
e	fr	10	1	40,67	27,25	38,39

TABLE 10 – EM des variantes `Llama3.1` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	6,46	3,75	6,00
d	en	0	1	—	—	—
d	en	1	0	15,43	6,00	13,83
d	en	1	1	—	—	—
d	en	3	0	20,15	8,00	18,08
d	en	3	1	—	—	—
d	en	5	0	22,21	7,75	19,75
d	en	5	1	—	—	—
d	en	7	0	22,51	8,25	20,09
d	en	7	1	—	—	—
d	en	10	0	27,23	8,75	24,09
d	en	10	1	—	—	—
d	fr	0	0	14,97	8,25	13,83
d	fr	0	1	—	—	—
d	fr	1	0	29,33	13,25	26,59
d	fr	1	1	—	—	—
d	fr	3	0	33,03	14,75	29,92
d	fr	3	1	—	—	—
d	fr	5	0	35,49	16,50	32,25
d	fr	5	1	—	—	—
d	fr	7	0	39,59	22,25	36,64
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**37,03**	**22,25**	**34,51**
d	fr	10	1	—	—	—
e	en	0	0	7,13	1,50	6,17
e	en	0	1	—	—	—
e	en	1	0	15,69	5,25	13,91
e	en	1	1	—	—	—
e	en	3	0	16,87	7,50	15,27
e	en	3	1	—	—	—
e	en	5	0	18,77	8,00	16,93
e	en	5	1	—	—	—
e	en	7	0	17,18	9,50	15,88
e	en	7	1	—	—	—
e	en	10	0	16,88	8,25	15,41
e	en	10	1	—	—	—
e	fr	0	0	14,26	5,75	12,81
e	fr	0	1	—	—	—
e	fr	1	0	27,13	14,25	24,94
e	fr	1	1	—	—	—
e	fr	3	0	31,59	14,50	28,68
e	fr	3	1	—	—	—
e	fr	5	0	33,03	20,25	30,85
e	fr	5	1	—	—	—
e	fr	7	0	34,46	21,25	32,21
e	fr	7	1	—	—	—
e	fr	10	0	35,59	22,50	33,36
e	fr	10	1	—	—	—

TABLE 11 – EM des variantes `Llama3.1` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	42,97	29,25	40,64
d	en	0	1	48,25	33,25	45,70
d	en	1	0	48,26	29,75	45,11
d	en	1	1	48,98	33,50	46,34
d	en	3	0	49,39	28,25	45,79
d	en	3	1	49,85	37,50	47,75
d	en	5	0	49,69	32,75	46,81
d	en	5	1	50,61	35,50	48,04
d	en	7	0	49,89	34,50	47,27
d	en	7	1	52,41	39,75	50,25
d	en	10	0	51,13	34,25	48,26
d	**en**	**10**	**1**	**53,59**	**37,75**	**50,89**
d	fr	0	0	46,87	30,00	44,00
d	fr	0	1	48,51	36,75	46,51
d	fr	1	0	47,80	29,00	44,60
d	fr	1	1	49,69	35,75	47,32
d	fr	3	0	51,49	30,50	47,92
d	fr	3	1	49,08	32,50	46,25
d	fr	5	0	47,18	34,00	44,94
d	fr	5	1	52,98	37,25	50,30
d	fr	7	0	52,16	32,75	48,85
d	fr	7	1	50,57	38,00	48,43
d	fr	10	0	50,26	20,50	45,19
d	fr	10	1	51,65	39,00	49,49
e	en	0	0	44,00	20,25	39,96
e	en	0	1	46,00	30,25	43,32
e	en	1	0	43,95	19,00	39,70
e	en	1	1	47,12	27,25	43,74
e	en	3	0	44,66	23,00	40,98
e	en	3	1	46,62	29,75	43,75
e	en	5	0	45,69	24,75	42,13
e	en	5	1	48,51	33,00	45,87
e	en	7	0	47,59	26,50	44,00
e	en	7	1	49,49	31,25	46,39
e	en	10	0	47,94	27,50	44,46
e	en	10	1	49,95	29,75	46,51
e	fr	0	0	38,77	26,25	36,64
e	fr	0	1	46,51	32,00	44,04
e	fr	1	0	45,54	25,50	42,13
e	fr	1	1	49,23	31,25	46,17
e	fr	3	0	48,87	25,25	44,85
e	fr	3	1	49,75	31,25	46,60
e	fr	5	0	48,57	27,50	44,98
e	fr	5	1	51,70	32,75	48,47
e	fr	7	0	50,98	27,00	46,90
e	fr	7	1	53,03	33,00	49,62
e	fr	10	0	50,88	31,25	47,54
e	fr	10	1	52,26	36,00	49,49

TABLE 12 – EM des variantes `Qwen` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	32,31	21,25	30,43
d	en	0	1	—	—	—
d	en	1	0	40,35	25,25	37,78
d	en	1	1	—	—	—
d	en	3	0	45,54	30,50	42,98
d	en	3	1	—	—	—
d	en	5	0	46,87	31,25	44,21
d	en	5	1	—	—	—
d	en	7	0	47,69	34,25	45,40
d	en	7	1	—	—	—
d	**en**	**10**	**0**	**48,56**	**34,00**	**46,08**
d	en	10	1	—	—	—
d	fr	0	0	31,33	17,00	28,89
d	fr	0	1	—	—	—
d	fr	1	0	40,71	26,00	38,21
d	fr	1	1	—	—	—
d	fr	3	0	44,87	30,25	42,38
d	fr	3	1	—	—	—
d	fr	5	0	46,05	26,75	42,77
d	fr	5	1	—	—	—
d	fr	7	0	47,23	29,75	44,25
d	fr	7	1	—	—	—
d	fr	10	0	47,75	30,50	44,81
d	fr	10	1	—	—	—
e	en	0	0	32,87	16,25	30,04
e	en	0	1	—	—	—
e	en	1	0	38,77	20,25	35,61
e	en	1	1	—	—	—
e	en	3	0	43,18	27,50	40,51
e	en	3	1	—	—	—
e	en	5	0	43,33	29,00	40,89
e	en	5	1	—	—	—
e	en	7	0	45,38	32,25	43,15
e	en	7	1	—	—	—
e	en	10	0	47,03	31,25	44,34
e	en	10	1	—	—	—
e	fr	0	0	32,66	17,00	30,00
e	fr	0	1	—	—	—
e	fr	1	0	39,65	20,00	36,30
e	fr	1	1	—	—	—
e	fr	3	0	44,36	26,25	41,28
e	fr	3	1	—	—	—
e	fr	5	0	44,41	26,50	41,36
e	fr	5	1	—	—	—
e	fr	7	0	46,41	27,75	43,23
e	fr	7	1	—	—	—
e	fr	10	0	47,34	28,75	44,17
e	fr	10	1	—	—	—

TABLE 13 – EM des variantes `Qwen` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	49,80	32,50	46,85
d	en	0	1	50,00	39,50	48,21
d	en	1	0	47,48	31,25	44,72
d	en	1	1	53,75	37,25	50,94
d	en	3	0	53,18	37,50	50,51
d	en	3	1	54,72	41,00	52,38
d	en	5	0	52,72	35,25	49,74
d	en	5	1	55,23	40,25	52,68
d	en	7	0	51,54	38,00	49,24
d	en	7	1	56,46	41,75	53,96
d	en	10	0	52,93	36,75	50,17
d	en	10	1	54,72	42,50	52,64
d	fr	0	0	46,82	33,75	44,60
d	fr	0	1	54,51	36,25	51,40
d	fr	1	0	49,90	33,50	47,11
d	fr	1	1	55,03	37,50	52,04
d	fr	3	0	53,13	34,25	49,91
d	fr	3	1	55,23	38,50	52,38
d	fr	5	0	54,56	33,00	50,89
d	fr	5	1	55,95	38,25	52,94
d	fr	7	0	54,92	34,50	51,45
d	fr	7	1	56,41	38,00	53,27
d	fr	10	0	54,87	34,75	51,44
d	**fr**	**10**	**1**	**56,83**	**40,25**	**54,00**
e	en	0	0	47,44	30,75	44,60
e	en	0	1	49,95	33,75	47,19
e	en	1	0	46,57	30,50	43,83
e	en	1	1	51,23	35,00	48,47
e	en	3	0	47,44	30,50	44,55
e	en	3	1	51,64	36,25	49,02
e	en	5	0	47,38	32,00	44,76
e	en	5	1	51,23	36,50	48,72
e	en	7	0	47,89	32,50	45,27
e	en	7	1	53,49	35,25	50,38
e	en	10	0	49,59	32,00	46,60
e	en	10	1	53,12	36,50	50,29
e	fr	0	0	47,95	30,00	44,89
e	fr	0	1	52,15	34,50	49,15
e	fr	1	0	50,05	30,00	46,64
e	fr	1	1	52,00	34,25	48,98
e	fr	3	0	53,59	28,75	49,37
e	fr	3	1	53,69	36,25	50,72
e	fr	5	0	49,65	31,00	46,47
e	fr	5	1	51,49	36,50	48,94
e	fr	7	0	52,00	29,75	48,21
e	fr	7	1	54,77	37,25	51,79
e	fr	10	0	51,02	31,00	47,62
e	fr	10	1	54,05	36,25	51,02

TABLE 14 – CM des variantes ChatGPT (Hist:0), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	40,51	19,50	36,93
d	en	0	1	—	—	—
d	en	1	0	46,77	23,50	42,81
d	en	1	1	—	—	—
d	en	3	0	49,90	25,00	45,66
d	en	3	1	—	—	—
d	en	5	0	51,59	29,00	47,74
d	en	5	1	—	—	—
d	en	7	0	51,74	31,50	48,30
d	en	7	1	—	—	—
d	en	10	0	53,70	32,00	50,00
d	en	10	1	—	—	—
d	fr	0	0	41,23	19,25	37,49
d	fr	0	1	—	—	—
d	fr	1	0	47,70	20,50	43,07
d	fr	1	1	—	—	—
d	fr	3	0	51,54	27,75	47,49
d	fr	3	1	—	—	—
d	fr	5	0	52,67	29,25	48,68
d	fr	5	1	—	—	—
d	fr	7	0	53,39	31,25	49,62
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**55,54**	**33,25**	**51,75**
d	fr	10	1	—	—	—
e	en	0	0	40,82	17,00	36,76
e	en	0	1	—	—	—
e	en	1	0	46,36	23,75	42,51
e	en	1	1	—	—	—
e	en	3	0	49,44	27,50	45,71
e	en	3	1	—	—	—
e	en	5	0	50,00	29,00	46,43
e	en	5	1	—	—	—
e	en	7	0	50,62	28,75	46,90
e	en	7	1	—	—	—
e	en	10	0	51,18	31,00	47,75
e	en	10	1	—	—	—
e	fr	0	0	42,21	15,25	37,62
e	fr	0	1	—	—	—
e	fr	1	0	48,31	21,25	43,70
e	fr	1	1	—	—	—
e	fr	3	0	50,98	24,75	46,51
e	fr	3	1	—	—	—
e	fr	5	0	51,54	26,50	47,28
e	fr	5	1	—	—	—
e	fr	7	0	52,31	29,00	48,34
e	fr	7	1	—	—	—
e	fr	10	0	52,31	31,00	48,68
e	fr	10	1	—	—	—

TABLE 15 – CM des variantes ChatGPT (Hist:1), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	52,87	28,00	48,64
d	en	0	1	56,25	36,25	52,85
d	en	1	0	53,49	26,75	48,94
d	en	1	1	56,98	39,50	54,00
d	en	3	0	53,70	27,00	49,15
d	en	3	1	57,59	37,75	54,21
d	en	5	0	55,08	30,50	50,90
d	en	5	1	58,00	36,50	54,34
d	en	7	0	54,31	27,50	49,74
d	**en**	**7**	**1**	**58,31**	**39,75**	**55,15**
d	en	10	0	53,85	28,50	49,53
d	en	10	1	57,69	41,00	54,85
d	fr	0	0	52,77	26,25	48,26
d	fr	0	1	56,26	32,00	52,13
d	fr	1	0	53,54	27,00	49,02
d	fr	1	1	57,84	34,00	53,78
d	fr	3	0	54,26	26,00	49,45
d	fr	3	1	57,85	34,75	53,92
d	fr	5	0	53,79	24,75	48,85
d	fr	5	1	57,75	35,00	53,87
d	fr	7	0	54,11	26,25	49,36
d	fr	7	1	57,70	35,50	53,92
d	fr	10	0	53,80	29,25	49,62
d	fr	10	1	57,49	37,75	54,13
e	en	0	0	52,92	26,00	48,34
e	en	0	1	56,00	35,50	52,51
e	en	1	0	53,79	25,50	48,98
e	en	1	1	56,67	35,50	53,07
e	en	3	0	54,05	24,50	49,02
e	en	3	1	57,59	34,75	53,71
e	en	5	0	54,87	24,50	49,70
e	en	5	1	57,85	32,75	53,57
e	en	7	0	54,72	24,25	49,53
e	en	7	1	56,98	37,75	53,70
e	en	10	0	55,44	25,50	50,34
e	en	10	1	58,36	35,00	54,38
e	fr	0	0	53,28	24,25	48,34
e	fr	0	1	56,00	35,75	52,56
e	fr	1	0	53,94	23,75	48,81
e	fr	1	1	56,31	35,25	52,73
e	fr	3	0	54,16	23,50	48,94
e	fr	3	1	56,67	32,25	52,51
e	fr	5	0	53,95	24,25	48,90
e	fr	5	1	57,08	32,50	52,90
e	fr	7	0	55,79	27,75	51,02
e	fr	7	1	56,46	34,00	52,64
e	fr	10	0	54,51	26,50	49,74
e	fr	10	1	57,28	33,25	53,19

TABLE 16 – CM des variantes `Llama3.0` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	15,85	11,00	15,02
d	en	0	1	—	—	—
d	en	1	0	22,36	14,75	21,06
d	en	1	1	—	—	—
d	en	3	0	24,46	12,25	22,38
d	en	3	1	—	—	—
d	en	5	0	24,57	12,50	22,51
d	en	5	1	—	—	—
d	en	7	0	25,79	15,00	23,96
d	en	7	1	—	—	—
d	en	10	0	26,21	14,75	24,26
d	en	10	1	—	—	—
d	fr	0	0	14,21	11,25	13,71
d	fr	0	1	—	—	—
d	fr	1	0	31,18	16,25	28,64
d	fr	1	1	—	—	—
d	fr	3	0	33,80	18,50	31,19
d	fr	3	1	—	—	—
d	fr	5	0	35,03	19,50	32,38
d	fr	5	1	—	—	—
d	fr	7	0	38,41	22,50	35,71
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**38,66**	**27,25**	**36,72**
d	fr	10	1	—	—	—
e	en	0	0	13,43	10,75	12,98
e	en	0	1	—	—	—
e	en	1	0	25,18	14,50	23,36
e	en	1	1	—	—	—
e	en	3	0	23,90	10,00	21,54
e	en	3	1	—	—	—
e	en	5	0	24,71	11,00	22,38
e	en	5	1	—	—	—
e	en	7	0	27,59	14,00	25,28
e	en	7	1	—	—	—
e	en	10	0	27,69	14,50	25,45
e	en	10	1	—	—	—
e	fr	0	0	12,62	13,00	12,68
e	fr	0	1	—	—	—
e	fr	1	0	29,59	16,00	27,27
e	fr	1	1	—	—	—
e	fr	3	0	32,61	18,50	30,21
e	fr	3	1	—	—	—
e	fr	5	0	32,67	23,00	31,02
e	fr	5	1	—	—	—
e	fr	7	0	34,97	24,50	33,19
e	fr	7	1	—	—	—
e	fr	10	0	34,97	27,50	33,70
e	fr	10	1	—	—	—

TABLE 17 – CM des variantes `Llama3.0` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	46,05	34,00	44,00
d	en	0	1	50,30	38,00	48,21
d	en	1	0	48,82	39,00	47,15
d	en	1	1	51,74	37,50	49,32
d	en	3	0	49,18	34,00	46,59
d	en	3	1	50,31	38,75	48,34
d	en	5	0	50,10	35,75	47,66
d	en	5	1	51,23	36,75	48,77
d	en	7	0	51,69	35,75	48,98
d	en	7	1	53,18	37,75	50,55
d	en	10	0	52,21	39,75	50,09
d	en	10	1	52,97	43,00	51,28
d	fr	0	0	47,28	30,75	44,47
d	fr	0	1	47,28	34,00	45,02
d	fr	1	0	51,23	34,50	48,39
d	fr	1	1	51,69	35,50	48,93
d	fr	3	0	46,87	26,75	43,45
d	fr	3	1	49,23	35,00	46,81
d	fr	5	0	46,66	34,50	44,59
d	fr	5	1	54,31	37,25	51,40
d	fr	7	0	53,64	36,25	50,68
d	fr	7	1	55,85	36,00	52,47
d	**fr**	**10**	**0**	**56,26**	**41,50**	**53,74**
d	fr	10	1	50,16	38,75	48,22
e	en	0	0	43,02	28,50	40,55
e	en	0	1	47,18	28,50	44,00
e	en	1	0	42,93	27,00	40,22
e	en	1	1	46,67	32,00	44,17
e	en	3	0	46,16	25,25	42,60
e	en	3	1	47,54	32,00	44,89
e	en	5	0	48,56	26,75	44,85
e	en	5	1	47,95	32,00	45,23
e	en	7	0	48,82	27,50	45,19
e	en	7	1	47,48	32,75	44,97
e	en	10	0	50,46	31,75	47,28
e	en	10	1	50,00	32,00	46,94
e	fr	0	0	39,08	27,00	37,02
e	fr	0	1	39,54	29,50	37,83
e	fr	1	0	47,13	30,25	44,26
e	fr	1	1	40,16	26,25	37,79
e	fr	3	0	49,03	31,75	46,09
e	fr	3	1	40,41	27,75	38,26
e	fr	5	0	48,93	32,00	46,05
e	fr	5	1	42,36	32,00	40,60
e	fr	7	0	50,35	32,75	47,36
e	fr	7	1	43,18	28,75	40,72
e	fr	10	0	51,65	30,75	48,09
e	fr	10	1	42,98	31,50	41,03

TABLE 18 – CM des variantes `Llama3.1` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	7,48	4,00	6,89
d	en	0	1	—	—	—
d	en	1	0	17,08	6,50	15,28
d	en	1	1	—	—	—
d	en	3	0	22,00	10,00	19,96
d	en	3	1	—	—	—
d	en	5	0	23,59	9,00	21,11
d	en	5	1	—	—	—
d	en	7	0	23,69	9,25	21,23
d	en	7	1	—	—	—
d	en	10	0	29,13	8,75	25,66
d	en	10	1	—	—	—
d	fr	0	0	17,23	9,50	15,92
d	fr	0	1	—	—	—
d	fr	1	0	32,41	14,00	29,27
d	fr	1	1	—	—	—
d	fr	3	0	35,54	16,75	32,34
d	fr	3	1	—	—	—
d	fr	5	0	38,25	18,00	34,81
d	fr	5	1	—	—	—
d	**fr**	**7**	**0**	**42,20**	**22,50**	**38,85**
d	fr	7	1	—	—	—
d	fr	10	0	39,43	22,75	36,59
d	fr	10	1	—	—	—
e	en	0	0	8,00	1,50	6,89
e	en	0	1	—	—	—
e	en	1	0	16,77	5,25	14,81
e	en	1	1	—	—	—
e	en	3	0	18,05	7,50	16,25
e	en	3	1	—	—	—
e	en	5	0	19,64	8,25	17,70
e	en	5	1	—	—	—
e	en	7	0	17,95	10,00	16,60
e	en	7	1	—	—	—
e	en	10	0	18,15	8,75	16,55
e	en	10	1	—	—	—
e	fr	0	0	15,74	6,25	14,12
e	fr	0	1	—	—	—
e	fr	1	0	29,95	15,00	27,40
e	fr	1	1	—	—	—
e	fr	3	0	34,10	16,75	31,15
e	fr	3	1	—	—	—
e	fr	5	0	35,54	22,25	33,28
e	fr	5	1	—	—	—
e	fr	7	0	37,07	22,75	34,64
e	fr	7	1	—	—	—
e	fr	10	0	37,95	23,75	35,53
e	fr	10	1	—	—	—

TABLE 19 – CM des variantes `Llama3.1` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	48,05	35,00	45,83
d	en	0	1	51,18	39,25	49,15
d	en	1	0	50,98	35,25	48,30
d	en	1	1	52,77	39,50	50,51
d	en	3	0	52,46	34,75	49,45
d	en	3	1	53,64	43,75	51,95
d	en	5	0	52,92	37,75	50,34
d	en	5	1	53,95	41,00	51,74
d	en	7	0	53,59	38,25	50,98
d	en	7	1	54,98	44,75	53,24
d	en	10	0	53,95	42,75	52,04
d	en	10	1	56,46	43,75	54,30
d	fr	0	0	52,82	34,75	49,75
d	fr	0	1	52,25	44,00	50,85
d	fr	1	0	51,79	32,75	48,55
d	fr	1	1	56,15	41,75	53,70
d	fr	3	0	55,28	38,50	52,43
d	fr	3	1	56,11	37,00	52,85
d	fr	5	0	55,54	38,75	52,68
d	fr	5	1	56,82	43,75	54,59
d	fr	7	0	56,20	36,75	52,89
d	fr	7	1	54,87	42,75	52,81
d	fr	10	0	56,97	41,75	54,38
d	**fr**	**10**	**1**	**57,07**	**47,75**	**55,49**
e	en	0	0	46,57	23,75	42,68
e	en	0	1	48,92	35,00	46,55
e	en	1	0	46,77	22,25	42,59
e	en	1	1	49,95	32,50	46,98
e	en	3	0	47,75	27,75	44,34
e	en	3	1	49,39	35,00	46,94
e	en	5	0	48,46	29,25	45,19
e	en	5	1	51,23	38,25	49,02
e	en	7	0	50,16	29,50	46,64
e	en	7	1	52,26	36,25	49,53
e	en	10	0	50,77	30,50	47,32
e	en	10	1	52,36	34,25	49,27
e	fr	0	0	44,26	32,25	42,21
e	fr	0	1	51,08	38,00	48,85
e	fr	1	0	48,56	30,75	45,53
e	fr	1	1	52,52	36,75	49,83
e	fr	3	0	51,59	31,25	48,13
e	fr	3	1	52,10	37,75	49,66
e	fr	5	0	51,13	33,50	48,13
e	fr	5	1	54,05	39,50	51,58
e	fr	7	0	53,64	33,50	50,21
e	fr	7	1	55,75	39,00	52,90
e	fr	10	0	53,79	35,75	50,72
e	fr	10	1	55,23	42,50	53,06

TABLE 20 – CM des variantes Qwen (Hist:0), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	36,20	23,25	34,00
d	en	0	1	—	—	—
d	en	1	0	44,92	28,00	42,04
d	en	1	1	—	—	—
d	en	3	0	50,00	33,00	47,10
d	en	3	1	—	—	—
d	en	5	0	51,53	33,50	48,46
d	en	5	1	—	—	—
d	en	7	0	51,90	37,00	49,37
d	en	7	1	—	—	—
d	**en**	**10**	**0**	**52,98**	**36,50**	**50,17**
d	en	10	1	—	—	—
d	fr	0	0	35,48	18,25	32,55
d	fr	0	1	—	—	—
d	fr	1	0	45,65	28,25	42,68
d	fr	1	1	—	—	—
d	fr	3	0	48,35	32,00	45,57
d	fr	3	1	—	—	—
d	fr	5	0	49,89	28,50	46,25
d	fr	5	1	—	—	—
d	fr	7	0	51,59	31,75	48,21
d	fr	7	1	—	—	—
d	fr	10	0	52,46	33,25	49,19
d	fr	10	1	—	—	—
e	en	0	0	36,52	17,50	33,28
e	en	0	1	—	—	—
e	en	1	0	42,93	22,25	39,41
e	en	1	1	—	—	—
e	en	3	0	47,64	30,00	44,64
e	en	3	1	—	—	—
e	en	5	0	47,59	32,50	45,02
e	en	5	1	—	—	—
e	en	7	0	49,75	34,50	47,15
e	en	7	1	—	—	—
e	en	10	0	51,13	33,50	48,13
e	en	10	1	—	—	—
e	fr	0	0	36,56	18,25	33,45
e	fr	0	1	—	—	—
e	fr	1	0	44,20	21,75	40,38
e	fr	1	1	—	—	—
e	fr	3	0	48,36	28,50	44,98
e	fr	3	1	—	—	—
e	fr	5	0	48,41	29,50	45,19
e	fr	5	1	—	—	—
e	fr	7	0	50,77	30,50	47,32
e	fr	7	1	—	—	—
e	fr	10	0	51,74	31,75	48,34
e	fr	10	1	—	—	—

TABLE 21 – CM des variantes Qwen (Hist:1), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	55,34	36,25	52,09
d	en	0	1	58,00	45,75	55,92
d	en	1	0	57,64	35,25	53,83
d	en	1	1	60,00	42,75	57,06
d	en	3	0	58,51	42,25	55,75
d	en	3	1	61,28	45,50	58,59
d	en	5	0	58,36	40,50	55,32
d	en	5	1	60,26	46,50	57,91
d	en	7	0	58,51	43,75	56,00
d	en	7	1	61,39	47,00	58,94
d	en	10	0	58,67	41,75	55,79
d	en	10	1	61,34	48,50	59,15
d	fr	0	0	57,39	37,25	53,96
d	fr	0	1	60,41	40,75	57,06
d	fr	1	0	56,87	36,50	53,40
d	fr	1	1	61,49	42,00	58,17
d	fr	3	0	58,16	39,25	54,94
d	fr	3	1	61,28	43,25	58,21
d	fr	5	0	59,03	39,00	55,62
d	fr	5	1	61,79	44,25	58,81
d	fr	7	0	59,64	40,50	56,38
d	fr	7	1	62,46	43,25	59,19
d	fr	10	0	59,80	41,00	56,60
d	**fr**	**10**	**1**	**62,30**	**46,75**	**59,66**
e	en	0	0	53,13	34,25	49,91
e	en	0	1	55,75	39,25	52,94
e	en	1	0	53,18	34,00	49,91
e	en	1	1	56,57	41,25	53,96
e	en	3	0	54,72	36,75	51,66
e	en	3	1	57,07	42,50	54,59
e	en	5	0	54,46	37,75	51,61
e	en	5	1	58,16	42,00	55,41
e	en	7	0	55,94	38,75	53,02
e	en	7	1	58,98	42,00	56,09
e	en	10	0	56,67	39,00	53,66
e	en	10	1	58,82	42,75	56,08
e	fr	0	0	53,48	34,00	50,17
e	fr	0	1	57,54	39,50	54,47
e	fr	1	0	54,56	35,00	51,23
e	fr	1	1	56,15	40,00	53,40
e	fr	3	0	57,54	33,50	53,44
e	fr	3	1	58,66	43,00	56,00
e	fr	5	0	56,77	36,50	53,32
e	fr	5	1	60,00	41,00	56,77
e	fr	7	0	57,02	34,75	53,23
e	fr	7	1	60,05	41,25	56,85
e	fr	10	0	57,33	35,50	53,62
e	fr	10	1	60,20	41,50	57,02

TABLE 22 – SM des variantes ChatGPT (Hist:0), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	37,65	19,00	34,47
d	en	0	1	—	—	—
d	en	1	0	44,26	21,00	40,30
d	en	1	1	—	—	—
d	en	3	0	47,18	23,50	43,15
d	en	3	1	—	—	—
d	en	5	0	48,87	26,00	44,98
d	en	5	1	—	—	—
d	en	7	0	49,08	29,50	45,75
d	en	7	1	—	—	—
d	en	10	0	50,72	30,50	47,27
d	en	10	1	—	—	—
d	fr	0	0	38,16	18,00	34,73
d	fr	0	1	—	—	—
d	fr	1	0	45,03	19,00	40,60
d	fr	1	1	—	—	—
d	fr	3	0	48,46	24,75	44,42
d	fr	3	1	—	—	—
d	fr	5	0	50,00	26,25	45,96
d	fr	5	1	—	—	—
d	fr	7	0	51,38	27,50	47,32
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**52,77**	**30,25**	**48,94**
d	fr	10	1	—	—	—
e	en	0	0	38,30	15,75	34,46
e	en	0	1	—	—	—
e	en	1	0	43,69	23,00	40,17
e	en	1	1	—	—	—
e	en	3	0	45,80	25,50	42,34
e	en	3	1	—	—	—
e	en	5	0	46,77	26,50	43,32
e	en	5	1	—	—	—
e	en	7	0	47,69	27,00	44,17
e	en	7	1	—	—	—
e	en	10	0	47,85	28,25	44,51
e	en	10	1	—	—	—
e	fr	0	0	39,08	14,50	34,90
e	fr	0	1	—	—	—
e	fr	1	0	45,12	19,75	40,81
e	fr	1	1	—	—	—
e	fr	3	0	47,49	23,00	43,32
e	fr	3	1	—	—	—
e	fr	5	0	48,41	24,00	44,25
e	fr	5	1	—	—	—
e	fr	7	0	49,59	27,25	45,79
e	fr	7	1	—	—	—
e	fr	10	0	49,13	28,75	45,66
e	fr	10	1	—	—	—

TABLE 23 – SM des variantes ChatGPT (Hist:1), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	48,72	26,00	44,85
d	en	0	1	53,03	33,25	49,66
d	en	1	0	50,31	25,25	46,05
d	en	1	1	53,74	35,50	50,64
d	en	3	0	50,56	25,25	46,25
d	en	3	1	54,36	34,50	50,98
d	en	5	0	52,30	28,75	48,29
d	en	5	1	54,92	33,25	51,23
d	en	7	0	51,13	25,50	46,77
d	en	7	1	54,98	34,50	51,49
d	en	10	0	50,10	26,25	46,04
d	**en**	**10**	**1**	**54,67**	**36,75**	**51,62**
d	fr	0	0	49,23	23,75	44,89
d	fr	0	1	53,08	29,00	48,98
d	fr	1	0	50,20	24,75	45,87
d	fr	1	1	53,75	30,50	49,79
d	fr	3	0	50,92	25,00	46,51
d	fr	3	1	54,51	31,00	50,51
d	fr	5	0	50,46	23,00	45,79
d	fr	5	1	54,41	32,25	50,64
d	fr	7	0	51,13	24,50	46,60
d	fr	7	1	53,85	31,75	50,08
d	fr	10	0	50,46	27,50	46,55
d	fr	10	1	54,00	34,25	50,64
e	en	0	0	49,23	24,00	44,94
e	en	0	1	52,05	32,25	48,68
e	en	1	0	50,36	23,75	45,83
e	en	1	1	52,87	32,75	49,44
e	en	3	0	50,30	23,50	45,74
e	en	3	1	54,30	32,75	50,63
e	en	5	0	51,54	22,25	46,56
e	en	5	1	54,05	30,00	49,96
e	en	7	0	51,44	22,75	46,55
e	en	7	1	54,05	34,00	50,64
e	en	10	0	51,54	24,75	46,98
e	en	10	1	55,23	32,75	51,40
e	fr	0	0	50,15	22,75	45,49
e	fr	0	1	52,72	32,50	49,28
e	fr	1	0	50,15	21,75	45,32
e	fr	1	1	52,72	32,50	49,28
e	fr	3	0	50,67	22,00	45,79
e	fr	3	1	53,08	29,75	49,11
e	fr	5	0	50,46	23,75	45,92
e	fr	5	1	53,90	29,00	49,66
e	fr	7	0	52,25	25,50	47,70
e	fr	7	1	53,75	29,75	49,66
e	fr	10	0	51,23	24,75	46,72
e	fr	10	1	54,21	30,25	50,13

TABLE 24 – SM des variantes `Llama3.0` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	14,26	10,50	13,62
d	en	0	1	—	—	—
d	en	1	0	20,36	14,50	19,36
d	en	1	1	—	—	—
d	en	3	0	22,51	12,00	20,72
d	en	3	1	—	—	—
d	en	5	0	23,08	12,00	21,19
d	en	5	1	—	—	—
d	en	7	0	24,36	14,50	22,68
d	en	7	1	—	—	—
d	en	10	0	24,62	14,75	22,94
d	en	10	1	—	—	—
d	fr	0	0	12,57	8,50	11,87
d	fr	0	1	—	—	—
d	fr	1	0	28,67	15,75	26,47
d	fr	1	1	—	—	—
d	fr	3	0	31,90	17,75	29,49
d	fr	3	1	—	—	—
d	fr	5	0	32,93	18,00	30,39
d	fr	5	1	—	—	—
d	fr	7	0	36,25	21,75	33,79
d	fr	7	1	—	—	—
d	**fr**	**10**	**0**	**36,87**	**26,00**	**35,02**
d	fr	10	1	—	—	—
e	en	0	0	12,66	9,00	12,04
e	en	0	1	—	—	—
e	en	1	0	23,23	13,75	21,62
e	en	1	1	—	—	—
e	en	3	0	22,21	10,50	20,21
e	en	3	1	—	—	—
e	en	5	0	22,92	10,75	20,85
e	en	5	1	—	—	—
e	en	7	0	25,39	11,50	23,02
e	en	7	1	—	—	—
e	en	10	0	25,59	13,25	23,49
e	en	10	1	—	—	—
e	fr	0	0	11,03	12,00	11,19
e	fr	0	1	—	—	—
e	fr	1	0	26,92	16,00	25,06
e	fr	1	1	—	—	—
e	fr	3	0	30,41	19,25	28,51
e	fr	3	1	—	—	—
e	fr	5	0	30,51	21,50	28,98
e	fr	5	1	—	—	—
e	fr	7	0	32,92	22,75	31,19
e	fr	7	1	—	—	—
e	fr	10	0	33,34	26,75	32,21
e	fr	10	1	—	—	—

TABLE 25 – SM des variantes `Llama3.0` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	43,28	18,75	39,11
d	en	0	1	47,13	23,25	43,07
d	en	1	0	45,95	34,00	43,91
d	en	1	1	48,21	34,75	45,92
d	en	3	0	46,52	32,00	44,04
d	en	3	1	48,00	36,50	46,04
d	en	5	0	47,64	25,25	43,83
d	en	5	1	48,15	32,50	45,49
d	en	7	0	49,02	32,75	46,25
d	en	7	1	48,62	34,00	46,13
d	en	10	0	49,64	25,25	45,49
d	en	10	1	49,03	40,75	47,62
d	fr	0	0	43,79	19,00	39,57
d	fr	0	1	40,98	25,25	38,30
d	fr	1	0	48,97	30,75	45,87
d	fr	1	1	48,72	31,75	45,83
d	fr	3	0	44,62	28,25	41,83
d	fr	3	1	47,28	32,25	44,72
d	fr	5	0	44,72	32,75	42,68
d	fr	5	1	51,90	34,00	48,85
d	fr	7	0	51,02	35,75	48,42
d	fr	7	1	53,13	33,50	49,79
d	**fr**	**10**	**0**	**53,70**	**39,00**	**51,19**
d	fr	10	1	47,89	35,50	45,78
e	en	0	0	40,82	24,75	38,08
e	en	0	1	44,51	15,00	39,49
e	en	1	0	40,46	23,75	37,61
e	en	1	1	44,46	28,25	41,70
e	en	3	0	43,59	13,00	38,38
e	en	3	1	45,13	16,75	40,30
e	en	5	0	46,31	14,75	40,94
e	en	5	1	45,54	28,75	42,68
e	en	7	0	46,57	13,00	40,85
e	en	7	1	43,80	30,00	41,45
e	en	10	0	47,84	27,50	44,38
e	en	10	1	47,38	16,25	42,08
e	fr	0	0	37,03	25,00	34,98
e	fr	0	1	38,05	26,00	36,00
e	fr	1	0	44,67	27,50	41,74
e	fr	1	1	39,13	26,25	36,94
e	fr	3	0	46,46	28,25	43,36
e	fr	3	1	39,03	27,75	37,11
e	fr	5	0	46,41	30,00	43,61
e	fr	5	1	41,23	30,25	39,36
e	fr	7	0	48,36	30,25	45,28
e	fr	7	1	42,36	28,00	39,92
e	fr	10	0	49,23	30,25	46,00
e	fr	10	1	42,00	29,50	39,88

TABLE 26 – SM des variantes `Llama3.1` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	6,66	4,75	6,34
d	en	0	1	—	—	—
d	en	1	0	15,75	6,25	14,13
d	en	1	1	—	—	—
d	en	3	0	20,36	8,50	18,34
d	en	3	1	—	—	—
d	en	5	0	22,46	8,50	20,09
d	en	5	1	—	—	—
d	en	7	0	23,08	8,50	20,60
d	en	7	1	—	—	—
d	en	10	0	27,49	10,25	24,55
d	en	10	1	—	—	—
d	fr	0	0	15,54	9,50	14,51
d	fr	0	1	—	—	—
d	fr	1	0	29,89	13,75	27,15
d	fr	1	1	—	—	—
d	fr	3	0	33,38	16,00	30,42
d	fr	3	1	—	—	—
d	fr	5	0	36,05	17,25	32,85
d	fr	5	1	—	—	—
d	**fr**	**7**	**0**	**40,05**	**22,50**	**37,06**
d	fr	7	1	—	—	—
d	fr	10	0	37,80	23,00	35,28
d	fr	10	1	—	—	—
e	en	0	0	7,29	1,50	6,30
e	en	0	1	—	—	—
e	en	1	0	15,90	5,25	14,09
e	en	1	1	—	—	—
e	en	3	0	17,03	7,50	15,41
e	en	3	1	—	—	—
e	en	5	0	18,87	8,00	17,02
e	en	5	1	—	—	—
e	en	7	0	17,49	9,75	16,17
e	en	7	1	—	—	—
e	en	10	0	17,08	8,50	15,62
e	en	10	1	—	—	—
e	fr	0	0	14,97	6,25	13,49
e	fr	0	1	—	—	—
e	fr	1	0	28,00	15,25	25,83
e	fr	1	1	—	—	—
e	fr	3	0	32,26	14,75	29,28
e	fr	3	1	—	—	—
e	fr	5	0	33,90	20,50	31,62
e	fr	5	1	—	—	—
e	fr	7	0	35,39	22,00	33,11
e	fr	7	1	—	—	—
e	fr	10	0	36,10	23,25	33,91
e	fr	10	1	—	—	—

TABLE 27 – SM des variantes `Llama3.1` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	43,54	30,50	41,32
d	en	0	1	48,72	33,50	46,13
d	en	1	0	48,57	30,50	45,49
d	en	1	1	49,65	34,25	47,03
d	en	3	0	50,41	29,00	46,76
d	en	3	1	50,67	38,25	48,56
d	en	5	0	50,87	33,50	47,92
d	en	5	1	51,33	36,00	48,72
d	en	7	0	50,61	35,50	48,04
d	en	7	1	53,13	40,25	50,94
d	en	10	0	51,59	35,50	48,85
d	en	10	1	54,31	38,75	51,66
d	fr	0	0	48,61	31,25	45,66
d	fr	0	1	49,33	37,50	47,32
d	fr	1	0	48,51	30,00	45,36
d	fr	1	1	50,56	37,00	48,25
d	fr	3	0	52,21	32,25	48,81
d	fr	3	1	50,10	33,25	47,23
d	fr	5	0	49,13	35,75	46,85
d	**fr**	**5**	**1**	**54,52**	**39,50**	**51,96**
d	fr	7	0	53,38	34,50	50,17
d	fr	7	1	51,39	39,50	49,36
d	fr	10	0	52,20	23,50	47,32
d	fr	10	1	53,02	41,25	51,02
e	en	0	0	44,41	21,25	40,47
e	en	0	1	46,25	30,50	43,57
e	en	1	0	44,41	19,75	40,21
e	en	1	1	47,54	27,50	44,13
e	en	3	0	45,18	23,50	41,49
e	en	3	1	47,03	30,00	44,13
e	en	5	0	46,41	26,25	42,98
e	en	5	1	48,97	33,50	46,34
e	en	7	0	47,85	26,75	44,26
e	en	7	1	50,05	31,75	46,93
e	en	10	0	48,71	28,00	45,19
e	en	10	1	50,26	30,25	46,85
e	fr	0	0	39,43	27,25	37,36
e	fr	0	1	47,34	32,25	44,77
e	fr	1	0	46,41	26,00	42,94
e	fr	1	1	50,00	31,50	46,85
e	fr	3	0	49,85	26,00	45,79
e	fr	3	1	50,46	31,75	47,27
e	fr	5	0	50,10	29,25	46,55
e	fr	5	1	52,56	33,75	49,36
e	fr	7	0	52,16	29,50	48,30
e	fr	7	1	53,90	34,25	50,55
e	fr	10	0	51,53	32,50	48,29
e	fr	10	1	53,18	37,00	50,43

TABLE 28 – SM des variantes `Qwen` (`Hist:0`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	32,88	21,75	30,98
d	en	0	1	—	—	—
d	en	1	0	41,13	25,50	38,47
d	en	1	1	—	—	—
d	en	3	0	46,36	30,50	43,66
d	en	3	1	—	—	—
d	en	5	0	47,74	31,50	44,98
d	en	5	1	—	—	—
d	en	7	0	48,61	35,25	46,34
d	en	7	1	—	—	—
d	**en**	**10**	**0**	**49,54**	**35,50**	**47,15**
d	en	10	1	—	—	—
d	fr	0	0	31,85	17,25	29,36
d	fr	0	1	—	—	—
d	fr	1	0	41,59	26,25	38,98
d	fr	1	1	—	—	—
d	fr	3	0	45,49	31,00	43,02
d	fr	3	1	—	—	—
d	fr	5	0	46,76	27,25	43,44
d	fr	5	1	—	—	—
d	fr	7	0	48,05	30,75	45,10
d	fr	7	1	—	—	—
d	fr	10	0	48,56	31,75	45,70
d	fr	10	1	—	—	—
e	en	0	0	33,08	16,25	30,21
e	en	0	1	—	—	—
e	en	1	0	39,54	20,50	36,30
e	en	1	1	—	—	—
e	en	3	0	44,00	28,25	41,32
e	en	3	1	—	—	—
e	en	5	0	44,11	29,00	41,54
e	en	5	1	—	—	—
e	en	7	0	46,20	32,50	43,87
e	en	7	1	—	—	—
e	en	10	0	47,75	31,75	45,02
e	en	10	1	—	—	—
e	fr	0	0	33,23	17,00	30,47
e	fr	0	1	—	—	—
e	fr	1	0	40,52	20,50	37,11
e	fr	1	1	—	—	—
e	fr	3	0	45,12	27,00	42,04
e	fr	3	1	—	—	—
e	fr	5	0	45,38	27,00	42,25
e	fr	5	1	—	—	—
e	fr	7	0	47,29	28,25	44,05
e	fr	7	1	—	—	—
e	fr	10	0	48,00	29,50	44,85
e	fr	10	1	—	—	—

TABLE 29 – SM des variantes `Qwen` (`Hist:1`), P=paradigmatique, S=syntagmatique, MG=moyenne générale

Pr	L	k	Fb	P	S	MG
d	en	0	0	50,46	33,50	47,57
d	en	0	1	51,64	41,25	49,87
d	en	1	0	48,67	31,75	45,79
d	en	1	1	54,92	38,25	52,09
d	en	3	0	53,80	38,00	51,11
d	en	3	1	55,49	43,00	53,36
d	en	5	0	53,90	36,50	50,94
d	en	5	1	56,72	41,00	54,04
d	en	7	0	52,57	39,00	50,26
d	en	7	1	57,69	42,75	55,15
d	en	10	0	54,00	38,00	51,28
d	en	10	1	55,85	43,75	53,79
d	fr	0	0	47,64	34,50	45,41
d	fr	0	1	55,33	38,25	52,43
d	fr	1	0	50,87	34,00	48,00
d	fr	1	1	56,25	38,25	53,19
d	fr	3	0	53,95	35,25	50,76
d	fr	3	1	56,41	40,25	53,66
d	fr	5	0	55,65	33,75	51,92
d	fr	5	1	56,93	39,75	54,00
d	fr	7	0	55,90	35,50	52,43
d	fr	7	1	57,59	39,25	54,47
d	fr	10	0	56,15	35,50	52,64
d	**fr**	**10**	**1**	**57,90**	**41,50**	**55,11**
e	en	0	0	48,46	31,50	45,58
e	en	0	1	50,77	34,50	48,00
e	en	1	0	47,18	31,50	44,51
e	en	1	1	51,94	35,25	49,10
e	en	3	0	48,25	31,50	45,40
e	en	3	1	52,46	37,25	49,87
e	en	5	0	48,31	33,00	45,70
e	en	5	1	52,31	37,25	49,75
e	en	7	0	49,33	33,00	46,55
e	en	7	1	54,31	36,25	51,23
e	en	10	0	50,41	33,00	47,45
e	en	10	1	54,21	38,00	51,45
e	fr	0	0	48,92	31,75	46,00
e	fr	0	1	52,82	35,25	49,83
e	fr	1	0	50,72	30,75	47,32
e	fr	1	1	52,61	35,50	49,70
e	fr	3	0	54,31	30,00	50,17
e	fr	3	1	54,25	37,75	51,45
e	fr	5	0	50,62	32,75	47,58
e	fr	5	1	52,41	37,25	49,83
e	fr	7	0	52,82	31,25	49,15
e	fr	7	1	55,80	38,75	52,90
e	fr	10	0	53,03	33,00	49,62
e	fr	10	1	54,62	37,25	51,66