

Les noms de marque dans les unités phraséologiques : enjeux méthodologiques d'un corpus franco-espagnol non normé. Quelle place pour l'intelligence artificielle ?

Aude Grezka

LIPN, Université Sorbonne Paris Nord

ORCID: 0000-0002-4582-3428

Natalia Soler Cifuentes

PLÉIADE, Université Sorbonne Paris Nord

CLESTHIA, Université Sorbonne Nouvelle

ORCID: 0009-0008-1128-5175

DOI: 10.54103/milanoup.305.c727

ABSTRACT

Brands are part of our lexiculture and their names are constantly circulating in our everyday discourses. The phenomenon of genericization of brand names leads to the emergence of new phraseological units (PU) that speakers establish to codify meanings and achieve linguistic economy or to bring more expressiveness (such as *avoir une tête de Playmobil* in French and *tener un pelo Pantene* in Spanish). To date, few researchers have focused on PU containing a brand name. The reason is undoubtedly the lack of available data, since the lexicographical treatment of these PU is almost nonexistent. These expressions mainly appear in oral communication or on social networks, grounds conducive to phraseological innovation. This study aims to analyze the methodological challenges involved in the constitution of a non-standardized corpus of PU containing a brand name in French and Spanish. The orthotypographical variability of brand names, coupled with the lack of available resources, makes this corpus a challenging winning combo for the linguist. The use of artificial intelligence has been explored to extract these PU (semi-)automatically, but these language models reveal their limitations when applied to such heterogeneous, non-standardized corpora.

Keywords: phraseological units; brand names; artificial intelligence; IA; French; Spanish.

RÉSUMÉ

Les marques commerciales font partie de notre lexiculture et leurs noms circulent en permanence dans nos discours quotidiens. Le phénomène de lexicalisation des noms de marque (NdM) conduit, à terme, à l'apparition d'unités phraséologiques (UP) que les locuteurs mettent en place afin de codifier des signifiés, dans le but de réaliser des économies linguistiques ou d'apporter plus d'expressivité (*avoir une tête de Playmobil* en français ; *tener un pelo Pantene* en espagnol). À l'heure actuelle, peu de chercheurs se sont intéressés aux UP comportant un NdM, sans doute en raison du manque de données disponibles, car le traitement lexicographique de ces UP est quasi inexistant. Ces UP se retrouvent principalement à l'oral ou sur les réseaux sociaux, des terrains propices à l'innovation phraséologique. L'objectif de cette étude est d'analyser les enjeux méthodologiques liés à la constitution d'un corpus non normé d'UP comportant un NdM en français et en espagnol. La variabilité orthotypographique des NdM, associée au manque de ressources disponibles constitue un véritable défi pour le linguiste. La voie de l'intelligence artificielle a été explorée afin d'extraire ces UP de manière (semi-)automatique, mais les modèles linguistiques montrent leurs limites lorsqu'ils sont appliqués à des corpus hétérogènes et non standardisés.

Mots-clés : unités phraséologiques ; noms de marque ; intelligence artificielle ; IA ; français ; espagnol.

1. Nom de marque, lexicalisation et unités phraséologiques

Le nom de marque (NdM) constitue une pièce lexicale intéressante pour le linguiste en raison de son statut polyphonique qui fait converger plusieurs voix à l'intérieur d'un même signe : celle du marketing, celle du droit et celle de la linguistique (Petit 2000, 2006 ; Vaxelaire 2014). Les NdM sortent de leur tanière publicitaire pour ensuite trouver refuge au plus profond de notre patrimoine lexiculturel (Tonti 2020). Ils sont un condensé culturel au même titre que les noms de fêtes, d'institutions, de personnages historiques, etc.

Lorsqu'un NdM se lexicalise dans la langue, il devient la dénomination générique de la catégorie de produits dans laquelle il s'inscrit : *Kleenex* pour « mouchoir jetable en papier », *Rimmel* pour « fard pour les cils », etc. Il perd ainsi sa fonction distinctive première — désigner une offre commerciale spécifique et la distinguer de celles des concurrents — et, par là même, le fondement de son existence juridique. L'enjeu est majeur pour son titulaire, qui encourt la déchéance de ses droits exclusifs¹.

En linguistique, la lexicalisation du NdM est décrite comme un phénomène graduel, qui peut aller du simple emploi métonymique (une [voiture] *Renault*) à une définition lexicale, attestée par son entrée dans le dictionnaire (Soler

1 « Le titulaire d'une marque devenue, de son fait, la désignation usuelle dans le commerce du produit ou du service encourt la déchéance de ses droits » (Article L. 714-6, a) du Code de la Propriété Intellectuelle).

Cifuentes 2021). Dans cette échelle, où le NdM oscille entre individualisation et catégorisation, certains emplois témoignent d'un haut degré de lexicalisation. Tel est le cas de la création de nouvelles unités phraséologiques (UP) dont les locuteurs se servent pour codifier des signifiés (*jeter comme un Kleenex*). Ces UP contenant un NdM ne sont pas des usages marginaux limités à un groupe social spécifique : une simple recherche sur Google permet de confirmer qu'elles sont assez répandues et peuvent même être considérées comme relevant d'un argot commun.

Malgré l'essor des études phraséologiques de ces dernières décennies, force est de constater que les UP avec NdM n'ont pas attiré l'attention des chercheurs. Cela peut s'expliquer pour plusieurs raisons, dont nous soulignons les contraintes juridiques du traitement lexicographique des NdM², ainsi que le fait que ces UP voyagent incognito dans nos discours quotidiens — surtout à l'oral. Il faut les saisir au vol et les répertorier, faute de quoi on se retrouve face au néant en termes de données, puisqu'à ce jour aucun ouvrage ne leur est consacré.

De tous ces constats est né le projet *Phrasmark* (Soler Cifuentes, Grezka et Garcia Flores 2024), un projet pluridisciplinaire dont l'objectif est de bâtir une base de données d'expressions figées contenant des NdM afin de :

1. Étudier la manière dont chaque communauté linguistique codifie ses référents culturels à travers les UP.
2. Analyser les processus d'encodage / décodage avec les marques en tant qu'élément central.
3. Développer un outil de travail pour traducteurs, interprètes, professeurs, lexicographes, etc.

2. Constitution d'un corpus non normé : enjeux méthodologiques et traitement lexicographique

Dans le cadre du présent article, nos langues de travail sont le français hexagonal et l'espagnol péninsulaire. Ce sont donc les marques présentes dans ces deux marchés qui font l'objet de notre recherche. Dans un premier temps, nous avons établi une liste de marques susceptibles de donner lieu à une UP. Le choix des marques repose sur deux critères : son potentiel phraséogénique et sa notoriété. Le potentiel phraséogénique de la marque correspond à sa capacité à véhiculer des sens figurés à travers ses attributs (matériels ou immatériels) et l'imaginaire qui s'est construit autour d'elle. À priori, toute marque peut subir

2 Conformément au droit de marques de l'Union européenne, il est requis qu'un NdM figurant dans un dictionnaire ou tout autre ouvrage académique soit assorti d'une mention indiquant qu'il s'agit d'une marque déposée. (RÈGLEMENT (UE) 2017/1001 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 juin 2017 sur la marque de l'Union européenne.) <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32017R1001>

d'un détournement de son sens initial sous l'effet de divers processus cognitifs. Le fait que seules certaines marques parviennent à figer ces significations secondaires dans une UP dépend de leur degré de circulation dans les discours. Les marques à forte notoriété sont davantage ancrées dans le système linguistico-culturel, ce qui se traduit par une circulation plus fréquente et intense, favorisant la création d'UP.

À partir d'une liste de plusieurs centaines de NdM, nous nous sommes lancées dans la collecte d'UP les contenant. Après plusieurs tentatives infructueuses de recherche dans des ouvrages et corpus de référence, nous avons fini par adopter une méthodologie d'extraction manuelle, basée sur deux méthodes complémentaires : d'une part, la collecte des données sur les réseaux sociaux, notamment sur *X* (anciennement *Twitter*) ; d'autre part, une collecte directe auprès de locuteurs natifs (tout âge et sexe confondus), qui ont permis d'attester l'existence et l'usage réel de ces UP. Il en résulte un corpus bilingue composé de 38 UP en langue française et 31 UP en langue espagnole.

Le principal obstacle de ce corpus tient à la nature non normée des NdM, qui entrave une automatisation de l'extraction des données. En effet, ces noms ne forment absolument pas une classe homogène avec des critères distinctifs, ce qui les rend spécialement sensibles à la manipulation orthotypographique par les locuteurs. Ainsi, elles fluctuent entre minuscule et majuscule initiales, entre adaptation et emprunt pur, entre trait d'union et espace. Cela donne lieu à plusieurs variantes d'un même NdM, aussi bien en français (*Flamby* et *Flanby*, le deuxième nom étant le NdM original qui défie les normes orthographiques habituelles et est « corrigé » par les locuteurs) qu'en espagnol (*clínex* par *Kleenex*, la première forme étant l'emprunt adapté préconisé par l'académie de la langue espagnole).

Outre les problèmes de variations graphiques, l'indisponibilité d'ouvrages de référence répertoriant des UP avec NdM complexifie davantage la constitution de notre corpus. Elles ne sont majoritairement pas dans les dictionnaires traditionnels, qu'ils soient normatifs ou descriptifs, pas non plus que dans les ouvrages spécialisés, comme les dictionnaires ou recueils phraséologiques. Ces UP, qui sont principalement employées à l'oral ou sur les réseaux sociaux, figurent rarement dans les ressources lexicographiques à cause de leur statut juridique et linguistique particulier. En tant que noms propres protégés par la loi, les NdM sont généralement absents des dictionnaires, à l'exception de ceux qui sont devenus génériques et, de ce fait, des noms communs à part entière.

À titre d'exemple, nous avons mis à l'épreuve un dictionnaire unilingue de référence pour chacune des langues (pour le français, *Le Grand Robert* ou LGR ; pour l'espagnol, le *Diccionario de la Lengua Española* ou DLE) et nous y avons cherché les UP qui composent notre corpus. Il faut préciser que certaines UP contiennent le même NdM (*être beurré comme un Petit Lu* et *tremper son Petit Lu*) et d'autres ont le même signifiant, mais une signification différente (*dar vaselina*,

lit. ‘donner de la vaseline’, dans le sens de ‘adoucir une critique par des compliments préalables’).

Les résultats sont sans appel : tandis que le LGR compte 6 UP de notre corpus français, soit 15 % des UP, le DLE n’en compte qu’une seule. Nous précisons que tout type de locution (*blanc comme un cachet d’aspirine*) ou sens métaphorique susceptible d’être utilisée en tant qu’UP (*être blanc comme un cachet d’aspirine*) a été comptabilisé.

Tableau 1 : UP avec NdM présentes dans le dictionnaire

Le Grand Robert (LGR)	Diccionario de la Lengua Española (DLE)
<ul style="list-style-type: none"> - <i>blanc comme un cachet d’aspirine</i> - <i>bébé cadum</i> - <i>passer comme une lettre à la poste</i> - <i>décrocher, attraper la queue du Mickey</i> - <i>beurré comme un petit Lu</i> - <i>être, rester scotché</i> 	<ul style="list-style-type: none"> - <i>estar alguien o algo más visto que el tebeo</i> (lit. ‘être plus vu que le tebeo’, signifie ‘être très répandu et ne plus susciter d’intérêt’)

Faute de sources disponibles, nous avons donc constitué notre corpus « à l’œil » et « à la main », en observant les différentes occurrences sur les réseaux sociaux, et en croisant ces données avec celles fournies par des locuteurs natifs de l’espagnol et du français. Or, la méthode de collecte manuelle s’avère longue et pas toujours fructueuse, car ces UP se fondent dans la masse des discours quotidiens et les marques qui les composent ne sont parfois pas reconnues comme telles. Ces difficultés nous ont amenées à réfléchir à la possibilité d’utiliser le potentiel de l’intelligence artificielle afin d’élargir notre corpus de façon automatique ou semi-automatique. Un tel outil peut-il réellement convenir à la création d’un corpus non normé ?

3. Vers une extraction supervisée d’UP contenant des NdM ?

L’intelligence artificielle (IA) est un terme générique qui désigne la capacité de machines ou de logiciels à simuler certaines fonctions cognitives humaines, telles que le raisonnement, l’apprentissage, la résolution de problèmes ou encore la compréhension du langage. Elle englobe plusieurs domaines technologiques, dont l’apprentissage automatique (*machine learning*), où des algorithmes apprennent à partir de données et s’améliorent au fil du temps sans être explicitement programmés pour chaque tâche.

Parmi les avancées les plus marquantes de ces dernières années figure l’IA générative, une branche spécifique de l’intelligence artificielle capable de produire automatiquement du contenu inédit, qu’il s’agisse de textes, d’images, de vidéos, de musique ou même de code informatique. Ces systèmes s’appuient

sur des modèles entraînés sur de vastes ensembles de données pour identifier des régularités, reproduire des styles, ou générer de nouvelles créations cohérentes et pertinentes. Popularisée fin 2022 par le lancement de ChatGPT, un agent conversationnel développé par OpenAI, reposant sur l'architecture GPT (*Generative Pre-trained Transformer*, soit « transformeur génératif pré-entraîné »), cette technologie a très rapidement trouvé un large public. ChatGPT s'est notamment distingué par sa capacité à produire du texte de qualité, à comprendre des requêtes formulées en langage naturel, et à simuler un dialogue fluide. L'engouement mondial pour ce type d'outil a marqué un tournant dans la perception et l'usage de l'IA générative, notamment dans les domaines de l'éducation, de la création de contenu et de la recherche linguistique.

Il était donc tout naturel de tester l'extraction d'UP avec NdM avec ChatGPT. Nous nous sommes appuyés sur la version gratuite et « grand public », plus simple à utiliser que le modèle de langage GPT-3³, mais dans lequel le modèle et les données d'entraînement sont les mêmes. La formulation du prompt est cruciale parce qu'elle détermine la qualité, la pertinence et la créativité des réponses obtenues, surtout quand il s'agit de recherches linguistiques fines comme les expressions idiomatiques avec des marques. Nous avons donc soumis le prompt suivant : « Donne-moi une liste d'unités phraséologiques utilisant des NdM (passées dans le langage courant), sur le modèle de “*beurré comme un petit Lu*” ou “*être, rester scotché*” et précise pour chaque expression ce qu'elle signifie ». Nous avons obtenu pour le français et l'espagnol les résultats suivants, accompagnés de leur signification :

Tableau 2 : UP contenant des NdM en français générées par ChatGPT

Amazon	<i>Être dans la zone Amazon</i>	Être totalement absorbé par une tâche ou une activité en ligne.
Blackberry	<i>Avoir un Blackberry dans la poche</i>	Être en mesure de rester en contact avec les autres à tout moment.
Coca-Cola	<i>Être le Coca-Cola de quelque chose</i>	Être le leader incontesté dans un domaine.
Colgate	<i>Avoir un sourire Colgate</i>	Avoir un sourire brillant et propre.
Facebook	<i>Avoir la boule de cristal de Facebook</i>	Avoir une grande capacité à prévoir l'avenir ou à prédire les tendances.

3 Pour l'instant, la version de base de ChatGPT s'appuie précisément sur le modèle de langage GPT-3.5. En mars 2023, OpenAI a sorti GPT-4, qui reste encore aujourd'hui réservé aux utilisateurs de ChatGPT Plus, un modèle bien plus puissant et performant, mais payant.

Ferrari	<i>Avoir la Ferrari des ordinateurs</i>	Posséder un ordinateur extrêmement puissant et performant.
Google	<i>Être sur la même longueur d'onde que Google</i>	Être d'accord avec quelqu'un, avoir la même compréhension ou la même opinion.
Nike	<i>Être le Nike de quelque chose (affaires)</i>	Être le leader ou le meilleur dans un domaine particulier.
Starbucks	<i>Faire un Starbucks run</i>	Aller chercher des boissons ou des collations chez Starbucks.

Tableau 3 : UP contenant des NdM en espagnol générées par ChatGPT

Coca-Cola	<i>Ser la Coca-Cola de los refrescos</i> (lit. 'être le Coca-Cola des sodas')	Être le leader du marché dans sa catégorie.
Colgate	<i>Tener una sonrisa Colgate</i> (lit. 'avoir un sourire Colgate')	Avoir un sourire éclatant et blanc.
Duracell	<i>Ponerse las pilas como Duracell</i> (lit. 'se mettre des piles comme Duracell')	Se mettre en action avec énergie et détermination.
Ferrari	<i>Ir como un Ferrari</i> (lit. 'aller comme une Ferrari')	Aller à grande vitesse ou avec beaucoup d'énergie.
Ikea	<i>Estar más perdido que un turista en el IKEA</i> (lit. 'être plus perdu qu'un touriste chez IKEA')	Être très confus ou désorienté.
McDonald's	<i>Hacer un McDonald's</i> (lit. 'faire un McDonald's')	Accomplir une tâche ou une action rapidement.
Red Bull	<i>Ir a toda pastilla como Red Bull</i> (lit. 'aller à toute vitesse comme Red Bull')	Aller à toute vitesse.
Renault	<i>Estar hecho un Renault</i> (lit. 'être une Renault')	Être en bon état, fonctionner correctement.
Telefónica	<i>Hablar por los codos como Telefónica</i> (lit. 'parler par les coudes comme Telefónica')	Parler beaucoup ou sans interruption.

Sur la forme, les réponses données par ChatGPT sont très convaincantes... mais en grande majorité fausses. Pour 89 % des expressions proposées par ChatGPT, soit 16 sur 18, nous n'avons pas pu attester d'un usage réel, ni sur le Web ni auprès des locuteurs. Il se pourrait qu'il s'agisse d'hapax, mais aucune trace n'a été retrouvée en ligne. Seules les expressions *avoir un sourire Colgate* et *tener una sonrisa Colgate* (lit. 'avoir un sourire Colgate') constituent de véritables UP, dont l'usage est attesté sur Internet.

ChatGPT fait ici de la « confabulation ». La confabulation est un phénomène psychologique où une personne remplit des lacunes dans sa mémoire en inventant inconsciemment des histoires ou des détails pour expliquer des événements ou des situations dont elle ne se souvient pas correctement. Ces créations peuvent sembler cohérentes et convaincantes pour la personne qui les énonce, mais elles ne sont basées sur aucun fait réel. Bien que ChatGPT ne fonctionne pas comme le cerveau humain, le terme « confabulation » est sans doute la meilleure métaphore pour caractériser ce principe créatif de comblement des lacunes⁴ (Huang *et al.* 2023 ; Shah 2023). Le modèle génère ici du texte basé sur des connaissances inexistantes ou des déductions incorrectes. Il génère du texte en fonction des modèles statistiques qu'il a appris et ne possède pas de compréhension ou de conscience réelles comme le font les humains.

Les chercheurs construisent, entraînent de grands modèles linguistiques comme GPT-3 en utilisant un processus appelé « apprentissage non supervisé », ce qui signifie que les données qu'ils utilisent pour former le modèle ne sont pas spécialement annotées ou étiquetées. Le modèle doit donc découvrir la structure inhérente des données par lui-même. Au cours de ce processus, le modèle est alimenté par un grand nombre de textes. Ces textes sont ensuite éclatés en fragments de mots ou « tokens » au cours d'un processus d'apprentissage complexe. Lorsqu'une phrase est soumise, les mots sont séparés. Chaque mot est caractérisé par sa proximité ou sa distance mesurée statistiquement avec d'autres mots, ce qui se traduira par une suite de nombres. Le modèle essaie à plusieurs reprises de prédire le mot suivant dans chaque séquence de mots. Si la prédiction du modèle est proche du mot suivant réel, le réseau neuronal met à jour ses paramètres pour renforcer les modèles qui ont conduit à cette prédiction. Inversement, si la prédiction est incorrecte, le modèle ajuste ses paramètres pour améliorer ses performances et réessaie (cela permet au modèle d'apprendre de ses erreurs et d'améliorer progressivement ses prédictions pendant le processus de formation). En conséquence, GPT apprend les associations statistiques entre les mots et les concepts connexes dans l'ensemble de données. C'est ce qui apparaît très clairement dans nos deux tableaux. Le modèle a ici construit ses UP sur les grandes propriétés qui caractérisent les marques (Ferrari = puissance ; Nike = leader ; etc.).

Il reste donc, à l'heure actuelle, délicat de s'appuyer sur un modèle de langage pour extraire des UP avec NdM (Soler Cifuentes, Grezka et García Flores 2024). Les modèles s'améliorent avec le temps et savent de mieux en mieux

4 Mikhaïl Parakhin, PDG de Advertising & Web Services chez Microsoft, qui travaille sur Bing Chat, a tweeté et reconnu la possibilité pour les grands modèles de langage (LLM) comme Bing « d'inventer des choses » : « This is what I tried to explain previously: hallucinations = creativity. It tries to produce the highest probability continuation of the string using all the data at its disposal. Very often it is correct. Sometimes people have never produced continuations like this » (X, 24 février 2023).

s'abstenir de répondre lorsqu'ils n'ont pas l'information. Cependant, en dehors des questions simples, un modèle de langage peut facilement inventer une réponse avec une explication tout aussi inventée et approximative (ici la définition de l'expression).

4. Marque et lexiculture : origine des UP contenant NdM

Si ChatGPT n'est pas capable de fournir des exemples réels d'UP contenant un NdM, ce n'est pas en raison de leur rareté, bien au contraire, ces UP appartiennent à la langue courante partagée par la majorité des locuteurs. En effet, la marque a vocation à être omniprésente et connue du plus grand nombre de consommateurs, et sa nature universelle se manifeste également dans la phraséologie. Les limites de ChatGPT concernent le type de diffusion de ces expressions, qui sont majoritairement utilisées à l'oral ou dans des productions écrites qui sont imprégnées d'oralité, particulièrement les productions sur les réseaux sociaux.

Bien que les UP avec NdM relèvent de nos discours les plus courants et quotidiens, elles passent inaperçues pour ChatGPT, car elles se propagent furtivement, souvent sans laisser des traces écrites. Elles font partie de notre lexiculture, terme que nous empruntons à Galisson (1998, 2000) pour rendre compte de « la culture en dépôt dans les mots », une culture implicite sous-jacente au lexique, connue et reconnue par les natifs, mais opaque et inaccessible pour étrangers et apprenants.

Le rôle de la publicité dans la création d'UP avec NdM est indéniable : slogans, spots publicitaires ou personnages de marque constituent autant de sources d'innovation phraséologique. Toutefois, l'analyse de notre corpus révèle que, dans la majorité des cas, ce sont les locuteurs eux-mêmes qui en sont à l'origine, soit par une création spontanée, soit par le détournement d'UP préexistantes. Cette tendance s'observe aussi bien en français qu'en espagnol, avec 84 % des unités en français (32 UP sur 38) et 74 % en espagnol (23 UP sur 31) attribuées aux locuteurs. Ainsi, on peut établir une première classification des UP en deux catégories en fonction de leur origine, selon que le moteur de la création phraséologique est le discours publicitaire ou bien les propres locuteurs.

4.1 Publicité

La publicité est un terrain fertile pour la création de nouvelles UP, et cela, en raison des slogans et d'autres énoncés qui, à force d'être répétés, s'implantent dans les esprits telle une comptine. Son pouvoir réside notamment dans sa capacité à créer une culture commune et inclusive. La publicité s'adresse le plus souvent à un large public et diffuse ainsi ses slogans, que chacun perçoit comme lui étant directement destinés.

Le français connaît la création d'UP directement de l'assimilation d'un slogan, comme dans *être comme le Port-Salut*. *Port-Salut* est une marque fromagère qui, à

partir des années soixante, va utiliser le slogan « C’est comme le Port-Salut, c’est écrit dessus » pour indiquer que les fromages produits par cette marque sont facilement identifiables, puisque le NdM est écrit sur l’emballage. Ce slogan est passé dans la langue courante pour indiquer l’évidence de quelque chose.

L’espagnol a fait de même avec *tener un pelo Pantene* (lit. ‘avoir des cheveux Pantene’, signifie ‘avoir de jolis cheveux, sains et brillants’). *Pantene* est une marque de produits capillaires qui a basé sa stratégie publicitaire en Espagne sur le slogan « pelo Pantene », en référence à des cheveux sains et brillants après l’utilisation des produits de la marque.

Dans d’autres cas, c’est le personnage de la marque qui contribue à la création d’une UP, comme c’est le cas pour *être un bébé Cadum* — dont le personnage de marque est un bébé joufflu, à la peau rose, lisse et douce —, et *durar más que el conejito de Duracell* (lit. ‘durer plus que le petit lapin de Duracell’, signifie ‘avoir beaucoup d’énergie’) — dont la mascotte est un lapin rose avec une énergie débordante.

4.2 Locuteurs

1. Propriétés physiques ou immatérielles de la marque

Les consommateurs peuvent interpréter les caractéristiques physiques ou symboliques que la marque véhicule pour créer de nouvelles UP. Ce sont les caractéristiques propres à chaque marque dont ChatGPT se sert pour faire de la confabulation, comme dans *être le Coca-Cola de quelque chose*. Cette UP, qui signifierait être le leader dans un domaine particulier, montre que ChatGPT base sa confabulation sur ses connaissances de *Coca-Cola* en tant que marque leader dans la catégorie des sodas. Ce qui est d’autant plus intéressant, c’est que les locuteurs espagnols ont codifié cette notion de succès commercial dans l’UP *triumfar como la Coca-Cola* (lit. ‘triumpher comme le Coca-Cola’, dans le sens de ‘rencontrer un grand succès’). C’est la preuve que ChatGPT n’était pas très loin dans ses suppositions, mais la différence réside dans le fait que l’expression figée proposée par le modèle de langage n’est pas issue d’un usage réel mais d’une invention.

La force créative des locuteurs se concrétise en français par l’emploi d’unités comme *avoir une tête de Playmobil* ou *avoir un corps d’Orangina*, dans lesquelles la caractéristique physique du produit que signe la marque est à l’origine des UP (la coupe carré de la figurine *Playmobil* ou l’allusion à la morphologie corporelle en forme de poire de la bouteille d’*Orangina*). En espagnol, nous retrouvons ce même mécanisme, mais cette fois-ci appliqué au personnage de la marque, comme dans *ser (un) Don Limpio* (lit. ‘être (un) Mr. Propre’, signifie ‘être chauve’), qui fait référence à un homme chauve tel que le personnage *Mr. Propre*.

2. UP existante

L’emploi des expressions figées est une spécificité de la publicité, à tel point que Berthelot-Guiet (2013, 89) affirme que « le message publicitaire aime la

formule figée pour mieux la détourner ou la “défiger” ». Des slogans célèbres comme *En avril, ne te découvre pas d’un Dim* ou encore *Les chiens aboient, les Lee Cooper passent* (cités par Grésillon et Maingueneau 1984, 117) en sont de bons exemples. Or, nous nous intéressons à un cas particulier d’UP existantes que les consommateurs détournent par substitution (ils remplacent une partie de l’expression par un NdM) ou par addition (ils ajoutent un NdM).

En langue française, l’expression *être mou comme un Chamallow*, variante d’*être mou comme une chiffre (molle)*, est employée pour décrire une personne de faible caractère ou dépourvue d’énergie. *Chamallow* est une marque de pâte de guimauve dont la consistance est moelleuse. *Chiffre* étant un vocable tombé en désuétude, il a été remplacé par une référence culturelle plus transparente pour des locuteurs du XXI^e siècle.

Le caractère pusillanime est également codifié en espagnol dans *tener Chufi en las venas* (lit. ‘avoir du Chufi dans les veines’, expression familière pour décrire quelqu’un de très calme, qui manque d’initiative), qui est une déformation de l’expression originale *tener borchata en las venas* (lit. ‘avoir de l’orgeat de souchet dans les veines’). L’orgeat de souchet est une boisson blanchâtre typique de la région de Valence, très consommée pendant la période estivale. Dès qu’une personne montre un caractère timoré, sans vie, sans énergie, on dit qu’elle n’a pas de sang dans les veines ou que ce sang est de l’orgeat. L’expression prend un nouveau tournant expressif quand le terme orgeat est remplacé par *Chufi*, marque espagnole leader de boissons d’orgeat de souchet.

5. Conclusions

L’influence de la marque transperce le domaine marchand et se glisse dans nos discours courants, en s’érigeant comme un messenger privilégié de notre lexiculture. Le phénomène de lexicalisation des NdM conduit en dernière instance à l’apparition de nouvelles UP qui font surface lors de nos échanges quotidiens.

À l’heure actuelle, les UP comportant un NdM restent peu étudiées, sans doute en raison du manque de données disponibles dans les ouvrages de référence (dictionnaires, recueils phraséologiques). Nous avons suivi une méthodologie d’extraction manuelle d’UP contenant NdM, en collectant des données sur les réseaux sociaux ou auprès de locuteurs natifs. Cette méthode reste néanmoins limitée par les nombreuses variations orthographiques que subissent les NdM, ainsi que par la non-reconnaissance de ces expressions qui passent inaperçues dans la masse d’échanges de la vie courante.

La constitution d’un corpus non normé constitue un défi pour le linguiste, défi que nous avons tenté de relever à l’aide de l’intelligence artificielle afin de constituer notre corpus de façon (semi-)automatique. Les résultats ont montré, pour le moment, les limites des modèles de langage dans ces cas de figure, car l’IA a recours à la confabulation.

Bibliographie

- Berthelot-Guiet, Karine. 2013. *Paroles de pub : la vie triviale de la publicité*. Paris : Éditions Non standard.
- Deval, Shah. 2023. *The Beginner's Guide to Hallucinations in Large Language Models, Lakera*, <https://www.lakera.ai/blog/guide-to-hallucinations-in-large-language-models>.
- Robert Galisson et Jean-Claude André. 1998. *Dictionnaire des noms de marques courants : essai de lexiculture ordinaire*. Paris : Didier érudition.
- Galisson, Robert. 2000. “La pragmatique lexiculturelle pour accéder autrement, à une autre culture, par un autre lexique”. *Mélanges CRAPEL, Une didactique des langues pour demain-En hommage au Professeur Henri Holec* 25: 47–73. <https://www.atilf.fr/wp-content/uploads/publications/MelangesCrapel/file-25-6-1.pdf>
- Grésillon, Almuth, et Dominique Maingueneau. 1984. “Polyphonie, proverbe et détournement, ou un proverbe peut en cacher un autre”. *Langages* 73: 112–125. <https://doi.org/10.3406/lgge.1984.1168>
- Huang, Lei, Weijiang Yu, Weitao Ma. 2023. *A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions*, CoRR abs/2311.05232. <https://doi.org/10.48550/arXiv.2311.05232>
- Petit, Gérard. 2000. “Un Hybride Sémiotique : Le Nom Déposé”. *Linguisticae Investigationes* 23 (1): 161-92. <https://doi.org/10.1075/li.23.1.06pet>
- Petit, Gérard. 2006. “Le nom de marque déposée : nom propre, nom commun et terme”. *Meta* 51 (4): 690-705. <https://doi.org/10.7202/014335ar>
- Soler Cifuentes, Natalia. 2021. “El nombre de marca: del nombre propio atípico al fenómeno de la lexicalización”. *Investigaciones léxicas. Estados, temas y rudimentos. Líneas de investigación del Seminario de Lexicografía Hispánica*, édité par M^a Águeda Moreno Moreno et Marta Torres Martínez, 498-520. Barcelona : Octaedro. <https://octaedro.com/libro/investigaciones-lexicas-estados-temas-y-rudimentos/>
- Soler Cifuentes, Natalia, Aude Grezka, et Jorge Garcia Flores. 2024. “Phraséologie et noms de marque : références culturelles en français et en espagnol dans le projet Phrasemark”. *Referencias culturales: Retos en la traducción de la fraseología y del lenguaje de especialidad*, édité par Mireia López-Simó, Pedro Mogorrón Huerta et Analía Cuadrado Rey, 1-19. Amsterdam : John Benjamins. <https://doi.org/10.1075/ivitra.43.01cif>
- Tonti, Michela. 2020. *Le nom de marque dans le discours au quotidien : Prisme lexiculturel et linguistique*. Paris : L'Harmattan.
- Vaxelaire, Jean-Louis. 2014. “Points de vue sur les noms déposés”. *Cahiers de linguistique. Revue de sociolinguistique et de sociologie de la langue française. Traduire aux confins du lexique : les nouveaux terrains de la terminologie* 40/2: 159-76.

Dictionnaires :

Le Grand Robert de la langue française - Version numérique [version 4.4 - février 2023]. <https://www.lerobert.com/dictionnaires/francais/dictionnaire-langue/dictionnaire-le-grand-robert-de-la-langue-francaise-edition-abonnes-3133099010289.html>

REAL ACADEMIA ESPAÑOLA: Diccionario de la lengua española, 23.^a ed., [version 23.7 - février 2023]. <https://dle.rae.es>

