



HAL
open science

Approximation et problèmes de dos

Silvia Adler, Iris Eshkol-Taravella

► **To cite this version:**

Silvia Adler, Iris Eshkol-Taravella. Approximation et problèmes de dos. Langue française, 2020, 207 (3), pp.41-56. 10.3917/lf.207.0041 . hal-03697988

HAL Id: hal-03697988

<https://hal.science/hal-03697988v1>

Submitted on 21 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approximation et problèmes de dos : aspects quantitatifs et qualitatifs relatifs aux stratégies d'approximation d'un corpus issu d'un forum de santé

1. INTRODUCTION

Dans les échanges portant sur des problèmes médicaux menés par des non-spécialistes, l'approximation et l'imprécision s'avèrent très saillantes. Elles se manifestent sous diverses formes – adverbes, tours négatifs, modificateurs prépositionnels, verbes modaux, intensifieurs et *downtowners*, prédicats évidentiels et concessions, toutes visant à « make things fuzzier or less fuzzy » (Lakoff, 1973 : 471), autrement dit, à modifier la prédication et sa relation potentielle à une catégorie prédéfinie ou l'appartenance même du référent à une catégorie désignée par un nom. L'identification et la description de la pathologie médicale, des caractéristiques de celle-ci ou de la catégorie à laquelle elle appartiendrait sont sous-tendues par un jugement subjectif relatif au degré d'appartenance du problème identifié ou décrit à une catégorie préalable.

Nous proposons d'étudier ce type d'échanges en nous fondant sur un corpus issu du Web et extrait du forum de santé *Doctissimo*, en particulier de la rubrique qui traite des douleurs de dos. Les questions et les réponses des internautes y abordent des thématiques variées comme, par exemple, le traitement et les procédures proposés et/ou subis, leur efficacité, les symptômes, les médecins rencontrés, les douleurs supportées, la couverture sociale, etc. Dans ce type de corpus, les notions d'*approximation* et de *flou* sont d'autant plus importantes qu'elles pourraient influencer sur la pertinence de la solution ou du conseil recherchés et fournis par les non-experts. Par l'approximation et le flou, les internautes

signalent que l'étiquette ou la description employées ne sont peut-être pas les plus correctes, adéquates, conformes ou rigides, mais que celles-ci s'approchent toutefois d'un prototype catégoriel, étant donné un degré plus ou moins marqué de ressemblance avec un prototype ou une catégorie envisagés (Mihatsch 2009 ; Adler & Asnes 2010, 2014 ; Gerhard-Krait & Vassiliadou 2014). Un énoncé contenant un approximateur ou un marqueur d'imprécision est donc « vériconditionnellement indécidable » (Lupu, 2003 : 296 ; Kleiber, 1987 : 161-162).

L'approximation et l'imprécision dévoilent parfois aussi la possibilité d'un écart entre le référent (ou la désignation) vulgarisé(é) par l'internaute non spécialiste et le phénomène médical tel qu'un professionnel l'aurait nommé ou décrit. Étant donné que l'approximation mesure le degré d'adéquation du référent par rapport à la catégorie ou au prototype, ou bien opère une extension de la catégorie ¹, elle permet, pragmatiquement parlant, de viser un champ référentiel plus vaste qui prend en compte aussi les marges d'erreur. Par l'imprécision et le flou, l'internaute va jusqu'à signaler que sa description ne correspond pas à une logique binaire et qu'entre le vrai et le faux il existe plusieurs degrés de vérité (Lupu 2003). Un dérivé de cette logique est qu'une entité peut appartenir « plus ou moins » à un ensemble englobant, à la différence d'une situation dichotomique d'appartenance ou d'exclusion totales. L'approximation et le flou peuvent être considérés comme deux degrés de l'imprécis (Adler & Asnes 2014) : alors que l'approximation peut être vue comme plus proche du précis, du fait d'une représentation inexacte d'une valeur précise, le flou consiste dans une impossibilité de fixer l'application référentielle de cette valeur.

Les approximateurs « classiques » comme, par exemple, les noms « métalinguistiques » (dits aussi « expressions d'enclosure ») *sorte, type, espèce* (Flaux & Van de Velde 2000 ; Mihatsch 2006 ; Schapira 2014), l'emploi métalinguistique de *comme* dans *comme un/e* (Moline 1996), des pronoms indéfinis tels *quelque chose (peut être que tu a quelque chose de déplace et il te le remettra en place – (sic) ²*, mais aussi des séquences verbales telles *ça ressemble à, on dirait*, visent tous à amplifier les chances de trouver un « matching » potentiel, sous forme d'une solution précise ou d'un conseil opératoire ³.

La méthodologie choisie pour l'étude de l'approximation et du flou s'inscrit dans le domaine de la linguistique outillée et se fonde sur l'annotation manuelle du corpus. Cette approche est inductive car elle repose sur les données présentes dans le corpus pour acquérir de nouvelles connaissances sur le phénomène étudié. Après avoir présenté le corpus (§ 2), nous développerons une typologie des marqueurs à partir de laquelle l'annotation manuelle a été réalisée (§ 3), puis nous décrirons le procédé d'annotation (§ 4). Enfin, nous présenterons de

1. Cf. approximation intracatégorielle et intercatégorielle (Gerhard-Krait & Vassiliadou, 2017 : 15).

2. Tous les exemples seront cités tels quels, pour rester fidèles aux propos tenus.

3. Voir Jayez & Tovenà (2008 : 8), pour l'« indiscernabilité » entre approximation et précision dans certains contextes.

façon détaillée les résultats quantitatifs et qualitatifs (§ 5). L'article conclura sur quelques perspectives envisagées pour de futurs travaux.

2. CORPUS

Le corpus sur lequel est fondé notre travail est extrait du Web. Il s'agit des échanges menés par des non-spécialistes autour de problèmes liés aux douleurs au dos ⁴. Le corpus étudié est composé de 504 posts extraits du forum de santé *Doctissimo*, de 10 607 mots et de 531 phrases, ce qui correspond à une phrase par post et à vingt mots par phrase en moyenne ⁵.

Les échanges sur des forums sont anonymes ou sous pseudonymes. L'auteur du message ne connaît pas son destinataire et, pour être compris, il doit adapter son message en y apportant assez d'information, avec un bon niveau de pertinence et de clarté.

L'approche outillée et le procédé d'annotation manuelle nous aideront non seulement à estimer la présence effective des marqueurs d'approximation dans le corpus mais aussi à apprécier leur valeur en contexte. L'annotation est fondée sur la typologie des marqueurs d'approximation présentée dans la section suivante.

3. TYPOLOGIE DES MARQUEURS D'APPROXIMATION ET DE SUBJECTIVITÉ

L'observation manuelle du corpus a permis de distinguer sept catégories d'approximation et de subjectivité, ainsi que deux types de contextes d'emploi.

3.1. Catégories des marqueurs d'approximation et de subjectivité

3.1.1. Indéfinis quantifieurs

Selon une définition large, les indéfinis quantifieurs expriment une quantité imprécise et un référent non identifiable (Barwise & Cooper 1981 ; Corblin 1997). Dans l'exemple (1), *une* pourrait s'opposer à *deux* et donc désigner une quantité précise. Mais, il est tout à fait probable que le locuteur communique en même temps une information approximative, liée à une impossibilité de désignation de la vertèbre en question dans un ensemble de vingt-quatre vertèbres. Il se peut fort bien aussi que le locuteur juge qu'une information précise ne soit pas pertinente pour les buts de sa requête ou qu'elle soit de trop. Si c'est en effet le

4. Le corpus a été constitué en mai 2013 et nous a été fourni par Natalia Grabar (CNRS, STL-UMR 8163).

5. Le nombre de phrases a été calculé en s'appuyant sur les signes de ponctuation. Or, dans le forum, les internautes n'utilisent pas toujours ces signes pour segmenter leur discours. Les chiffres présentés doivent donc être relativisés.

cas, l'information procurée suivrait la maxime de quantité ou celle de pertinence (Grice 1975) :

- (1) il m'a remis **une vertèbre** en place et m'a dit que si la douleur persistait je devais aller faire une radio

3.1.2. Localisateurs spatiaux

La localisation est une relation asymétrique, en ce sens qu'elle implique un site (un point d'ancrage stable) et une cible dont le repérage positionnel se fait par rapport à ce site (Vandeloise 1986). Quoique les localisateurs spatiaux relevés dans le corpus soient de deux types – ceux dont la vocation dénotationnelle est spatiale par excellence (régions, quartiers, etc., comme dans (2)) et ceux qui fonctionnent accidentellement comme site de localisation –, dans le corpus étudié, c'est souvent les sites accidentels, à savoir les parties du corps, qui servent de repère spatial (3) :

- (2) nath **du 83**
- (3) avec douleur **dans la jambe droite**

Dans (2), même si l'internaute donne le numéro du département, il est impossible de déterminer le lieu exact de son habitat, d'où le caractère approximatif de l'énoncé. Il est vrai que, normalement, on ne donne pas une mention précise de l'adresse postale sur les forums, mais dans un cas où l'on cherche un spécialiste proche de son domicile, l'information locale procurée aurait pu être plus restrictive. Dans (3), même si la métonymie intégrée (Kleiber 1999) est assez efficace pour caractériser la totalité du membre inférieur rien que par la seule partie affectée, l'information *dans la jambe droite* sera considérée comme approximative, surtout qu'une désignation plus précise de la partie affectée (cuisse, genou, cheville, pied) aurait pu conduire à un «matching» plus efficace. Il est capital de souligner aussi que, quoique les descriptions localisantes comptent parmi les stratégies d'approximation, on ne pourra pas dire qu'elles opèrent une catégorisation. Tout procédé d'approximation n'est donc pas forcément catégoriel dans ce type de corpus.

3.1.3. Localisateurs temporels

Le moment de l'énonciation dans le corpus est le moment d'écriture du message. Par conséquent, le temps proche de ce moment a été considéré comme plus ou moins précis (*il y a 3 jours, hier, aujourd'hui*) et donc écarté de la notation (le seuil a été fixé à 5 jours). En revanche, les expressions *il y a une semaine et demie, dans 10 jours*, etc., sont plus sujettes à un arrondissement et donc il est plus difficile de les situer sur une échelle temporelle. Autrement dit, l'approximation découle ici d'une estimation temporelle relative au commencement, à la durée ou à la fin d'un événement et peut donc ne pas être marquée par des approximateurs tels que *environ, autour de*, etc. Les localisateurs temporels annotés comprennent également les cas où l'internaute indique le temps d'une manière associative comme dans (4) :

- (4) **un début de** tendinite au coude droit

3.1.4. Intensifieurs

[L'intensité est] une intervention du locuteur qui consiste à amplifier ou à amoindrir le degré d'une certaine propriété. L'intensification présuppose donc un jugement subjectif vis-à-vis d'une propriété ou d'une quantité et une comparaison entre deux ou plusieurs degrés possibles et relatifs à la propriété intensifiée. (Adler & Asnes, 2013 : 10)

d'où son rapport au flou ou à l'approximation subjective.

L'annotation manuelle nous a permis de constater que l'intensification se manifeste à travers une grande variété d'éléments lexicaux, mais elle a révélé en même temps à quel point il est capital de bien évaluer les observables. Ainsi, *beaucoup*, dans (5), est un intensifieur mais, dans (6), il est quantifieur ; par conséquent, il a été annoté dans la catégorie des indéfinis quantifieurs :

- (5) Un Iplikator peut **beaucoup** t'aider à soulager les douleurs.
(6) Tu sais c'est vrai que l'on l'impression d'être un numéro quand on va chez le neurochir, mais il n' y en a pas **beaucoup**

Par ailleurs, les observables démontrent que l'intensification peut être marquée aussi par des signes de ponctuation (7), ce qui rend encore plus difficile la délimitation de la souffrance ou de la douleur :

- (7) il souffre ++++

3.1.5. Noms métalinguistiques

Selon N. Flaux et D. Van de Velde (2000), les noms métalinguistiques tels que *sorte, genre, espèce, manière*, etc. qualifient d'autres noms et permettent au locuteur non seulement « d'introduire dans le GN un commentaire sur la pertinence du N utilisé » (*ibid.* : 26) mais aussi de « signaler une adéquation imparfaite du N à l'objet nommé, ou au moins une incertitude quant à sa pertinence » (*ibid.* : 27). Voici, par exemple :

- (8) Le canal rachidien a l'air important pour **ce genre de douleurs** ?

3.1.6. Comparaisons et métaphores

Les comparaisons et les métaphores sont un moyen économique pour décrire et catégoriser. Dans (9), le locuteur a recours à une comparaison pour décrire l'intensité de la douleur, mais cette intensité – subjective – ne peut pas être mesurée avec précision :

- (9) douloureux et dur **comme de la pierre**

Ce procédé sert à indiquer que c'est une douleur très intense, si l'on admet que l'organe en question a perdu sa souplesse et sa flexibilité. Dans (10), le locuteur se compare à son interlocuteur et va jusqu'à établir une équation, mais

cette équation ne peut être qu'approximative parce que les descriptions détaillées dévoilent des éléments de nature différentielle :

(10) moi c'était **comme toi** et ça à marcher du tonnerre de Dieu !

3.1.7. Autres

La dernière catégorie réunit des éléments hétérogènes, dont :

– les adverbes *peut-être* et *probablement* désignant une probabilité ou un doute :

(11) Alors je ne sais pas **peut être** que cela viens de là

– les fins d'énumération (*etc.* et d'autres expressions) :

(12) posture à avoir, à éviter, **patati, patata**

– les marqueurs évidentiels :

(13) **Il me semble** qu'il y a un délais de 24h pour déclarer un accident du travail.

(14) ce n'est que **mon avis**

– les exemplifications

(15) tape neurochirurgien orthopédique ou chirurgien spécailliste du rachis dans les départements que tu as nommés, **par exemple** dans ggogle et tu auras la liste.

– les extrapredications (sur le choix de la formulation)

(16) Le spécialiste ma découvert un processus épineux cassé (enfin **si ça s'appelle comme ça** je ne suis plus sur.)

(17) c'est avec toi **entre autre** que j'en avais parlé sur ce forum.

– les arrondissements (*rounders*, Adler & Asnes 2010), les cas difficiles à catégoriser, etc. :

(18) en effet l'examen scanner ne révèle rien de grave pour une personne **de 40 ans**

(19) un problème de coagulation ou **autre**

Ces éléments sont associés à l'approximation du fait d'une tentative d'intégration des incréments ou des marges d'erreur avec le N ou la prédication indiqués.

3.2. Contextes d'emploi

Les marqueurs d'approximation, d'imprécision ou de subjectivité peuvent aussi se distinguer selon la nature du contexte dans lequel ils sont employés. Deux types de contextes sont proposés dans ce travail : un contexte général et un autre médical, qui déclenche l'intervention dans le forum. Observons l'exemple (20) :

(20) J'ai 15 ans je suis en apprentissage en Pâtisserie et **il y a quelques mois** je m'étais fais mal **au dos**, je me suis bloqué le dos en m'étirant, mais la douleur ne revenait que si je forçais **beaucoup, il y a une semaine et demi** j'ai été au

Lasergame avec **des amis** et pour résumé c'est **une espère de labyrinthe** où l'ont doit se tirer dessus avec **des lasers** ;

Plusieurs marqueurs d'approximation et d'imprécision peuvent y être relevés : localisateurs temporels (*il y a quelques mois, il y a une semaine et demi*), localisateurs spatiaux (*au dos*), intensifieurs (*beaucoup*), indéfinis quantifieurs (*des amis, des lasers*), noms métalinguistiques (*une espèce de labyrinthe*). Cependant, tous ne portent pas sur des problèmes médicaux : la description des activités dans *Lasergame* et la définition de ce jeu proviennent d'un domaine extra-médical. Par conséquent, on peut distinguer entre deux localisateurs temporels : *il y a quelques mois*, employé pour désigner le début des douleurs au dos, sera associé au contexte médical tandis que *il y a une semaine et demi*, indiquant la date où l'auteur du message s'est rendu au *Lasergame*, sera considéré comme une description générale, périphérique, sans rapport de conditionnement avec le problème de santé.

En d'autres termes, toute occurrence d'un approximateur dans un corpus pareil ne sert pas nécessairement à la description du phénomène médical, de ses symptômes ou de son traitement. Plus généralement dit, les approximateurs du corpus ne sont pas toujours activés dans une tentative de catégorisation du phénomène médical par rapport à un parangon (donc une tentative d'inclusion du référent dans une classe prédéfinie), ni, par ailleurs, dans une tentative de comparaison du référent à un prototype ou à une catégorie d'accueil, suivant une ressemblance saillante. Les approximateurs peuvent donc être relatifs aux circonstances locales, temporelles, budgétaires, etc. : *il est temps de faire quelque chose* ne cherche pas à catégoriser ou à nommer, à la différence de *Pas un coussin, à moins d'un mou, quelque chose qui se moule à l'endroit douloureux*.

4. ANNOTATION MANUELLE

La typologie de marqueurs d'approximation issue de l'observation du corpus a permis d'élaborer les conventions d'annotation grâce auxquelles le corpus a été annoté manuellement. Cette annotation a été réalisée sous format XML (les éléments annotés sont entourés de balises), ce qui permet, par la suite, de réaliser leurs analyses quantitative et qualitative.

L'annotation a été effectuée par deux annotateurs humains. La première version de la typologie de marqueurs a été testée à travers l'essai d'annotation réalisé par deux étudiantes du Master TAL⁶. L'accord inter-annotateur calculé en utilisant une mesure Kappa de J. Cohen (1960) a obtenu 0,92, ce qui est considéré comme un accord presque parfait selon l'échelle proposée par J. R. Landis et G. Koch (1977). Cette version a été corrigée par les autrices de l'article et elles ont proposé une seconde version de la typologie de marqueurs (celle présentée

6. Clémence Laurent & Lucie Warckol, M1 Plurital, 2019, Université Paris Nanterre.

supra et qui a servi à l'annotation du corpus). Le nouvel accord inter-annotateur a obtenu une mesure égale à 0,85. La différence entre les deux accords résulte d'un raffinement de la typologie proposée.

L'exemple (21) montre le résultat d'annotation du message présenté dans la section précédente :

- (21) J'ai 15 ans je suis en apprentissage en Pâtisserie et **<loc_temporel type="med">il y a quelques mois</loc_temporel>** je m'étais fais mal **<loc_spatial type="med">au dos</loc_spatial>**, je me suis bloqué le dos en m'étirant, mais la douleur ne revenait que si je forçais **<intens type="med">beaucoup</intens>**, **<loc_temporel type="gen">il y a une semaine et demi</loc_temporel>** j'ai été au Lasergame avec **<ind_quant type="gen">des amis</ind_quant>** et pour résumé c'est **<nom_meta type="gen">une espère de labyrinthe</nom_meta>** où l'ont doit se tirer dessus avec **<ind_quant type="gen">des lasers</ind_quant>**⁷ ;

Les éléments en gras sont les marqueurs d'approximation relevés dans ce message. Les balises qui entourent ces marqueurs indiquent leurs frontières, leurs catégories (*loc_temporel*, *loc_spatial*, *ind_quant*, *intens*, *noms_meta*, *compar*, *autres*) et le contexte de leur emploi (*gen*, *med*).

D'une manière générale, 572 marqueurs d'approximation ont été annotés, ce qui représente 5 % de tous les mots du corpus. Le corpus annoté a servi de modèle pour les analyses quantitative et qualitative réalisées.

5. ANALYSE

5.1. Distribution des marqueurs d'approximation

La première observation de nature quantitative sur les marqueurs d'approximation relevés concerne leur distribution dans le corpus. Les indéfinis quantifieurs représentent la majorité (34.6 %) parmi tous les approximateurs, ce qui n'est pas surprenant vu le rapport de ces marqueurs avec la détermination ou la pronominalisation. Les intensifieurs et les localisateurs temporels occupent une deuxième place dans la distribution, représentant 20 % en moyenne des approximateurs annotés. Les localisateurs spatiaux et les autres marqueurs (*autres*) sont employés avec une fréquence relative tournant autour de 10.5 %. Enfin, les structures comparatives, les métaphores et les noms métalinguistiques sont peu représentatifs dans le corpus et varient entre 0.5 % et 1.6 % des emplois. Cela montre que l'approximation, dans ce type de corpus, peut largement dépasser les tentatives de la catégorisation claire.

7. Cet exemple comprend d'autres marqueurs d'approximation comme, par exemple, *pour résumé*, qui est un marqueur énonciatif. Cette étude n'a sélectionné que quelques stratégies d'approximation et d'imprécision présentes dans le corpus.

5.1.1. Indéfinis quantifieurs

Comme mentionné *supra*, les indéfinis quantifieurs sont les marqueurs d'approximation les plus fréquents. Ils représentent le tiers des marqueurs annotés. *Tous*, quantifieur exhaustif, se distingue comme l'un des approximateurs indéfinis les plus utilisés, apparaissant dans 5.9 % des cas. Deux autres éléments fréquents sont le pronom *quelqu'un* et le quantifieur *un peu de*. Ils représentent environ 3 % des cas d'indéfinis quantifieurs.

Le choix entre deux éléments désignant une quantité amène également à l'approximation :

- (22) <ind_quant>1 ou 2 mm</ind_quant> de renfort à certains endroits modifie l'équilibre du corps

Le fait que l'internaute hésite entre deux quantités ne permet pas de définir la taille du renfort.

Enfin, l'approximation peut être observée dans les messages où les internautes s'adressent directement à leurs lecteurs :

- (23) Bonne journée et joyeuses paques à <ind_quant>tous</ind_quant>

Ce type de message reflète la nature-même de la communication sur les forums, menée dans le cadre d'une « discussion asynchrone par écrit en groupe restreint » (Colin & Mourlhon-Dallies 2004). Le scripteur simule en quelque sorte un contact direct ; par conséquent, il utilise les procédés discursifs permettant de maintenir ce cadre conversationnel. Dans l'exemple (23), il souhaite de bonnes fêtes à un public inconnu.

5.1.2. Localisateurs temporels

Les localisateurs temporels sont fréquents autant que les indéfinis quantifieurs dans le corpus et représentent 19.4 % des approximateurs annotés.

L'adverbe *souvent* représente à lui seul 5.5 % des localisateurs temporels. *Longtemps* et *à bientôt* représentent 2.7 % des cas. Ce dernier, une formule typique pour prendre congé, indique un futur proche mais non déterminé. Les structures les plus fréquentes sont celles introduites par *il y a*, *depuis*, *dans*, *pendant* ; elles représentent respectivement 10 %, 9 %, 8 %, 5.4 % des approximateurs temporels annotés.

Toutes les périodes de temps sont mentionnées dans le corpus. On y retrouve des expressions renvoyant vers le passé (*il y a 10 jours*, *une semaine et demie*, *depuis des mois*), le futur (*dans 80 prochaines années*, *à bientôt*, *dans 10 jours*) ou au temps présent (*aujourd'hui*). Concernant les mesures du temps, ce sont les *mois* et les *jours* qui sont les plus mentionnés. Elles apparaissent respectivement dans 8 % et 6.3 % des cas de localisateurs temporels annotés.

D'autres observations de nature qualitative ont été constatées. L'adverbe *aujourd'hui* peut avoir un emploi ambigu dans le corpus : il peut indiquer un

présent indéterminé, voire une actualité, et donc, dans un cas pareil, il a été annoté, comme en (24), contrairement à l'exemple (25), où l'adverbe désigne un jour bien précis, correspondant au jour d'écriture du message (cette occurrence n'a donc pas été annotée) :

- (24) <loc_temporel>Aujourd'hui</loc_temporel> le scanner ne montre pas de séquelle de cette hernie
(25) j'ai reçu ce résultat **aujourd'hui**

Enfin, les internautes relient parfois deux mentions de temps, comme dans (26). On assiste ainsi à une forme de reformulation où le scripteur remplace un segment (*depuis l'intervention*) par un autre (*en fin octobre*) afin de préciser ses propos. Ce type de reformulations est au centre des études d'I. Eshkol-Taravella et N. Grabar (2018, 2019).

- (26) <loc_temporel>depuis l'intervention en fin octobre </loc_temporel>

5.1.3. Localisateurs spatiaux

Quant aux localisateurs spatiaux, ils représentent 10.1 % des approximateurs annotés. Ils renvoient le plus souvent aux parties du corps relatives à la source du problème médical ou à la douleur.

Les parties du corps les plus mentionnées sont *jambe(s)* et *dos* représentant respectivement 16.4 % et 26.2 % des cas annotés, ce qui est tout à fait pertinent pour un corpus composé de messages portant sur les douleurs du dos (les nerfs issus de la colonne lombaire vont aux jambes). La structure la plus récurrente est le groupe prépositionnel [*dans + déterminant + jambe*] (*dans la jambe, dans une jambe, dans les jambes, dans ma jambe*) représentant 13 % des cas des localisateurs spatiaux annotés.

Les internautes ont recours à différents moyens pour décrire la localisation de leurs douleurs :

- (27) une grosse perte de sensibilité <loc_spatial>du coté gauche de la face au bas de la jambe</loc_spatial>
(28) la sensation d'endormissement<loc_spatial> au niveau des jambes</loc_spatial>

mais malgré tous ces efforts, il est difficile d'en définir la source exacte.

Tous les localisateurs spatiaux n'ont pas été annotés. Ainsi, dans (29)-(30), les internautes indiquent un emplacement précis (qui fait probablement écho à un diagnostic médical et à une terminologie anatomique) :

- (29) au niveau des angles supérieurs de **L3-L4**.
(30) Le neurochir m'a fait 2 infiltrations au niveau de **la naissance du nerf d'arnold**

5.1.4. Intensifieurs

Les intensifieurs représentent 22.3 % des approximateurs annotés. La distribution de quelques adverbess d'intensité représentatifs figure dans le Tableau 1. Ces cas représentent 74.2 % des intensifieurs présents dans le corpus.

Tableau 1 : Distributions de 5 adverbess intensifieurs

Intensifieurs	<i>beaucoup</i>	<i>moins</i>	<i>peu</i>	<i>plus</i>	<i>très</i>
Fréquence relative	5.4 %	1.5 %	22.4 %	29.4 %	15.5 %

On constate que l'adverbe *moins* est très peu présent, contrairement aux autres intensifieurs. Notons que *un peu* et *beaucoup* ont un double emploi. Ils peuvent désigner la quantité indéfinie, comme dans les exemples (31)-(32) :

- (31) Tu sais c'est vrai que l'on l'impression d'être un numéro quand on va chez le neurochir, mais il n'y en a pas <ind_quant>beaucoup</ind_quant>
 (32) j'aimerais bien que l'on éclaire<ind_quant> un peu</ind_quant>.

ou l'intensité comme dans (33)-(34) :

- (33) Un Iplikator peut <intens>beaucoup t'aider</intens> à soulager les douleurs.
 (34) mais avec le long week-end qu'il y a eu je m'étais <intens>un peu embroilée</intens> dans la date.

De même, certains localisateurs temporels peuvent indiquer également l'intensité lorsqu'ils désignent l'itération appliquée à la douleur comme dans (35) :

- (35) <ind_quant>un de mes muscles</ind_quant> du tricep droit s'est mit a faire <ind_quant>des convulsions</ind_quant> <intens>sans arrêt</intens> <intens>plusieurs fois par minutes</intens> et ce <intens>pendant plus de 48 heures</intens>.

Dans ces cas, ils ont été annotés comme intensifieurs.

Rappelons aussi que l'intensité peut être marquée, dans le corpus, à travers les signes de ponctuation, comme en atteste l'exemple (7) *supra*.

5.1.5. Noms métalinguistiques

Il s'avère que les approximateurs les moins fréquents sont les noms métalinguistiques, avec seulement 0.5 % d'apparition dans le corpus. Observons quelques exemples :

- (36) Le canal rachidien a l'air important pour <noms_meta>ce genre de douleurs</noms_meta>?
 (37) j'ai été au Lasergame avec des amis et pour résumé c'est <noms_meta>une espère de labyrinthe</noms_meta> où l'ont doit se tirer dessus avec des lasers
 (38) j'ai <noms_meta>toutes sortes de traitements</noms_meta>: cortisone, AINS, paracétamol, [...] et j'en passe surment

- (39) <noms_meta>une sorte de gel</noms_meta> qui sort du disque lorsque celui-ci est fissuré
- (40) <noms_meta>une sorte de disque artificiel</noms_meta> pour éviter de bloquer les 2 vertèbres
- (41) contrairement au scanner tu es dans <noms_meta>une sorte de "tunnel"</noms_meta>, je suis un peu claustrophobe et c'est vrai que ce n'est pas facile

Il est évident que les internautes ont utilisé les expressions à enclosure pour opérer une identification approximative, et non pas pour procurer une sous-catégorisation scientifique, comme ce serait le cas dans *Seboraya est une sorte de maladie de peau*, que l'on trouverait peut-être dans un corpus de spécialistes répondant aux questions d'internautes non-spécialistes, et dont la lecture privilégiée serait alors sous-catégorisante. Comme le précise l'introduction du présent volume (Capin, Benninger & Vladimirska 2020), le rôle d'approximateurs de ces expressions varie : au-delà du compromis terminologique, définitoire ou référentiel, on peut lier leur activation à une tentative d'atténuation relative à la validité de la désignation ou de l'identification du référent.

5.1.6. Structures comparatives

Les structures comparatives que nous avons annotées sous la balise **compar** représentent 1.6 % des approximateurs. Elles sont introduites surtout par la préposition *comme* et variantes (*tout comme*, *comme si*) et la construction : [*dans le même* + N] (*dans le même cas*, *dans le même sens*).

5.1.7. «Autres»

Enfin, la fréquence relative de la catégorie «autres» dans le corpus est de 9.7 %. Il s'agit surtout de l'adverbe *peut-être* dont la graphie varie beaucoup dans ce corpus (*peut-être*, *peut être*, *peut etre*, *pr*) et de l'adverbe *probablement*. Leur fréquence relative atteint 33.8 % des cas annotés. Les expressions désignant l'avis du scripteur (*je pense*, *je crois*, *je trouve*, *il me semble*, etc.) font aussi partie de cette catégorie et représentent 30 % des cas. On y retrouve également des remarques et des commentaires comme *entre autre*, *j'en passe surment*, *comme on dit*, etc. Ces cas représentent 7.8 %. Ces chiffres ne comptent pas les commentaires placés entre parenthèses. Enfin, les fins d'énumération marquées par des unités lexicales telles que *etc.*, *patati patata*, *et tout le monde*, mais aussi par la ponctuation, comme dans (42), représentent 12.3 % des occurrences.

- (42) On peut avoir des paresthésies en cas de problèmes au niveau des nerfs (compression, atteinte du nerf, irritation, <autres> . . . </autres>) et vaisseaux (compression).

5.2. Distribution des contextes d'emploi

Comme il a été mentionné dans la section 3.2 *supra*, nous distinguons deux types d'emplois des moyens utilisés pour noter les différents types d'approximation :

selon qu'ils portent sur des éléments appartenant au domaine médical ou sur un contexte de nature générale.

L'analyse quantitative montre que leur distribution dans le corpus est quasi équivalente et tourne autour de 50 % (le contexte médical représente 54 %, le contexte général 46 %). Les internautes semblent donc ne pas se concentrer uniquement sur le domaine médical : ils utilisent l'approximation et le flou pour décrire également des circonstances périphériques, pour soulever d'autres thématiques ou pour accomplir diverses fonctions discursives, comme celle d'interpeller l'interlocuteur, par exemple.

L'analyse de ces deux contextes d'emploi selon le type d'approximation est présentée dans la Figure 1. On constate que les internautes utilisent surtout les localisateurs spatiaux, les localisateurs temporels et les intensifieurs pour décrire leurs problèmes médicaux. La présence des noms métalinguistiques est surtout réservée au contexte médical.

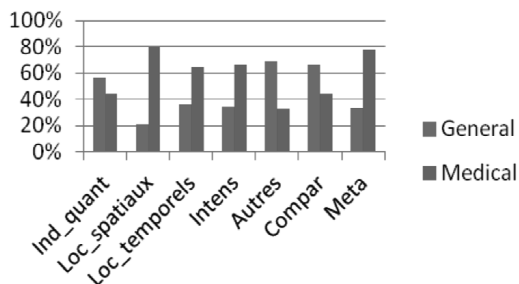


Figure 1 : Distribution de contextes d'emploi selon le type d'approximateur (à gauche *general*, à droite *medical*)

Le domaine général est plutôt réservé aux approximateurs appartenant à la catégorie <autres>. Il s'agit surtout d'éléments relevant du discours conversationnel. Les internautes s'adressent aux destinataires et communiquent avec eux : *Bonjour à tous, à bientôt*, etc. Les avis sont aussi très présents dans le corpus : *je pense, je crois, il me semble que*, etc. Les structures comparatives tendent à être employées dans un contexte général, mais leur faible fréquence ne permet pas de l'affirmer avec certitude. Enfin, les indéfinis quantifieurs, très fréquents dans le corpus, sont employés d'une manière quasi équivalente dans les deux domaines : 56 % dans le domaine général et 44 % dans le domaine médical.

5.3. Contextes d'approximation

Le corpus étudié a permis également d'observer les contextes propices à l'approximation. En premier lieu, il s'agit des cas où le scripteur cherche une alternative ou hésite entre deux possibilités, comme dans les exemples suivants :

- (43) <loc_temporel type="med">3 ou 4 minutes</loc_temporel> penché et je me sens <intens type="med">tout bonnement</intens> barré dans cette position.
- (44) <ind_quant type="med">1 ou 2 mm</ind_quant> de renfort à <loc_spatial type="med">certaines endroits </loc_spatial>modifie l'équilibre du corps
- (45) si ce n'est "qu'une pointe" <ind_quant type="med">1 ou 2 séances</ind_quant> d'ostéopathie

En deuxième lieu, l'approximation s'observe dans la numération non finalisée, comme dans (12) et (42) (cf. p. ex. l'emploi de *etc.*, de *patati patata* ou de la ponctuation en trois points).

Les commentaires et les avis constituent aussi une stratégie d'approximation et révèlent des degrés de prise en charge des propos émis par le scripteur.

Ces contextes présentent un dénominateur commun : tous dévoilent que les scripteurs incluent dans la dénotation non seulement ce à quoi ils se réfèrent mais aussi les marges d'erreur qui puissent refléter une adéquation maximale et permettre aux destinataires d'envisager un <matching> au contenu affiché.

6. CONCLUSION ET PERSPECTIVES

Cet article présente une étude des marqueurs de catégorisation par approximation, des marqueurs de catégorisation subjective et des marqueurs du flou, relevés et annotés dans un corpus portant sur les douleurs au dos extrait d'un forum de santé. L'étude se concentre sur sept catégories observées dans le corpus désignées par les étiquettes <indéfinis quantifieurs>, <localisateurs temporels>, <localisateurs spatiaux>, <intensifieurs>, <noms métalinguistiques>, <comparaison>, <autres>.

Les indéfinis quantifieurs représentent la majorité des marqueurs annotés (34.6 %) et sont employés d'une manière quasi équivalente dans les domaines général (56 %) et médical (44 %). Les intensifieurs et les localisateurs temporels obtiennent une fréquence relative de 20 % en moyenne. Les localisateurs temporels peuvent indiquer aussi l'intensité lorsqu'ils désignent l'itération ou une longue durée appliquée à la <douleur>. Ces cas sont représentatifs du corpus étudié. Les localisateurs spatiaux et les marqueurs de la catégorie <autres> sont employés avec une fréquence relative de 10.5 %. Les autres marqueurs sont peu présents dans le corpus.

Le corpus révèle que les internautes ne cherchent pas seulement des solutions à leurs problèmes médicaux : mises à part les formules de salutation en début et fin de message, ils demandent des nouvelles de leurs destinataires, ils émettent également leurs avis ou ajoutent des commentaires. Ce sont surtout les localisateurs spatiaux, les localisateurs temporels et les intensifieurs qui sont utilisés pour décrire les problèmes médicaux. En fait, cela suggère que les internautes ne font pas qu'utiliser des structures spécifiques que la langue a développées

pour marquer explicitement l'approximation. Ils se servent aussi de stratégies discursives et pragmatiques qui n'ont pas l'approximation – catégorielle ou non – comme vocation première et qui, en contexte, contribuent à un discours médical approximatif.

Diverses perspectives sont envisagées. Les tendances quantitatives observées dans ce corpus d'échanges entre personnes non spécialistes peuvent être comparées avec celles d'un corpus d'échanges entre médecins et patients. Une autre piste prometteuse s'est dégagée pendant l'annotation du corpus : elle concerne l'approximation issue des formes d'exagération. Dans l'exemple annoté *Mais je viens d'apprendre que l'entreprise veut me muter à 250 km de chez moi*, le chiffre peut bel et bien être le résultat d'un arrondissement et même d'une exagération (v. Adler & Asnes 2010 pour des expressions, précises en apparence, qui peuvent générer une interprétation approximative).

Références bibliographiques

- [DOCTISSIMO] *Forum de santé et bien être*. [<https://www.doctissimo.fr/>]
- ADLER S. & ASNES M. (2010), « Autour de la précision », *L'Information grammaticale* 125, 36-43.
- ADLER S. & ASNES M. (2013), « Qui sème la quantification récolte l'intensification », *Langue française* 177, 9-22.
- ADLER S. & ASNES M. (2014), « Quantification imprécise et quantification floue : essai de précision », dans H. Bat-Zeev Shyldkrot, S. Adler & M. Asnes (éds), *Précis et imprécis : étude sur l'approximation et la précision*, Paris, Honoré Champion, 25-42.
- BARWISE J. & COOPER R. (1981), "Generalized quantifiers and natural language", *Linguistics and Philosophy* 4, 159-219.
- CAPIN D., BENNINGER C. & VLADIMIRSKA E. (2020), « De la catégorisation subjective : procédés et pratiques linguistiques », *Langue française* 207. (ce volume)
- COHEN J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement* 20 (1), 37-46.
- COLIN J.-Y. & MOURLHON-DALLIES F. (2004), « Du courrier des lecteurs aux forums de discussion sur l'internet : retour sur la notion de *genre* », *Les Carnets du Cediscor* 8, 113-140. [en ligne]
- CORBLIN F. (1997), « Les indéfinis : variables et quantificateurs », *Langue française* 116, 8-32.
- ESHKOL-TARAVELLA I. & GRABAR N. (2018), « Reformulations avec et sans marqueurs : étude de trois entretiens de l'oral », dans F. Neveu et alii (éds), *Actes du 6^e Congrès Mondial de Linguistique Française – CMLF 2018* (Mons, Belgique), Les Ulis, EDP Sciences. [en ligne]
- ESHKOL-TARAVELLA I. & GRABAR N. (2019), « Reformulation en tant que procédé multidimensionnel », dans H. Landolsi, C. Norén & M. Svensson (éds), *La reformulation : à la recherche d'une frontière*, Uppsala, Acta Universitatis Upsaliensis, 59-78.
- FLAUX N. & VAN DE VELDE D. (2000), *Les noms en français : esquisse de classement*, Paris/Gap, Ophrys.
- GERHARD-KRAIT F. & VASSILIADOU H. (2014), « Lectures taxinomique et floue appliquées aux noms : quelques réflexions... », *Travaux de linguistique* 69, 57-75.
- GERHARD-KRAIT F. & VASSILIADOU H. (2017), « *Clapotis, murmures* et autres manifestations sonores : les méandres de l'approximation catégorielle », *Syntaxe et Sémantique* 18, 19-43.

De la catégorisation subjective

- GRICE P. (1975), "Logic and conversation", in P. Cole & J. L. Morgan, (eds.), *Syntax and Semantics*, vol. 3: *Speech Acts*, New York, Academic Press, 41-58.
- JAYEZ J. & TOVENA L. M. (2008), "Presque and almost: How argumentation derives from comparative meaning", in O. Bonami & P. Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics 7*, 217-239. [en ligne]
- KLEIBER G. (1987), « Quelques réflexions sur le vague dans les langues naturelles », dans *Études de linguistique générale et de linguistique latine offertes en hommage à Guy Serbat*, Paris, Société pour l'Information Grammaticale, 157-172.
- KLEIBER G. (1999), *Problèmes de sémantique : la polysémie en questions*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- LAKOFF G. (1973), "Hedges: A study in meaning criteria and the logic of fuzzy concepts", *Journal of Philosophical Logic* 2 (4), 458-508.
- LANDIS J. R. & KOCH G. (1977), "The measurement of observer agreement for categorical data", *Biometrics* 33 (1), 159-174.
- LUPU M. (2003), « Concepts vagues et catégorisation », *Cahiers de linguistique française* 25, 291-304.
- MIHATSCH W. (2006), « *Espèce de, genre de, sorte de* : des marqueurs d'approximation entre sémantique et pragmatique », *Revue de Sémantique et Pragmatique* 19-20, 229-248.
- MIHATSCH W. (2009), « L'approximation entre sens et signification : un tour d'horizon », dans D. Verbeke (éd.), *Entre sens et signification. Constitution du sens : points de vue sur l'articulation sémantique-pragmatique*, Paris, L'Harmattan, 100-116.
- MOLINE E. (1996), « *Y'a comme un problème* : un emploi métalinguistique de *comme* ? », *Champs du signe* 6, Toulouse, Presses Universitaires du Midi, 249-277.
- SCHAPIRA C. (2014), « L'identification approximative : *une sorte de, un genre de, une espèce de + N* », dans H. Bat-Zeev Shyldkrot, S. Adler & M. Asnes (éds), *Précis et imprécis : étude sur l'approximation et la précision*, Paris, Honoré Champion, 145-158.
- VANDELOISE C. (1986), *L'espace en français : sémantique des prépositions spatiales*, Paris, Éditions du Seuil.