École Doctorale Informatique, Télécommunications et Eléctronique (Paris)

"How much is 'about' ?" Modélisation computationnelle de l'interprétation cognitive des expressions numériques approximatives

THÈSE

présentée et soutenue publiquement le 19 septembre 2017

pour l'obtention du

Doctorat de l'Université Pierre et Marie Curie

(mention intelligence artificielle)

par

Sébastien Lefort

Composition du jury

Jean Baratgin	Rapporteur
Olivier Pivert	Rapporteur
Jean-François Bonnefon	Examinateur
Jean-Gabriel Ganascia	Examinateur
Marie-Jeanne Lesot	Directrice de thèse
Elisabetta Zibetti	Codirectrice de thèse
Marcin Detyniecki	Codirecteur de thèse
Charles Tijus	Codirecteur de thèse

Laboratoire d'Informatique de Paris 6 - UMR7606

Laboratoire CHArt-LUTIN - EA4004

Résumé

Nos travaux portent sur les Expressions Numériques Approximatives (ENA), définies comme des expressions linguistiques impliquant des valeurs numériques et illustrées par des exemples tels que "*environ 100*".

Nous nous intéressons d'abord à l'interprétation des ENA non contextualisées, définie comme la détermination de la signification de la portée du modificateur numérique *environ*, à la fois dans ses aspects humain et computationnel. Plus spécifiquement, après avoir défini et formalisé les dimensions arithmétiques et cognitive pertinentes permettant de caractériser les ENA, nous avons conduit une étude empirique pour collecter les intervalles correspondant aux plages de valeurs dénotées de 24 ENA. Nous montrons ainsi que les dimensions que nous proposons sont impliquées dans l'interprétation des ENA et qu'elles rendent mieux compte des intervalles collectés que les modèles existant dans la littérature.

Ces résultats nous conduisent, dans un deuxième temps, à proposer un principe général d'interprétation des ENA impliquant des nombres naturels, fondé sur l'hypothèse d'un compromis entre la saillance cognitive des bornes des intervalles et leur distance à la valeur numérique de référence impliquée dans l'ENA considérée. Ce principe général est ensuite instancié dans deux types de modèles computationnels. Le premier est conçu pour estimer l'intervalle de valeurs dénotées par une ENA alors que le second génère un intervalle flou représentant l'imprécision qui lui est associée. Ces modèles sont validés expérimentalement à partir des données collectées et nous montrons qu'ils offrent de meilleures performances que les modèles d'interprétation existants. Nous montrons également l'intérêt du modèle flou d'interprétation des ENA que nous proposons en l'implémentant dans le cadre de moteurs de requêtes flexibles. Enfin, nous proposons une extension des modèles permettant d'interpréter des ENA impliquant des nombres décimaux.

Nous avons également conduit une étude pour examiner l'effet d'un contexte sémantique et les différences entre interprétations explicite et implicite d'une ENA. Nous montrons que, bien que le contexte et le type d'interprétation considérés aient peu d'effet sur les intervalles collectés, les dimensions arithmétiques et cognitive que nous proposons sont bien impliquées dans l'interprétation des ENA.

Nous nous intéressons ensuite à la combinaison d'ENA dans le calcul arithmétique imprécis, par exemple dans l'évaluation de la surface d'une pièce mesurant "*environ 10 mètres*" et "*environ 20 mètres*" de côtés. Plus précisément, nous considérons deux opérations arithmétiques, l'addition et la multiplication. Nous avons mené une étude empirique

pour collecter les résultats de calculs imprécis. Nous montrons que les imprécisions liées aux opérandes ne semblent pas prises en compte lors de l'estimation de l'imprécision du résultat ; en particulier, nous montrons que les résultats sont sensibles aux paradoxes sorites et qu'ils ne correspondent pas aux résultats fournis par l'arithmétique des intervalles et l'arithmétique floue.

Mots-clés: Expressions numériques approximatives, Imprécisions, Interprétation cognitive, Approximateurs, Calcul imprécis, Propagation de l'imprécision, Logique floue, Requêtes approximatives

Abstract

Approximate Numerical Expressions (ANE) are imprecise linguistic expressions implying numerical values, illustrated by examples such as "*about 100*" and pervasive in everyday communication between human beings. Our work focuses on two major tasks related to ANE, namely their interpretation and their combination in arithmetical calculation.

We first focus on the interpretation of uncontextualised ANEs, defined as the determination of the signification of the numerical modifier *about*, both in its human and computational aspects. More specifically, after defining and formalising the relevant arithmetical dimensions that allow to characterise ANEs, we conducted an empirical study to collect intervals corresponding to range of values denoted by 24 ANEs. We show that the dimensions we propose are involved in the interpretation of ANEs and that they better account for the collected intervals than existing models.

These results lead us, in a second step, to propose a general principle to interpret ANEs involving natural numbers, based on the assumption of a compromise between the cognitive salience of the endpoints and their distance to the numerical value of the considered ANE. This general principle is then instantiated in two types of computational models. The first one is designed to estimate the interval of values denoted by an ANE. The second one generates a fuzzy interval representing the imprecision it conveys. Both models are experimentally validated using the collected data and we show that they offer better performances than existing models. We also show the relevance of the proposed fuzzy model by implementing it in the framework of flexible queries of numerical databases. Finally, we propose an extension of the models to interpret ANEs involving decimal numbers.

We also conducted an empirical study to examine the effect of a semantic context and the differences between implicit and explicit interpretations of ANEs. We show that, even if the context and the type of interpretation have a weak effect on the collected intervals, the arithmetic and cognitive dimensions we propose are involved in ANEs interpretation.

Finally, we consider the task of ANE combination, in imprecise arithmetic calculations, for instance in the estimation of the area of a room whose are "*about 10 meters*" and "*about 20 meters*" long. More specifically, we consider two arithmetic operations, additions and products. We conducted an empirical study to collect the results of imprecise calculations. We show that the imprecision associated with the operands does not seem to be taken into account during the estimation of the final imprecision. We show that the results of the calculations are sensitive to the sorites paradox and that they do not correspond to the operations performed in the formal frameworks of interval arithmetic nor of fuzzy arithmetic.

Keywords: Approximate numerical expressions, Imprecisions, Cognitive interpretation, Approximators, Imprecise calculation, Imprecision propagation, Fuzzy logic, Approximative queries

Les travaux présentés dans cette thèse ont fait l'objet des publications suivantes :

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C. & Detyniecki, M. (2016). How much is "about" ? Fuzzy interpretation of approximate numerical expressions. *Proc. of the 16th Int Conf on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'16)* (pp. 226-237). Eindhoven : Springer.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C. & Detyniecki, M. (2016). Interprétation floue des expressions numériques approximatives. *Proc. of Rencontres Francophones sur la Logique Floue et ses Applications (LFA'16)*. La Rochelle, France.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C. & Detyniecki, M. (2017). Interpretation of approximate numerical expressions : Computational model and empirical study. *International Journal of Approximate Reasoning, 82*, 193-209.

Lefort, S., Zibetti, E., Lesot, M.-J., Detyniecki, M. & Tijus, C. (2017). Dimensions for automatic interpretation of approximate numerical expressions : An empirical study. *Proc. of the 22nd Int Conf on Intelligent User Interfaces* (pp. 107-117). Cyprus : ACM.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C. & Detyniecki, M. (2017). How arithmetically fuzzy are we ? An empirical comparison of human imprecise calculation and fuzzy arithmetic. *FUZZ-IEEE - 2017 IEEE Int Conf on Fuzzy Systems*. Naples, Italy.

Cette thèse a été réalisée dans le cadre d'une codirection entre l'Université Pierre et Marie Curie et l'Université Paris 8, au Laboratoire d'Informatique de Paris 6 et au laboratoire CHArt-LUTIN.

Ce travail été effectué dans le cadre du Labex SMART (ANR-11-LABX-65) et a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'Avenir portant la référence ANR-11-IDEX-0004-02.

Sommaire

Chapitre 1

Introduction générale

> On n'écrit qu'à la pointe de son savoir, à cette pointe extrême qui sépare notre savoir et notre ignorance, *et qui fait passer l'un dans l'autre.*
>
> Gilles DELEUZE
> *Différence et Répétition*

Fraîchement débarqué à Paris, l'envie me prend de visiter et de me perdre dans les ruelles labyrinthiques du quartier Saint-Michel. L'heure tourne et cela finit par arriver, les ruelles s'assombrissent et je suis réellement perdu. Sans plan, sans smartphone, je demande à un passant de m'indiquer le chemin menant à la station de métro la plus proche. Il me répond : "*Continuez sur 100 mètres puis tournez à droite et vous apercevrez la bouche de métro*". Je le remercie, nous nous séparons avant que je ne prenne la direction que le passant m'a indiquée.

A Saint-Michel l'organisation des rues relève de l'anarchie. Celle où je marche n'échappe pas à la règle, sur ma droite, une ruelle toutes les dizaines de mètres. Pourtant, je n'ai aucun mal à trouver la rue que je dois prendre. Non, bien sûr, elle ne se trouve pas à exactement 100 mètres du lieu où j'ai interrogé le passant. A vue de nez, 90-95 mètres. La précédente rue est quant à elle à 65-70 mètres, peut-être, et la suivante à 140-150 mètres.

Que s'est-il passé ? Comment ai-je pu trouver mon chemin malgré une description imprécise ? Intuitivement, de manière tout à fait inconsciente, deux connaissances ont été mobilisées. Tout d'abord, je n'ai pas pris au pied de la lettre l'expression *100 mètres*. Je savais que celle-ci n'exprime pas une exactitude mais plutôt une indication, entourée d'un halo d'imprécision. Dans un second temps, j'ai interprété et construit une représentation de cette imprécision. Je n'ai pas tourné dans la rue située à 65-70 mètres car elle n'entrait pas dans cette représentation alors que celle que j'ai prise y entrait, elle.

Cette situation, toute personne ayant un tant soit peu voyagé en a fait l'expérience.

D'autres exemples similaires peuvent être évoqués. Lorsque je prétends que j'arriverai à 18h, personne ne s'attend à ce que j'arrive à exactement 18h, mais plutôt entre 17h50 et 18h10 par exemple. Ou encore, si je me fixe un budget de 100 euros, j'accepterais certainement toute marchandise répondant à mes attentes et moins chères. Mais j'accepterais également certains objets coûtant un peu plus cher.

Dans l'exemple introductif de la visite du quartier Saint-Michel, un robot mis à la place du narrateur et dépourvu de tout algorithme traitant les imprécisions aurait cherché une rue à exactement 100 mètres et n'aurait pas considéré la bonne rue comme pertinente. C'est que le robot, en tant que système informatique, est conçu pour traiter des informations précises.

Le passant informant le robot aurait pu rendre explicite la nature imprécise de l'expression 100 mètres en lui adjoignant un approximateur comme "*environ*" ou "*approximativement*" : "*Continuez sur* ***environ 100 mètres*** *puis tournez à droite et vous apercevrez la bouche de métro*". Ainsi, il aurait informé le robot que l'expression ne devrait pas être considérée comme exacte. Toutefois, même si le système sait que l'information est imprécise, la question de la détermination de la plage de distances acceptables pour "*environ 100 mètres*" demeure problématique.

Ces exemples introduisent la notion d'Expression Numérique Approximative (ENA), de la forme "*environ* x", où x est un nombre, et la problématique, centrale dans les travaux que nous présentons dans cette thèse, liée à leur interprétation, à la fois dans ses aspects humain et informatique.

Les Expressions Numériques Approximatives

Dans cette thèse, nous nous intéressons à des expressions vagues, impliquant des valeurs numériques : les Expressions Numériques Approximatives (ENA). Nous définissons celles-ci comme des expressions linguistiques imprécises dont la forme générale est "*environ* x", où x est un nombre. Elles sont couramment utilisées dans la langue naturelle pour désigner une plage imprécise de valeurs, par exemple "*environ 100 mètres*". Plus précisément, nos travaux portent principalement sur des ENA impliquant des nombres entiers, i.e., $x \in \mathbb{N}$.

Trois champs : linguistique, informatique et psychologie cognitive L'étude des expressions numériques approximatives implique trois champs disciplinaires : la linguistique, l'informatique et la psychologie cognitive.

En linguistique, les ENA sont considérées comme faisant partie du champ plus vaste des expressions vagues (Solt, 2014). Elles sont abordées différemment selon la manière considérée de rendre compte de l'aspect vague du langage naturel et sont traitées sous

l'angle de leur interprétation dans un cadre purement théorique (Krifka, 2007; Sauerland & Stateva, 2011; Solt, 2014) : il s'agit de déterminer la plage de valeurs qu'elles dénotent.

En informatique, et plus particulièrement dans le champ de la représentation des connaissances, les ENA sont considérées comme faisant partie du domaine des imprécisions, par opposition aux incertitudes, qui sont deux formes que peut prendre une connaissance imparfaite (Smets, 1991). L'incertitude porte sur le statut d'une proposition, le locuteur ne peut attester de sa véracité, par exemple, "*Il va peut-être pleuvoir demain*". L'imprécision porte sur le contenu de la proposition : le locuteur est sûr quant à sa véracité mais le contenu est imprécis, par exemple "*Il est né à Paris ou à Lyon*" ou "*La distance entre Paris et Berlin est approximativement de 900km*". Des formalismes ont été proposés pour représenter et combiner les imprécisions, en particulier les intervalles (Moore, 1966) et la logique floue (Zadeh, 1965), qui sont mis en œuvre dans cette thèse. En revanche, à notre connaissance, ces formalismes ne spécifient pas comment quantifier les imprécisions. Nous définissons la quantification de l'imprécision comme la détermination des bornes de la plage de valeurs dénotées dans la représentation par intervalles et des fonctions d'appartenance dans la représentation par sous-ensembles flous.

En psychologie cognitive, la majorité des travaux portant sur les imprécisions numériques s'intéressent à des quantités représentées non-symboliquement, par exemple des points dessinés dans un tableau (voir par ex., Gelman & Gallistel, 2004; Dehaene, 2003; DeWind et al., 2015). En revanche, les imprécisions exprimées symboliquement, qui incluent les ENA, sont à ce jour très peu explorées et font en partie l'objet de cette thèse.

Approche considérée : psychologie cognitive et informatique L'approche que nous défendons dans cette thèse est la croisée des chemins entre psychologie cognitive et informatique. Elle consiste, dans un premier temps, à comprendre et formaliser, à l'aide d'études empiriques, la manière dont l'être humain traite les ENA. Dans un second temps, nous proposons de modéliser computationnellement ces processus. Enfin, dans un dernier temps, nous cherchons à valider expérimentalement les modèles computationnels à partir de données réelles. Cette approche nous permet de proposer des modèles cohérents à la fois avec la manière dont les individus se représentent les ENA et avec les formalismes existants.

Au delà de l'aspect méthodologique global, nous avons à cœur d'intégrer ces deux champs à chacune de nos propositions : s'appuyer sur des modèles formels pour représenter et analyser des données empiriquement collectées et formaliser des notions relevant de la cognition dans la modélisation computationnelle. Ainsi, nous cherchons à instiller du raisonnement humain dans les systèmes intelligents de manière à ce que leur comportement s'adapte intelligemment à ce qui est attendu par les utilisateurs.

Contextualisation des expressions numériques approximatives Nos travaux se veulent comme une approche systématique et générique de la quantification des imprécisions. Dans la mesure où, à notre connaissance, le traitement par le système cognitif humain des imprécisions exprimées verbalement a été très peu étudié jusqu'à présent, nous nous sommes fixé comme objectif de mettre en lumière les facteurs qui sont impliqués quel que soit le contexte envisagé. Mettre en œuvre un ou plusieurs contextes particuliers rendrait les résultats obtenus spécifiques à ceux-ci (Carraher et al., 1985; Lasersohn, 1999; Sadock, 1977; Williamson, 1994) et limiterait, selon nous, leur généralisation. Par conséquent, nous avons fait le choix de ne pas contextualiser les expressions numériques approximatives que nous utilisons, ni sémantiquement, ni pragmatiquement. Les modèles que nous proposons, ainsi que la plupart des résultats que nous reportons, sont donc indépendants du domaine sémantique considéré et peuvent donc être implémentés dans un large éventail d'applications impliquant le traitement d'ENA.

Deux types de traitement des ENA, en partie illustrés par les exemples introductifs, sont abordés dans cette thèse : leur interprétation, c'est-à-dire la détermination de l'imprécision qu'elles communiquent, et leur combinaison arithmétique dans le calcul. Les problématiques que ces deux traitements soulèvent et que nous proposons d'explorer sont successivement détaillées dans les prochaines sections.

Interprétation des expressions numériques approximatives

Une importante partie de nos travaux traite de l'interprétation automatique des expressions numériques approximatives dans une situation de communication en langue naturelle allant de l'utilisateur vers le système. Nous définissons l'interprétation d'une expression numérique approximative comme la détermination de la signification de la portée du modificateur numérique *environ*.

Pouvoir interpréter automatiquement les expressions numériques approximatives d'une manière qui soit cohérente avec la représentation cognitive que s'en font les êtres humains est une problématique importante à l'heure où les systèmes intelligents dont le mode d'interaction repose sur la langue naturelle se disséminent dans notre environnement quotidien. Les domaines d'application incluent, entre autres, les systèmes d'information géographique (Bennett, 2010; Delboni et al., 2007), par exemple, pour rechercher une zone dont la surface est d'*environ 1000m*^{2} ; les systèmes experts, en particulier médicaux (Straszecka, 2006), par exemple, pour représenter l'information d'un patient affirmant qu'il se sent fiévreux depuis *à peu près une semaine* ; ou encore les assistants personnels tels que Siri, Cortana ou Alexa, par exemple, pour chercher un hôtel à *environ 30km*. Dans chaque cas,

et comme pour l'exemple introductif du robot, il s'agit de déterminer la plage de valeurs que dénote l'ENA.

L'approche que nous mettons en œuvre pour étudier l'interprétation des ENA s'inscrit dans celle, générale que nous défendons, à l'intersection entre psychologie cognitive et informatique. Nous proposons, dans un premier temps, de définir formellement les caractéristiques arithmétiques et cognitive qui peuvent rendre compte de l'interprétation d'une ENA. Dans un second temps, nous conduisons une étude empirique pour valider l'implication de ces caractéristique dans l'interprétation d'une ENA. Enfin, nous proposons des modèles computationnels, expérimentalement validés, de l'interprétation des ENA.

Combinaison des expressions numériques approximatives

Au delà de l'interprétation des expressions numériques approximatives, nous nous intéressons à leur combinaison arithmétique dans le calcul, par exemple, pour estimer la surface d'une pièce dont les côtés mesurent "*environ 10*" en largeur et "*environ 20*" mètres en longueur, ou l'heure à laquelle peut finir un après-midi avec trois réunions durant chacune "*environ 1 heure*".

Nous étudions la manière dont l'être humain réalise des calculs impliquant des opérandes imprécis et le comparons aux formalismes mathématiques. Comme pour les autres domaines du raisonnement humain (voir, par ex. Kahneman, 2003a), des biais peuvent apparaître pendant la combinaison de données numériques imprécises. Par conséquent, examiner la manière dont l'être humain traite et propage les imprécisions lors de telles combinaisons est important pour identifier et prévenir ces biais.

L'incertitude a fait l'objet d'un nombre considérable d'études (voir, par ex., Samson et al., 2009), dans des domaines appliqués impliquant des processus de prise de décision, comme l'économie (Kahneman, 2003a), la gestion de projet (Loch et al., 2011), la santé (Hunink et al., 2014) et dans des domaines plus théoriques, comme la linguistique (Teigen, 1988) ou la psychologie cognitive (Kahneman, 2003b; Cosmides & Tooby, 1996; Politzer & Baratgin, 2016). Plus spécifiquement, à la suite des travaux féconds de Tversky et Kahneman (1974), de nombreuses études ont mis en évidence des biais dans le raisonnement humain sur des données ou dans des conditions incertaines (Kahneman, 2003a; Cosmides & Tooby, 1996).

En revanche, la propagation des imprécisions dans le raisonnement et dans le calcul impliquant des expressions numériques approximatives chez l'être humain n'a, à notre connaissance, pas été traitée. Or, il se peut que, à l'instar du traitement des incertitudes, le traitement des imprécisions par le système cognitif humain présente de nombreux biais.

Nous proposons dans cette thèse une étude qui se veut être une première approche de la

propagation des imprécisions dans le calcul imprécis, dans laquelle nous mettons en œuvre une méthodologie originale d'analyse des données.

Structure de la thèse

La thèse est structurée selon les deux problématiques traitées. Le chapitre 2 est dédié à une présentation générale du contexte de nos travaux. Nous y traitons des concepts et expressions vagues d'un point de vue linguistique et plus particulièrement la façon dont les imprécisions sont exprimées dans le langage naturel. Dans ce chapitre, nous présentons également la manière dont le système cognitif humain traite et représente les nombres et quantités, imprécises ou non.

Le chapitre 3 décrit l'état de l'art portant sur la représentation formelle, l'interprétation d'un point de vue computationnel, et le calcul des expressions numériques approximatives.

Dans le chapitre 4, nous proposons une formalisation des dimensions arithmétiques et cognitive caractérisant les ENA. Nous détaillons ensuite l'étude empirique que nous avons conduite pour montrer l'implication de ces dimensions dans l'interprétation des expressions numériques approximatives.

Nos propositions de modèles computationnels d'interprétation, conçus à partir des résultats de l'étude empirique, sont décrites dans le chapitre 5. Nous y exposons d'abord le principe général que nous proposons. Nous instancions ensuite ce principe dans deux types de modèles computationnels d'interprétation des ENA : le premier consiste à estimer l'intervalle de valeurs dénotées alors que le second produit un sous-ensemble flou représentant l'imprécision exprimée par une ENA. Enfin, pour chacun des modèles, nous réalisons une validation expérimentale basée sur des données collectées empiriquement.

Le chapitre 6 est dédié à l'implémentation du modèle d'interprétation produisant des intervalles flous dans un système de requêtes flexibles dans des bases de données, ainsi qu'à une extension des modèles que nous proposons, qui permet d'interpréter des ENA impliquant des nombres décimaux.

Le chapitre 7 présente la seconde étude empirique de l'interprétation des expressions numériques approximatives que nous avons conduite pour étendre et généraliser les résultats de la première étude, selon deux axes : étudier l'effet de la présence d'un contexte sémantique et comparer l'interprétation explicite d'une ENA avec son interprétation implicite.

Le chapitre 8 est dédié à l'étude empirique de la propagation des imprécisions lors de calculs imprécis, en particulier l'addition et la multiplication, réalisés par des être humains.

Enfin, le chapitre 9 présente les conclusions et les perspectives ouvertes par cette thèse.

Chapitre 2

Contexte

Dans ce chapitre, nous présentons le contexte théorique dans lequel nos travaux s'inscrivent. Les Expressions Numériques Approximatives du type "*environ x*", où x est un entier, sont des expressions vagues impliquant des valeurs numériques, elles se rattachent à deux champs théoriques : les expressions linguistiques vagues et la représentation des nombres et des quantités dans le système cognitif humain.

La section 2.1 introduit le champ des expressions vagues, d'un point de vue linguistique d'abord et d'un point de vue informatique ensuite, dans leurs aspects général puis spécifiquement appliqué aux expressions numériques. La manière dont sont encodés et représentés les nombres dans le système cognitif humain est présentée dans la section 2.2.

2.1 Aspects du vague dans le langage naturel

Dans une certaine mesure, tous les termes, toutes les expressions de la langue naturelle sont vagues (Channell, 1994; Egré & Klinedist, 2011). Cela paraît évident pour certains termes, comme *jeune*, *grand*, *rouge* ou certaines expressions, telles que "*j'arrive souvent en retard*" ou "*il va sûrement pleuvoir demain*". Cela est moins évident pour d'autres termes, comme on peut l'illustrer pour celui de chaise : à première vue, il est bien défini : il s'agit d'un meuble sur lequel on peut s'asseoir, composé de quatre pieds, d'un fessier et d'un dossier. Il paraît donc aisé de décider si, oui ou non, un objet donné est une chaise. Néanmoins, des designers inventifs peuvent créer des fauteuils qui tendent à être des chaises, ou vice-versa, des chaises qui tendent vers des fauteuils. Dans certains cas extrêmes, décider si la création est une chaise ou un fauteuil constitue un jugement subjectif, qui ne fait pas l'unanimité : pour certains, il s'agit d'un fauteuil, pour d'autres, d'une chaise. Un concept apparemment bien défini, comme celui de chaise, peut donc se révéler vague dans certains contextes, quoique dans une moindre mesure que d'autres, *jeune* par exemple.

L'aspect vague du langage naturel concerne différents domaines théoriques, en particu-

lier la linguistique et l'informatique, qui peuvent traiter différemment cet aspect. Ainsi, en informatique, on distingue classiquement les imprécisions des incertitudes (Smets, 1997), sans que ces deux notions soient incluses dans celle, plus large, de vague. Au contraire, en linguistique, les incertitudes constituent l'une des formes que peut prendre une expression vague (Jucker et al., 2003). Dans cette section, nous présentons les éléments théoriques relatifs au vague dans ces deux domaines. Dans un premier temps, nous nous intéressons aux concepts vagues et leur définition philosophique et linguistique. Ensuite, au-delà des concepts, les différentes formes que peuvent prendre les expressions linguistiques vagues sont exposées. La section 2.1.3 est focalisée sur les imprécisions numériques comme cas particuliers du vague. Enfin, la section 2.1.4 présente les formalismes mathématiques classiquement utilisés pour représenter les concepts et expressions vagues, en particulier la théorie des sous-ensembles flous (Zadeh, 1965).

2.1.1 Concepts vagues

D'un point de vue philosophique et linguistique, un concept est dit vague lorsque son extension est difficile à spécifier intégralement (Williamson, 1994). Dans cette section, nous détaillons en premier lieu les propriétés d'un concept vague. Dans un deuxième temps, nous présentons les deux perspectives, épistémique et sémantique respectivement, proposées dans la littérature pour rendre compte du vague.

2.1.1.1 Propriétés des concepts vagues

Un concept est considéré comme vague lorsqu'il présente trois caractéristiques liées entre elles (Keefe, 2000; Kennedy, 2007; Smith, 2008) : (i) l'existence de cas limites, (ii) l'absence de frontières nettes à son extension et (iii) la susceptibilité aux paradoxes sorites. Nous détaillons tour à tour dans les prochains paragraphes chacune de ces trois caractéristiques.

L'existence de cas limites Les cas limites sont ceux pour lesquels on ne peut pas dire si le concept s'applique clairement ou ne s'applique clairement pas, comme la chaise-fauteuil créée par le designer inventif de l'exemple introductif ou l'application du concept de jeunesse à une personne de 35 ans.

Ces cas limites posent un problème vis-à-vis de la logique classique. En effet, le principe de non contradiction pose que si p est une proposition logique, p et $\neg p$ ne peuvent être vraies en même temps. Or, dire que la création du designer inventif est à la fois un fauteuil et une chaise peut être interprété comme le fait qu'il s'agit à la fois d'une chaise et d'une non chaise.

L'absence de frontières nettes Les cas limites mettent également en évidence l'absence de frontières nettes à l'extension du concept considéré. Cette caractéristique est particulièrement évidente dans le cas d'adjectifs graduels. Considérons l'exemple de *grand* appliqué à un adulte. Il ne fait aucun doute qu'une personne mesurant deux mètres peut être qualifiée de *grande*, de même que quelqu'un mesurant 1m85. Il ne fait pas non plus de doute qu'une taille de 1m60 ne correspond pas à *grand*. En revanche, il existe une plage de tailles limites pour lesquelles l'application de *grand* est moins évidente et systématique. Soyons plus spécifiques et considérons que l'on parle d'un homme. Peut-on considérer qu'un homme dont la taille est 1m75 est *grand* ? Et qu'en est-il pour 1m70 ? Notons, et nous reviendrons sur ce point, que le contexte joue ici un grand rôle puisque dans le cas où l'on considère les femmes, *grand* serait encore applicable pour une personne mesurant 1m75. De même son applicabilité peut être différente selon la région du monde où l'on se place. Quoi qu'il en soit, on voit qu'outre le contexte dans lequel est employé le terme *grand*, l'extension de son application n'est pas claire et nette.

Les cas limites du paragraphe précédent sont ceux situés à proximité de la frontière imprécise d'un concept. Il faut souligner que celle-ci pose également problème à la logique classique. En effet, considérons cette fois *non grand* pour un homme adulte. Il est évident que *non grand* est applicable à une personne mesurant 1m60 mais pas à une personne dont la taille est 1m90. Et, comme pour *grand*, il existe une plage de valeurs pour lesquelles l'application de *non grand* n'est pas évidente. Il se peut donc, pour cette plage de valeurs, que *grand* et *non grand* soient applicables tous deux ou qu'aucun des deux ne soit applicable, ce qui viole encore une fois le principe de non contradiction.

On pourrait penser qu'en pratique, ce n'est pas le cas et que lorsqu'on interroge quelqu'un sur le fait qu'un homme adulte mesurant 1m75 est grand ou non, il soit cohérent et ne considère pas ces deux options comme simultanément vraies. Or, des études (Alxatib & Pelletier, 2011; Bonini et al., 1999; Serchuk et al., 2011) ont montré qu'une telle cohérence n'est pas systématique. En particulier, Bonini et al. (1999) ont demandé à des participants à partir de quelle taille quelqu'un peut être considéré comme grand. Ils ont également demandé, aux mêmes participants, en dessous de quelle taille quelqu'un ne peut pas être considéré comme grand. Les résultats obtenus montrent qu'il existe bien une plage de valeurs dans laquelle un individu n'est considéré ni comme *grand* ni comme *non grand*. Par conséquent, les participants de cette étude ne semblent pas respecter le principe du tiers exclu, équivalent en creux du principe de non contradiction, dans l'attribution de *grand*.

Susceptibilité aux paradoxes sorites La troisième propriété est la plus citée dans la littérature comme caractérisant les concepts vagues (Alxatib & Pelletier, 2011; Douven

et al., 2013; Egré & Klinedist, 2011; Egré, 2013; Fisher, 2000; Keefe, 2000; Williamson, 1994).

Ce paradoxe, qui a été formulé par Eubulide au IV$^{\text{ème}}$ siècle avant notre ère, décrit un raisonnement par lequel on conclut à l'impossibilité d'un grand changement par accumulation de petits changements. Classiquement, il prend la forme du paradoxe du tas (*soros*, en grec) :

1. Un grain de sable ne constitue pas un tas de sable
2. L'ajout d'un unique grain ne modifie pas l'ensemble, qui ne constitue toujours pas un tas
3. L'ajour d'un grain aux deux précédents ne provoque pas de changement et on n'obtient toujours pas de tas
4. etc.

Par généralisation, on obtient l'argument suivant : si n grains de sable ne forment pas un tas de sable, alors, comme l'ajout d'un unique, minuscule grain ne change pas la nature de l'ensemble, $n+1$ grains ne forment pas non plus de tas. Par conséquent, les tas de sable n'existent pas.

La prémisse de ce raisonnement est également appelée *principe de tolérance*, qui, pour certains auteurs (Dummett, 1975; Wright, 1976) est, seul, caractéristique d'un concept vague. Sans aller jusqu'à la formation d'un paradoxe, ce principe peut prendre la forme simple suivante : un prédicat P, par exemple *n'est pas un tas de sable* dans le cas du paradoxe du tas, est dit tolérant par rapport à Φ, *l'ensemble de n grains de sable*, s'il existe un changement positif possible mais insuffisant dans Φ, *l'ajout d'un grain de sable*, pour que l'application de P ne change pas (Wright, 1976). L'affranchissement d'une lettre constitue une autre illustration du principe de tolérance : il est courant de considérer que l'ajout d'un timbre à une lettre ne modifie pas assez le poids total pour devoir changer le montant de l'affranchissement de celle-ci.

Le principe de tolérance peut prendre deux formes. La première est subjective et intra-individuelle : une même personne *tolère* une faible variation sans remettre en cause l'applicabilité du prédicat ou du concept. La seconde forme, objective et inter-individuelle, est une conséquence de la première : puisqu'une même personne est *tolérante*, son jugement peut varier au cours du temps. Par conséquent, plusieurs personnes interrogées peuvent avoir des jugement différents quant à l'applicabilité du prédicat ou du concept.

Ce paradoxe met en évidence le problème qu'il peut y avoir, pour les concepts vagues, à articuler deux niveaux de granularité (Egré, 2013), le premier à l'échelle locale et incrémentale (chaque grain de sable pris individuellement), le second à l'échelle globale et catégorielle (la catégorie du tas de sable).

2.1.1.2 Deux perspectives du vague

Deux points de vue, successivement présentés dans les deux paragraphes suivants, ont été proposés en linguistique et en philosophie pour rendre compte des paradoxes sorites : le point de vue épistémique et le point de vue sémantique (Alxatib & Pelletier, 2011; Egré & Klinedist, 2011; Egré, 2013; Fisher, 2000; Sauerland & Stateva, 2011). Ils s'articulent autour du principe de non contradiction et correspondent, d'un point de vue formel, à la différence entre incertitude et imprécision, telle que discutée dans la théorie des données imparfaites et évoquée dans la section 2.1.4.

Point de vue épistémique Du point de vue épistémique du vague (Sainsbury, 1989; Sorensen, 2001a; Williamson, 1994), les concepts possèdent une extension clairement limitée et des frontières nettes, mais celles-ci ne sont pas connues [1]. Ainsi, il existe bien une valeur n seuil à partir de laquelle on passe de *non tas* à *tas*, mais celle-ci est inconnue.

Dans cette perspective, l'aspect vague d'un concept a son origine dans un défaut de connaissance (Williamson, 1994). Pour un individu, à un moment donné, il existe une conception claire de ce qui est un tas et de ce qui n'en est pas. Toutefois, d'un individu à un autre et d'un moment à un autre, le seuil de passage entre *non tas* à *tas* peut varier.

Ce point de vue entraîne deux conséquences principales. D'abord, le principe de tolérance est considéré comme faux (Sorensen, 2001b). En effet, il existe au moins un n, le seuil entre *non tas* et *tas*, tel que ce principe n'est pas applicable. Deuxièmement, le problème lié à la modélisation des paradoxes sorites ne porte plus sur la logique classique, puisque le principe de non contradiction est préservé, mais sur la manière de modéliser le manque de connaissance à propos du seuil de passage entre p et $\neg p$.

Point de vue sémantique Le point de vue sémantique du vague (Keefe, 2000; Lewis, 1986) repose sur une notion de gradualité et de transition progressive : il réfute l'idée d'un unique seuil servant de pivot entre p et $\neg p$ mais permet un passage plus doux entre les deux.

L'absence de transition nette peut être formalisée par deux approches : le supervaluationisme, proposé par Van Fraassen (1966) et appliqué au vague par Fine (1975), Kamp (1975) et Varzi (2007) ; et la théorie des sous-ensembles flous, proposée par Zadeh (1965) (Lakoff, 1973; Zadeh, 1975).

Dans cette perspective, le principe de non-contradiction n'est pas maintenu : p et $\neg p$ peuvent, dans une certaine mesure, être simultanément vrais. En effet, la véracité est pondérée : p s'applique à un certain degré, $\neg p$ aussi.

1. Ce lien avec la connaissance justifie le qualificatif *épistémique* associé à cette perspective.

Dans le cas des paradoxes sorites, on peut dire que $n + 1$ grains forment *un peu plus* un tas, même s'ils ne forment toujours pas clairement un tas, que n grains. Cette approche permet par conséquent de passer graduellement de *non tas* à *tas*, préservant ainsi le principe de tolérance. Contrairement au point de vue épistémique donc, le point de vue sémantique considère qu'il n'existe pas de n tel que n grains ne forment absolument pas un tas mais que $n + 1$ en forment un absolument. En revanche, quel que soit le nombre n de grains, ils forment toujours, à un certain degré, même minime, un tas (Lakoff, 1973).

2.1.2 Expressions linguistiques vagues

La dimension vague du langage se retrouve non seulement au niveau des concepts et prédicats mais également au niveau des propositions. La définition de l'aspect vague d'une proposition est moins consensuelle : alors que certains auteurs distinguent clairement le vague, en particulier l'imprécision, de l'incertitude (Bennett, 2010; Smets, 1991; Zadeh, 1975), d'autres considèrent que l'incertitude est l'une des formes que peut prendre le vague (Jucker et al., 2003).

Vague et incertitude On peut considérer que le point de vue à adopter dépend de l'application du critère de véracité vis-à-vis d'une proposition. En effet, dans l'incertitude ce critère s'applique au statut de la proposition : le locuteur n'est pas en mesure d'attester de la véracité de la proposition. Celle-ci ne peut donc être déclarée absolument vraie ou absolument fausse du point de vue du locuteur (par ex., "*je crois que son nom est Jean*").

Dans le vague, distingué de l'incertitude, le critère de véracité est au contraire appliqué au contenu (Bennett, 2010; Smets, 1991; Smets, 1997). En effet, le locuteur exprimant une proposition vague est sûr quant à sa véracité, c'est-à-dire qu'elle est considérée comme vraie même si elle est ambiguë (par ex., "*son prénom est Jean ou Paul*") ou si l'information transmise n'est pas assez précise (par ex., "*Berlin se trouve à environ 900km de Paris*"). De ce point de vue, le vague recouvre la notion d'imprécision (Jucker et al., 2003; Williamson, 1994).

Toutefois, il peut être difficile en pratique de distinguer les deux. Ainsi, la proposition "*il pleuvra certainement demain*" peut être considérée comme à la fois incertaine, car le locuteur n'est pas totalement sûr qu'il pleuve demain, et vague, car son degré de confiance en la véracité de la proposition, exprimé par le terme "*certainement*", est lui-même imprécis.

Pragmatique du vague D'un point de vue pragmatique, les expressions vagues sont définies par leur manque de précision (Jucker et al., 2003; Williamson, 1994). L'utilisation du vague en général répond à un besoin d'économie d'effort. En effet, chercher à être très précis peut résulter en une perte de temps et en une inflexibilité alors qu'en étant

vague, un locuteur peut atteindre son but en minimisant l'effort requis (Grice, 1991). Une proposition vague transmet davantage d'information contextuelle pertinente que ne le ferait une expression précise pourvu que le contexte culturel soit partagé entre les deux interlocuteurs (Shapiro, 2006). L'exemple paradigmatique à cet égard est la communication de quantités. En effet, au quotidien, il est plus courant d'employer une approximation qu'une valeur précise pour dénoter une quantité (Moxey & Sanford, 1997; Zhang, 1998).

Dans une perspective pragmatique, puisqu'il existe une infinité de façons de faire référence à une entité, une propriété, un événement ou à un état mental, le choix d'une expression plus ou moins vague dépend du but du locuteur et tient lieu de stratégie interactionnelle. En effet, la variation du degré de vague d'une proposition permet au locuteur d'espérer produire les représentations attendues chez l'auditeur (Jucker et al., 2003; Shapiro, 2006).

Selon l'usage souhaité, six différents types d'expressions vagues peuvent être employés (Jucker et al., 2003). Pour chaque type, un certain nombre de termes spécifiques sont utilisés pour exprimer le caractère vague. Nous évoquons d'abord les termes exprimant une imprécision, puis ceux associés à une incertitude. Pour chacune de ces deux catégories, nous les analysons selon leur dépendance au contexte.

Considérons d'abord les catégories d'incertitude et d'imprécision. Trois types de termes communiquent une imprécision plutôt qu'une incertitude. Du moins dépendant au plus dépendant au contexte, on trouve les *quantificateurs*, les *approximateurs* et les *adaptateurs*.

Les *quantificateurs*, comme "*certains*" ou "*la plupart*", et les *approximateurs*, par exemple "*environ*" ou "*au moins*", portent tous deux sur des quantités. Leur distinction réside dans le fait que les *quantificateurs* sont relatifs et expriment des proportions là où les *approximateurs* sont absolus et expriment, de manière vague, les quantités proprement dites (Channell, 1994). Par conséquent, les *quantificateurs* sont moins dépendants du contexte que les *approximateurs* (Moxey & Sanford, 1997) : les proportions restent équivalentes, quels que soient l'ensemble ou le domaine sur lesquels elles portent. Au contraire, les approximations dépendent du domaine considéré. Par exemple, dans le cas de la température corporelle, le domaine possible est grossièrement compris chez l'être humain entre 36 ° et 43 ° . Par conséquent, une expression telle que "*environ 40 °*" exprime une imprécision moins importante que la même expression utilisée pour décrire la température estivale moyenne d'un pays. Le cas des approximateurs, qui intéresse particulièrement nos travaux, est traité de manière plus détaillée dans la section 2.1.3. Ajoutons que les *quantificateurs* dénotent un ensemble qui peut être représenté sur une échelle et qu'ils donnent lieu à des *implicatures scalaires* (Grice, 1975). Celles-ci sont définies comme des inférences réalisées par l'interlocuteur résultant non pas d'un raisonnement logique mais d'un processus pragmatique. Dans le cas des quantificateurs, l'implicature scalaire dénote le fait qu'ils expriment ex-

plicitement la proportion considérée et, en même temps, qu'ils expriment implicitement la négation de cette proportion (Yule, 1996). Par exemple, l'usage de "*certains*" exprime implicitement "*mais pas tous*". Dans le domaine informatique, les quantificateurs ont fait l'objet de nombreux travaux, notamment dans le cadre de la logique floue et des résumés linguistiques (Zadeh (1983), voir par ex. Moyse et Lesot (2016) pour une application concrète).

Les *adaptateurs*, comme "*en quelque sorte*" ou "*quelque chose comme ça*", sont des termes permettant d'accentuer l'imprécision communiquée par une expression (Prince et al., 1982). Ils marquent également une distance plus forte entre la pensée que le locuteur cherche à exprimer et ce qu'il ou elle exprime effectivement. Les *adaptateurs* sont imprécis et très dépendants du contexte. En effet, une expression comme "*quelque chose comme ça*" renvoie à un "*ça*" dénotant une entité spécifiée dans le contexte et ne saurait être interprétée sans cette spécification.

Les trois autres types de termes renvoient plus à une incertitude qu'à une imprécision. Dans cette catégorie, on trouve les *boucliers*, les *adverbes de probabilité* et les *adverbes de fréquence*. Tous trois dépendent relativement peu du contexte.

Les *boucliers*, comme "*je pense que*" ou "*il me semble que*" expriment l'attitude propositionnelle, c'est-à-dire l'attitude du locuteur vis-à-vis de la proposition qu'il communique (Blakemore, 1992; Prince et al., 1982; Stubbs, 1986). Ils permettent de faire varier le degré d'engagement du locuteur vis-à-vis de la véracité du contenu exprimé et dépendent donc très peu du contexte.

Les *adverbes de probabilité*, comme "*probablement*" ou "*certainement*", expriment directement et clairement une incertitude sur l'occurrence d'un événement (Teigen, 1988). D'un point de vue pragmatique, ces termes ne communiquent pas seulement un degré de probabilité d'occurrence, leur rôle est également de gérer le focus de l'auditeur de manière à ce que celui-ci renforce ou au contraire affaiblisse ses hypothèses quant à un événement (Jucker et al., 2003).

Enfin, les *adverbes de fréquence*, comme "*souvent*" ou "*rarement*", communiquent une fréquence d'occurrence d'événements d'une manière absolue là où les adverbes de probabilité les expriment de manière relative. Du point de vue de la représentation, les *adverbes de fréquence* ont des caractéristiques proches de celles des *quantificateurs* (Channell, 1994; Moxey & Sanford, 1997). Comme eux, ils peuvent être représentés sur une échelle et donnent lieu à des implicatures scalaires. Par exemple, "*souvent*" exprime également de manière implicite "*mais pas toujours*".

2.1.3 Les approximateurs

Après ce rapide tour d'horizon des termes pouvant rendre une proposition vague, nous détaillons le cas des approximateurs, dont fait partie le terme "*environ*" qui fait l'objet de nos travaux.

Pragmatique des approximateurs Comme indiqué dans la section précédente, les approximateurs, également appelés *rounders* (Prince et al., 1982), sont utilisés pour exprimer de manière vague des quantités, par exemple "*environ 100*", ou des plages de quantités, par exemple, "*au moins 100*" (Channell, 1980; Wierzbicka, 1986).

L'emploi d'un approximateur pour exprimer une quantité n'est pas toujours nécessaire et permet surtout d'expliciter l'aspect vague d'une proposition. En effet, le plus souvent, un nombre rond est interprété comme une approximation d'une valeur plus précise (Krifka, 2007; Wachtel, 1980). Par exemple, la différence entre "*Cette chaise coûte 100 euros*" et "*Cette chaise coûte* ***environ*** *100 euros*" réside en ce que la nature vague du coût, 100 euros, est explicitement communiquée dans la seconde proposition, même si la première ne prétend pas donner une information exacte.

Du point de vue de la pragmatique du langage, une approximation peut avoir deux fonctions. La première permet au locuteur d'exprimer un ordre de grandeur relatif à une quantité dont il ignore la valeur exacte. Par exemple, s'il sait que la population de Paris est d'environ 2 millions d'habitants et qu'il ignore que cette population est d'exactement 2 220 445 habitants [2], il n'a pas d'autre choix que d'utiliser une approximation pour exprimer l'ordre de grandeur.

La deuxième fonction porte sur l'effort nécessaire à la production et à l'interprétation d'une valeur numérique (Jucker et al., 2003; Krifka, 2007). Les approximations impliquent le plus souvent des nombres ronds, qui sont plus simples à la fois au niveau représentationnel et au niveau verbal (Krifka, 2007). Par exemple, il est plus aisé de se représenter 100 que 118. De plus, verbalement, 100 ne contient qu'une seule syllabe, là où 118 en comprend trois. Les simplicités représentationnelle et verbale ne se recouvrent toutefois pas nécessairement. En effet, dans le cadre temporel, par exemple, *une demi-heure*, comprenant quatre syllabes est moins simple verbalement que *10 minutes*, comprenant trois syllabes. A l'inverse, elle est plus simple représentationnellement car elle correspond à la moitié d'une heure (Krifka, 2007). Dans ce cas, la préférence du locuteur pour une expression simple peut varier entre les dimensions représentationnelle et verbale, selon le contexte.

2. D'après le recensement de l'INSEE réalisé en 2013.

Perspectives épistémique et sémantique des approximateurs Comme pour les expressions vagues, les approximateurs peuvent être traités dans la perspective épistémique ou dans la perspective sémantique (voir section 2.1.1).

Du point de vue d'une interprétation purement sémantique, et en accord avec la perspective sémantique du vague, une proposition impliquant une approximation ne peut être dite absolument vraie ou absolument fausse pour certaines valeurs qui se trouvent à proximité de la frontière de son extension (Lakoff, 1973). Pour surmonter cette difficulté, Lakoff (1973) propose d'adopter une sémantique basée sur la logique floue. Dans ce cas, une proposition impliquant une approximation peut prendre des valeurs de vérité graduelles. Par exemple, la population de Paris étant de 2 220 445 habitants en 2013, la proposition "*la population de Paris est d'environ 2 millions d'habitants*" est associée à une valeur de vérité plus élevée que la proposition "*la population de Paris est d'environ 3 millions d'habitants*".

Du point de vue purement pragmatique, en revanche, une approximation ne peut en soi recevoir une valeur de vérité, même graduelle. A la limite, une proposition impliquant une approximation peut être infalsifiable (Sadock, 1977). En effet, une approximation est sensible au contexte dans lequel elle est exprimée. La nature de l'entité pour laquelle une approximation est réalisée et le but du locuteur ont tous deux un effet sur la manière dont l'approximation est estimée (Sadock, 1977). La fonction qui fait correspondre à l'approximation une plage de valeurs dénotées doit donc prendre en paramètres les caractéristiques du contexte. On retrouve donc ici la perspective épistémique sur le vague. En effet, du point de vue pragmatique, il existe une plage de valeurs dénotées clairement définie mais dont les bornes sont trop dépendantes au contexte pour être connues.

A la suite de Wachtel (1980), un compromis entre ces deux points de vue peut être trouvé : on peut d'une part considérer que la valeur de vérité décroît graduellement avec la distance à la valeur numérique impliquée dans l'approximation. D'autre part, on peut également considérer que la manière dont décroît, et jusqu'où décroît cette fonction dépend de paramètres liés au contexte.

Enfin, notons que dans cette thèse, nous nous intéressons en particulier à l'approximateur "*environ*". Toutefois, il a été montré (Channell, 1980; Ferson et al., 2015) que l'interprétation d'expressions impliquant des approximateurs sémantiquement très proches, comme "*approximativement*" ou "*autour de*", est très similaire à celle impliquant "*environ*". Sous réserve de confirmation empirique, les résultats que nous présentons, sont donc généralisables aux approximateurs sémantiquement proches de "*environ*".

2.1.4 Représentations formelles du vague

Dans le domaine de la représentation des connaissances dans les systèmes informatiques, qui traitent par nature des données précises, le vague est l'une des formes que peut prendre une connaissance imparfaite (Smets 1991; 1997).

Dans ce champ théorique, on distingue essentiellement deux types de vague, comme nous l'avons vu dans la section 2.1.2 : les imprécisions et les incertitudes. Alors que l'incertitude renvoie au statut de la proposition, l'imprécision porte sur le contenu de celle-ci (Bennett, 2010 ; Smets 1991; 1997).

Deux types d'imprécisions peuvent à leur tour être distingués (Smets, 1997). Premièrement, l'ambiguïté fait référence à une imprécision sémantique et qualitative, comme dans la proposition "*ce livre est très bon*", où "*très bon*" peut renvoyer à "*bien écrit*" ou à "*captivant*" par exemple. Dans ce cas, l'imprécision porte sur le sens même du terme.

Deuxièmement, l'approximation fait référence à une imprécision quantitative, lorsque des expressions numériques sont impliquées. Dans ce cas, la valeur numérique exacte n'est pas nécessairement connue par le locuteur bien que la plage ou l'intervalle de valeurs dans laquelle elle se trouve le soit.

Il existe différentes manières de formaliser les incertitudes et les imprécisions, parmi lesquelles la théorie des probabilités (Degroot, 1970; Fine, 1973), la théorie des possibilités (Zadeh, 1978; Dubois & Prade, 1988) et la théorie des sous-ensembles flous (Zadeh 1965; 1975 ; Dubois & Prade, 1980).

La théorie des sous-ensembles flous (Zadeh 1965; 1975) a été développé dans le but de formaliser les imprécisions. Dans cette perspective liée au point de vue sémantique du vague, l'extension d'un concept vague n'est pas nettement limitée. Dans le cadre de la théorie des sous-ensembles flous, le degré d'appartenance d'une entité à un ensemble, n'est pas limité à vrai ou faux, ou 0 ou 1, mais peut prendre une valeur réelle dans tout l'intervalle $[0, 1]$ (voir annexe A p. 185). Par conséquent, dans le cadre de la théorie des sous-ensembles flous, un concept est plus ou moins applicable à une entité.

2.2 Cognition humaine des nombres

Le deuxième champ théorique auquel peuvent être rattachées les Expressions Numériques Approximatives, en plus de la notion intrinsèque de vague, est la manière dont les individus se représentent et traitent cognitivement les nombres. Nous présentons dans cette section les aspects cognitifs relatifs aux nombres. Dans un premier temps, nous proposons un rapide tour d'horizon des usages des nombres dans la langue naturelle, par exemple le dénombrement et la mesure. Dans un second temps, nous exposons l'état de l'art quant

à la manière dont les nombres et les quantités sont traités et représentés par le système cognitif humain.

2.2.1 L'usage des nombres

Dans la langue naturelle, les nombres peuvent être utilisés dans différents contextes et à différentes fins. Dans la mesure où le contexte d'une approximation numérique, comme une ENA, peut influencer son interprétation (Lasersohn, 1999; Sadock, 1977; Williamson, 1994), il nous paraît pertinent de présenter brièvement les différents usages des nombres et de préciser quels sont ceux qui peuvent faire intervenir l'utilisation d'ENA.

Parmi l'ensemble des usages possibles d'un nombre, les plus courants sont la numérotation, l'ordonnancement, la localisation, la mesure, le dénombrement et le calcul (Groza, 1968) : la numérotation consiste à affecter un nombre unique à chaque élément d'un ensemble d'éléments. L'ordonnancement vise à classer les éléments d'un ensemble suivant un nombre qui leur est affecté, du plus petit au plus grand ou du plus grand au plus petit, par exemple. La mesure consiste à déterminer la valeur d'une grandeur physique, généralement au moyen d'un instrument. Le dénombrement peut être rapproché de la mesure puisqu'il s'agit de déterminer la quantité d'éléments présents dans un ensemble. Enfin, le calcul consiste à effectuer des opérations mathématiques.

Intuitivement, les ENA peuvent s'appliquer dans le cadre de deux usages : le dénombrement et la mesure. En effet, en tant qu'approximations, les ENA portent sur des quantités et seuls ces deux usages en impliquent. On peut toutefois noter que ces deux types de quantités ne sont pas représentés dans les mêmes ensembles de nombres : le dénombrement est réalisé dans l'ensemble des entiers naturels, $\mathbb{N}$, alors que la mesure peut être réalisée dans l'ensemble des réels, $\mathbb{R}$.

2.2.2 Représentation et traitement cognitifs des nombres

Deux sous-systèmes cognitifs différents sont impliqués dans la cognition des nombres et des quantités (Dehaene 2003; 2009) : le système approximatif et le système exact des nombres. Nous les présentons dans un premier temps avant d'exposer la notion de saillance cognitive des nombres qui renvoie aux préférences des individus quant à certains nombres.

Deux sous-systèmes cognitifs Le premier sous-système cognitif responsable du traitement des nombres, appelé *système approximatif des nombres* (*Approximate Number System*) repose sur des représentations approximatives, analogiques et non-symboliques (Dehaene, 2003). Les nombres y sont représentés sur une ligne mentale logarithmiquement compressée, sur laquelle deux quantités x_1 et x_2 peuvent être distinguées si la valeur ab-

solue de leur différence absolue est supérieure à une fraction c, appelée fraction de Weber, de la plus grande des deux, formellement :

$$\frac{|x_1 - x_2|}{\max(x_1, x_2)} > c \tag{2.1}$$

Cette indistinguabilité entre quantités proches conduit à des imprécisions lorsque des estimations sont réalisées. Par exemple, en estimant, sans compter, le nombre de points dans un tableau présenté visuellement, on ne peut distinguer s'il y en a 98 ou 101. La valeur de la fraction de Weber c varie d'un individu à l'autre et selon l'âge (Halberda et al., 2008). En moyenne, elle serait de $0,11$ (Halberda & Feigenson, 2008) ou $0,12$ (Pica et al., 2004).

Les représentations issues du système approximatif des nombres sont abstraites et indépendantes de la modalité de présentation (Holyoak & Walker, 1976; Dehaene, 2003; Walsh, 2003). En effet, le rôle de ce sous-système est d'encoder la magnitude de la qualité considérée (nombre, espace, temps, luminosité, volume sonore, etc.). Enfin, on peut noter que la représentation des quantificateurs (voir section 2.1.2 p. 12) est réalisée sur la ligne mentale des nombres du système approximatif des nombres (Coventry et al., 2010).

Les représentations du deuxième sous-système, le *système exact des nombres*, sont exactes et symboliques (Dehaene, 2003; Gelman & Butterworth, 2005). Basées sur le langage, elles ne souffrent d'aucune imprécision. Le système exact des nombres est responsable du traitement des faits arithmétiques.

Ces deux systèmes ne sont pas totalement indépendants l'un de l'autre. En effet, d'une part, le système exact des nombres est, d'un point de vue évolutif, plus tardif (Gelman & Gallistel, 2004; Gelman & Butterworth, 2005) et son développement est fondé sur le système approximatif des nombres (Berteletti et al., 2010).

Enfin, la sémantique des nombres repose sur le système approximatif des nombres (Dehaene & Cohen, 1995). Autrement dit, une quantité exprimée verbalement ou communiquée par écrit, bien que traitée dans un premier temps par le système exact des nombres, ne prend sens que lorsque sa magnitude est encodée dans le système approximatif des nombres. Aussi, mettre en relation deux nombres, par exemple, dans une tâche de comparaison, ne peut se produire qu'au niveau des représentations analogiques du système approximatif des nombres.

Saillance cognitive des nombres Du point de vue représentationnel, certains nombres sont plus saillants que d'autres, c'est-à-dire sont plus facilement évoqués par le système cognitif que d'autres. En effet, il a été observé, dans des corpus de textes et de journaux quotidiens, que certains nombres apparaissent plus fréquemment que d'autres dans la langue naturelle (Dehaene & Mehler, 1992; Jansen & Pollmann, 2001).

Il a été montré que dans le système numérique arabe, deux caractéristiques des nombres influencent leur fréquence d'apparition. D'abord, plus un nombre possède de chiffres significatifs, plus il est complexe à générer et moins il apparaît fréquemment (Dehaene & Mehler, 1992). D'un point de vue cognitif, les nombres fréquents seraient stockés en mémoire sous forme d'information lexicale alors que les nombres moins fréquents seraient générés par une procédure de type pas à pas (Dehaene & Mehler, 1992). Deuxièmement, les nombres dont le dernier chiffre significatif est 5 et, dans une moindre mesure, 2, apparaissent plus fréquemment (Jansen & Pollmann, 2001).

Ces nombres plus fréquents sont appelés *nombres référents* (Dehaene & Mehler, 1992) et sont ceux qui couvrent une large part de la ligne mentale du système approximatif des nombres, sur laquelle sont représentées les quantités. Ils sont donc particulièrement adaptés pour servir de valeurs approchées dans le cadre de l'approximation de valeurs précises (Dehaene & Mehler, 1992).

2.2.3 Les ENA à la lumière de la cognition des nombres

Du modèle de la cognition humaine des nombres, deux perspectives dans le traitement des expressions numériques approximatives peuvent être considérées. Premièrement, interpréter une ENA consiste à estimer l'imprécision, c'est-à-dire la part couverte par la valeur de référence de l'ENA sur la ligne mentale représentant les quantités dans le système approximatif des nombres. Cette perspective conduit à considérer la valeur numérique de l'ENA comme pertinente puisqu'elle correspond à la magnitude encodée.

Deuxièmement, les ENA étant des expressions linguistiques et numériques, leur traitement implique également le système exact des nombres. Cette perspective conduit à considérer ce traitement comme un problème formel, entendu comme un raisonnement arithmétique, comme celui proposé par le modèle basé sur des échelles décrit dans le prochain chapitre (voir section 3.2.2 p. 30). En effet, formellement, estimer l'imprécision dénotée par une ENA relativement précise, par exemple "*environ 4730*", uniquement à la lumière de sa magnitude conduirait à une plage de valeurs beaucoup trop large, car comparable à celle de "*environ 4700*" ou "*environ 5000*". Or intuitivement, l'imprécision associée à "*environ 4730*" devrait être nettement inférieure à celle associée à "*environ 5000*".

Il nous semble par conséquent raisonnable de penser que l'attribution d'une imprécision à une ENA résulte d'un compromis entre ces deux perspectives. Par exemple, dans le cas de "*environ 8150*", l'imprécision liée à sa représentation analogique devrait être importante, proche de celle correspondant à "*environ 8000*" car l'ordre de grandeur de leur magnitude est similaire. Au contraire, pour le système exact des nombres et en tant que fait numérique, "*environ 8150*" est une description beaucoup plus précise de ce qui est évalué ou compté que

"*environ 8000*". Par conséquent, l'imprécision assignée à la première devrait être nettement plus faible que celle assignée à la seconde.

Traiter une ENA nécessite donc de tenir compte à la fois de l'imprécision liée à sa représentation analogique dans le système approximatif des nombres et de l'imprécision logiquement liée à son niveau de précision, telle qu'elle peut être évaluée par le système exact des nombres. On peut s'interroger sur la nature de ce compromis et dans quelle mesure l'un des deux systèmes prend le pas sur l'autre.

2.3 Bilan

Dans ce chapitre, nous nous sommes intéressé au contexte théorique dans lequel s'inscrivent nos travaux portant sur l'interprétation et le calcul des expressions numériques approximatives.

De l'étude du champ théorique relatif au vague, il ressort que deux perspectives principales peuvent être envisagées. Du point de vue épistémique, la difficulté à établir clairement l'extension d'un concept vague prend sa source dans la variabilité des opinions, à la fois à travers les individus, le temps et les contextes. Du point de vue sémantique, les frontières de l'extension d'un concept vague sont par nature graduelles. Dans cette perspective, on se propose d'abandonner le principe de non contradiction de la logique classique en s'appuyant par exemple sur des logiques multivaluées, comme la logique floue. Ces deux perspectives nous conduisent à considérer deux approches, détaillées dans le chapitre suivant et mises en œuvre dans l'ensemble de nos travaux, pour représenter l'imprécision liée à une ENA : une approche par intervalles et une approche par sous-ensembles flous, correspondant respectivement aux deux perspectives.

Dans un second temps, nous avons présenté la manière dont les nombres et quantités sont encodés et représentés dans le système cognitif humain. Il apparaît que deux sous-systèmes en sont responsables, le système approximatif et le système exact des nombres, ce qui nous conduit à considérer que la représentation et le traitement d'une ENA peuvent résulter d'un compromis entre eux.

Chapitre 3

Etat de l'art : représentation, interprétation et calcul d'ENA

Ce chapitre propose un état de l'art portant sur trois aspects des Expressions Numériques Approximatives : la section 3.1 présente la problématique de leur représentation ainsi que les approches formelles qui ont été proposées pour les représenter. La section 3.2 considère la tâche d'interprétation des ENA et présente les trois modèles computationnels que nous avons identifiés dans la littérature. La section 3.3 est dédiée aux études portant sur le calcul imprécis. Enfin, nous proposons un bilan de cet état de l'art dans la dernière section.

3.1 Représentation formelle des ENA

Dans cette section, nous présentons successivement deux formalismes utilisés dans la littérature pour représenter les imprécisions sur des univers numériques : l'approche par intervalles et l'approche par sous-ensembles flous. Nous les interprétons selon les deux perspectives du vague présentées dans le chapitre précédent (voir section 2.1.1 p. 8), les perspectives épistémique et sémantique respectivement. De plus, nous présentons brièvement la manière de construire ces représentations à partir de données collectées.

3.1.1 L'approche par intervalles

La notion de *halo pragmatique* a été introduite par Lasersohn (1999) pour formaliser les expressions vagues en général, au delà des expressions numériques. Le halo pragmatique d'une expression vague est défini comme un ensemble d'entités résultant de l'application d'une fonction de dénotation à un terme vague. Parmi ces entités, l'une est explicitement dénotée par l'expression vague alors que les autres le sont implicitement (Lasersohn, 1999).

Par exemple, dans la proposition "*je serai là autour de 18h*", l'expression vague "*autour de 18h*" dénote explicitement l'heure précise 18h00 et implicitement un ensemble d'horaires autour de 18h, par exemple, l'ensemble des instants compris entre 17h45 et 18h15.

D'un point de vue logique, la valeur de vérité d'une expression vague est vraie si l'entité à laquelle elle fait référence est comprise dans son halo pragmatique (Krifka, 2007; Lasersohn, 1999). Par conséquent, si j'arrive à 17h49, la proposition "*je serai là autour de 18h*" est vraie car 17h49 est compris dans le halo pragmatique de "*autour de 18h*". En revanche, si j'arrive à 18h27, cette même proposition est considérée comme fausse.

Cette approche, pragmatique, peut être interprétée comme une formalisation de la perspective épistémique du vague. En effet, l'extension de l'expression vague est clairement délimitée par les frontières du halo pragmatique et l'expression ne peut se voir affecter que deux valeurs de vérité : vrai ou faux.

Appliquée aux ENA, la notion de halo pragmatique peut être formalisée par un intervalle. Plus spécifiquement, la fonction de dénotation prend en entrée une valeur numérique x et renvoie un intervalle, $[x^-, x^+]$. Dans ce cadre, le rôle des approximateurs est de moduler la forme et la taille des halos pragmatiques (Lasersohn, 1999). Par exemple, *exactement* réduit le halo pragmatique à la seule quantité dénotée alors que *environ* l'élargit à des quantités qui peuvent être éloignées de celle explicitement dénotée. A chaque approximateur correspond donc une fonction de dénotation particulière. Ainsi dans l'exemple précédent, le halo pragmatique de "*autour de 18h*" peut être formalisé par l'intervalle [17h45, 18h15].

Du point de vue de la méthodologie de la collecte des données, la représentation par intervalles présente l'avantage d'être très intuitive même pour les non-experts (Chameau & Santamarina, 1987). En effet, il suffit de demander la plage de valeurs, ou ses bornes, correspondant à l'expression, par exemple, "*selon vous, entre quelle heure et quelle heure peut-on dire qu'on arrive autour de 18h ?*"

3.1.2 L'approche par sous-ensembles flous

Dès son introduction par Zadeh (1965), l'un des objectifs de la théorie des sous-ensembles flous, dont un bref rappel est donné en annexe A p. 185, est de proposer un formalisme permettant de modéliser le vague (voir par ex., Hersh & Caramazza, 1976; Lakoff, 1973).

Principe de l'approche par sous-ensembles flous Dans la théorie des sous-ensembles flous, l'appartenance d'une entité à un ensemble n'est pas binaire, oui/non, mais elle est représentée par une valeur réelle comprise dans l'intervalle $[0, 1]$, appelée degré d'appartenance, où 0 correspond à la non-appartenance classique et 1 à l'appartenance classique à un ensemble. Aussi, l'applicabilité d'un concept ou d'une expression, représenté par un

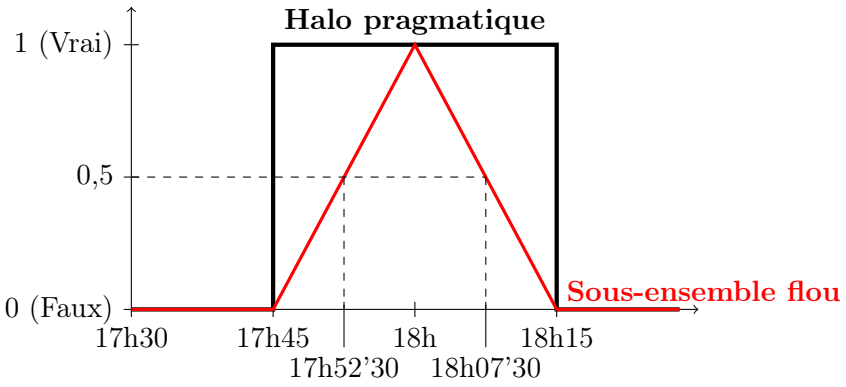


FIGURE 3.1 – Illustration d'un halo pragmatique (en noir) et d'un sous-ensemble flou (en rouge) représentant l'imprécision liée à l'expression "*autour de 18h*". Les lignes en pointillés montrent les valeurs pour lesquelles le degré d'appartenance au sous-ensemble flou est 0,5, c'est-à-dire les valeurs qui rendent l'expression autant vraie que fausse.

sous-ensemble flou, à une entité est graduelle. Ce formalisme entre donc dans le cadre de la perspective sémantique du vague, selon laquelle les limites de l'extension d'un concept sont graduelles (voir section 2.1.1 p. 8 et Hersh & Caramazza, 1976; Lakoff, 1973).

Les sous-ensembles flous généralisent la notion d'ensembles. Lorsque l'univers considéré est l'ensemble des réels $\mathbb{R}$, les sous-ensembles flous généralisent les intervalles et la notion de halo pragmatique : les entités comprises dans le halo pragmatique ont pour degré d'appartenance 1 et toutes les entités qui en sont exclues ont un degré d'appartenance de 0. La théorie des sous-ensemble flous permet donc de nuancer les halos pragmatiques : les entités qui ne sont pas totalement exclues du halo pragmatique peuvent y être incluses à différents degrés.

Ceci est illustré par la figure 3.1, représentant un halo pragmatique et un sous-ensemble flou correspondants à l'expression "*autour de 18h*". Deux observations peuvent être faites. D'abord, on voit que dans le cadre des halos pragmatiques, le fait d'arriver à 17h45 rend la proposition vraie alors que le fait d'arriver à 17h44 la rend fausse. Au contraire, dans le cadre de la théorie des sous-ensembles flous, arriver 17h45 rend la proposition un peu plus vraie que le fait d'arriver à 17h44. De même, si l'heure d'arrivée est 17h46, alors la proposition est encore un peu plus vraie, et ainsi de suite, jusqu'au cas où le fait d'arriver à 18h rende la proposition absolument vraie.

Deuxièmement, la figure 3.1 illustre le comportement de la représentation par intervalles et de la représentation par sous-ensembles flous vis-à-vis du principe de tolérance et du principe de contradiction. L'approche par halos pragmatiques conserve le principe de contradiction et abandonne le principe de tolérance. Une entité, ici un horaire, appartient ou n'appartient pas au halo pragmatique et le passage de l'un à l'autre est brutal. Ainsi, le

petit changement entre 17h44 et 17h45 fait passer de la non applicabilité à l'applicabilité de l'expression "*autour de 18h*". Au contraire, l'approche par sous-ensembles flous abandonne le principe de non-contradiction mais conserve le principe de tolérance. En effet, le passage de 17h44 à 17h45 ne change pas fondamentalement le degré d'applicabilité de l'expression "*autour de 18h*". On peut observer que cette expression est considérée comme aussi vraie que fausse quand on considère l'horaire 17h52min30, pour lequel le degré d'appartenance est 0,5, à mi-chemin entre le vrai absolu et le faux absolu.

Appliquée aux ENA, l'imprécision est formalisée par un sous-ensemble flou dont l'univers est l'ensemble des réels $\mathbb{R}$. Plus précisément, les ENA sont représentées par des intervalles flous. Ceux-ci se définissent par deux caractéristiques (Zadeh, 1978) : (i) le noyau, c'est-à-dire l'ensemble des valeurs dont le degré d'appartenance est 1, forme un intervalle ; et (ii) de part et d'autre du noyau, le degré d'appartenance est une fonction décroissante de la distance au noyau jusqu'à atteindre 0.

Dans cette approche par sous-ensembles flous, la fonction de dénotation d'une ENA prend donc en entrée une valeur numérique x et renvoie une fonction d'appartenance $\mu_x(y)$ respectant les contraintes d'un intervalle flou, où y est une valeur de l'univers sur lequel est défini le sous-ensemble flou, $\mathbb{R}$.

Construction de sous-ensembles flous Il existe cinq méthodes de construction de sous-ensembles flous à partir de données collectées, correspondant à cinq interprétations des degrés d'appartenance (Bilgiç & Türkşen, 2000). Nous illustrons ces méthodes dans le cas de l'ENA "*environ 100*" en considérant la valeur 95 dont nous fixons arbitrairement le degré d'appartenance à 0,6. Chaque interprétation ne permet de capturer qu'un seul des deux types de tolérance, intra ou inter-individuel (voir section 2.1.1 p. 8), précisé dans chaque cas ci-dessous. Cette limite n'est pas due au type d'interprétation considérée mais à la nature même de la représentation par sous-ensembles flous, qui ne permet qu'un seul degré de liberté.

Les degrés d'appartenance peuvent d'abord être interprétés comme décrivant le degré de vraisemblance d'une valeur candidate vis-à-vis de l'ENA (Hisdal, 1985). Le degré d'appartenance de 95, valant 0,6, est interprété comme "*60% des personnes interrogées déclarent que 95 entre dans la description d'*environ 100". Pour construire un sous-ensemble flou correspondant à cette interprétation, il faut donc proposer plusieurs valeurs, 80, 85, 90, 95 par exemple, à chaque personne interrogée. Dans cette perspective, les deux types de tolérance sont collectés : plusieurs personnes sont interrogées pour chaque ENA, permettant de capturer la tolérance inter-individuelle et, dans la mesure où chaque personne se voit proposer plusieurs valeurs, on capture la tolérance intra-individuelle. Toutefois, suite à l'agrégation des données, une tolérance composite émerge dont il est difficile de déterminer

les parts intra et inter-individuelle.

Dans la seconde interprétation, *random set*, un sous-ensemble flou est vu comme un ensemble d'intervalles, les α-coupes, imbriqués (Dubois & Prade, 1989). Dans cette perspective, le degré d'appartenance d'une valeur correspond à la fréquence relative cumulée de personnes interrogées qui incluent cette valeur dans l'intervalle correspondant à "*environ 100*". Le degré d'appartenance de 95 est donc interprété comme "*60% des personnes interrogées décrivent* environ 100 *par un intervalle contenant 95*". Par conséquent, les personnes ne sont pas interrogées sur une valeur précise, ici 95, mais doivent chacune donner un intervalle qui leur semble correspondre à l'ENA considérée pour construire le sous-ensemble flou la représentant. Dans la mesure où chaque personne interrogée fixe les bornes de l'intervalle correspondant à une ENA, cette perspective ne capture que la tolérance inter-individuelle.

La troisième interprétation, de *similarité*, considère le degré d'appartenance comme une distance à un prototype représentant la catégorie (Rosch & Mervis, 1975; Lakoff, 1987). Dans le cas des ENA, le prototype est la valeur de référence, ici 100. Le degré d'appartenance de 95 est par conséquent interprété comme "*95 est éloigné de* environ 100 *à un degré 0,4*". Comme dans l'interprétation de *vraisemblance*, construire un sous-ensemble flou dans cette perspective suppose de proposer plusieurs valeurs à chaque personne interrogée : "*selon vous, à quel degré 95 est proche de environ 100* ?" Par conséquent, on capture dans cette perspective la tolérance intra-individuelle.

La quatrième interprétation, d'*utilité*, consiste à interpréter le degré d'appartenance comme "*0,6 est le degré d'utilité de dire que 95 entre dans* environ 100" (Bilgiç & Türkşen, 2000). L'adaptation de cette interprétation aux ENA présente des difficultés car elle est plus appropriée aux prédicats. Dans cette perspective également, les personnes doivent chacune être interrogée sur plusieurs valeurs et seule la tolérance intra-individuelle est capturée.

Enfin, dans la dernière interprétation, de *mesure*, le degré d'appartenance est déterminé en prenant pour référence d'autre valeurs (Krantz et al., 1971). L'adaptation de cette interprétation aux ENA présente également des difficultés car il faut pouvoir déterminer à quelles autres valeurs celle qui est considérée doit être comparée. Le degré d'appartenance de 95 peut toutefois être interprété comme "*comparé aux autres, 95 est plus près de* environ 100 *et cela peut être encodé sur une échelle par la valeur 0,6*" (Bilgiç & Türkşen, 2000). Une telle interprétation peut poser de réelles difficultés quand à la manière d'interroger les personnes.

Ainsi, parmi les cinq interprétations présentées, la perspective *random set* nous paraît être la plus appropriée pour construire le sous-ensemble flou représentant une ENA et elle présente un avantage indéniable en terme de méthode de collecte de donnée : si l'on souhaite interroger des non-experts, fournir un intervalle par le biais de ses bornes est in-

tuitivement une manière naturelle pour décrire une imprécision (Chameau & Santamarina, 1987). D'autre part, contrairement aux autres interprétations, elle n'exige pas d'interroger plusieurs fois la même personne à propos de la même ENA. Elle permet donc de rendre la tâche moins fastidieuse pour le participant et d'éviter les biais liés à la répétition des questions (Gideon, 2012). Toutefois, si une collecte de données réalisée dans la perspective *random set* capture la tolérance inter-individuelle (voir section 2.1.1 p. 8), elle ne permet pas de capturer la tolérance intra-individuelle puisqu'une personne n'est interrogée qu'une seule fois à propos d'une ENA.

Sous-ensembles flous de type 1 et de type 2 Les sous-ensembles flous que nous avons évoqués jusqu'à présent sont dits de type 1 car ils ne donnent lieu qu'à un seul degré de liberté pour une valeur, son degré d'appartenance. Par conséquent, l'utilisation de ce type de sous-ensembles flous conduit à choisir entre capturer la tolérance intra-individuelle ou la tolérance inter-individuelle.

Les sous-ensembles flous de type 2 permettent de surmonter cette difficulté (Zadeh, 1975). En effet, ceux-ci sont une généralisation des sous-ensembles flous de type 1 en ce qu'ils donnent lieu à un deuxième degré de liberté pour une même valeur : le degré d'appartenance d'une valeur est lui-même représenté par un sous-ensemble flou. On peut alors représenter à la fois les deux types de tolérances, intra-individuelle et inter-individuelle, pour une même ENA. Il existe deux méthodes de construction des sous-ensembles flous de type 2 par collecte de données (Mendel, 2007a).

La première, *par personne*, consiste à demander aux individus interrogés de fournir un sous-ensemble flou de type 1 représentant l'imprécision liée au terme vague considéré. On capture ainsi dans un premier temps la tolérance intra-individuelle. Ces sous-ensembles flous de type 1 sont ensuite agrégés, de manière à capturer la tolérance inter-individuelle, pour construire un sous-ensemble flou de type 2 (Mendel, 2007b).

La seconde méthode, *par bornes d'intervalles*, se déroule en trois étapes. D'abord, il s'agit de collecter auprès des personnes interrogées les intervalles correspondant à une ENA, à la manière de la perspective *random set* de construction des sous-ensembles flous de type 1 décrite précédemment. Dans un deuxième temps, des valeurs statistiques sont établies pour les bornes des intervalles ainsi collectées. Enfin, ces valeurs statistiques sont insérées dans des sous-ensembles flous de type 2 pré-paramétrés (Mendel, 2007b).

Ces deux méthodes présentent chacune des limites. Si l'on s'intéresse aux termes vagues de la langue naturelle et que l'on souhaite interroger des personnes non-expertes, il peut être difficile pour elles de comprendre ce que sont des sous-ensembles flous et la manière de les construire. L'approche *par personne* peut par conséquent d'une part être fastidieuse pour les participants et d'autre part conduire à des biais dus à une mauvaise compréhension

des consignes.

L'approche *par bornes d'intervalles* permet de surmonter cette difficulté en ce qu'un intervalle est une manière naturelle de représenter l'imprécision d'un terme, en particulier numérique. Toutefois, elle repose sur un certain nombre d'hypothèses a priori sur la forme que doit prendre le sous-ensemble flou ainsi que sur les statistiques à appliquer aux bornes des intervalles.

3.2 Interprétation des ENA

La section précédente présente les différents formalismes de représentation de l'imprécision liée à une ENA. Cette section considère plus spécifiquement le problème de leur interprétation, dont l'objectif est de quantifier l'imprécision qu'elles dénotent. Elle présente les trois modèles d'interprétation des ENA que nous avons pu identifier dans la littérature.

Les deux premiers modèles, le modèle proportionnel (*Ratio Model*, RM Krifka, 2007) et le modèle basé sur des échelles (*Scale-Based Model*, SBM Sauerland & Stateva, 2007; Solt, 2014), se situent dans une approche théorique et linguistique alors que le troisième, le modèle régressif (*REGressive Model*, REGM Ferson et al., 2015), se situe dans une approche empirique. Tous trois prennent le parti de l'approche par intervalles (voir section 3.1.1) : étant donnée une ENA "*environ x*", ils construisent un intervalle, noté $I(x)$. Dans chaque cas, nous soulignons les caractéristiques de x qui sont prises en compte. Les sous-sections suivantes détaillent les principes et formalismes de ces trois modèles d'interprétation des ENA. La dernière sous-section présente une synthèse de ces modèles.

3.2.1 Modèle proportionnel RM : la magnitude

Le *Ratio Model*, RM (Krifka, 2007) consiste à modéliser l'interprétation des ENA comme des intervalles en définissant la taille de ceux-ci comme étant un pourcentage fixe, défini par l'utilisateur, de la magnitude de la valeur de référence de l'ENA (Krifka, 2007, p. 116). En notant s ce pourcentage, l'intervalle est formellement défini comme :

$$I_{RM}(x) = [x - x \cdot s, x + x \cdot s] \tag{3.1}$$

Par exemple, en fixant $s = 10\%$, "*environ* 300" est interprété comme l'intervalle $[270, 330]$.

Ce modèle proportionnel est inspiré par la loi de Weber-Fechner impliquée dans la cognition humaine des nombres (voir section 2.2 p. 17). En effet, en conséquence de cette loi, pour une valeur de référence de magnitude x, toutes les valeurs v dont la différence

absolue à x est inférieure à $s \cdot x$, c'est-à-dire telles que $|v - x| \leq s \cdot x$, où s est lié à la fraction de Weber c, sont indistinguables de x.

On voit donc que ce modèle, en ne tenant compte que de la magnitude de la valeur de référence des ENA, prend le parti d'une interprétation uniquement fondée sur la représentation des nombres dans le système approximatif des nombres (voir section 2.2 p. 17) et ne considère pas le système exact des nombres comme impliqué dans ce traitement.

Il peut produire des résultats contre intuitifs pour des ENA dont la valeur de référence possède plusieurs chiffres significatifs. Par exemple, dans le cas où $x = 8150$, avec $s = 10\%$, on obtient $I_{RM}(8150) = [7335, 8965]$: la taille de cet intervalle (1630) peut être considérée comme trop élevée par rapport au degré de précision de x, représenté par son dernier chiffre significatif et les zéros qui le suivent (50). Cette limite, qui n'apparaît pas pour les ENA dont la valeur de référence ne possède qu'un seul chiffre significatif, est due au fait que l'information communiquée par des nombres avec plusieurs chiffres significatifs est plus précise que celle communiquée par les nombres avec un seul chiffre significatif. Le modèle basé sur des échelles, présenté dans la sous-section suivante, traite de cette limite.

3.2.2 Modèle basé sur des échelles SBM : la granularité

L'approche pragmatique a été développée et formalisée par Sauerland et Stateva (2007) et Solt (2014) dans le modèle d'interprétation des ENA basé sur des échelles (SBM).

Principe général Dans ce modèle, un système d'échelles est défini comme un ensemble d'échelles exprimant différents niveaux de granularité de mesure. Par exemple, dans le domaine temporel, le système d'échelles peut être, de la granularité la plus fine à la plus grossière : minutes, quarts d'heure, demi-heures, heures, jours, etc. Dans le système décimal, les échelles peuvent être les unités, dizaines, centaines, etc.

Formellement, le système d'échelles est défini comme $S = \{s_1, \ldots, s_n\}$, où s_i est le $i$ème niveau de granularité, avec $s_{i+1} > s_i$. Un système d'échelles est dit optimal si $s_{i+1} = a_i \cdot s_i$, avec $a_i \in \mathbb{N}^*$ (par ex., le système décimal $S = \{1, 10, 100, \ldots\}$, où $\forall i, a_i = 10$). L'ensemble des niveaux de granularité $Grans(x) = \{g_1(x), \ldots, g_m(x)\}$ auxquels appartient un nombre x est défini comme :

$$Grans(x) = \bigcup_{\{s_i \in S | x \mod s_i = 0\}} \{s_i\} \tag{3.2}$$

Par exemple, dans le système décimal, le nombre 100 appartient à trois niveaux : 1, 10 et 100, alors que 112 n'appartient qu'au niveau de granularité 1.

L'intervalle de valeurs dénotées par une expression numérique impliquant un nombre x dépend alors du niveau de granularité considéré : il contient les valeurs plus proches de x

que de tout autre point au niveau de granularité $g_i(x)$, formellement :

$$I^i_{SBM}(x) = \left[x - \frac{g_i(x)}{2}, x + \frac{g_i(x)}{2}\right] \tag{3.3}$$

L'interprétation d'une expression numérique peut être réalisée à tout niveau de granularité auquel elle appartient (Solt, 2014).

Considérons, par exemple, le cas $x = 200$. Les niveaux de granularité auxquels il appartient dans le système décimal sont $Grans(200) = \{1, 10, 100\}$. On peut penser comme raisonnable que l'interprétation se fasse au niveau des dizaines, correspondant à $I^2_{SBM}(200) = [195, 205]$, ou au niveau des centaines, donnant lieu à l'intervalle $I^3_{SBM}(200) = [150, 250]$.

Cas des ENA Dans le cas particulier des expressions numériques, les approximateurs sont utilisés pour modifier leur halo pragmatique (voir section 3.1.1 p. 23). Le locuteur peut s'en servir pour exprimer explicitement le niveau de granularité à utiliser pour l'interprétation (Solt, 2014). Dans le cas des ENA, pour lesquelles l'approximateur est "*environ*", le niveau de granularité le plus grossier (*Coarsest*) $Gran_C(x)$ est utilisé (Solt, 2014). Formellement, $Gran_C(x)$ est calculé comme :

$$Gran_C(x) = \sup_{\{s_i \in S | x \mod s_i = 0\}} s_i \tag{3.4}$$

L'intervalle de valeurs dénotées par une ENA "*environ x*" est alors être formellement défini comme :

$$I_{SBM}(x) = \left[x - \frac{Gran_C(x)}{2}, x + \frac{Gran_C(x)}{2}\right] \tag{3.5}$$

Par exemple, $I_{SBM}(300) = [250, 350]$ et $I_{SBM}(340) = [335, 345]$.

Par conséquent, selon le modèle basé sur des échelles, la taille de l'intervalle correspondant à une ENA est égale au niveau de granularité le plus grossier auquel appartient sa valeur de référence.

Caractéristiques Cette approche théorique présente l'avantage de tenir compte de la granularité d'une ENA à travers un système d'échelles. On voit donc que ce modèle prend le parti de considérer que l'interprétation d'une ENA est réalisée au niveau du système exact des nombres (voir section 2.2 p. 17) et n'implique pas la magnitude de la valeur de référence et donc le système approximatif des nombres. Toutes les ENA appartenant au même niveau de granularité, quelle que soit la magnitude de la valeur de référence, conduisent à la même taille d'intervalle. Par exemple, $|I_{SBM}(8150)| = |I_{SBM}(150)| = |I_{SBM}(50)| = 10$. Or, si le système approximatif des nombres est également impliqué dans l'interprétation des ENA, ces intervalles devraient être différents les uns des autres dans la mesure où la magnitude de la valeur de référence varie.

De plus, le modèle SBM est théorique et, à notre connaissance, n'a pas fait l'objet d'une validation empirique.

3.2.3 Modèle régressif REGM : magnitude, granularité et *fiveness*

A notre connaissance, seuls Ferson et al. (2015) ont empiriquement collecté des données pour proposer un modèle définissant quantitativement l'imprécision communiquée par des approximateurs.

Leur étude a consisté à demander à des participants de donner les valeurs des bornes des intervalles correspondant à différentes expressions numériques contenant des approximateurs, comme : "*Greece enjoys* ***more than 250 days*** *of sunshine a year*" ou "*Bats make up* ***about 20%*** *of all classified mammal species globally*".

Le but des auteurs est de tester la pertinence de plusieurs dimensions arithmétiques de la valeur de référence des ENA comme prédicteurs de la taille des intervalles, en se basant sur des régressions linéaires.

Définitions Trois dimensions caractéristiques des nombres sont introduites pour rendre compte de l'interprétation des ENA, dont les effets sont considérés individuellement, deux à deux et tous ensemble : la *rondeur* (*roundness*), l'ordre de magnitude et la *fiveness*. La rondeur $R(x)$ est définie comme la position décimale du dernier chiffre significatif (par ex., $R(13) = 1$ et $R(130) = 2$) et est liée à la granularité définie dans la section précédente : $R(x) = \log_{10}(Gran_C(x))$.

L'ordre de magnitude $z(x)$ est lié à la valeur de référence x de l'ENA : $O_m(x) = \log_{10}(x)$.

Enfin, la *fiveness* $f(x)$ dépend de la valeur du dernier chiffre significatif : elle vaut 1 si le dernier chiffre significatif est 5 (par ex., 150) et 0 sinon (par ex., 140).

Dans ce modèle régressif REGM, l'intervalle correspondant à une ENA est formellement défini comme :

$$I_{REGM}(x) = \left[x - \frac{10^{L(x)}}{2}, x + \frac{10^{L(x)}}{2}\right] \tag{3.6}$$

où $L(x)$ est calculé comme :

$$\begin{aligned} L(x) = {} & \omega_0 + \omega_1 \cdot O_m(x) + \omega_2 \cdot R(x) + \omega_3 \cdot f(x) \\ & + \omega_4 \cdot O_m(x) \cdot R(x) + \omega_5 \cdot O_m(x) \cdot f(x) + \omega_6 \cdot R(x) \cdot f(x) \\ & + \omega_7 \cdot O_m(x) \cdot R(x) \cdot f(x) \end{aligned} \tag{3.7}$$

où ω_0 à ω_7 sont des paramètres empiriquement déterminés en réalisant une régression linéaire sur un ensemble de données collectées.

Les auteurs mettent en évidence que la combinaison de ces trois dimensions arithmétiques est un bon prédicteur de la taille des intervalles correspondant à des ENA. De plus,

ils montrent qu'elles sont impliquées dans l'interprétation des ENA à une échelle logarithmique : la rondeur et l'ordre de magnitude sont calculés comme les logarithmes de la granularité et de la magnitude de la valeur de référence.

Caractéristiques Ces résultats sont cohérents avec la manière dont sont traités les nombres et les quantités par le système cognitif humain (voir section 2.2 p. 17). En effet, l'ordre de magnitude correspond à la manière dont le système approximatif des nombres encode les quantités et la *rondeur*, liée à la granularité, renvoie au traitement opéré par le système exact des nombres en tant qu'interpréter une ENA est considéré comme un problème formel. Ce modèle appuie par conséquent l'hypothèse d'un compromis réalisé entre les deux sous-systèmes cognitifs responsables du traitement des nombres dans l'interprétation des ENA. De plus, ces dimensions relatives à la valeur de référence semblent impliquées à une échelle logarithmique, ce qui est cohérent avec le fait que la ligne mentale des nombres du système approximatif des nombres est logarithmiquement compressée (voir section 2.2 p. 17).

Toutefois, d'un point de vue méthodologique, une limite de l'étude de Ferson et al. (2015) tient au fait que les auteurs ne contrôlent pas le contexte sémantique des ENA. En effet, bien que le type d'unité auquel est rapportée chaque expression (discrète, longueur, temps, etc.) soit contrôlé, le questionnaire utilisé mélange plusieurs contextes sémantiques, ce qui peut conduire à différentes interprétations et donc à différents intervalles pour une même valeur de référence (Lasersohn, 1999; Sadock, 1977; Williamson, 1994). Par conséquent, une étude dont le but est d'identifier les dimensions arithmétiques impliquées dans l'interprétation des ENA devrait utiliser comme matériel des ENA non contextualisées ou un contexte contrôlé.

3.2.4 Modèles d'interprétation des ENA : synthèse

De l'examen des trois modèles existants, RM (Krifka, 2007), SBM (Sauerland & Stateva, 2007; Solt, 2014) et REGM (Ferson et al., 2015), il ressort que la magnitude, la granularité (liée à la *rondeur* dans Ferson et al., 2015), et le dernier chiffre significatif de la valeur de référence d'une ENA sont des dimensions clés en jeu dans son interprétation. Cette analyse est cohérente avec la manière dont le système cognitif humain traite les nombres et les quantités (voir section 2.2 p. 17). En effet, la magnitude renvoie à l'encodage des quantités dans le système approximatif des nombres alors que la granularité et le dernier chiffre significatif renvoient au traitement opéré par le système exact des nombres lors de l'interprétation des ENA.

On peut également noter que les trois modèles de la littérature considèrent implicitement que les intervalles correspondant aux ENA sont symétriques, centrés autour de la

valeur de référence (par ex., $I(100) = [95, 105]$).

Enfin, ces modèles se situent dans le cadre de l'approche épistémique du vague et ne rendent pas compte du principe de tolérance lié aux expressions vagues. Par conséquent, les intervalles qu'ils définissent sont supposés correspondre à l'interprétation de tout un chacun. Or, qu'on se place dans la perspective épistémique ou la perspective sémantique du vague, un intervalle peut sembler correct pour une personne dans un contexte donné mais incorrect pour une autre personne ou dans un contexte.

3.3 Combinaison des ENA dans le calcul

Au delà de la question de la représentation et de l'interprétation des ENA, se pose celle de leur combinaison dans le calcul. Nous examinons ici les études issues de la psychologie cognitive qui traitent du calcul imprécis.

Les adultes, les enfants et les animaux dépourvus de langage verbal sont capables de réaliser des calculs arithmétiques avec des opérandes approximatives (Barth et al. 2006; 2008 ; Gilmore et al., 2010; Knops et al., 2009), en particulier l'addition et la soustraction.

Ces habilités cognitives sont examinées dans le cadre d'études portant sur le système approximatif des nombres (voir section 2.2 p. 17). Les méthodes mises en œuvre dans les études portant sur le calcul approximatif se déroulent en deux étapes. La première consiste à présenter deux tableaux de points aux sujets, correspondant à une représentation non-symbolique des opérandes considérées. Dans la seconde étape, un ou plusieurs tableaux de points sont présentés, correspondant au résultat du calcul considéré. Dans le premier cas, les sujets doivent dire si le nombre de points présents dans le dernier tableau est inférieur ou supérieur à la somme ou à la soustraction des deux opérandes (Barth et al. 2006; 2008 ; Gilmore et al., 2010). Dans le second cas, le sujet doit sélectionner parmi plusieurs tableaux de points celui qui correspond au résultat de l'opération (Knops et al., 2009).

Ainsi, dans ces études, le problème de l'imprécision et de sa propagation n'est pas directement traité. En effet, à aucun moment, l'imprécision associée aux opérandes n'est évaluée, par exemple, en demandant au participant d'estimer la plage de quantités de points que contient un tableau présenté. De plus, il ne lui est pas demandé de produire lui-même le résultat d'un calcul imprécis. Les données ainsi collectées ne permettent donc pas de mettre en relation l'imprécision associée aux opérandes et celle associée au résultat. Pourtant, l'étude de la propagation des imprécisions dans le calcul imprécis, que nous nous proposons de traiter dans cette thèse, ne peut faire l'économie d'une estimation, de la part des participants, des imprécisions liées aux opérandes et au résultat d'un calcul.

De plus, les opérandes utilisées dans ces études sont représentées non-symboliquement, contrairement aux ENA considérées dans cette thèse. A notre connaissance, aucune étude

du calcul impliquant des opérandes symboliques, telles que les ENA, n'a été publiée.

Bien que la problématique de ces études diffère sensiblement de celle que nous nous proposons de traiter, on peut relever que la loi de Weber-Fechner (voir section 2.2 p. 17) apparaît comme valable pour les calculs imprécis (Knops et al., 2009) : ceci indique que le système approximatif des nombres est impliqué dans le calcul imprécis, bien que le calcul arithmétique soit une tâche qui relève typiquement du système exact des nombres (Dehaene, 2003).

3.4 Bilan

Dans ce chapitre, nous avons présenté un état de l'art portant sur le traitement des ENA, selon trois axes : leur représentation formelle, leur interprétation et leur combinaison dans le calcul.

En premier lieu, nous avons présenté deux approches formelles classiquement utilisées pour représenter les imprécisions numériques : l'approche par intervalles, correspondant à la perspective épistémique du vague et l'approche par sous-ensembles flous, correspondant à la perspective sémantique. Ces deux approches sont celles qui sont mises en œuvre dans nos travaux pour représenter l'imprécision dénotée par une ENA (voir chapitre 5).

Dans un deuxième temps, nous avons examiné trois modèles d'interprétation des ENA identifiés dans la littérature. Cet examen suggère que trois dimensions arithmétiques de la valeur de référence d'une ENA, formalisées différemment selon le modèle considéré, seraient impliquées dans son interprétation : la magnitude, la granularité et le dernier chiffre significatif. Il nous a également permis de souligner l'hypothèse implicite de ces modèles, selon laquelle la plage de valeurs dénotées par une ENA est centrée sur sa valeur de référence. Cet examen motive nos travaux présentés dans le chapitre 4, portant sur une définition formelle des dimensions arithmétiques de la valeur de référence des ENA et sur l'étude de leur implication dans l'interprétation des ENA.

Enfin, dans la dernière section de ce chapitre, nous avons présenté les études qui peuvent être mises en rapport avec la problématique du calcul imprécis. Celles-ci portent essentiellement sur l'arithmétique appliquée à des quantités imprécises représentées visuellement. A notre connaissance, aucune étude portant sur le traitement et la propagation des imprécisions exprimées verbalement n'a été publiée à ce jour. Cette lacune théorique motive nos travaux décrits dans le chapitre 8.

Chapitre 4

Interprétation des ENA : définition et validation empirique des dimensions

Ce chapitre présente notre contribution à l'étude de l'interprétation des Expressions Numériques Approximatives par l'être humain, selon deux axes. Le premier consiste à définir et formaliser les dimensions arithmétiques et cognitive rendant compte de l'interprétation des ENA, à partir de la revue de littérature et en tenant compte de la manière dont l'être humain traite et représente cognitivement les nombres.

Le second axe est une étude empirique au cours de laquelle des données réelles sont collectées auprès de non-experts, qui nous permettent de valider les dimensions arithmétiques que nous proposons comme impliquées dans l'interprétation des ENA. Cette étude constitue la première étape vers la modélisation computationnelle de l'interprétation : elles seront ensuite combinées dans le cadre de modèles computationnels décrits et expérimentalement validés dans le chapitre 5.

Dans la section 4.1, nous définissons et formalisons les dimensions arithmétiques et cognitive des ENA que nous proposons pour rendre compte de leur interprétation. La problématique et les hypothèses de cette étude sont détaillées dans la section 4.2. La section 4.3 décrit les méthodes mises en œuvre pour collecter, nettoyer, pré-traiter et analyser les données. La section 4.4 présente les résultats obtenus. Ceux-ci sont discutés à la lumière de la littérature dans la section 4.5.

Plusieurs parties de ce chapitre ont fait l'objet d'une publication, Lefort et al. (2017c).

Dimension	Définition formelle	Exemple
Magnitude	x	8150
Granularité	$Gran(x) = 10^{i^*}$ où $i^* = \min\{i \mid a_i \neq 0\}$	10
Dernier chiffre significatif	$LSD(x) = a_{i^*}$	5
Magnitude relative	$RelMag(x) = a_{i^*} \cdot 10^{i^*}$	50
Nombre de chiffres significatifs	$NSD(x) = q - i^* + 1$	3
Complexité	$Cpx(x) = NSD(x) - B(x)$	2.5

TABLEAU 4.1 – Les six dimensions que nous proposons d'un entier positif $x = \sum_{i=0}^{q} a_i \cdot 10^i$, illustrées pour $x = 8150$ dans la dernière colonne. $B(x)$, utilisé pour le calcul de la complexité, est défini dans l'équation (4.1).

4.1 Définition et formalisation des dimensions des ENA

Cette section introduit les définitions et notations des dimensions caractérisant les ENA, que nous avons conceptualisées à partir des caractéristiques sur lesquelles sont basés les modèles de la littérature (voir section 3.2 p. 29) ainsi qu'en les rapportant à la manière dont l'être humain traite et représente cognitivement les nombres et quantités (voir section 2.2 p. 17).

Les ENA considérées dans cette étude sont de la forme "*environ* x", où $x \in \mathbb{N}^*$.

Pour isoler les dimensions liées à la valeur de référence d'une ENA, nous proposons de définir six caractéristiques d'un entier positif, résumées dans le tableau 4.1. Deux types, faisant respectivement l'objet des sections 4.1.1 et 4.1.2 sont distingués : les dimensions arithmétiques, en référence aux propriétés formelles des nombres et une dimensions cognitive, en référence à la représentation des nombres dans le système cognitif humain.

4.1.1 Dimensions arithmétiques

Dans le système décimal, un entier positif x s'écrit comme une somme pondérée de puissances de 10, $x = \sum_{i=0}^{q} a_i \cdot 10^i$, où $a_i \in [\![0, 9]\!]$ étant ses chiffres.

Nous proposons de considérer cinq dimensions arithmétiques de x dont les définitions formelles sont données dans la deuxième colonne du tableau 4.1, avec un exemple pour le cas $x = 8150$. Les trois premières, la magnitude, la granularité et le dernier chiffre significatif sont celles dont nous supposons qu'elles sont impliquées dans l'interprétation des ENA (voir section 4.2 p. 41). Les deux autres, le nombre de chiffres significatifs et la magnitude relative sont utilisés dans la partie modélisation computationnelle (voir chapitre 5) comme caractéristiques des ENA.

Nous définissons la *magnitude* comme la valeur globale de x. Cette dimension est à rapprocher de l'ordre de grandeur proposé par Ferson et al. (2015) : $O_m(x) = log_{10}(x)$.

La *granularité* est définie comme l'ordre de grandeur auquel le dernier chiffre significatif de x est associé dans le système décimal. Cette définition de la granularité peut être rapprochée de celle issue du modèle SBM $Gran_C(x)$ (voir section 3.2.2 p. 30). En effet, $Gran_C(x) = Gran(x)$ lorsque le système d'échelles sélectionné est le système décimal. La granularité telle que définie ici est également liée à la notion de rondeur proposée par Ferson et al. (2015) qui est égale à son logarithme en base 10 : $Gran(x) = 10^{R(x)}$.

Le *dernier chiffre significatif*, noté $LSD(x)$ (*Last Significant Digit*) est une dimension plus générale que la *fiveness* proposée par Ferson et al. (2015). En effet, les auteurs considèrent le chiffre 5 comme un cas particulier de dernier chiffre significatif par rapport aux autres, alors que nous proposons de considérer toutes ses valeurs possibles comme des cas distincts.

La *magnitude relative* $RelMag(x)$ est le produit de la granularité par le dernier chiffre significatif. Enfin, le *nombre de chiffres significatifs* de x est noté $NSD(x)$ (*Number of Significant Digits*).

Parmi ces cinq dimensions, trois nous paraissent particulièrement pertinentes quant à la représentation et à l'interprétation des ENA : la magnitude, la granularité et le dernier chiffre significatif. En effet, d'une part les travaux de Ferson et al. (2015) suggèrent qu'elles sont impliquées dans l'interprétation des ENA. D'autre part, elles renvoient aux deux sous-systèmes responsables de la cognition humaine des nombres (voir section 2.2 p. 17) : la magnitude correspond à la manière dont les quantités sont encodées dans le système approximatif des nombres, qui ne tient compte que de la valeur de référence x. La granularité et le dernier chiffre significatif renvoient quant à eux au système exact des nombres dans la mesure où l'interprétation des ENA est traitée par ce système comme un problème formel, à la manière du modèle SBM (voir section 3.2.2 p. 30).

4.1.2 Dimension cognitive : la complexité

Au-delà des dimensions arithmétiques, nous proposons d'introduire une mesure de complexité, notée $Cpx(x)$, pour capturer la notion de saillance cognitive des nombres, c'est-à-dire le fait que certains nombres sont plus facilement évoqués que d'autres par les êtres humains (voir section 2.2 p. 17).

Dans le système numérique arabe, deux caractéristiques des nombres influencent leur fréquence d'apparition. Premièrement, plus un nombre possède de chiffres significatifs et moins il apparaît fréquemment (Dehaene & Mehler, 1992). Deuxièmement, les nombres dont le dernier chiffre significatif est 5 et, dans une moindre mesure, 2, apparaissent plus

fréquemment (Jansen & Pollmann, 2001).

En prenant en compte ces observations, nous proposons de définir une mesure de complexité pour les entiers positifs exprimés dans le système décimal qui dépend de leur nombre de chiffres significatifs $NSD(x)$ et de la valeur de leur dernier chiffre significatif $LSD(x)$. Plus précisément, nous proposons de formaliser la complexité d'un entier positif comme son nombre de chiffres significatifs auquel est soustrait un bonus si la valeur de son dernier chiffre significatif est 2 ou 5 et son nombre de chiffres significatifs supérieur ou égal à 2 (voir tableau 4.1 page précédente et équation 4.1 ci-dessous).

Pour des raisons de symétrie autour des multiples de 10, nous proposons de considérer le cas des entiers dont le dernier chiffre significatif est 8 de la même manière que le cas des entiers dont le dernier chiffre significatif est 2 : tous les entiers tels que $LSD(x) = 2$ sont de la forme $(10 \cdot k + 2) \cdot 10^{i^*}$ (par ex., $320 = (30+2) \cdot 10$). Nous proposons de traiter de la même manière, dans la fonction de bonus $B(x)$, leurs symétriques par rapport à $(10 \cdot k) \cdot 10^{i^*}$, c'est-à-dire les entiers tels que $LSD(x) = 8$, qui sont de la forme $(10 \cdot k - 2) \cdot 10^{i^*}$ (par ex., $280 = (30 - 2) \cdot 10$). Cet argument de symétrie n'a pas besoin d'être appliqué aux entiers dont le dernier chiffre est significatif est 5 puisque leurs symétriques vérifient la même propriété : $LSD((10 \cdot k - 5) \cdot 10^{i^*}) = LSD((10 \cdot k + 5) \cdot 10^{i^*}) = 5$.

La fonction de bonus $B(x)$ distingue par conséquent trois catégories d'entiers et prend trois valeurs différentes, selon la valeur du dernier chiffre significatif $LSD(x)$, en respectant l'ordre de la fréquence d'apparition des entiers : $B(x_1) > B(x_2) > B(x_3)$, pour $x_1, x_2, x_3 \in \mathbb{N}^*$ tels que $LSD(x_1) = 5$, $LSD(x_2) \in \{2, 8\}$ et $LSD(x_3) \notin \{2, 5, 8\}$.

Comme l'exploitation proposée de la mesure de complexité est ordinale (voir section 5.1 p. 62), les valeurs choisies sont uniquement contraintes par l'ordre induit. Nous proposons de les fixer arbitrairement à 0,5, 0,25 et 0. La fonction de bonus est dès lors formalisée comme :

$$B(x) = \begin{cases} 0,5 & \text{si } LSD(x) = 5 \text{ et } NSD(x) > 1 \\ 0,25 & \text{si } LSD(x) \in \{2, 8\} \text{ et } NSD(x) > 1 \\ 0 & \text{sinon} \end{cases} \tag{4.1}$$

La figure 4.1 illustre les valeurs de la complexité $Cpx(x) = NSD(x) - B(x)$ pour tous les entiers x compris entre 400 et 500. La valeur minimale est 1, pour les entiers ne possédant qu'un seul chiffre significatif, soit 400 et 500. Sur cette plage, la valeur maximale de la complexité est 3, pour les entiers possédant trois chiffres significatifs et donnant lieu à un bonus nul. On peut remarquer que 450 est l'unique valeur se voyant attribuer une complexité égale à 1,5 car il s'agit de la seule de cette plage qui possède deux chiffres significatifs, dont le dernier est 5, donnant lieu à un bonus de 0,5.

Un entier dont la complexité est moins élevée est considéré comme plus saillant qu'un autre dont la complexité est plus élevée. Par exemple, 400 est plus saillant que 401 car

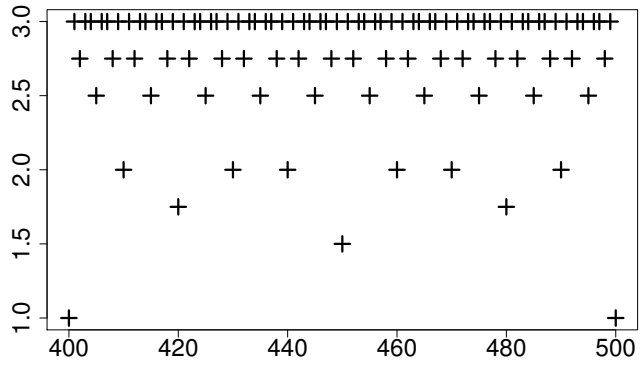


FIGURE 4.1 – Complexité $Cpx(x)$ des valeurs entières comprises entre 400 et 500.

$Cpx(400) = 1$ et $Cpx(401) = 3$.

4.2 Validation des dimensions arithmétiques : problématique et hypothèses

L'objectif de l'étude présentée dans ce chapitre consiste à valider empiriquement l'implication des dimensions arithmétiques proposées dans la section précédente dans l'interprétation d'ENA sémantiquement non contextualisées. Plus précisément, le but est triple :

1. Valider empiriquement l'implication de la magnitude, de la granularité et du dernier chiffre significatif dans l'interprétation des ENA (hypothèses H1 et H2).
2. Tester empiriquement l'hypothèse implicite des modèles de la littérature (Ferson et al., 2015; Krifka, 2007; Sauerland & Stateva, 2007; Solt, 2014) de symétrie des intervalles correspondant aux ENA (hypothèse H3).
3. Collecter explicitement les heuristiques de production déployées par des êtres humains pour estimer les bornes des intervalles correspondant à des ENA (hypothèse H4).

Quatre hypothèses principales correspondant aux buts indiqués sont décrites ci-dessous, avec les hypothèses dérivées.

Formellement, nous notons l'intervalle correspondant à l'ENA "*environ x*" $I(x) = [x - \Delta^-(x), x + \Delta^+(x)]$.

H1 - L'état de l'art (voir chapitre 3) suggère que la magnitude, la granularité et le dernier chiffre significatif sont des dimensions clés de la valeur de référence d'une ENA en jeu dans

son interprétation, on devrait donc observer l'implication individuelle de chacune de ces trois dimensions :

H1.1 - Magnitude : Ferson et al. (2015) ont montré que la magnitude d'une ENA est l'un des prédicteurs de la taille de l'intervalle correspondant (voir section 3.2.3 p. 32). De plus, la magnitude correspond à la quantité qui est encodée dans le système approximatif des nombres. Par conséquent, si ce système cognitif est impliqué dans la représentation et l'interprétation des ENA, nous devrions observer un effet de la magnitude sur les intervalles collectés. Nous prédisons donc qu'à granularité et dernier chiffre significatif constants, des magnitudes différentes conduisent à des distances $\Delta(x)$ différentes entre les bornes des intervalles et la valeur de référence x des ENA.

Plus précisément, plus la magnitude est élevée, plus la distance est importante, formellement :

$$\text{si } x > y, Gran(x) = Gran(y) \text{ et } LSD(x) = LSD(y), \text{ alors } \Delta(x) > \Delta(y).$$

Par exemple, $8150 > 50$, $Gran(8150) = Gran(50) = 10$ et $LSD(8150) = LSD(50) = 5$, par conséquent $\Delta(8150) > \Delta(50)$.

H1.2 - Granularité : selon le modèle SBM et le modèle REGM, la granularité est l'une des dimensions en jeu dans l'interprétation des ENA. Plus spécifiquement, selon le modèle SBM, une ENA est interprétée à son niveau de granularité le plus élevé et dénote l'ensemble des valeurs qui sont plus proches de la valeur de référence de l'ENA que d'un autre point situé à ce niveau de granularité.

De plus, du point de vue cognitif, cette dimension est celle qui correspond à l'interprétation des ENA par le système exact des nombres.

Par conséquent, nous nous attendons à observer un effet de la granularité sur les intervalles : à dernier chiffre significatif constant, des granularités différentes donnent lieu à des distances différentes entre les bornes de l'intervalle et la valeur de référence de l'ENA.

Plus précisément, l'hypothèse postule que plus la granularité est élevée, plus la distance doit être importante, formellement :

$$\text{si } Gran(x) > Gran(y) \text{ et } LSD(x) = LSD(y), \text{ alors } \Delta(x) > \Delta(y).$$

Par exemple, $Gran(5000) = 1000$, $Gran(500) = 100$, $LSD(5000) = LSD(500) = 5$, par conséquent $\Delta(5000) > \Delta(500)$.

H1.3 - Dernier chiffre significatif : bien que cette dimension ne soit pas prise en compte dans le modèle SBM, la manière dont le système exact des nombres traite

l'interprétation des ENA comme un problème formel suggère qu'à granularité constante, un dernier chiffre significatif différent devrait donner lieu à une distance différente entre les bornes de l'intervalle et la valeur de référence d'une ENA (voir section 2.2 p. 17).

Plus spécifiquement, l'hypothèse postule que plus le dernier chiffre significatif est élevé, plus la distance doit être importante, formellement :

$$\text{si } LSD(x) > LSD(y) \text{ et } Gran(x) = Gran(y), \text{ alors } \Delta(x) > \Delta(y).$$

Par exemple, $LSD(80) = 8$, $LSD(50) = 5$, $Gran(80) = Gran(50) = 10$ donc $\Delta(80) > \Delta(50)$.

H2 - Pertinence d'un modèle tridimensionnel à l'échelle logarithmique pour l'interprétation des ENA.

H2.1 - Pertinence d'un modèle tridimensionnel : nous postulons qu'un modèle qui tient compte à la fois de la magnitude, de la granularité et du dernier chiffre significatif devrait mieux rendre compte des intervalles observés que les modèles basés sur les dimensions proposées dans la littérature.

H2.2 - Pertinence d'un modèle à l'échelle logarithmique : les travaux de Ferson et al. (2015) suggèrent que la relation entre les bornes des intervalles et les dimensions arithmétiques des ENA se produit à une échelle logarithmique (par ex., $\log_{10}(\Delta(x)) = \alpha \cdot \log_{10}(Gran(x))$). De plus, la ligne mentale du système approximatif des nombres, sur laquelle sont encodés les nombres et quantités, est logarithmiquement compressée (voir section 2.2 p. 17). Par conséquent, les modèles devraient mieux rendre compte des intervalles collectés lorsque leurs variables sont à une échelle logarithmique.

H3 - Symétrie des intervalles Comme détaillé au chapitre 3, les modèles de la littérature postulent que les intervalles correspondant aux ENA sont centrés autour de leur valeur de référence.

Cela nous conduit à considérer l'hypothèse de symétrie des intervalles, c'est-à-dire que la distance entre la borne inférieure de l'intervalle et la valeur de référence de l'ENA devrait être égale à la distance entre la borne supérieure et la valeur de référence, formellement : $\Delta^-(x) = \Delta^+(x)$. Par exemple, $I(100) = [100 - 10, 100 + 10] = [90, 110]$.

H4 - Heuristiques de production des intervalles Si la magnitude, la granularité, et la valeur du dernier chiffre significatif sont les dimensions clés de l'interprétation des ENA, les participants de l'étude empirique devraient évoquer explicitement celles-ci dans

la description des heuristiques qu'ils mettent en place lorsque l'on leur demande s'ils ont mis au point une stratégie pour élaborer leurs réponses.

Plus spécifiquement, nous nous attendons à ce que des termes comme "*magnitude*", "*valeur globale*" ou "*précision*" ou des synonymes apparaissent plus fréquemment dans les descriptions que d'autres propriétés des nombres. Cependant, nous ne nous attendons pas à ce que les participants emploient les termes techniques *granularité* ou *dernier chiffre significatif*.

4.3 Méthodes

Nous décrivons dans cette section la méthodologie que nous avons mise en œuvre pour collecter, pré-traiter, nettoyer et analyser les données nous permettant de tester les quatre hypothèses considérées dans cette étude.

4.3.1 Population

146 adultes se sont portés volontaires pour cette étude, 102 femmes et 44 hommes, âgés entre 20 et 70 ans ($M = 38,6$; $\sigma = 14,2$). Tous ont pour langue maternelle le français et ont été recrutés via une liste de diffusion sur internet.

4.3.2 Matériel

Un questionnaire a été conçu pour collecter les intervalles correspondant à 24 ENA, dont les valeurs de référence sont listées dans le tableau 4.2. Ce questionnaire comporte trois parties.

La première vise à collecter les intervalles correspondant aux 24 ENA. Les valeurs de référence de celles-ci sont sélectionnées pour couvrir différentes combinaisons de dimensions, de manière à tester les hypothèses considérées : plusieurs valeurs du dernier chiffre significatif à granularité constante (dizaines : 20/30/40/50/80 ; centaines : 100/200/400/500/600/800 ; milliers : 1000/2000/6000/8000) ; plusieurs granularités à dernier chiffre significatif constant (20/200/2000 ; 40/400 ; 50/500 ; 600/6000 ; 80/800/8000) ; plusieurs magnitudes à granularité et dernier chiffre significatif constants (40/440 ; 100/1100 ; 500/1500 ; 30/4730 ; 50/150/8150).

Les ENA ne sont pas sémantiquement contextualisées, aucun indice n'est donné aux participants quant à ce qui est mesuré ou compté. Il se peut que les participants apportent leur propre contexte aux ENA présentées. Toutefois, il est raisonnable de penser que ce contexte est le même pendant toute la durée de l'expérience, si bien que l'on peut détecter les différences d'une ENA à l'autre. Lorsque les différences observées entre ENA sont si-

x	$Gran(x)$	$LSD(x)$	$RelMag(x)$	$NSD(x)$
20	10	2	20	1
30	10	3	30	1
40	10	4	40	1
50	10	5	50	1
80	10	8	80	1
100	100	1	100	1
110	10	1	10	2
150	10	5	50	2
200	100	2	200	1
400	100	4	400	1
440	10	4	40	2
500	100	5	500	1
560	10	6	60	2
600	100	6	600	1
800	100	8	800	1
1000	1000	1	1000	1
1100	100	1	100	2
1500	100	5	500	2
2000	1000	2	2000	1
4700	100	7	700	2
4730	10	3	30	3
6000	1000	6	6000	1
8000	1000	8	8000	1
8150	10	5	50	3

TABLEAU 4.2 – Valeurs de référence des ENA utilisées dans le questionnaire et leurs dimensions arithmétiques : magnitude (x), granularité ($Gran$), dernier chiffre significatif (LSD), magnitude relative ($RelMag$) et nombre de chiffres significatifs (NSD), tels que définis dans la section 4.1 p. 38.

milaires d'un participant à l'autre, elles suggèrent l'implication d'une dimension spécifique à la valeur de référence des ENA, telle que celles que nous proposons, indépendamment donc du contexte.

La consigne donnée aux participants est : "*Selon vous, entre quelles valeurs MINIMALE et MAXIMALE se trouve environ x ?*", x étant la valeur de référence de l'ENA considérée.

La seconde partie du questionnaire comporte deux questions dont le but est de contrôler

la variabilité inter-individuelle quant :

- à l'usage des mathématiques, de manière à contrôler si un usage quotidien des mathématiques au travail ou dans les études affecte l'interprétation des ENA.
- au niveau subjectif en calcul mental : comme information complémentaire à la précédente question, le participant estime, sur une échelle de type Likert graduée de 1 à 5, son niveau en calcul mental.

La troisième et dernière partie du questionnaire porte sur la ou les stratégie(s) mise(s) en place par les participants pour estimer les bornes des intervalles correspondant aux ENA. La consigne de cette partie est : "*Pensez-vous avoir utilisé une stratégie pour répondre aux questions qui vous ont été posées ? Si oui, pouvez-vous la décrire ci-dessous ?*"

4.3.3 Procédure

L'ordre de présentation des ENA du questionnaire est aléatoirement tiré pour chaque participant. Celui-ci répond en donnant les bornes inférieure et supérieure de l'intervalle qu'il estime être dénoté en tapant leur valeur dans l'espace blanc dédié, en une seule fois.

Une fois qu'un participant a donné tous les intervalles, il lui est demandé de répondre aux deux questions portant sur l'usage des mathématiques et le niveau subjectif en calcul mental.

Enfin, il est demandé aux participants s'ils ont mis en place une ou des stratégie(s) pour estimer les intervalles correspondant aux ENA proposées. Dans le cas d'une réponse positive, le participant est invité à décrire librement cette ou ces stratégie(s).

4.3.4 Pré-traitement des données

L'intervalle donné par le participant p pour l'ENA "*environ* x" est noté $I_p(x) = [I_p^-(x), I_p^+(x)]$. Les intervalles sont utilisés pour calculer deux distances absolues $\Delta P_p^b(x)$ entre la valeur de référence x de l'ENA et les bornes $b \in \{-, +\}$ de l'intervalle : $\Delta P_p^b(x) = |I_p^b(x) - x|$. Par conséquent, $I_p(x) = [x - \Delta P_p^-(x), x + \Delta P_p^+(x)]$.

Les analyses rapportées dans les sections suivantes sont réalisées sur les distances absolues $\Delta P_p^b(x)$. Cette variable est plus pertinente que la taille de l'intervalle car elle permet de comparer différents intervalles sans perdre l'information de symétrie. En effet, deux tailles d'intervalle peuvent être égales alors que les valeurs des bornes sont différentes (par ex., $|[95, 110]| = |[90, 105]|$).

4.3.5 Nettoyage des données

Pour exclure les paires de données aberrantes (bornes inférieure et supérieure) de l'ensemble des données, celles-ci sont traitées selon une procédure en trois étapes :

1. Les intervalles sont considérés comme aberrants si, soit :
   — ils sont inadéquats ($[0, \text{infini}[$)
   — la borne supérieure est en dessous de la valeur de référence ou la borne inférieure est au dessus de la valeur de référence, formellement : $I_p^+(x) < x$ ou $I_p^-(x) > x$ (par ex., $I_P(800) = [700, 750]$ ou $I_P(800) = [810, 850]$)
   — une borne est dix fois supérieure ou dix fois inférieure à la valeur de référence, formellement : $I_p^+(x) > 10 \cdot x$ ou $I_p^-(x) < \frac{x}{10}$ (par ex., $I_P(100) = [9, 1101]$)
2. Pour chaque borne de chaque ENA, la moyenne et l'écart-type des données restantes après l'étape 1 sont calculés. Toute borne au-delà de trois écart-types de la moyenne est considérée comme aberrante et l'intervalle auquel elle appartient est exclu.
3. 10 participants sont considérés comme non fiables et tous leurs intervalles sont exclus car 70% de ceux-ci sont manquants ou considérés comme aberrants.

Sur les 3504 intervalles collectés, 3177 (91%) sont utilisés dans les analyses statistiques.

4.3.6 Analyses statistiques

Quatre types d'analyses statistiques, détaillés ci-dessous, sont réalisés, selon l'hypothèse considérée. Le seuil de significativité est fixé à $p = 0,01$ dans chaque cas.

H1 - Magnitude, granularité et dernier chiffre significatif Pour tester les hypothèses d'un effet individuel de la magnitude, de la granularité et du dernier chiffre significatif, nous comparons la distribution des distances $\Delta P_p^b(x)$ de plusieurs ENA (par ex., "*environ 20*" vs. "*environ 30*").

Dans la mesure où les distributions observées ne sont pas normales, les comparaisons entre deux ENA sont réalisées en utilisant le test des rangs signés de Wilcoxon, conçu pour des mesures répétées.

Lorsque les comparaisons impliquent plus de deux ENA, le test de Friedman est employé. En cas d'effet significatif détecté par celui-ci, des analyses post-hoc sont réalisées en utilisant le test de Nemenyi, conçu pour des comparaisons multiples deux à deux.

H2 - Modèle tridimensionnel à l'échelle logarithmique Pour tester la pertinence d'une combinaison des trois dimensions arithmétiques que nous proposons, la magnitude, la granularité et le dernier chiffre significatif, dans l'interprétation des ENA, nous proposons de construire un modèle combinant linéairement ces trois dimensions, noté MGLSD.

Nous comparons ce modèle linéaire à trois modèles correspondant aux approches de la littérature détaillées dans le chapitre 3. Pour chacune d'entre elles, un modèle combinant

linéairement les dimensions qu'elles prennent respectivement en compte est construit. La comparaison est réalisée sur la base d'analyses bayésiennes.

Le modèle proportionnel RM (Krifka, 2007) est uniquement basé sur la magnitude, il est donc de la forme : $\Delta P_p^b(x) = \alpha_0 + \alpha_1 \cdot x$.

Le modèle basé sur des échelles SBM (Sauerland & Stateva, 2007; Solt, 2014) ne tient compte que de la granularité $Gran(x)$, il est donc de la forme :
$\Delta P_p^b(x) = \beta_0 + \beta_1 \cdot Gran(x)$.

Le modèle régressif REGM (Ferson et al., 2015) est basé sur une combinaison de la magnitude, de la granularité et de la *fiveness* (voir section 3.2.3 p. 32), formellement :
$\Delta P_p^b(x) = \gamma_0 + \gamma_1 \cdot x + \gamma_2 \cdot Gran(x) + \gamma_3 \cdot f(x)$.

Enfin, l'approche MGLSD que nous proposons comme référence tient compte de la magnitude, de la granularité et du dernier chiffre significatif, formellement :
$\Delta P_p^b(x) = \epsilon_0 + \epsilon_1 \cdot x + \epsilon_2 \cdot Gran(x) + \epsilon_3 \cdot LSD(x)$.

Pour chaque modèle, le facteur de Bayes résultant de sa comparaison à l'absence de relation avec la variable dépendante, $\Delta P_p^b(x)$ est rapporté. Les facteurs obtenus par comparaison entre les modèles eux-mêmes sont ensuite présentés.

Pour tester l'hypothèse H2.2, prédisant que lorsque les modèles sont portés à l'échelle logarithmique, ils rendent mieux compte des intervalles observés, une seconde itération de la procédure décrite ci-dessus est réalisée, en utilisant le logarithme des variables.

On peut toutefois noter que le logarithme de la variable *fiveness* proposée par Ferson et al. (2015) ne peut être calculé car la fonction ne peut prendre que deux valeurs : 0 et 1.

Pour cette hypothèse, les modèles prennent respectivement les formes suivantes :

- modèle proportionnel : $\log_{10}(\Delta P_p^b(x)) = \alpha_0 + \alpha_1 \cdot \log_{10}(x)$
- modèle basé sur des échelles : $\log_{10}(\Delta P_p^b(x)) = \beta_0 + \beta_1 \cdot \log_{10}(Gran(x))$
- modèle régressif :
  $\log_{10}(\Delta P_p^b(x)) = \gamma_0 + \gamma_1 \cdot \log_{10}(x) + \gamma_2 \cdot \log_{10}(Gran(x)) + \gamma_3 \cdot f(x)$
- modèle MGLSD proposé :
  $\log_{10}(\Delta P_p^b(x)) = \epsilon_0 + \epsilon_1 \cdot \log_{10}(x) + \epsilon_2 \cdot \log_{10}(Gran(x)) + \epsilon_3 \cdot \log_{10}(LSD(x))$

H3 - Symétrie des intervalles Pour tester l'hypothèse de symétrie des intervalles, nous proposons de ne pas utiliser le test des rangs signés de Wilcoxon. En effet, ce test donne un résultat significatif lorsque la différence moyenne des rangs entre les distributions est significativement différente de zéro. Par conséquent, il n'est pas adapté pour détecter des différences individuelles entre valeurs de bornes s'il n'y a pas de tendance claire de ces différences dans un sens ou dans l'autre.

Nous proposons donc d'évaluer l'égalité entre la distribution des distances des bornes inférieures à la valeur de référence x et la distribution des distances des bornes supérieures

à la valeur de référence en comptant le nombre de celles-ci qui sont égales.

Formellement, nous proposons que l'égalité Eq entre des deux distances de bornes à la valeur de référence, respectivement Δ_1 et Δ_2, tienne compte d'une erreur relative autorisée de 10%, et soit calculée comme :

$$Eq(\Delta_1, \Delta_2) = \begin{cases} 1 & \text{si } \dfrac{|\Delta_1 - \Delta_2|}{\min(\Delta_1, \Delta_2)} \leq 0,1 \\ 0 & \text{sinon} \end{cases} \tag{4.2}$$

Le score de symétrie $Sym(x)$ pour une ENA "*environ x*", qui agrège l'égalité entre distances de bornes sur l'ensemble des participants $\mathcal{P}(x)$ pour lesquels l'intervalle correspondant à "*environ x*" n'est pas considéré comme aberrant, est calculé comme :

$$Sym(x) = \frac{1}{|\mathcal{P}(x)|} \sum_{p \in \mathcal{P}(x)} Eq(\Delta P_p^-(x), \Delta P_p^+(x)) \tag{4.3}$$

Heuristiques mises en place par les participants Le dernier item du questionnaire en ligne est dédié aux heuristiques mises en place par les participants pour estimer les intervalles correspondant aux ENA proposées. Les réponses à cette question ouverte sont codées et catégorisées selon les trois dimensions arithmétiques évoquées par les participants. La magnitude relative étant la combinaison de la granularité et du dernier chiffre significatif, nous considérons que l'emploi de termes se référant à cette dimension, comme "*précision*" (8150 étant plus précis que 8000 ou 8100 par exemple), renvoie à ces deux dimensions.

Les réponses sont classées dans l'une des quatre catégories suivantes, nommées selon la dimension correspondante :

1. **Constante :** cette heuristique consiste à placer les bornes de l'intervalle à une distance constante de la valeur de référence, quelle que soit l'ENA considérée. Par exemple, "*fourchette systématique entre -5 et +5*".
2. **Magnitude :** les participants ayant rapporté cette heuristique lient la taille de l'intervalle à la magnitude de la valeur de référence de l'ENA, par exemple, "*marges plus ou moins en proportion des grandeurs proposées*".
3. **Magnitude relative :** la magnitude relative, qui combine la granularité et le dernier chiffre significatif, apparaît comme la dimension clé dans cette heuristique : "*je pense que j'avais tendance à faire + ou -2 unités, dizaines, ou centaines, selon si les nombres étaient arrondis à l'unité, la dizaine, la centaine, ou au millier*".
4. **Autre :** bien qu'il ne s'agisse pas à proprement parler d'une heuristique, cette catégorie regroupe toutes les réponses qui ne peuvent être incluses dans les trois précédentes. Ces réponses concernent les heuristiques qui ne sont rapportées que par un seul participant, par exemple : "*pas vraiment, mais il me semblait que le*

> *nombre maximal devait être plus proche du nombre exact, comme s'il ne fallait pas trop dépasser*".

Variabilité inter-individuelle : compétences arithmétiques Deux items du questionnaires sont dédiés aux compétences en arithmétique. L'objectif est d'examiner si l'usage régulier des mathématiques ou le niveau en calcul mental affectent l'interprétation des ENA.

L'effet d'un usage régulier des mathématiques est étudié en utilisant le test de la somme des rangs de Wilcoxon. Nous comparons les valeurs des bornes données par les participants pour chaque ENA séparément pour examiner la différence entre les intervalles donnés par les participants qui font un usage régulier des mathématiques et les intervalles donnés par les autres.

L'effet du niveau en calcul mental est testé en utilisant le même test. Pour ce faire, dans un premier temps, deux groupes sont créés, selon le niveau en calcul mental rapporté : les participants donnant un score compris entre 1 et 3 inclus sont affectés au groupe "faible niveau" ; le groupe "haut niveau" est composé de participants ayant rapporté un score de 4 ou 5.

4.4 Résultats

Dans cette section, nous présentons les résultats obtenus après la collecte et l'analyse des données. Dans un premier temps, nous proposons quelques statistiques descriptives et observations préliminaires. Ensuite, nous détaillons les résultats obtenus, pour chacune des quatre hypothèses considérées dans cette étude, décrites dans la section 4.2 p. 41.

4.4.1 Statistiques descriptives et observation préliminaire

Le tableau B.1 p. 188 de l'annexe B présente les moyennes et médianes des valeurs données par les participants comme bornes des intervalles correspondant aux 24 ENA de l'étude. Les deux dernières colonnes du tableau donnent le nombre de valeurs observées différentes pour chaque borne.

Alors que les modèles de la littérature présentés dans la section 3.2 p. 29 partent de l'hypothèse que l'interprétation d'une ENA donne lieu à un unique intervalle, nous observons une forte variabilité dans ceux donnés par les participants. En effet, de 9 à 22 ($M = 15,4$) valeurs différentes sont données par borne d'intervalle d'une ENA.

Cette observation nous semble cohérente avec les deux perspectives du vague, sémantique et épistémique (voir section 2.1.1 p. 8). En effet, dans la perspective sémantique, le principe de tolérance est maintenu. Méthodologiquement, le protocole mis en œuvre

dans cette étude ne permet pas d'examiner ce principe au niveau intra-individuel puisque chaque participant ne fournit qu'un seul intervalle par ENA. En revanche, la tolérance inter-individuelle est observée puisque les participants ne sont pas d'accord entre eux et que des valeurs proches semblent acceptées pour certains mais pas pour d'autres.

La variabilité des intervalles est également cohérente avec la perspective épistémique du vague. En effet, dans cette perspective, il existe des limites nettes à l'extension d'une ENA. Néanmoins, ces limites sont soumises à variation, dans le temps et à travers les contextes et les individus, ce qui est observé.

4.4.2 Différence entre les intervalles selon l'usage des mathématiques et le niveau en calcul mental

Sur les 136 participants inclus dans les analyses, 66 ont rapporté utiliser régulièrement les mathématiques au travail ou dans leurs études et 70 ont rapporté ne pas les utiliser régulièrement. Aucune différence significative entre les deux groupes n'est observée, quelle que soit la borne de l'ENA considérée (W allant de 1867,5 à 2414,5 ; $p = N.S.$).

78 participants sont inclus dans le groupe "faible niveau" en calcul mental et 58 dans le groupe "haut niveau". Les analyses ne révèlent aucune différence entre les deux groupes, quelle que soit la borne de l'ENA considérée (W allant de 1767,5 à 2614 ; $p = N.S.$).

Ces résultats indiquent donc qu'un usage régulier des mathématiques ou que le niveau subjectif en calcul mental n'ont pas d'effet sur l'interprétation des ENA. Par conséquent, les analyses qui suivent portent sur l'ensemble des intervalles collectés auprès de tous les participants, sans distinction de niveau en calcul mental ou d'usage régulier des mathématiques.

4.4.3 Magnitude, granularité et dernier chiffre significatif (H1)

4.4.3.1 Magnitude

Pour tester l'effet de la magnitude sur les intervalles collectés, c'est-à-dire l'hypothèse H1, nous comparons quatre couples et un triplet de distributions de bornes d'ENA dont la magnitude diffère et dont la granularité et le dernier chiffre significatif sont constants.

Les analyses montrent un effet significatif de la magnitude sur la distance des bornes à la valeur de référence des ENA à toutes les comparaisons (voir tableau 4.3). Les analyses post-hoc réalisées pour le triplet révèlent quant à elles des différences significatives entre les trois ENA du triplet (voir tableau B.2 p. 188 de l'annexe B).

Ces résultats nous invitent à accepter l'hypothèse H1.1 : à granularité et dernier chiffre significatif constants, les intervalles d'ENA dont la magnitude diffère sont différents. Plus

Comparaisons	Test (V : Wilcoxon ; χ^2 : Friedman)
40/440	$V = 115,5^*$
100/1100	$V = 87^*$
500/1500	$V = 237,5^*$
30/4730	$V = 325^*$
50/150/8150	$\chi^2 = 297,8^*$

TABLEAU 4.3 – Effet de la magnitude : résultats des comparaisons entre ENA de même granularité et dernier chiffre significatif et de magnitude différente.
* : différence significative observée à $p < 0,01$.

Comparaisons	Test (V : Wilcoxon ; χ^2 : Friedman)
40/400	$V = 92^*$
50/500	$V = 252,5^*$
100/1000	$V = 265,5^*$
600/6000	$V = 62^*$
20/200/2000	$\chi^2 = 388,5^*$
80/800/8000	$\chi^2 = 407,7^*$

TABLEAU 4.4 – Effet de la granularité : résultats des comparaisons entre ENA de même dernier chiffre significatif et de granularité différente.
* : différence significative observée à $p < 0,01$.

précisément, plus la magnitude est élevée, plus les bornes sont éloignées de la valeur de référence.

4.4.3.2 Granularité

Pour tester l'effet de la granularité, les distributions des bornes d'ENA à un seul chiffre significatif dont la granularité diffère et dont le dernier chiffre significatif est le même sont comparées (voir tableau 4.4) : quatre couples (40/400 ; 50/500 ; 100/1000 ; 600/6000) et deux triplets (20/200/2000 ; 80/800/8000) sont considérés.

Les résultats montrent des différences significatives à toutes les comparaisons. De plus, les analyses post-hoc réalisées pour les deux triplets révèlent que, dans tous les cas, tous les niveaux de granularité diffèrent les uns des autres (voir tableaux B.3 et B.4, p. 189 de l'annexe B).

Ces résultats appuient donc l'hypothèse H1.2, à savoir que la granularité a un effet sur les intervalles correspondant aux ENA. Plus précisément, ils montrent que plus la

Niveaux de granularité	Test de Friedman
Dizaines (20/30/40/50/80)	$\chi^2 = 275,6^*$
Centaines (100/200/400/500/600/800)	$\chi^2 = 542,5^*$
Milliers (1000/2000/6000/8000)	$\chi^2 = 373,2^*$

TABLEAU 4.5 – Effet du dernier chiffre significatif : résultats des comparaisons entre ENA de même niveau de granularité et dont le dernier chiffre significatif diffère.
* : différence significative observée à $p < 0,01$.

granularité est élevée, plus l'intervalle de valeurs dénotées est large.

4.4.3.3 Dernier chiffre significatif

Les distributions des bornes d'ENA à un seul chiffre significatif, dont celui-ci diffère et dont la granularité est constante, sont comparées pour tester l'hypothèse d'un effet du dernier chiffre significatif : dizaines (20 / 30 / 40 / 50 / 80), centaines (100 / 200 / 400 / 500 / 600 / 800) et milliers (1000 / 2000 / 6000 / 8000).

Les tests de Friedman montrent un effet significatif du dernier chiffre significatif à chaque niveau de granularité (voir tableau 4.5). Toutefois, les analyses post-hoc ne révèlent pas de différence significative à chaque comparaison (voir tableaux B.5 à B.7, p. 189 de l'annexe B). En effet, dans le cas des centaines, par exemple, on observe que les distances des bornes aux valeurs de référence correspondant à "*environ 500*", "*environ 600*" et "*environ 800*" ne diffèrent pas significativement les unes des autres. En revanche, on observe bien des différences significatives entre "*environ 200*" et "*environ 400*", par exemple. De même, aucune différence significative n'est observée au niveau des dizaines entre "*environ 50*" et "*environ 80*", ni au niveau des milliers entre "*environ 6000*" et "*environ 8000*". Autrement dit, l'imprécision assignée à une ENA ne semble pas différer significativement à chaque incrémentation du dernier chiffre significatif, bien que cette dimension ait un effet global sur les intervalles.

Ces résultats semblent ambigus vis-à-vis de la troisième hypothèse dérivée. D'un côté, l'effet du dernier chiffre significatif détecté par les tests de Friedman tend à l'appuyer. D'un autre côté, les analyses post-hoc suggèrent la présence de paliers.

4.4.4 Modèle tridimensionnel à l'échelle logarithmique (H2)

Le tableau 4.6 présente les facteurs de Bayes obtenus par chacun des quatre modèles considérés : modèle proportionnel RM, modèle basé sur des échelles SBM, modèle régressif REGM et MGLSD. Plus ce score est élevé, mieux le modèle considéré rend compte des

	Facteurs de Bayes	
Modèle	Echelle linéaire	Echelle logarithmique
Proportionnel	$8,99 \cdot 10^{208}$	$3,07 \cdot 10^{374}$
Basé sur des échelles	$1,63 \cdot 10^{238}$	$9,60 \cdot 10^{283}$
Régressif	$1,16 \cdot 10^{319}$	$6,20 \cdot 10^{445}$
MGLSD	$1,40 \cdot 10^{329}$	$4,52 \cdot 10^{446}$

TABLEAU 4.6 – Facteurs de Bayes obtenus par analyses bayésiennes des quatre modèles, à l'échelle linéaire (colonne de gauche) et à l'échelle logarithmique (colonne de droite).

Facteurs de Bayes à l'échelle linéaire (ligne vs. colonne)			
Modèle	Proportionnel	Basé sur des échelles	Régressif
Basé sur des échelles	$1,81 \cdot 10^{29}$	-	-
Régressif	$1,29 \cdot 10^{110}$	$7,12 \cdot 10^{80}$	-
MGLSD	$1,55 \cdot 10^{120}$	$8,59 \cdot 10^{90}$	$1,21 \cdot 10^{10}$

TABLEAU 4.7 – Facteurs de Bayes obtenus par comparaison des quatre modèles entre eux, à l'échelle linéaire.

données collectées, validant l'implication des dimensions qu'il prend en compte.

On observe que MGLSD, le modèle que nous proposons, présente les facteurs de Bayes les plus élevés. Cette observation est appuyée par le fait que les comparaisons des quatre modèles entre eux (tableaux 4.7 et 4.8) montrent que le modèle MGLSD, qui tient compte de la magnitude, de la granularité et du dernier chiffre significatif de la valeur de référence des ENA, rend mieux compte des intervalles collectés.

Ces résultats sont valables aussi bien à l'échelle linéaire que logarithmique, appuyant ainsi les dimensions dont nous proposons de tenir compte. Toutefois, à l'échelle logarithmique, l'écart entre le modèle MGLSD et le modèle régressif est plus faible (facteur de Bayes : 7,28, voir tableau 4.8). Le modèle basé sur des échelles obtient les facteurs de Bayes les moins élevés.

Enfin, les résultats des analyses bayésiennes montrent que les quatre modèles rendent mieux compte des intervalles observés lorsqu'ils sont portés à l'échelle logarithmique. En effet, on observe que les facteurs de Bayes sont plus élevés, quel que soit le modèle considéré, lorsque les dimensions sont portées à l'échelle logarithmique (voir tableau 4.6, colonne de droite) que lorsqu'elles sont à l'échelle linéaire (voir tableau 4.6, colonne de gauche).

Dans l'ensemble, ces résultats valident notre seconde hypothèse H2. La magnitude, la granularité et du dernier chiffre significatif de la valeur de référence d'une ENA sont simultanément impliqués dans son interprétation. Un modèle tenant compte de ces trois

Facteurs de Bayes à l'échelle logarithmique (ligne vs. colonne)			
Modèle	Proportionnel	Basé sur des échelles	Régressif
Basé sur des échelles	$3,12 \cdot 10^{-91}$	-	-
Régressif	$2,02 \cdot 10^{71}$	$6,46 \cdot 10^{161}$	-
MGLSD	$1,47 \cdot 10^{72}$	$4,71 \cdot 10^{162}$	$7,28$

TABLEAU 4.8 – Facteurs de Bayes obtenus par comparaison des quatre modèles entre eux, à l'échelle logarithmique.

dimensions arithmétiques rend mieux compte des intervalles correspondant aux ENA, en particulier lorsqu'elles sont portées à l'échelle logarithmique.

4.4.5 Symétrie des intervalles (H3)

En nous basant sur le critère défini dans la section 4.3.6 p. 47, nous testons dans quelle mesure l'hypothèse de symétrie posée par les modèles de la littérature (Ferson et al., 2015; Krifka, 2007; Sauerland & Stateva, 2007; Solt, 2014) est valable pour les intervalles collectés.

Sur l'ensemble des données, 78,7% des intervalles sont symétriques. Ce score varie de 50,7% à 89,6%, selon l'ENA considérée (voir tableau B.8 p. 190 de l'annexe B pour les scores détaillés).

Les scores bas, inférieurs à 70%, sont principalement observés pour les ENA dont la valeur de référence possède au moins deux chiffres significatifs. Plus précisément, dans le cas de deux chiffres significatifs, une asymétrie est observée si le dernier chiffre significatif n'est ni 1, ni 5 : "*environ 440*" (61,5%), "*environ 560*" (67,7%) et "*environ 4700*" (65,7%). Les deux ENA dont la valeur de référence possède trois chiffres significatifs, "*environ 4730*" et "*environ 8150*", donnent lieu à des intervalles plus souvent asymétriques que les autres, 50,7% pour "*environ 4730*" et 64,2% pour "*environ 8150*", bien que le dernier chiffre significatif de cette dernière soit 5.

Ces résultats suggèrent donc que l'hypothèse de symétrie des intervalles ne tient pas pour toutes les valeurs de référence. En effet, celles possédant plusieurs chiffres significatifs conduisent à des intervalles plus souvent asymétriques.

4.4.6 Heuristiques mises en place par les participants (H4)

Le dernier item du questionnaire en ligne est dédié aux heuristiques mises en place par les participants pour estimer les intervalles correspondant aux ENA. Cet item est optionnel. 90 participants n'ont pas répondu ou ont rapporté ne pas avoir utilisé de stratégie. Nous

Heuristique	% de participants
Constante	6,5% ($N = 3$)
Magnitude	**45,6%** ($N = 21$)
Magnitude relative	**26,0%** ($N = 12$)
Autre	21,9% ($N = 10$)

TABLEAU 4.9 – Pourcentages et, entre parenthèses, nombres de participants ayant rapporté l'utilisation de l'une des quatre heuristiques pour estimer les bornes des intervalles correspondant aux ENA.

avons donc analysé les données qualitatives fournies par 46 participants selon la méthode décrite dans la section 4.3.6 p. 47. Le tableau 4.9 présente les pourcentages de réponses incluses dans chacune des quatre catégories définies.

Les résultats montrent que les participants qui explicitent leurs heuristiques ont tendance à considérer soit la magnitude, soit la magnitude relative de la valeur de référence d'une ENA pour estimer les bornes de l'intervalle correspondant. Ces observations appuient donc l'hypothèse H4 qui postule une implication de la magnitude, de la granularité et du dernier chiffre significatif dans l'interprétation des ENA. En effet, la magnitude relative étant le produit de la granularité par la valeur du dernier chiffre significatif (voir section 4.1 p. 38), ces deux dimensions, ajoutées à la magnitude, semblent être des dimensions clés de l'interprétation des ENA, pour les participants eux-mêmes.

4.5 Discussion

Dans cette section, nous discutons des résultats présentés dans la section précédente, selon deux axes : (i) les dimensions arithmétiques impliquées dans l'interprétation des ENA et (ii) la symétrie de l'imprécision dénotée par ENA par rapport à sa valeur de référence.

Dimensions arithmétiques impliquées dans l'interprétation des ENA La littérature suggère que la magnitude, la granularité et le dernier chiffre significatif de la valeur de référence des ENA sont les trois dimensions clés dans leur interprétation (voir chapitre 3).

Plus spécifiquement, le modèle d'interprétation des ENA basé sur des échelles (Sauerland & Stateva, 2007; Solt, 2014) définit la granularité comme la seule dimension arithmétique à prendre en compte. Les travaux empiriques de Ferson et al. (2015) suggèrent quand à eux que l'intervalle correspondant à une ENA est également affecté par sa magnitude ainsi que par son dernier chiffre significatif, en particulier si sa valeur est 5.

Par ailleurs, la manière dont les nombres et les quantités sont traités et représentés

dans le système cognitif humain suggère également l'implication des ces dimensions dans l'interprétation des ENA (voir section 2.2 p. 17). En effet, de ce point de vue, la magnitude correspondrait à la représentation des quantités dans le système approximatif des nombres alors que la granularité et le dernier chiffre significatif seraient les deux dimensions prises en compte par le système exact dans la mesure où l'interprétation d'une ENA est traitée comme un problème formel.

Conformément à ce que nous attendions, nous observons l'implication de la magnitude, de la granularité et du dernier chiffre significatif dans l'interprétation. Plus précisément, trois types d'analyses réalisées appuient ce résultat. D'abord, nous observons un effet séparé de chacune de ces trois dimensions arithmétiques sur les intervalles collectés. Ensuite, un modèle qui combine les trois dimensions à l'échelle logarithmique rend mieux compte des intervalles collectés que des modèles basés sur les dimensions proposées dans la littérature. Enfin, en interrogeant les participants sur les stratégies qu'ils ont pu mettre en place, leurs réponses évoquent majoritairement la magnitude, la granularité et le dernier chiffre significatif de la valeur de référence d'une ENA.

Bien que nous mettions nettement en évidence l'effet de la magnitude et de la granularité, le cas du dernier chiffre significatif est moins clair. En effet, en comparant plusieurs ENA de même granularité (dizaines, centaines, milliers), nous avons trouvé des différences significatives entre ENA dont le dernier chiffre significatif est différent. Toutefois, les résultats obtenus suggèrent que la relation entre la distance des bornes à la valeur de référence et la valeur du dernier chiffre significatif n'est pas linéaire. En effet, nous avons observé que la distance n'augmente pas significativement à chaque incrémentation du dernier chiffre significatif (par ex., entre 30 et 40, ou entre 400 et 500).

En outre, ces résultats portant sur le dernier chiffre significatif ne sont pas en accord avec les conclusions de Ferson et al. (2015) en ce qui concerne la pertinence de la dimension *fiveness* comme prédicteur de la taille des intervalles. Selon les auteurs, les ENA dont le dernier chiffre significatif est 5 donnent lieu à des intervalles plus larges que les ENA dont le dernier chiffre significatif est différent de 5 à granularité constante. Les analyses que nous avons réalisées ne montrent pas de différence significative entre 50 et 80, ni entre 500 et 800. De plus, nos résultats montrent que des différences significatives apparaissent entre des ENA dont le dernier chiffre significatif diffère. Par conséquent, la valeur 5 comme dernier chiffre significatif ne semble pas être plus spécifique que les autres valeurs.

Enfin, le fait que les participants aient tendance à prendre en compte à la fois la magnitude, la granularité et le dernier chiffre significatif peut être interprété comme un compromis entre les deux systèmes cognitifs en jeu dans la représentation et le traitement des nombres (voir section 2.2 p. 17).

Symétrie de l'imprécision dénotée par une ENA Le deuxième objectif de cette étude est d'examiner la pertinence cognitive de la symétrie des intervalles par rapport à la valeur de référence des ENA, supposée par les modèles de la littérature (Ferson et al., 2015; Krifka, 2007; Sauerland & Stateva, 2007; Solt, 2014). Nous avons testé cette symétrie sur les intervalles collectés.

Bien que le score moyen de symétrie sur l'ensemble des données soit élevé, cinq ENA donnent lieu à de faibles scores : "*environ 440*", "*environ 560*", "*environ 4700*", "*environ 4730*" et "*environ 8150*". Les valeurs de référence de ces ENA peuvent être caractérisées par deux propriétés : (i) elles possèdent au moins deux chiffres significatifs ; (ii) quand elles possèdent deux chiffres significatifs (par ex., "*environ 440*" et "*environ 4700*"), le dernier d'entre eux n'est ni 1, ni 5 alors que lorsqu'elles en possèdent trois, la valeur du chiffre significatif ne semble pas avoir d'effet.

Partant de ces deux observations, nous proposons une explication fondée sur la saillance cognitive des nombres (voir section 4.1 p. 38). En effet, les ENA qui donnent lieu à des scores de symétrie faibles sont proches de nombres plus saillants que leur valeur de référence. Par exemple, 4700 est proche de 4500 et de 5000. De même, 8150 est proche de 8000 et de 8200. Puisque les nombres saillants sont plus facilement évoqués, il se pourrait que les participants soient biaisés à leur endroit. Autrement dit, les participants auraient tendance à donner des valeurs plutôt rondes, avec peu de chiffres significatifs, donc peu complexes, comme bornes des intervalles. Dans le cas où ces nombres saillants ne sont pas situés à la même distance de la valeur de référence de l'ENA (par ex., 4700 et 4750 pour "*environ 4730*"), les intervalles sont asymétriques.

4.6 Bilan

Ce chapitre a présenté l'étude empirique originale que nous avons conduite dans le but d'examiner l'implication des dimensions arithmétiques des ENA dans leur interprétation.

Dans un premier temps, nous avons proposé une définition formelle de ces dimensions sur la base de l'état de l'art, aussi bien dans le domaine de l'interprétation des ENA que dans celui portant sur la cognition humaine des nombres. Dans un second temps, nous avons analysé des données collectées au moyen d'un questionnaire en ligne pour examiner l'implication de ces dimensions dans l'interprétation des ENA.

Nous avons ainsi pu mettre en évidence les effets de la magnitude, de la granularité et du dernier chiffre significatif, en particulier lorsque ces dimensions sont portées à l'échelle logarithmique, alors qu'aucun des modèles de la littérature ne tient compte de toutes les trois à la fois. Leur implication dans l'interprétation des ENA nous invite à considérer l'étendue de la plage de valeurs dénotées par une ENA comme résultant d'un compromis

entre le système approximatif des nombres et le système exact des nombres.

De plus, bien que les modèles de la littérature fassent l'hypothèse que la plage de valeurs dénotée par une ENA est centrée sur sa valeur de référence, nous avons pu mettre en évidence le fait que cette hypothèse n'est pas toujours valable et que la symétrie dépend des dimensions arithmétiques caractéristiques de l'ENA considérée et de la distribution des nombres cognitivement saillants autour de sa valeur de référence.

Enfin, en accord avec les perspectives sémantique et épistémique du vague, la variabilité observée dans les intervalles donnés par les participants nous conduit à interroger la pertinence d'un modèle ne prédisant qu'un seul intervalle.

L'ensemble des résultats et observations issus de l'étude présentée dans ce chapitre nous ont conduit à proposer d'autres modèles d'interprétation des ENA, décrits dans le chapitre suivant.

Chapitre 5

Interprétation des ENA : modélisation computationnelle

Ce chapitre présente les modèles computationnels de l'interprétation des Expressions Numériques Approximatives que nous proposons, qui tiennent compte de l'étude des modèles de la littérature, détaillée dans le chapitre 3, et des résultats de l'étude empirique décrite dans le chapitre précédent.

Interpréter une ENA consiste à délimiter, nettement dans la perspective épistémique du vague et de manière graduelle dans la perspective sémantique, la plage de valeurs qu'elle dénote. Nous proposons dans ce chapitre deux types de modèles, correspondant aux deux perspectives du vague : le premier s'inscrit dans la perspective épistémique et estime un intervalle représentant la plage de valeurs dénotées ; le second type s'inscrit dans la perspective sémantique et représente l'imprécision dénotée par une ENA par un sous-ensemble flou.

Les modèles d'interprétation des ENA que nous proposons sont fondés sur un principe général tenant compte de la saillance cognitive des nombres. Ce principe est décrit dans la section 5.1. Nous développons deux modèles de l'approche par intervalles dans la section 5.2. La section 5.3 décrit le modèle que nous proposons dans l'approche par sous-ensembles flous. Ces deux approches font chacune l'objet d'une validation expérimentale montrant leur intérêt par rapport aux modèles identifiés dans la littérature.

Plusieurs parties de ce chapitre ont été publiées dans les articles, Lefort et al. (2016a; 2016b; 2017b).

5.1 Principe général : compromis entre saillance cognitive et plage de valeurs dénotées

Les résultats de l'étude empirique présentée au chapitre précédent suggèrent que toutes les valeurs n'ont pas la même pertinence cognitive pour être les bornes de la plage de valeurs que dénote une ENA "*environ x*". Par conséquent, la première étape de la modélisation computationnelle de l'interprétation des ENA consiste à sélectionner les valeurs qui sont pertinentes du point de vue du traitement cognitif humain. Les valeurs ainsi sélectionnées sont par la suite utilisées comme candidates pour être les bornes des intervalles dans l'approche par intervalles, décrite dans la section 5.2, et les bornes des α-coupes des intervalles flous dans l'approche par sous-ensembles flous, décrite dans la section 5.3.

Dans cette section, nous détaillons en premier lieu le principe général sur lequel repose la sélection des valeurs cognitivement pertinentes. La mise en œuvre du principe proposé est ensuite décrite dans la section 5.1.2.

5.1.1 Principe

La motivation du principe que nous proposons est que pour être une bonne candidate, une borne doit simultanément satisfaire deux contraintes. La première est qu'elle doit correspondre à la quantité d'imprécision qu'un individu se représente comme dénotée par une ENA.

La seconde contrainte est que la valeur de la borne doit être la plus saillante possible. Cette contrainte s'appuie sur le principe de tolérance (voir section 2.1.1 p. 8), selon lequel modifier légèrement la valeur d'une borne de l'intervalle considérée par un individu comme acceptable pour une ENA donnée ne devrait pas modifier le jugement d'acceptabilité. Par exemple, si 132 est jugé comme borne acceptable pour "*environ 150*", alors 131 ou 130 devraient l'être également. Nous pouvons donc considérer que les valeurs de bornes que donne un individu pour une ENA sont des indications et qu'il existe une zone d'acceptabilité autour de ces valeurs. Dans une perspective pragmatique, les individus tendent à préférer une expression simple à une expression complexe pour dénoter la même entité (Grice, 1991). Appliqué aux ENA, cela implique que les individus tendent à préférer des nombres plus saillants à des nombres moins saillants pour exprimer une plage de valeur ou une indication. Par conséquent, du point de vue pragmatique, nous pouvons nous attendre à ce que les individus fournissent des valeurs plutôt saillantes comme bornes des intervalles correspondant à une ENA. Ainsi, selon cette contrainte, les individus préfèrent 130 comme valeur de borne à 131 ou 132 pour "*environ 150*", car 130 est plus saillant, donc plus simple, et que la plage de valeurs dénotées par l'ENA "*environ 150*" n'est que très peu affectée par

ce décalage.

Toutefois, les valeurs les plus saillantes, respectant la seconde contrainte, peuvent être fortement éloignées de la valeur de référence de l'ENA. Par exemple, pour "*environ 150*", 100 et 200 sont les deux valeurs les plus saillantes à proximité de la valeur de référence 150. Or, si un individu considère la plage de valeurs entre ces deux bornes comme trop large pour correspondre à l'imprécision dénotée par "*environ 150*", il doit trouver un meilleur compromis entre la saillance des bornes et la taille de la plage de valeurs dénotées. Ainsi, les intervalles $[120, 180]$ ou $[145, 155]$ peuvent représenter un bon compromis pour certains individus. En revanche, pour des plages de valeurs dénotées similaires, $[121, 181]$ ou $[146, 154]$ représentent de mauvais compromis. En effet, d'un point de vue pragmatique, ces deux intervalles sont des mauvais choix puisqu'ils dénotent des plages de valeurs très similaires aux deux premiers intervalles, $[120, 180]$ et $[145, 155]$, tout en étant moins saillants, donc moins simples.

Ces exemples illustrent que les deux contraintes, à savoir que les bornes des intervalles dénotés par une ENA doivent correspondre à la représentation de l'imprécision qu'elle dénote, et que ces bornes doivent être autant que possible saillantes, peuvent être contradictoires.

Par conséquent, un compromis doit être considéré : le principe général que nous proposons repose sur l'hypothèse que lors de l'interprétation d'une ENA "*environ* x", un compromis est réalisé entre la saillance des valeurs des bornes et leur distance à la valeur de référence x.

Cette hypothèse implique que pour une distance donnée, la saillance des valeurs des bornes est maximisée ; de même, pour une saillance donnée, la distance des valeurs des bornes à la valeur de référence x est minimisée. Par exemple, étant donnée l'ENA "*environ 500*", des intervalles tels que $[499, 501]$, $[490, 510]$ ou $[450, 550]$ sont de bons candidats. En effet, leurs bornes sont les plus proches de la valeur de référence $x = 500$ quand la complexité des candidats est respectivement 3, 2 et 1,5. Au contraire, $[497, 503]$ ou $[460, 540]$ ne sont pas considérés comme de bons candidats car ils sont dominés par d'autres intervalles, c'est-à-dire que d'autres intervalles existent qui représentent un meilleur compromis entre saillance cognitive et distance des bornes à la valeur de référence : ainsi, $[497, 503]$ est dominé par $[499, 501]$, dont les valeurs des bornes sont aussi saillantes tout en étant plus proches de la valeur de référence 500. Pour la même raison, $[460, 540]$ est dominé par $[490, 510]$, dont les valeurs des bornes sont à la fois aussi saillantes et plus proches de la valeur de référence.

Les valeurs qui optimisent le compromis entre saillance cognitive et distance à la valeur de référence, c'est-à-dire qui ne sont pas dominées par d'autres valeurs, sont de meilleures candidates pour être les bornes de l'intervalle correspondant à l'ENA. La formalisation de

ce principe de sélection est décrite dans la section suivante.

5.1.2 Mise en œuvre : construction de fronts de Pareto

Les bonnes valeurs candidates, optimisant le compromis entre saillance cognitive et distance à la valeur de référence, sont celles qui ne sont pas dominées par d'autres valeurs. Formellement, la notion de valeurs non dominées, pour au moins deux critères quelconques, est capturée par le concept de front de Pareto. Cette section formalise les deux critères qui définissent l'opérateur de comparaison proposé, puis la construction des deux fronts de Pareto considérés, $P^-(x)$ pour la borne inférieure et $P^+(x)$ pour la borne supérieure de l'intervalle dénoté par une ENA.

Pour construire un front de Pareto, nous considérons comme candidates toutes les valeurs $v \in V^b(x) \subset \mathbb{N}^*$, avec $b \in \{-,+\}$ et $V^-(x) = [1, x[$ et $V^+(x) =]x, +\infty[$. Ces valeurs sont comparées entre elles selon deux critères : (i) leur distance absolue à la valeur de référence de l'ENA : $d_x(v) = |v-x|$; et (ii) leur complexité $Cpx(v)$, capturant leur saillance cognitive. Nous proposons de calculer la complexité $Cpx(x)$ en utilisant la fonction définie dans le tableau 4.1 p. 38.

La notion de dominance est un ordre partiel : formellement, v' domine v (noté $v' \succ v$) si et seulement si $d_x(v') < d_x(v) \wedge Cpx(v') \leq Cpx(v)$, c'est-à-dire si v' est à la fois plus proche et pas plus complexe que v.

L'ensemble des valeurs non dominées constitue le front de Pareto, formellement défini pour $b \in \{-,+\}$ comme :

$$P^b(x) = \{v \in V^b(x) / \forall v' \neq v \in V^b(x) \quad v' \not\succ v\} \tag{5.1}$$

De plus, nous imposons un ordre sur ces valeurs et considérons la liste ordonnée par distance croissante à la valeur de référence, que nous notons aussi $P^b(x)$ abusivement : $P^b(x) = [y_1^b, \ldots, y_n^b]$, avec $b \in \{-,+\}$ et $\forall i, d_x(y_i) < d_x(y_{i+1})$.

On peut noter que $P^b(x)$ est fini car la complexité d'un candidat est bornée : d'une part, elle ne peut être inférieure à 1, valeur obtenue pour les candidats ne possédant qu'un seul chiffre significatif ; d'autre part, elle ne peut être supérieure au nombre de chiffres, en incluant les zéros à droite du dernier chiffre significatif, de la valeur de référence de l'ENA. Par exemple, dans le cas de "*environ 560*", la complexité des candidats est comprise entre 1 (par ex., 500 ou 600) et 3 (par ex., 559 ou 561). Bien qu'il existe une infinité de candidats dont la complexité est 1 ou 3, seul l'un d'entre eux, le plus proche de x, est sélectionné car il domine tous les autres.

Enfin, on peut remarquer que pour une ENA "*environ x*", les deux fronts de Pareto comprennent toujours $x-1$ et $x+1$, valeurs qui sont les plus proches de x et qui optimisent par conséquent toujours le compromis. De même, les deux fronts de Pareto comprennent

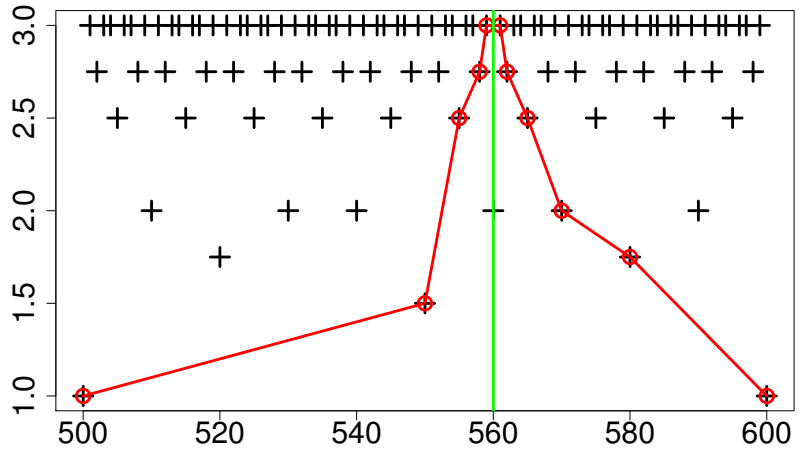


FIGURE 5.1 – Complexité $Cpx(v)$ (plus) des nombres entiers compris entre 500 et 600, ainsi que les fronts de Pareto (lignes rouges) pour la borne inférieure (à gauche de la ligne verticale verte) et supérieure (à droite de la ligne verticale) de l'ENA "*environ 560*".

toujours les nombres à un seul chiffre significatif les plus proches, par exemple 500 et 600 pour "*environ 560*", car leur complexité est minimale et par conséquent optimisent toujours le compromis.

La figure 5.1 illustre les deux fronts de Pareto correspondant à l'ENA "*environ 560*" : les signes + représentent la distribution des couples $(v, Cpx(v))$ pour toutes les valeurs v comprises entre 500 et 600.

L'axe horizontal représentant les candidats v représente également, de manière indirecte, leur distance $d_x(v)$: cette dernière croît linéairement depuis la ligne verticale verte représentant $x = 560$. Par conséquent, la partie située à droite de la ligne verticale peut également être vue comme représentant la distribution des $(d_x(v) + x, Cpx(v))$; la partie à gauche de la ligne verticale représente quant à elle la distribution des $(x - d_x(v), Cpx(v))$.

$P^-(560)$ et $P^+(560)$ sont représentés par les points connectés encerclés en rouge, respectivement à gauche et à droite de la ligne verticale. Comme résultat, on obtient $P^-(560) = [559, 558, 555, 550, 500]$ et $P^+(560) = [561, 562, 565, 570, 580, 600]$.

Dans le cas de $x = 560$, 540 n'est pas considérée comme une bonne valeur candidate pour la borne inférieure car elle est dominée par 550, qui est à la fois plus proche et plus saillante : 550 représente un meilleur compromis entre saillance cognitive et distance à la valeur de référence.

On peut remarquer que les fronts de Pareto $P^-(x)$ et $P^+(x)$ construits selon la méthode que nous proposons, en tenant compte de la complexité des nombres, capturent naturel-

lement l'asymétrie observée empiriquement (voir section 4.4 p. 50). En effet, dans le cas où x possède plusieurs chiffres significatifs, les valeurs saillantes ne sont pas nécessairement distribuées symétriquement autour de x. Ceci est illustré sur la figure 5.1 pour $x = 560$, où 500 est plus éloigné de 560 que 600.

5.1.3 Validation préliminaire

Comme première validation du principe général du compromis entre saillance cognitive et distance à la valeur de référence, nous examinons dans quelle mesure les bornes des intervalles collectés lors de l'étude empirique décrite dans le chapitre précédent réalisent ce compromis.

Sur l'ensemble des intervalles collectés, 84,4% des bornes sont situées sur les fronts de Pareto. Le taux de capture, variant entre 74,1%, pour "*environ 80*" et 91,0%, pour "*environ 1100*", de chacune des 24 ENA considérées est détaillé dans le tableau C.1 p. 192 de l'annexe C.

Le principe général que nous proposons rend donc compte, dans leur variété, des intervalles correspondant aux plages de valeurs dénotées par les 24 ENA, telles que données par les participants de l'étude empirique.

Les deux exploitations décrites et validées expérimentalement dans les deux sections suivantes considèrent les valeurs comprises dans les front de Pareto comme des candidats pour être les bornes des intervalles dénotés par les ENA (section 5.2) et les bornes des α-coupes des intervalles flous représentant les ENA (section 5.3).

5.2 Exploitation 1 : modèle d'interprétation à intervalles

Cette section présente la première approche de modélisation computationnelle de l'interprétation des ENA que nous proposons, exploitant le principe général du compromis entre saillance cognitive et plage de valeurs dénotée. Cette approche vise à estimer l'intervalle représentant l'imprécision dénotée par une ENA, noté $I(x) = [x - \Delta^-(x), x + \Delta^+(x)]$, où $\Delta^-(x)$ et $\Delta^+(x)$ sont les distances à la valeur de référence x des bornes inférieure et supérieure, respectivement. Cette approche par intervalles s'inscrit dans le cadre de la perspective épistémique du vague, telle que décrite dans la section 2.1.1 p. 8.

Plus précisément, deux modèles, successivement décrits dans les deux sous-sections suivantes, sont proposés pour sélectionner les valeurs des bornes de l'intervalle parmi celles qui sont situées sur les fronts de Pareto : le premier, appelé LLM, est basé sur la discrétisation d'une régression préliminaire, le second, appelé RKM, exploite les rangs des candidats. La validation expérimentale de ces deux modèles est décrite dans la sous-section 5.2.3.

5.2.1 Modèle log-linéaire LLM

Comme le modèle régressif REGM (Ferson et al., 2015, voir section 3.2.3 p. 32), le modèle log-linéaire LLM que nous proposons est basé sur une régression linéaire. Outre des dimensions arithmétiques prises en compte différentes, son originalité réside dans l'utilisation d'une étape finale de discrétisation qui arrondit l'estimation fournie par la fonction de régression à la valeur la plus proche située sur les fronts de Pareto $P^-(x)$ et $P^+(x)$.

Comme REGM, et contrairement au modèle basé sur des échelles SBM (Sauerland & Stateva, 2007; Solt, 2014, voir section 3.2.2 p. 30), le modèle que nous proposons est adaptable à différents contextes sémantiques : les coefficients de la fonction de régression sont appris à partir d'un ensemble d'intervalles empiriquement collectés.

Le principe du modèle est dans un premier temps introduit, avant que la phase d'apprentissage ne soit présentée.

5.2.1.1 Principe

Le modèle estime les bornes de l'intervalle à partir d'une fonction de régression. Comme suggéré dans la section 4.1 p. 38 et dans la section 3.2 p. 29, la magnitude x et la magnitude relative $RelMag(x)$, qui combine la granularité et le dernier chiffre significatif, de la valeur de référence d'une ENA sont impliquées dans son interprétation.

De plus, comme suggéré par la manière dont l'être humain encode les nombres et les quantités (voir section 2.2 p. 17) et appuyé par les résultats de l'étude empirique de l'interprétation des ENA (voir section 4.4 p. 50), nous proposons d'utiliser le logarithme de ces deux dimensions comme variables d'une fonction de régression.

Les valeurs réelles fournies par cette régression sont ensuite discrétisées au candidat le plus proche situé sur les fronts de Pareto correspondant à l'ENA considérée.

Formellement, la fonction de régression estime $\Delta_C(x)$ qui représente la distance absolue des bornes de l'intervalle à la valeur de référence x de l'ENA, elle est calculée comme :

$$\log(\Delta_C(x)) = \alpha_0 + \alpha_1 \cdot \log(x) + \alpha_2 \cdot \log(RelMag(x)) \tag{5.2}$$

où α_0, α_1 et α_2 sont les paramètres de la régression. Leur apprentissage est décrit dans la sous-section suivante.

A cette étape, un intervalle symétrique dont les valeurs des bornes sont réelles peut être construit comme $I = [x_R^-, x_R^+] = [x - \Delta_C(x), x + \Delta_C(x)]$. Puisque les valeurs des bornes qui sont saillantes sont plus pertinentes que celles qui prennent une valeur réelle, nous proposons de les discrétiser au candidat le plus proche sur les fronts de Pareto, selon la fonction suivante :

$$D(v) = \arg\min_{y_i \in P(x)} |v - y_i| \tag{5.3}$$

où $P(x)$ est l'union des deux fronts de Pareto $P(x) = P^-(x) \cup P^+(x)$. L'intervalle final correspondant à l'ENA "*environ x*" est alors :

$$I_{LLM}(x) = [D(x_R^-), D(x_R^+)] \tag{5.4}$$

On peut remarquer que l'intervalle $I_{LLM}(x)$ peut être asymétrique puisque la distribution des candidats sur les fronts de Pareto peut être différente selon qu'on considère la borne inférieure ou la borne supérieure, comme illustré sur la figure 5.1.

5.2.1.2 Paramètres de la régression

Les paramètres de la régression α_0, α_1, et α_2 sont appris à partir d'une base empiriquement collectée d'intervalles correspondant à des ENA, comme celle décrite dans la section 4.3 p. 44.

Pour chaque ENA "*environ x*", les exemples collectés $I_p(x) = [x - \Delta P_p^-(x), x + \Delta P_p^+(x)]$ sont agrégés sur l'ensemble des participants en $\Delta^-(x)$ et $\Delta^+(x)$ puis donnent deux triplets $(x, RelMag(x); \Delta^b(x))$, où $b \in \{-, +\}$. Dans la perspective épistémique du vague, le choix de l'opérateur d'agrégation dépend de l'interprétation souhaitée de l'intervalle à retourner (voir section 2.1.1 p. 8). Le choix que nous proposons de cet opérateur est spécifié dans la section 5.2.3.2 p. 71.

Dans un second temps, une régression linéaire est réalisée selon la méthode des moindres carrés pour trouver les valeurs optimales de α_0, α_1, et α_2 (voir équation 5.2).

5.2.2 Modèle des rangs RKM

Le deuxième modèle que nous proposons, appelé RKM (*RanK Model*), exploite les valeurs situées sur les fronts de Pareto d'une manière différente : il consiste à estimer directement la position de la valeur retenue des bornes sur les fronts de Pareto. Pour ce faire, le modèle sélectionne comme borne la valeur $y_{r_P(x)} \in P^b(x) = [y_1, \ldots, y_n]$.

La question est dès lors d'estimer le rang $r_P(x)$ qui est le plus pertinent. Comme le modèle LLM présenté ci-dessus, RKM dépend à la fois de la magnitude relative $RelMag(x)$ et de la magnitude de la valeur de référence x de l'ENA considérée. Toutefois, au lieu d'utiliser directement la magnitude x, nous proposons d'utiliser son nombre de chiffres significatifs $NSD(x)$: à magnitude relative constante, un entier dont le nombre de chiffres significatifs est supérieur à 1 présente une magnitude plus élevée. Par exemple, $RelMag(8150) = RelMag(150) = RelMag(50) = 50$; $NSD(8150) > NSD(150) > NSD(50)$ et $8150 > 150 > 50$. L'avantage d'utiliser le nombre de chiffres significatifs ici est d'obtenir directement un nombre entier, puisque le rang doit en être un lui-même.

Le rang de la valeur sélectionnée pour la borne est calculé selon l'équation suivante :

$$r_P(x) = round\left(\log(RelMag(x)) - 1 + \sum_{k=1}^{NSD(x)} k\right) \tag{5.5}$$

L'estimation de l'intervalle correspondant à l'ENA "*environ x*" est donc :

$$I_{RKM}(x) = [y^-_{r_P(x)}, y^+_{r_P(x)}] \tag{5.6}$$

où $r_P(x)$ est limité dans $[1, n^b]$, c'est-à-dire :

$$r_P(x) = \min(\max(1, r_P(x)), n^b) \tag{5.7}$$

avec $b \in \{-, +\}$, où n^- et n^+ sont respectivement la longueur de $P^-(x)$ et de $P^+(x)$.

5.2.3 Validation expérimentale : protocole

Les performances des deux modèles que nous proposons, LLM et RKM, sont comparées à celles obtenues par les trois modèles de la littérature, présentés dans la section 3.2 p. 29 : RM, SBM et REGM. Les données utilisées sont celles qui ont été collectées lors de l'étude empirique de l'interprétation des ENA, décrite dans la section 4.3 p. 44.

Les deux premières sous-sections décrivent les quatre critères de qualité utilisés pour évaluer les performances des cinq modèles : les deux premiers portent directement sur l'estimation des bornes fournies par les participants de l'étude empirique tandis que les deux autres se focalisent sur l'estimation des médianes des bornes, calculées sur l'ensemble des participants. Les deux sous-sections suivantes sont dédiées à la procédure expérimentale mise en œuvre pour comparer entre eux les modèles ainsi que leur paramétrisation. Enfin, les résultats sont présentés et discutés dans la section 5.2.4.

5.2.3.1 Critères de prédiction de bornes

Nous proposons deux critères de qualité pour évaluer dans quelle mesure les intervalles estimés par les modèles d'interprétation correspondent aux données collectées auprès des participants de l'étude empirique : le taux de bonnes prédictions de bornes PA (*Prediction Accuracy*) et la distance relative RD (*Relative Distance*).

$\mathcal{X}$ représente l'ensemble des ENA considérées et $\mathcal{P}(x)$ l'ensemble des participants dont les intervalles sont pris en compte, c'est-à-dire ne sont pas considérés comme aberrants, pour l'ENA "*environ x*" (voir section 4.3 p. 44). Pour un participant p et un modèle m, nous notons respectivement $\Delta P^b_p(x)$ et $\Delta M^b_m(x)$ la distance absolue de la borne qu'ils produisent, inférieure ou supérieure pour $b = -$ et $b = +$ respectivement, à la valeur de référence x de l'ENA.

Taux de bonnes prédictions de bornes Le taux de bonnes prédictions de bornes PA est un critère conçu pour évaluer dans quelle mesure un modèle estime correctement les intervalles fournis par les participants, en calculant la proportion d'intervalles bien prédits.

Une prédiction est considérée comme correcte lorsque la distance entre l'estimation et la valeur de référence x de l'ENA est égale à celle du participant, en autorisant une erreur relative de 10% [3].

Formellement, l'égalité Eq est calculée selon l'équation 4.2 (voir section 4.3.6, p. 49), rappelée ici pour faciliter la lecture :

$$Eq(\Delta_1, \Delta_2) = \begin{cases} 1 & \text{si } \dfrac{|\Delta_1 - \Delta_2|}{\min(\Delta_1, \Delta_2)} \leq 0,1 \\ 0 & \text{sinon} \end{cases}$$

Elle est appliquée à $\Delta M_m^b(x)$ et $\Delta P_p^b(x)$ pour évaluer si le modèle m effectue une bonne prédiction pour le participant p. Le score global du modèle est obtenu en moyennant ce score d'égalité sur l'ensemble des ENA $x \in \mathcal{X}$ et des participants $p \in \mathcal{P}(x)$. Le taux de bonnes prédictions, à maximiser, est donc défini comme :

$$PA(m) = \frac{1}{2 \cdot |\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{b \in \{-,+\}} \frac{1}{|\mathcal{P}(x)|} \sum_{p \in \mathcal{P}(x)} Eq(\Delta M_m^b(x), \Delta P_p^b(x)) \tag{5.8}$$

Il faut noter que les modèles considérés ne produisent qu'une seule estimation par borne. Or, nous avons pu observer lors de l'étude empirique de l'interprétation des ENA (voir section 4.4 p. 50) une variabilité dans les bornes données par les participants pour une même ENA. Par conséquent, les modèles ne peuvent atteindre un taux de bonnes de prédictions de 100%. Il est limité à la fréquence relative de la borne la plus fréquemment donnée par les participants, autrement dit au mode des valeurs de bornes observées.

Distance relative Nous proposons de plus une mesure de distance afin d'évaluer la proximité des bornes estimées par les modèles avec celles données par les participants. En effet, en cas de mauvaise prédiction de la part d'un modèle, on souhaite pouvoir évaluer le degré d'erreur, c'est-à-dire la distance entre l'estimation produite et la valeur de borne observée.

Pour tenir compte de l'erreur en considérant les dimensions de l'ENA, nous proposons de la définir relativement à la magnitude relative $RelMag(x)$ de la valeur de référence. En effet, une erreur de 10 unités par exemple apparaît comme plus significative pour $x = 100$ que pour $x = 1000$. Cependant, la même erreur doit être considérée comme aussi significative pour $x = 8150$ que pour $x = 50$. Ceci implique que le facteur de normalisation ne doit

3. Nous avons également effectué des tests avec une erreur relative de 5%, mais nous avons pu observer que cette valeur pénalise fortement les modèles qui produisent comme estimations des nombres réels : le modèle proportionnel RM et le modèle régressif REGM.

pas dépendre de la magnitude mais de la magnitude relative de l'ENA, nous conduisant à proposer :

$$RD(x, m, p, b) = \frac{|\Delta M_m^b(x) - \Delta P_p^b(x)|}{RelMag(x)} \tag{5.9}$$

En moyennant sur les deux bornes $b \in \{-, +\}$ et tous les participants $p \in \mathcal{P}(x)$, comme pour PA, nous proposons de définir cette distance relative RD, à minimiser, pour le modèle m comme :

$$RD(m) = \frac{1}{2 \cdot |\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{b \in \{-,+\}} \frac{1}{|\mathcal{P}(x)|} \sum_{p \in \mathcal{P}(x)} RD(x, m, p, b) \tag{5.10}$$

5.2.3.2 Critères de prédiction de médianes

Comme détaillé dans la section 4.4 p. 50, l'étude empirique montre qu'il n'y a pas de consensus concernant l'interprétation des ENA et qu'une grande variabilité apparaît dans les intervalles collectés pour une même ENA à travers les participants. Cette observation est cohérente avec la perspective épistémique du vague (voir section 2.1.1 p. 8) dans laquelle l'extension d'une ENA possède des limites nettes mais non nécessairement connues. Il nous semble donc pertinent de discuter de la définition de l'intervalle qu'un modèle d'interprétation doit estimer.

Le premier paragraphe détaille les motivations de l'approche proposée, qui se focalise sur la médiane des bornes observées. Les deux paragraphes suivants introduisent les critères de qualité conçus pour évaluer dans quelle mesure ce but est atteint.

Motivation de l'intervalle médian Nous proposons de modéliser les valeurs des bornes des intervalles correspondant à une ENA comme une distribution statistique pour tenir compte de la variabilité observée dans les intervalles collectés à travers les participants. La quantité de données nécessaire à l'apprentissage d'une telle distribution pouvant être très importante, nous proposons d'estimer des paramètres de cette distribution, tels que la moyenne ou la médiane.

Estimer l'intervalle médian plutôt que l'intervalle moyen en tant qu'indicateur central des distributions des bornes présente deux avantages. En premier lieu, la médiane est un indicateur qui prend une valeur observée dans la distribution. Comme remarqué dans l'étude empirique de l'interprétation des ENA (voir section 4.4 p. 50), les participants tendent à préférer les nombres saillants, c'est-à-dire plutôt ronds, comme valeurs de bornes. Estimer la médiane revient à produire de tels nombres qui semblent naturels pour les utilisateurs. Au contraire, la moyenne est une valeur qui n'est pas nécessairement observée dans la distribution. De plus, elle prend le plus souvent une valeur qui n'est pas ronde et qui peut, par conséquent, apparaître moins naturelle pour les utilisateurs. En second lieu,

la médiane est un paramètre robuste vis-à-vis des valeurs extrêmes alors que la moyenne y est sensible.

Taux de bonnes prédictions de médianes Dans un premier temps, nous proposons d'utiliser le taux de bonnes prédictions de médianes comme critère de qualité vis-à-vis de l'estimation de médiane. Nous notons $\Delta Med^b(x)$ la médiane sur l'ensemble des participants $p \in \mathcal{P}(x)$ des distances $\Delta P_p^b(x)$.

L'égalité entre la distance $\Delta M_m^b(x)$ de la borne $b \in \{-,+\}$ estimée par le modèle m à la valeur de référence b et la médiane des distances observée $\Delta Med^b(x)$ est calculée, comme pour PA, selon l'équation 4.2 (voir section 4.3.6, p. 49). Le score global du modèle m est obtenu en moyennant ce score d'égalité sur l'ensemble des ENA $x \in \mathcal{X}$. Le taux de bonnes prédictions de médianes MA, à maximiser, est donc défini comme :

$$MA(m) = \frac{1}{2 \cdot |\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{b \in \{-,+\}} Eq(\Delta M_m^b(x), \Delta Med^b(x)) \tag{5.11}$$

Score d'erreur de médiane En cas d'estimation incorrecte de la médiane, on souhaite pouvoir évaluer le degré d'erreur, c'est-à-dire la distance entre l'estimation et la médiane observée. Pour ce faire, nous proposons un critère de qualité qui évalue l'équilibre entre la quantité de valeurs données par les participants qui sont en dessous et au dessus de l'estimation de la borne considérée.

Ces quantités peuvent être respectivement définies comme

$$N_-(x,b) = |\{p \in \mathcal{P}(x) | \Delta P_p^b(x) < \Delta M_m^b(x)\}| \tag{5.12}$$

$$\text{et } N_+(x,b) = |\{p \in \mathcal{P}(x) | \Delta P_p^b(x) > \Delta M_m^b(x)\}|. \tag{5.13}$$

Le modèle m devrait produire une estimation telle que $N_+(x,b) = N_-(x,b)$ pour toute borne $b \in \{-,+\}$ de toute ENA $x \in \mathcal{X}$.

Toutefois, dans la mesure où les bornes des intervalles fournies par les participants ne sont pas uniformément distribuées mais plutôt réparties sur quelques valeurs, un équilibre parfait peut ne pas être atteignable. Par exemple, pour l'ENA "*environ 50*", les valeurs des bornes inférieures données par les participants pourraient être 40 (donnée par 20 participants), 45 (donnée par 30 participants) et 48 (donnée par 10 participants). La médiane de cette distribution est $\Delta Med^-(50) = 50 - 45 = 5$. Même si le modèle prédit correctement cette médiane, l'équilibre parfait n'est pas atteint car 20 bornes observées sont en dessous et 10 sont au dessus. L'inégalité entre $N_-(x,b)$ et $N_+(x,b)$ ne reflète donc pas le fait que la médiane est correctement estimée.

Pour surmonter cette faiblesse, nous proposons que le score d'équilibre de l'estimation

tienne compte de l'équilibre vis-à-vis de la médiane observée, c'est-à-dire des deux quantités

$$N_-^*(x,b) = |\{p \in \mathcal{P}(x) | \Delta P_p^b(x) < \Delta Med^b(x)\}| \quad (5.14)$$

$$\text{et } N_+^*(x,b) = |\{p \in \mathcal{P}(x) | \Delta P_p^b(x) > \Delta Med^b(x)\}|. \quad (5.15)$$

Le score du modèle dépend dès lors de la différence absolue entre $N_+(x,b)$ et $N_+^*(x,b)$ ainsi que de la différence absolue entre $N_-(x,b)$ et $N_-^*(x,b)$. En moyennant sur les deux bornes de toutes les ENA, le score d'erreur de médiane, à minimiser, est défini comme :

$$MErr(m) = \frac{1}{2 \cdot |\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{b \in \{-,+\}} \frac{1}{|\mathcal{P}(x)|} \cdot (|N_+(x,b) - N_+^*(x,b)| + |N_-(x,b) - N_-^*(x,b)|) \quad (5.16)$$

5.2.3.3 Procédure expérimentale

En utilisant les quatre critères de qualité PA, RD, MA et $MErr$ décrits ci-dessus, ainsi que les intervalles collectés empiriquement décrits dans la section 4.4 p. 50, nous comparons les performances des deux modèles que nous proposons (voir section 5.1 p. 62), le modèle log-linéaire LLM et le modèle des rangs RKM, à celles des trois modèles de la littérature discutés dans la section 3.2 p. 29 : RM, SBM et REGM.

Pour les modèles utilisant une phase d'apprentissage (i.e., REGM et LLM), une procédure de validation croisée est réalisée sur deux benchmarks de 1000 exécutions des étapes d'apprentissage et de test. Pour chaque exécution, tous les modèles, qu'ils nécessitent ou non un apprentissage, sont évalués sur le même ensemble de test :

1. **Participant** : l'ensemble d'apprentissage est composé de tous les intervalles de 75% des participants, aléatoirement sélectionnés. Les intervalles des 25% participants restants forment l'ensemble de test. Le but de ce benchmark est d'évaluer dans quelle mesure les modèles nécessitant un apprentissage sont capables de généraliser sur les participants.
2. **ENA** : l'ensemble d'apprentissage est composé des intervalles donnés par tous les participants à 17 (66,7%) des 24 ENA, aléatoirement tirées. Les intervalles de tous les participants correspondant aux 7 ENA restantes forment l'ensemble de test. Ce benchmark est conçu pour évaluer la robustesse de LLM car les paramètres de régression y sont estimés à partir de seulement $17 \cdot 2 = 34$ triplets. Il a également pour but d'évaluer la capacité de généralisation des modèles quant aux ENA.

Pour éviter tout biais vers les modèles qui sont avantagés par les valeurs de référence à un seul ou à plusieurs chiffres significatifs, une contrainte est ajoutée à la sélection aléatoire dans le benchmark ENA : les ensembles d'apprentissage et de test doivent tous deux inclure à la fois des nombres à un seul et à plusieurs chiffres significatifs.

Pour déterminer quel modèle offre les meilleures performances à chaque benchmark, des tests ANOVA avec le modèle comme facteur sont réalisés. En cas d'effet significatif observé, des tests post-hoc HSD de Tukey (Winer et al., 1991) sont réalisés. Le niveau de significativité est fixé pour toutes ces analyses à $p = 0,01$.

5.2.3.4 Paramétrisation des modèles

Plusieurs combinaisons de paramètres ont été testées pour les modèles paramétrés. Les résultats présentés dans la section 5.2.4 sont ceux obtenus avec les paramètres qui conduisent aux meilleurs scores pour les quatre critères de qualité. En effet, il a pu être observé qu'une modification des paramètres d'un modèle avait toujours pour conséquence un changement dans la même direction de tous les scores.

RM offre les meilleures performances lorsque le paramètre $s = 5\%$. Pour SBM, le système décimal $S = \{1, 10, 100, \ldots\}$ est celui qui permet au modèle de mieux correspondre aux données.

Dans la mesure où REGM évalue la taille des intervalles et ne donne aucune information sur leur position ou leur symétrie, nous faisons par défaut l'hypothèse que ceux-ci sont centrés autour de la valeur de référence des ENA.

Lors de la phase d'apprentissage, LLM requiert le choix d'un représentant des intervalles collectés correspondant à une ENA. Pour tenir compte de la variabilité observée (voir section 4.4 p. 50), nous proposons d'utiliser la médiane comme opérateur d'agrégation des intervalles pour constituer sa base de triplets d'apprentissage (voir section 5.2.1 p. 67).

5.2.3.5 Examen préliminaire du modèle régressif REGM

Parmi les trois modèles identifiés dans la littérature, seul REGM est basé sur une régression linéaire. Nous proposons, dans un premier temps, un examen préliminaire de ce modèle en testant la fonction de régression proposée par Ferson et al. (2015) (voir équation 3.6 section 3.2.3 p. 32) sur l'ensemble des données considérées (voir section 4.4 p. 50).

Le tableau 5.1 présente les coefficients de régression obtenus et ceux rapportés par les auteurs (Ferson et al., 2015). Quatre variables de la régression sur sept apparaissent comme statistiquement significatives sur les données collectées : l'ordre de magnitude ($O_m(x)$, associée à ω_1), la rondeur ($R(x)$, associée à ω_2), le produit de l'ordre de magnitude par la rondeur ($O_m(x) \cdot R(x)$, associée à ω_4), et le produit de la rondeur par la *fiveness* ($R(x) \cdot f(x)$, associée à ω_6). D'autre part, le coefficient de détermination est nettement plus faible sur les données collectées ($R^2 = 0,272$) que celui rapporté par les auteurs ($R^2 = 0,741$).

Ces résultats suggèrent que le modèle régressif proposé par Ferson et al. (2015) ne

Paramètre	Variables	Rapporté (Ferson et al., 2015)	Obtenu sur les données collectées
ω_0	-	**-0,208**	0,375
ω_1	$O_m(x)$	**0,428**	**0,220**
ω_2	$R(x)$	**0,281**	**-0,00485**
ω_3	$f(x)$	**0,0940**	0,977
ω_4	$O_m(x) \cdot R(x)$	**0,0147**	**0,0807**
ω_5	$O_m(x) \cdot f(x)$	**-0,0640**	-0,244
ω_6	$R(x) \cdot f(x)$	**-0,0102**	**-0,496**
ω_7	$O_m(x) \cdot R(x) \cdot f(x)$	**0,0404**	0,138

TABLEAU 5.1 – Coefficients de la fonction de régression de REGM (voir équation 3.6 section 3.2.3 p. 32), rapportés par Ferson et al. (2015) et obtenus à partir des données collectées. Les valeurs en gras sont celles qui apparaissent comme significatives.

rend pas bien compte des intervalles que nous avons collectés. Ceci peut être dû à des différences de méthodologie de collecte des intervalles : alors que Ferson et al. (2015) incluent les ENA dans différents contextes sémantiques, celles de ce travail ne sont pas contextualisées. De plus, les participants de l'étude empirique ont pour langue maternelle le français alors que les participants de l'étude de Ferson et al. (2015) parlent anglais. Une étude plus approfondie sur les raisons de ces différences constitue l'une des perspectives de l'examen de REGM pour confirmer ces pistes.

Lors de l'exécution de notre protocole de validation expérimentale, il apparaît que certains tirages aléatoires dans le benchmark ENA conduisent à des coefficients de régression aberrants pour REGM. Cette observation peut s'expliquer par la variable *fiveness* f, une propriété qui n'apparaît que cinq fois dans les ENA de ce travail : certaines bases d'apprentissage ne possèdent pas assez d'ENA présentant cette propriété pour estimer correctement les coefficients relatifs à f.

Les exécutions lors desquelles des coefficients aberrants sont observés, conduisant à des bornes estimées au delà de 10 000 unités de la valeur de référence de l'ENA, ne sont pas incluses dans les analyses et ne sont donc pas prises en compte dans les résultats présentés dans la section suivante. Sur 1000 exécutions dans le benchmark ENA, des coefficients aberrants sont observés 88 fois. Par conséquent, les moyennes et écart-types des performances de REGM au benchmark ENA sont calculés sur 912 exécutions.

	Scores au benchmark Participant			
Modèle	Prédiction de borne PA (%)	Distance relative RD	Prédiction de médiane MA (%)	Erreur de médiane $MErr$
RM	8,9 (1,1)	0,78 (0,05)	18,4 (7,2)	0,76 (0,11)
SBM	19,5 (2,4)	0,43 (0,08)	28,0 (6,9)	0,76 (0,08)
REGM	7,9 (1,7)	**0,39 (0,09)**	20,0 (7,2)	0,67 (0,18)
LLM	**22,4 (2,6)**	**0,38 (0,08)**	54,1 (9,6)	0,38 (0,11)
RKM	**22,2 (2,6)**	**0,39 (0,08)**	**58,3 (8,9)**	**0,35 (0,12)**

TABLEAU 5.2 – Moyennes et, entre parenthèses, écart-types des scores obtenus au benchmark Participant par chaque modèle aux quatre critères de qualité. Les scores en gras sont statistiquement les meilleurs, selon les tests ANOVA et post-hoc réalisés. Plusieurs scores en gras pour un même critère indiquent une absence de différence significative lors des analyses post-hoc.

5.2.4 Validation expérimentale : résultats

Les tableaux 5.2 et 5.3 présentent les performances des cinq modèles aux quatre critères de qualité proposés, alors que les figures 5.2 et 5.3 les illustrent pour les deux benchmarks respectivement (Participant : tableau 5.2 et figure 5.2, ENA : tableau 5.3 et figure 5.3). On peut observer que les modèles LLM et RKM que nous proposons offrent les meilleures performances. Comme les résultats aux deux benchmarks sont similaires, ils ne sont pas commentés séparément. Les sous-sections suivantes discutent les résultats obtenus pour chacun des quatre critères de qualité séparément.

5.2.4.1 Taux de bonnes prédictions de bornes, PA

Statistiquement, les scores PA obtenus par LLM et RKM sont les meilleurs et ne diffèrent pas significativement l'un de l'autre selon les tests post-hoc réalisés. RM et REGM offrent de faibles performances en prédiction de borne. On peut définir une valeur de référence, calculée comme la moyenne des fréquences relatives du mode des bornes (voir section 5.2.3.1 p. 69). Sur les données considérées, elle vaut 28,1% ($\sigma = 6,9$), score dont s'approchent LLM et RKM. On observe que LLM et RKM sont comparables, alors que les performances des autres modèles sont significativement plus faibles.

Les scores élevés de LLM et RKM peuvent être dus au fait qu'ils produisent des intervalles qui ne sont pas nécessairement symétriques alors que les trois autres modèles produisent des intervalles toujours centrés autour de la valeur de référence des ENA.

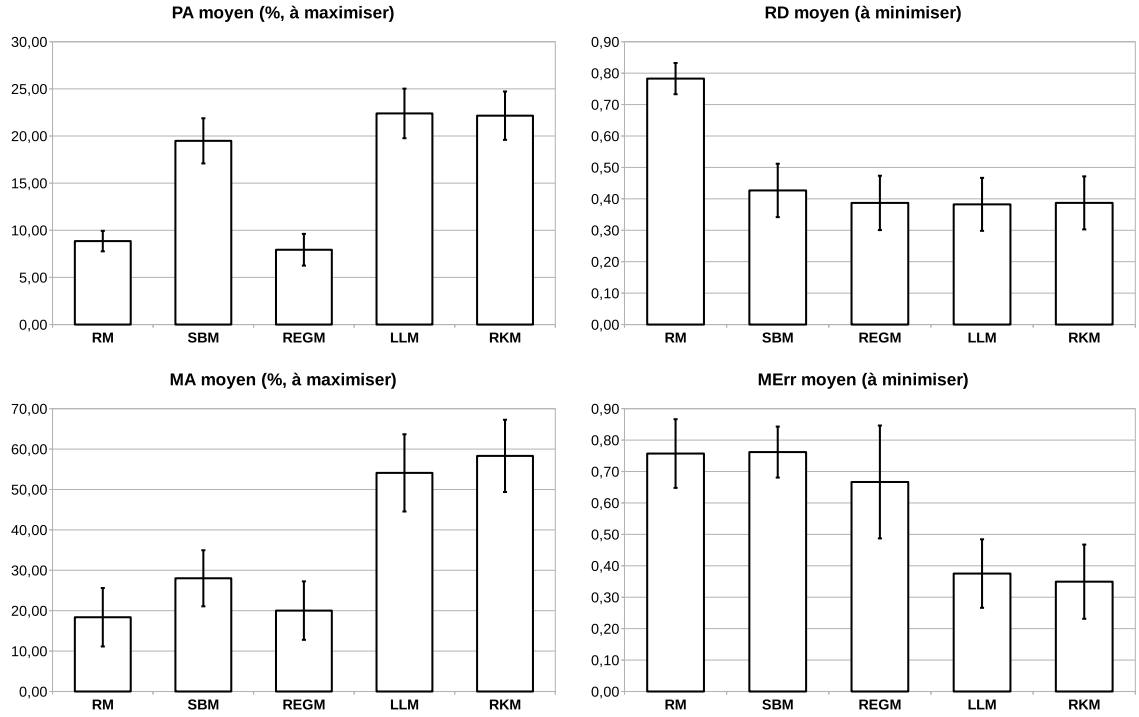


FIGURE 5.2 – Moyennes des scores de chaque modèle au benchmark Participant : prédiction de borne PA (en haut, à gauche), distance relative RD (en haut, à droite), prédiction de médiane MA (en bas, à gauche), et erreur de médiane $MErr$ (en bas, à droite). Les deux modèles que nous proposons, LLM et RKM, offrent les meilleures performances aux quatre critères de qualité.

Les scores faibles de RM et REGM peuvent s'expliquer par le fait qu'ils ont tendance à produire des valeurs réelles comme estimations de borne alors que les participants ont tendance à donner des nombres entiers. De plus, RM est particulièrement adapté aux ENA dont la valeur de référence ne possède qu'un seul chiffre significatif, conduisant à de faibles performances sur les ENA dont la valeur de référence en possède plusieurs.

5.2.4.2 Distance relative, RD

Les tests ANOVA réalisés montrent un effet significatif du facteur modèle sur les scores RD. Les tests post-hoc montrent quant à eux que cet effet est surtout dû à RM puisque les scores des autres modèles ne différent pas significativement les uns des autres.

Comme pour la prédiction de bornes, le modèle RM offre de faibles performances car il est peu adapté aux valeurs de référence à plusieurs chiffres significatifs. Ses performances sont d'autant plus basses que la différence entre la magnitude et la magnitude relative de la valeur de référence des ENA est élevée. En effet, les distances relatives les plus hautes

	Scores au benchmark ENA			
Modèle	Prédiction de borne PA (%)	Distance relative RD	Prédiction de médiane MA (%)	Erreur de médiane $MErr$
RM	8,7 (3,0)	0,84 (0,65)	24,4 (13,0)	0,70 (0,17)
SBM	19,3 (2,7)	**0,44 (0,23)**	24,9 (14,5)	0,79 (0,18)
REGM	8,1 (4,1)	**0,45 (0,64)**	15,7 (13,2)	0,65 (0,14)
LLM	**21,9 (3,2)**	**0,41 (0,25)**	56,6 (15,3)	0,32 (0,14)
RKM	**22,0 (3,1)**	**0,40 (0,22)**	**63,8 (14,0)**	**0,27 (0,16)**

TABLEAU 5.3 – Moyennes et, entre parenthèses, écart-types des scores obtenus au benchmark ENA par chaque modèle aux quatre critères de qualité. Les scores en gras sont statistiquement les meilleurs, selon les tests ANOVA et post-hoc réalisés. Plusieurs scores en gras pour un même critère indiquent une absence de différence significative lors des analyses post-hoc.

sont observées pour "*environ 4730*" ($RD = 6,92$) et "*environ 8150*" ($RD = 7,46$), les deux ENA avec la plus grande différence entre l'ordre de grandeur de la magnitude et celui de la magnitude relative, alors que les scores obtenus avec les autres ENA varient entre 0,06 et 0,73.

Une double explication peut rendre compte de l'absence de différence significative au score de distance relative entre les quatre autres modèles. Premièrement, REGM et LLM sont des modèles basés sur des fonctions de régression. Bien qu'ils ne tiennent pas compte des mêmes dimensions des ENA, ils tendent tous les deux à minimiser la distance entre les données d'apprentissage et la fonction apprise, conduisant à des estimations proches des intervalles collectés.

Deuxièmement, pour SBM et REGM, les résultats montrent des différences entre des ENA spécifiques, selon les dimensions prises en compte par le modèle : SBM offre de faibles performances pour des ENA dont la magnitude et la magnitude relative sont très différentes (par ex., $x = 8150$) car il ne considère pas la magnitude d'une ENA. L'absence de différence significative entre le score de ce modèle et celui des autres modèles peut donc s'expliquer par le faible nombre d'ENA présentant cette caractéristique dans notre corpus.

5.2.4.3 Taux de bonnes prédictions de médianes, MA, et erreur de médiane, $MErr$

Les analyses statistiques montrent que RKM offre les meilleures performances, à la fois en prédiction et en erreur de médiane, validant ainsi empiriquement ce modèle ainsi que

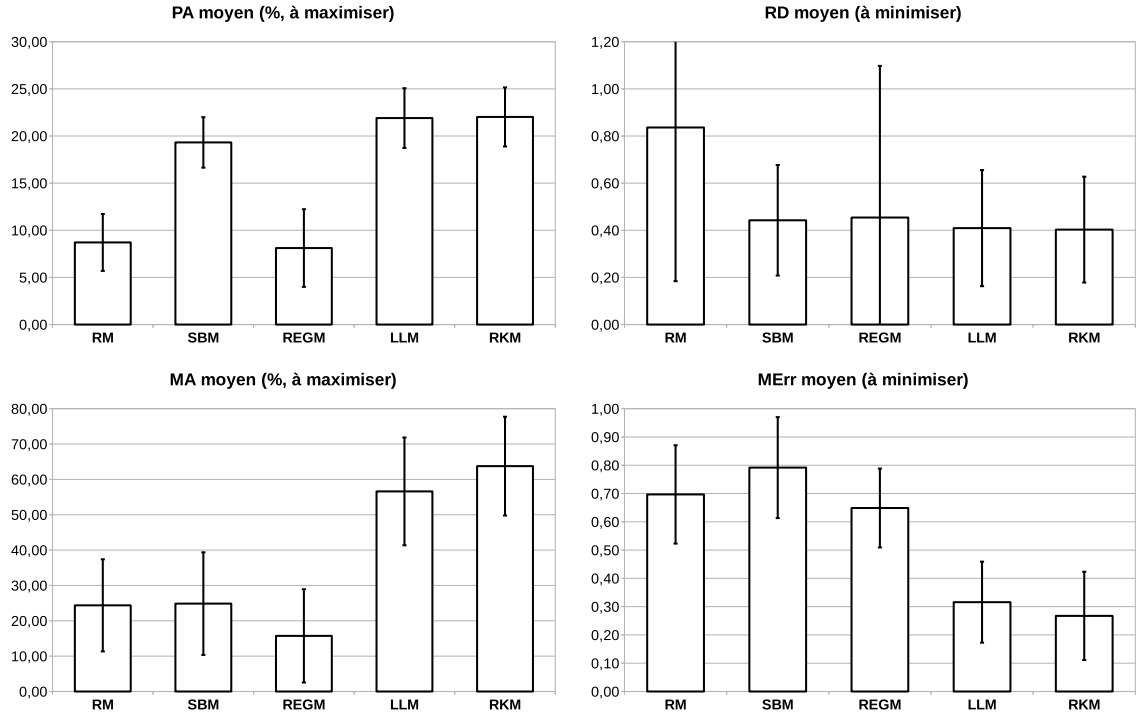


FIGURE 5.3 – Moyennes des scores de chaque modèle au benchmark Participant : prédiction de borne PA (en haut, à gauche), distance relative RD (en haut, à droite), prédiction de médiane MA (en bas, à gauche), et erreur de médiane $MErr$ (en bas, à droite). Les deux modèles que nous proposons, LLM et RKM, offrent les meilleures performances aux quatre critères de qualité.

le principe général du compromis entre saillance cognitive et taille de la plage de valeurs dénotées que nous proposons.

REGM obtient un bon score en prédiction mais un score moyen en erreur. Comme pour la prédiction de bornes, ceci peut être dû au fait que le modèle produit des valeurs réelles comme estimations alors que les participants ont tendance à donner des entiers. Toutefois, le score d'erreur montre que ces estimations réelles sont plus proches des médianes observées que les estimations fournies par SBM et RM. A l'inverse, et en cohérence avec les observations précédentes, SBM est meilleur que REGM en taux de bonnes prédictions alors que ses erreurs sont plus importantes.

5.2.4.4 Différences entre ENA à un seul et à plusieurs chiffres significatifs

Nous avons réalisé des analyses complémentaires pour examiner dans quelle mesure les performances des modèles sur l'ensemble des ENA commentées dans dans les pages précédentes sont également valables lorsqu'une distinction est faite entre ENA dont la

valeur de référence ne possède qu'un seul chiffre significatif (par ex., "*environ 50*") et les ENA dont la valeur de référence en possède plusieurs (par ex., "*environ 8150*"). Les tableaux C.2, p. 192 et C.3, p. 193 en annexe C présentent les résultats détaillés. Ils montrent que les scores obtenus pour ces deux types d'ENA sont très similaires pour LLM, RKM et REGM, au benchmark Participant.

RM offre de bonnes performances pour les ENA dont la valeur de référence ne possède qu'un seul chiffre significatif et de mauvaises performances pour celles à plusieurs chiffres significatifs, SBM présente le profil inverse. Ces résultats sont en accord avec les caractéristiques de ces deux modèles : RM ne considère que la magnitude tandis que SBM ne tient que compte que de la granularité.

Quant au benchmark ENA, on peut observer une différence entre les deux types d'ENA pour tous les modèles, bien que cette différence soit moins marquée pour RKM. En effet, les modèles offrent de moins bonnes performances avec des ENA dont la valeur de référence possède plusieurs chiffres significatifs par rapport à celle n'en possédant qu'un seul. De plus, les écart-types sont plus élevés pour les ENA à plusieurs chiffres significatifs.

Ces différences entre les deux benchmarks peuvent être dues au tirage aléatoire des ENA dans le benchmark ENA. En effet, pour ce dernier, les modèles sont évalués sur seulement 7 ENA : chaque erreur d'estimation contribue à hauteur de 14% du score final, ce qui peut rapidement conduire à de faibles scores.

De plus, comme observé pour le benchmark Participant, RM et SBM sont plus ou moins adaptés selon le nombre de chiffres significatifs de la valeur de référence de l'ENA considérée. Leurs scores dépendent dès lors beaucoup du tirage aléatoire des ENA dans l'ensemble de test. Concernant REGM et LLM, qui requièrent une phase d'apprentissage, l'ensemble d'entraînement est réduit à 34 triplets au benchmark ENA alors qu'il comporte 48 triplets dans le benchmark Participants. Les paramètres de régression appris peuvent donc être moins robustes dans le benchmark ENA.

Il peut être conclu de ces résultats distinguant ENA à un seul et à plusieurs chiffres significatifs que la capacité de généralisation des modèles est sensiblement meilleure sur les participants que sur les ENA. Par conséquent, l'ensemble d'apprentissage doit être attentivement constitué de manière à couvrir une large plage de combinaisons entre les dimensions des ENA.

5.3 Exploitation 2 : modèle flou d'interprétation fRKM

L'approche des modèles présentés dans la section précédente s'inscrit dans la perspective épistémique du vague et consiste à estimer les bornes de l'intervalle correspondant à la plage de valeurs dénotées par une ENA. Une autre approche, orientée vers la perspective

sémantique du vague, consiste à modéliser l'aspect graduel de l'interprétation des bornes d'une ENA sous la forme d'intervalles flous (Zadeh, 1983).

L'objectif de cette seconde exploitation du principe général du compromis entre saillance cognitive et taille de la plage de valeurs dénotées est de caractériser le support, le noyau et la 0,5-coupe des intervalles flous représentant les ENA, à partir des valeurs situées sur les fronts de Pareto (voir section 5.1 p. 62).

La première sous-section décrit le modèle flou fRKM (*fuzzy RanK Model*) que nous proposons. Dans un deuxième temps, nous détaillons le protocole de validation expérimentale mis en œuvre. Enfin, nous présentons et discutons les résultats de cette validation.

5.3.1 Modèle proposé fRKM

L'objectif du modèle proposé est de caractériser le support, le noyau et la 0,5-coupe d'un intervalle flou représentant une ENA "*environ x*", à partir des valeurs situées sur les fronts de Pareto (voir section 5.1 p. 62). Dans la mesure où le principe général du compromis entre saillance cognitive et taille de la plage de valeurs dénotées rend compte de l'asymétrie observée des intervalles correspondant aux ENA, le support, le noyau et la 0,5 coupe ne sont pas nécessairement symétriques par rapport à la valeur de référence d'une ENA.

Toute valeur en dehors de l'intervalle du support, noté $I_S(x)$, est considérée comme absolument non dénotée par l'ENA. Nous proposons de définir les valeurs des fronts de Pareto les plus éloignées de x comme bornes de cet intervalle, formellement :

$$I_S(x) = [y^-_{n^-}, y^+_{n^+}] \tag{5.17}$$

Toute valeur à l'intérieur de l'intervalle du noyau, noté $I_K(x)$, est considérée comme pleinement dénotée par l'ENA. Nous proposons de définir les valeurs des fronts de Pareto les plus proches de x comme bornes de cet intervalle, formellement :

$$I_K(x) = [y^-_1, y^+_1] \tag{5.18}$$

Dans la perspective *random set* d'élicitation de sous-ensembles flous proposée par Bilgiç et Türkşen (2000), la 0,5-coupe correspond à l'intervalle médian (voir section 3.1.2 p. 24). Comme bornes de l'intervalle correspondant à la 0,5-coupe, noté $I_M(x)$, nous proposons donc d'utiliser les estimations produites par le modèle des rangs RKM dont nous avons montré qu'elles correspondent bien aux médianes observées (voir section 5.2.2 p. 68), formellement :

$$I_K(x) = [y^-_{r_P(x)}, y^+_{r_P(x)}] \tag{5.19}$$

Pour construire le sous-ensemble flou à partir du support, de la 0,5-coupe et du noyau estimés, nous proposons d'interpoler linéairement ceux-ci, comme illustré dans la figure 5.4.

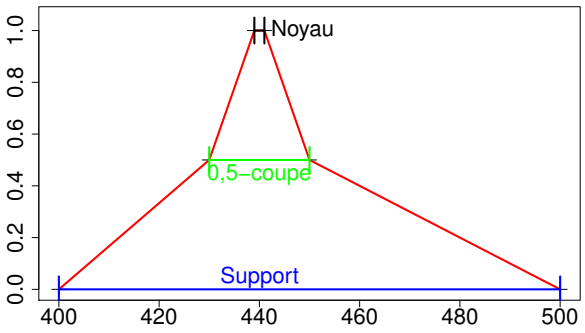


FIGURE 5.4 – Support, noyau et 0,5-coupe du sous-ensemble flou représentant l'ENA "*environ 440*", à partir des valeurs situées sur les fronts de Pareto, reliés par interpolation linéaire.

Par exemple, pour "*environ 440*", illustré par la figure 5.4, on obtient :

$$\begin{aligned} I_S(440) &= [400, 500] \\ I_K(440) &= [439, 441] \\ I_M(440) &= [430, 450] \end{aligned}$$

5.3.2 Validation expérimentale : protocole

Cette section présente l'étude expérimentale réalisée pour évaluer la qualité des trois paramètres des intervalles flous correspondant aux ENA : le support, le noyau et la 0,5-coupe. Dans la mesure où l'estimation de la 0,5-coupe correspond à celle de la médiane de la distribution des bornes données par les participants, nous ne reproduisons pas ici la validation expérimentale de ce paramètre (voir section 5.2.4 p. 76).

Evaluer la qualité des estimations du support et du noyau de la même manière que pour la 0,5-coupe pose le problème des valeurs extrêmes. En effet, dans la perspective *random set*, le support correspond à l'intervalle le plus large donné par les participants et le noyau correspond à l'intervalle le plus réduit. Par conséquent, la présence d'une seule réponse extrême aboutit à un support ou à un noyau aberrant. Le taux de bonnes prédictions ou la distance à la valeur observée sont donc deux critères manquant de robustesse vis-à-vis de ces réponses extrêmes.

Pour surmonter ce problème, nous proposons de construire une fonction d'appartenance de référence, $f^E_{\widetilde{x}}(y)$, élicitée à partir des données collectées. Cette dernière est construite dans la perspective *random set*, définissant $f^E_{\widetilde{x}}(y)$ comme la fréquence relative cumulée des participants incluant y dans l'intervalle correspondant à "*environ* x". La fonction d'ap-

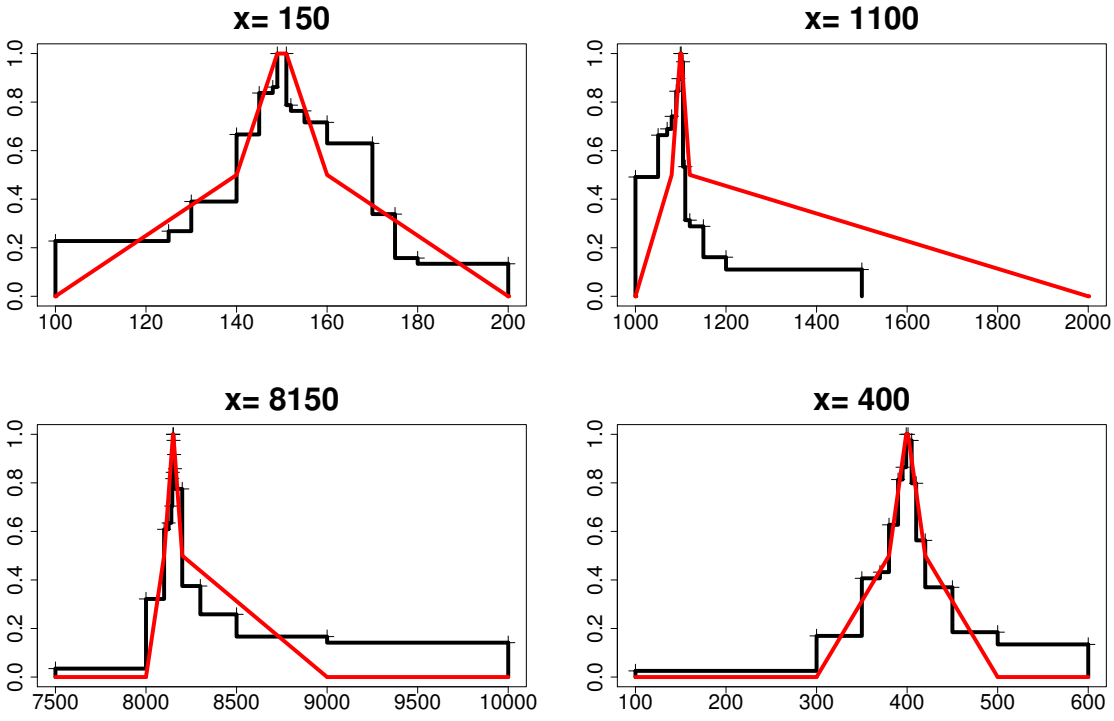


FIGURE 5.5 – Fonctions d'appartenance générées (en rouge) et élicitées (en noir) de quatre des 24 ENA considérées dans cette étude : "*environ 150*" (en haut, à gauche), "*environ 1100*" (en haut, à droite), "*environ 8150*" (en bas, à gauche) et "*environ 400*" (en bas, à droite).

partenance générée par le modèle, notée $f^G_{\widetilde{x}}(y)$, est ensuite comparée à cette fonction de référence.

Nous proposons de comparer $f^G_{\widetilde{x}}(y)$ à $f^E_{\widetilde{x}}(y)$ en utilisant l'aire de leur différence, relativement à l'aire sous la fonction d'appartenance de référence $f^E_{\widetilde{x}}(y)$. Ce critère, à minimiser et noté MFQ (*Membership Function Quality*), peut être formalisé de la manière suivante :

$$MFQ(x) = \frac{\int_y |f^G_{\widetilde{x}}(y) - f^E_{\widetilde{x}}(y)|}{\int_y f^E_{\widetilde{x}}(y)} \tag{5.20}$$

5.3.3 Validation expérimentale : résultats

La figure 5.5 illustre quatre exemples de fonctions d'appartenance générées et élicitées (voir figure C.1 p. 194 en annexe C pour les résultats détaillés de l'ensemble des 24 ENA). On peut remarquer que la fonction générée s'accorde bien avec la fonction élicitée pour l'ensemble des 24 ENA. De plus, l'asymétrie de $f^E_{\widetilde{x}}(y)$ semble bien capturée par le modèle.

Les scores obtenus pour chacune des 24 ENA considérées sont présentés dans le tableau C.4 p. 193 de l'annexe C. En moyenne, le score de qualité est de 0,502, allant de

0,211 pour "*environ 150*" à 0,950 pour "*environ 1100*".

A partir de l'analyse visuelle de la correspondance entre les fonctions d'appartenance, nous proposons de fixer un seuil de $MFQ = 0,6$ pour considérer une estimation comme bonne. Selon ce critère, 17 des 24 fonctions d'appartenance générées, soit 70,1%, peuvent être considérées comme correctes, résultat qui tend à valider le modèle d'interprétation flou proposé.

Comme attendu, la présence de valeurs extrêmes (par ex., 7500 et 10000 pour "*environ 8150*" ; 100 et 600 pour "*environ 400*") tendent à faire baisser le score de certaines ENA. Dans le cas particulier de "*environ 1100*" (figure 5.5, en haut, à droite), le faible ajustement observé et la faible performance ($MFQ = 0,951$) peuvent s'expliquer par le fait que la dernière valeur du front de Pareto de la borne supérieure est 2000, une valeur de borne qui n'a pas été donnée par les participants.

En détaillant les différences entre ENA dont la valeur de référence ne possède qu'un seul chiffre significatif et celles qui en possèdent plusieurs, il apparaît que le score moyen obtenu par les premières ($MFQ = 0,488$) est similaire à celui obtenu par les secondes ($MFQ = 0,524$). Toutefois, les écart-types montrent une variabilité nettement plus importante pour les ENA dont la valeur de référence possède plusieurs chiffres significatifs ($\sigma = 0,272$) que pour celles n'en possédant qu'un seul ($\sigma = 0,087$), indiquant que certaines ENA sont mieux capturées que d'autres. En particulier, les scores de "*environ 1100*" ($MFQ = 0,951$, voir figure 5.5) et de "*environ 4730*" ($MFQ = 0,776$, voir figure C.1 p. 194 de l'annexe C) sont plus éloignés de la moyenne que les scores d'autres ENA.

5.4 Bilan

Dans ce chapitre, nous avons présenté notre contribution à la modélisation computationnelle de l'interprétation des ENA. Dans un premier temps, nous avons décrit un principe général basé sur l'optimisation d'un compromis entre la saillance cognitive d'une valeur de borne et sa distance à la valeur de référence. Ce principe permet de capturer une grande majorité des données collectées lors de l'étude empirique de l'interprétation des ENA décrite au chapitre précédent.

Nous avons ensuite décrit trois modèles d'interprétation des ENA basé sur ce principe général. Les deux premiers, log-linéaire LLM et des rangs RKM, s'inscrivent dans la perspective épistémique du vague et permettent d'estimer l'intervalle de valeurs dénotées par une ENA. Les résultats de la comparaison entre les deux modèles que nous proposons et ceux de la littérature ont validé la pertinence du point de vue cognitif de nos propositions : ils rendent mieux compte des intervalles correspondant à des ENA collectées auprès de non-experts.

Le troisième modèle que nous proposons, fRKM s'inscrit dans la perspective sémantique du vague et est basé sur la représentation par sous-ensembles flous des ENA. Il permet de générer des intervalles flous. La validation expérimentale que nous avons réalisée montre que fRKM rend bien compte des intervalles collectés lors de l'étude empirique décrite au chapitre précédent.

Nous présentons, dans le prochain chapitre, la mise en œuvre concrète du modèle flou d'interprétation des ENA que nous avons réalisée dans le cadre d'une application de requêtes flexibles dans des bases de données. Nous y décrivons également une extension applicable aux trois modèles que nous avons proposés dans ce chapitre.

Chapitre 6

Interprétation des ENA : application et extension des modèles

Ce chapitre regroupe deux exploitations à caractère exploratoire, qui se situent à des niveaux différents, des modèles proposés et décrits dans le chapitre précédent pour interpréter des expressions numériques approximatives.

La première exploitation, présentée dans la section 6.1, est de type applicatif : elle vise à mettre en œuvre le modèle flou d'interprétation des ENA, fRKM, dans le cadre de moteurs de requêtes flexibles. Dans le contexte d'interrogation de bases de données, ces dernières augmentent l'expressivité des moteurs classiques en autorisant des requêtes imprécises : par exemple, pour une base de données décrivant des voitures d'occasion, ils permettent à l'utilisateur de rechercher les voitures ayant un kilométrage *faible*. L'exploitation du modèle fRKM dans ce cadre permet de combiner cette expressivité accrue avec une formulation d'expressions intuitives pour l'utilisateur, pour modéliser de façon pertinente une requête visant, par exemple, à identifier les voitures dont le kilométrage est d'*environ 140 000km*.

La seconde exploitation, à un niveau plus théorique, vise à étendre les modèles proposés au-delà du cadre initial des entiers naturels, en considérant deux généralisations décrites dans la section 6.2. Il s'agit d'une part de proposer une méthode d'interprétation pour des ENA impliquant des nombres décimaux, comme "*environ 3,14*". D'autre part, la généralisation au-delà des entiers naturels vise aussi à ne pas imposer que les bornes des intervalles produits soient des entiers. En effet, dans le cas par exemple de "*environ 16*", celle-ci permet d'envisager $[15, 17]$ ou $[14, 18]$, mais elle exclut la possibilité d'interpréter l'expression avec $[15, 5, 16, 5]$, qui peut néanmoins paraître pertinent.

6.1 Application aux moteurs de requêtes flexibles

Cette section décrit une exploitation pratique du modèle flou d'interprétation des ENA fRKM présenté dans le chapitre précédent, dans le contexte applicatif des moteurs de requêtes flexibles : après avoir présenté le domaine et les problématiques associées, elle décrit et illustre la mise en œuvre proposée ainsi que son implémentation dans le système ReqFlex (Smits et al., 2013).

Nous remercions Gregory Smits et Olivier Pivert pour leur collaboration qui a permis l'implémentation de cette application.

6.1.1 Contexte

Dans les systèmes classiques de bases de données, les requêtes sont exprimées en logique binaire. Le résultat renvoyé est l'ensemble des données de la base qui rendent la requête vraie. A titre d'exemple, considérons le cas d'une base de données gérée par les ressources humaines d'une entreprise, décrivant ses salariés. Une requête typique serait de lister les employés dont le salaire est compris entre 1900 et 2100 euros. Ce type de bases de données nécessite que les connaissances quant à ce qui doit être trouvé soient précises et bien connues.

Les requêtes flexibles reposent sur l'utilisation de la logique floue plutôt que de la logique classique pour autoriser des degrés de vérité et donc identifier les résultats qui rendent la requête vraie à différents degrés (Bosc & Pivert, 1992). Elles permettent de considérer des réponses qui ne respectent pas, de peu, les contraintes exprimées par des requêtes classiques, par exemple un employé dont le salaire est 1897 euros. Plus généralement, les requêtes flexibles autorisent une gradualité dans le degré de vérité : plus on s'éloigne des contraintes strictes, plus celui-ci diminue. Ce degré de vérité peut être interprété en terme de préférences pondérées du côté utilisateur. Ainsi, dans l'exemple ci-dessus, le degré de vérité des employés gagnant entre 1900 et 2100 euros peut être égal à 1, reflétant le fait qu'ils correspondent parfaitement aux préférences de l'utilisateur ; le degré de vérité d'un employé dont le salaire est 1850 peut être 0,7, reflétant le fait qu'il correspond un peu moins à ses préférences.

Trois problématiques sont associées aux requêtes flexibles, à plusieurs niveaux. Au niveau théorique, elles impliquent la formulation d'un langage de requête spécifique. SQLf (Bosc & Pivert, 1995) est un langage étendant le langage classique d'interrogation de bases de données SQL qui répond à cet objectif. Il permet d'utiliser, au delà des prédicats classiques, des prédicats flous exprimés en langage naturel, tels que *jeune* et *grand*, plutôt que des prédicats booléens, tels que *inférieur à 30 ans* et *supérieur à 1,75m*. Chaque prédicat flou est représenté par une fonction d'appartenance. Une requête flexible peut comporter

plusieurs prédicats, combinés selon les principes de la logique floue, par exemple *jeune et grand*.

Au niveau algorithmique et pratique, ensuite, se pose la question d'une implémentation efficace d'un moteur d'inférence (Bosc & Pivert, 2000). La *dérivation* (Bosc & Pivert, 1993) est l'une des méthodes classiquement mises en œuvre pour y répondre. Elle comporte trois étapes. La première consiste à traduire la requête flexible en requête classique. Celle-ci est ensuite envoyée à un moteur de requête classique. Enfin, dans la troisième étape, le degré de vérité des résultats renvoyés par le moteur sont calculés. Par exemple, PostgreSQLf (Smits et al., 2012) a été développé dans cette optique en tant qu'extension du moteur de requêtes PostgreSQL.

La troisième problématique concerne l'ergonomie. Les requêtes flexibles autorisent une expressivité accrue pour l'utilisateur, mais celle-ci doit être formulable simplement et intuitivement. Trois approches existent pour permettre à l'utilisateur de définir les prédicats flous. La première approche consiste à lui soumettre un ensemble de prédicats prédéfinis, parmi lesquels il peut choisir. Cette approche est illustrée dans la partie droite de la figure 6.1, qui illustre l'interface ReqFlex (Smits et al., 2013) : l'utilisateur peut sélectionner un prédicat parmi ceux qui sont prédéfinis par le concepteur ou qui ont été enregistrés par lui-même ou d'autres utilisateurs. Les prédicats peuvent également être construits dans une approche guidée par les données pour veiller à la cohérence entre les souhaits de l'utilisateur et le contenu de la base : les sous-ensembles flous sont construis dynamiquement à partir des distributions des données dans la base, l'utilisateur attribuant dans un second temps des étiquettes linguistiques décrivant les prédicats (Smits et al., 2014).

La seconde approche pour définir un prédicat flou consiste à donner une liberté absolue à l'utilisateur : il peut choisir aussi bien la forme générale du sous-ensemble flou, triangulaire, trapézoïdale ou gaussienne par exemple, que ses paramètres, support, noyau, etc. Cette approche présente toutefois l'inconvénient d'être peu intuitive pour un non-expert. En effet, elle implique un minimum de connaissance en logique floue que le tout-venant ne possède généralement pas.

La troisième approche de définition d'un prédicat flou est une hybridation entre les deux premières : le sous-ensemble flou représentant le prédicat est soumis à certaines contraintes, restreignant ainsi les possibilités de l'utilisateur. Celui-ci peut dès lors se focaliser sur les paramètres importants de la définition. Par exemple, comme illustré dans la partie gauche de la figure 6.1, dans ReqFlex, les sous-ensembles flous représentant les prédicats numériques sont tous de forme trapézoïdale. L'utilisateur n'a qu'à définir les valeurs qui lui semblent acceptables et celles qui lui semblent idéales, correspondant respectivement au support et au noyau du sous-ensemble flou trapézoïdal, qui interpole linéairement entre les bornes de ces derniers.

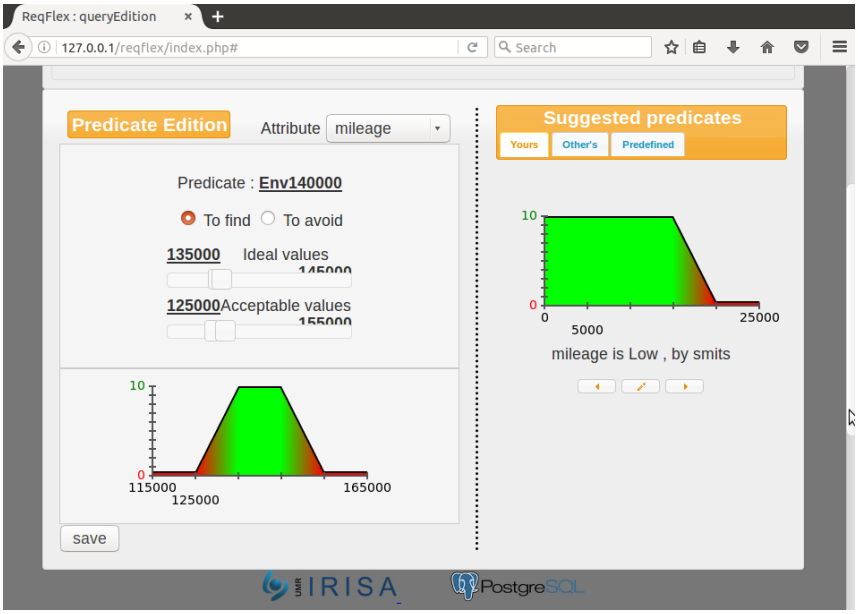


FIGURE 6.1 – Interface originale de ReqFlex (Smits et al., 2013). La partie gauche permet de définir le sous-ensemble flou correspondant à un prédicat (*environ 140 000km*) d'un attribut numérique, ici le kilométrage (*mileage*) de véhicules d'occasion. Le noyau (*Ideal values*) est compris dans l'intervalle $[135000, 145000]$ et le support (*Acceptable values*) dans l'intervalle $[125000, 155000]$. La partie droite de l'interface fournit à l'utilisateur un ensemble de prédicats prédéfinis ou précédemment sauvegardés.

6.1.2 Mise en œuvre du modèle flou fRKM proposé

La mise en œuvre du modèle flou d'interprétation des ENA, fRKM, que nous avons décrit au chapitre précédent s'inscrit dans la problématique de la définition des prédicats flous. Nous proposons définir un nouveau type de prédicat prenant la forme "*environ* x" : partant d'une valeur de référence x, entrée par l'utilisateur, le modèle renvoie le support, le noyau et la 0,5-coupe du sous-ensemble flou représentant le prédicat. Le support correspond aux valeurs acceptables, le noyau aux valeurs idéales et la 0,5-coupe peut être interprétée comme les valeurs intermédiaires entre acceptables et idéales. La validation expérimentale du modèle fRKM détaillée dans le chapitre précédent (voir section 5.3.3 p. 83) permet de s'assurer que les bornes générées par le modèle sont pertinentes d'un point de vue cognitif pour l'utilisateur.

La définition d'un nouveau type de prédicat "*environ* x" est utile pour éviter la définition des sous-ensembles flous correspondant aux prédicats, qui peut rapidement devenir fastidieuse pour l'utilisateur. En effet, dans l'exemple de la base de données de salariés, il peut à un moment donné rechercher les employés dont le salaire est "*environ 2000 euros*" puis ceux dont le salaire est "*environ 3000 euros*", puis ceux dont le salaire est "*environ 1500*

euros", etc. Pour chaque requête, il doit définir le sous-ensemble flou correspondant, ce qui résulte en une perte de temps non négligeable.

6.1.3 Implémentation de fRKM dans ReqFlex

Nous remercions chaleureusement Gregory Smits et Olivier Pivert pour nous avoir fourni le code source de ReqFlex et avec qui nous avons eu l'occasion de collaborer en vue de cette implémentation. Nous avons encadré deux étudiants, Anita Caron et Hassane Belkacem, pour un projet logiciel de Master 1 Données, Apprentissage et Connaissances (DAC), qui ont fourni les bases de cette implémentation.

ReqFlex (Smits et al., 2013) est une interface graphique permettant de définir et d'exécuter des requêtes flexibles de manière intuitive et naturelle pour l'utilisateur. Cette interface, qui permet de générer automatiquement des requêtes SQLf (Bosc & Pivert, 1995), est conçue comme une surcouche du moteur de requêtes étendu PostgreSQLf (Smits et al., 2012).

Notre contribution au développement de ReqFlex a été réalisée selon deux axes. Le premier a consisté à implémenter un nouveau type d'éditeur de prédicat de type "*environ x*", où un seul champ est à remplir par l'utilisateur, correspondant à la valeur de référence de l'ENA. Le second axe repose sur l'implémentation du modèle fRKM décrit dans le chapitre précédent, qui renvoie le sous-ensemble flou représentant le prédicat. Celui-ci est exploité de deux manières : d'une part, sa visualisation dans l'interface et, d'autre part, son utilisation pour l'interrogation de la base de données.

La figure 6.2 illustre l'interface de l'implémentation que nous avons réalisée, dans le cas d'une base de données de voitures d'occasions. La requête considérée est la recherche de véhicules dont le kilométrage est d'"*environ 140 000km*". On peut observer que le sous-ensemble flou représentant le prédicat "*environ 140 000km*" n'est plus trapézoïdal comme dans la figure 6.1. En effet, dans le cas d'attributs numériques, l'utilisateur donne le support (valeurs acceptables) et le noyau (valeurs idéales) pour construire un sous-ensemble flou de forme trapézoïdale. Le modèle fRKM spécifie davantage le sous-ensemble flou, en considérant la 0,5-coupe comme un troisième paramètre de celui-ci. On voit que l'utilisateur se contente d'entrer la valeur de référence de l'ENA et n'a plus à définir lui-même le support et le noyau du sous-ensemble flou. On peut également observer que celui-ci, contrairement à ce qui aurait pu être intuitivement réalisé par l'utilisateur, n'est pas symétrique.

6.2 Interprétation des ENA à nombres décimaux

Cette section présente et illustre une extension des modèles proposés dans le chapitre précédent pour l'interprétation d'ENA impliquant des nombres décimaux, i.e., avec un

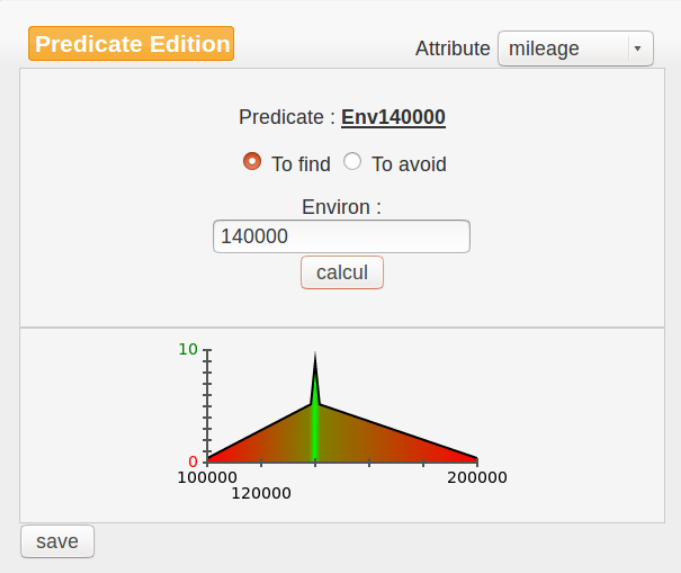


FIGURE 6.2 – Fenêtre d'édition des prédicats du type "*environ x*", que nous avons développée, illustrée dans le cas de "*environ 140 000km*".

nombre fini de chiffres non nuls (par ex. "*environ 3,14*"), au-delà des entiers considérés précédemment.

Elle montre ensuite comment employer cette extension pour s'affranchir de la contrainte d'intervalles à bornes entières pour des ENA dont la valeur de référence est entière et de granularité 1 (par ex., "*environ 16*").

6.2.1 Principe

Nous considérons ici des nombres dits décimaux, c'est-à-dire avec un nombre fini de chiffres non nuls après la virgule. Formellement, il s'écrivent $x = \sum_{i=r}^{s} a_i \cdot 10^i$, où $a_i \in [\![0, 9]\!]$, s est la puissance de 10 du premier chiffre significatif, formellement $s = \max\{i | a_i \neq 0\}$ et r est la puissance de 10 du dernier chiffre significatif, formellement $r = \min\{i | a_i \neq 0\}$. Dans le cas des nombres décimaux, r et s peuvent être négatifs, là où les entiers considérés dans les chapitres précédents correspondent au cas où r et $s \geq 0$.

On peut remarquer que cette méthode peut être appliquée aux nombres réels, si leur valeur de référence est tronquée : pour une valeur $x \in \mathbb{R}$, r peut être égal à moins l'infini, la troncature permet de le rendre fini.

Cette section considère la tâche d'interprétation d'ENA contenant de tels nombres, par exemple "*environ 3,5*".

La méthode que nous proposons pour tenir compte des nombres décimaux ne nécessite pas de généralisation de la notion de complexité et compte trois étapes. La première consiste à transformer la valeur de référence, en la multipliant par une puissance de dix, pour

obtenir un entier. Cet entier est dans un deuxième temps traité par les modèles présentés au chapitre précédent. Enfin, la dernière étape consiste à appliquer la transformée inverse aux estimations renvoyées par les modèles : les valeurs des bornes des intervalles (dans le cas de LLM et RKM) et des α-coupes (dans le cas de fRKM) sont divisées par la puissance de dix utilisée dans la première étape.

Formellement, la valeur de référence décimale, notée x_D, est multipliée par une puissance de 10 pour obtenir valeur de référence entière, notée x_N si $r \leq 0$:

$$x_N = \begin{cases} x_D \cdot 10^{-r+\omega} & \text{si } r \leq 0 \\ x_D & \text{sinon} \end{cases} \tag{6.1}$$

où $\omega \in \mathbb{N}^* \geq 1$ est un paramètre défini par l'utilisateur. Ce paramètre permet de contrôler la richesse d'interprétation. En effet, plus ω est important, plus la puissance de 10 correspondant au dernier chiffre significatif de x_N est élevée et plus il y a de candidats pour être les bornes de l'intervalle dénoté par "*environ* x_D". Par exemple, pour "*environ 0,5*", avec $\omega = 1$, $x_N = 50$. Si $\omega = 2$, $x_N = 500$.

La seconde étape consiste à interpréter x_N selon l'un des modèles proposés dans le chapitre précédent. L'ensemble de valeurs renvoyées par l'interprétation est noté $V_N(x_N) = \{v_{N1}, \ldots, v_{Nn}\}$. V_N peut d'abord être égal au fronts de Pareto formalisant le compromis entre saillance cognitive et distance à la valeur de référence : $V_N(x_N) = P(x_N)$. Dans le cas d'une interprétation par intervalles, par LLM ou RKM (voir section 5.2 p. 66), $n = 2$ et les valeurs correspondent aux bornes de l'intervalle estimé. Dans le cas d'une interprétation floue par fRKM (voir section 5.3 p. 80), $n = 6$ et les valeurs correspondent aux bornes du support, de la $0,5$-coupe et du noyau de l'intervalle flou représentant l'ENA.

La troisième étape, enfin, consiste à appliquer la transformée inverse de la première étape aux valeurs entières contenues dans V_N pour obtenir les résultats attendus en valeurs décimales V_D, formellement :

$$v_{Di} = \begin{cases} v_{Ni} \cdot 10^{r-\omega} & \text{si } r \leq 0 \\ v_{Ni} & \text{sinon} \end{cases} \tag{6.2}$$

6.2.2 Résultats illustratifs

Pour illustrer la méthode d'interprétation des ENA impliquant des nombres décimaux, nous proposons de considérer trois exemples : "*environ 0,068*", "*environ 0,4*" et "*environ 3,14*". Pour chacun des exemples, ω est fixé à 1.

La figure 6.3 illustre les valeurs candidates identifiées dans les fronts de Pareto, ainsi que les fonctions d'appartenance générées par le modèle fRKM pour ces ENA. Le tableau 6.1 présente les intervalles estimés par le modèle des rangs RKM après application de la méthode décrite dans la section précédente.

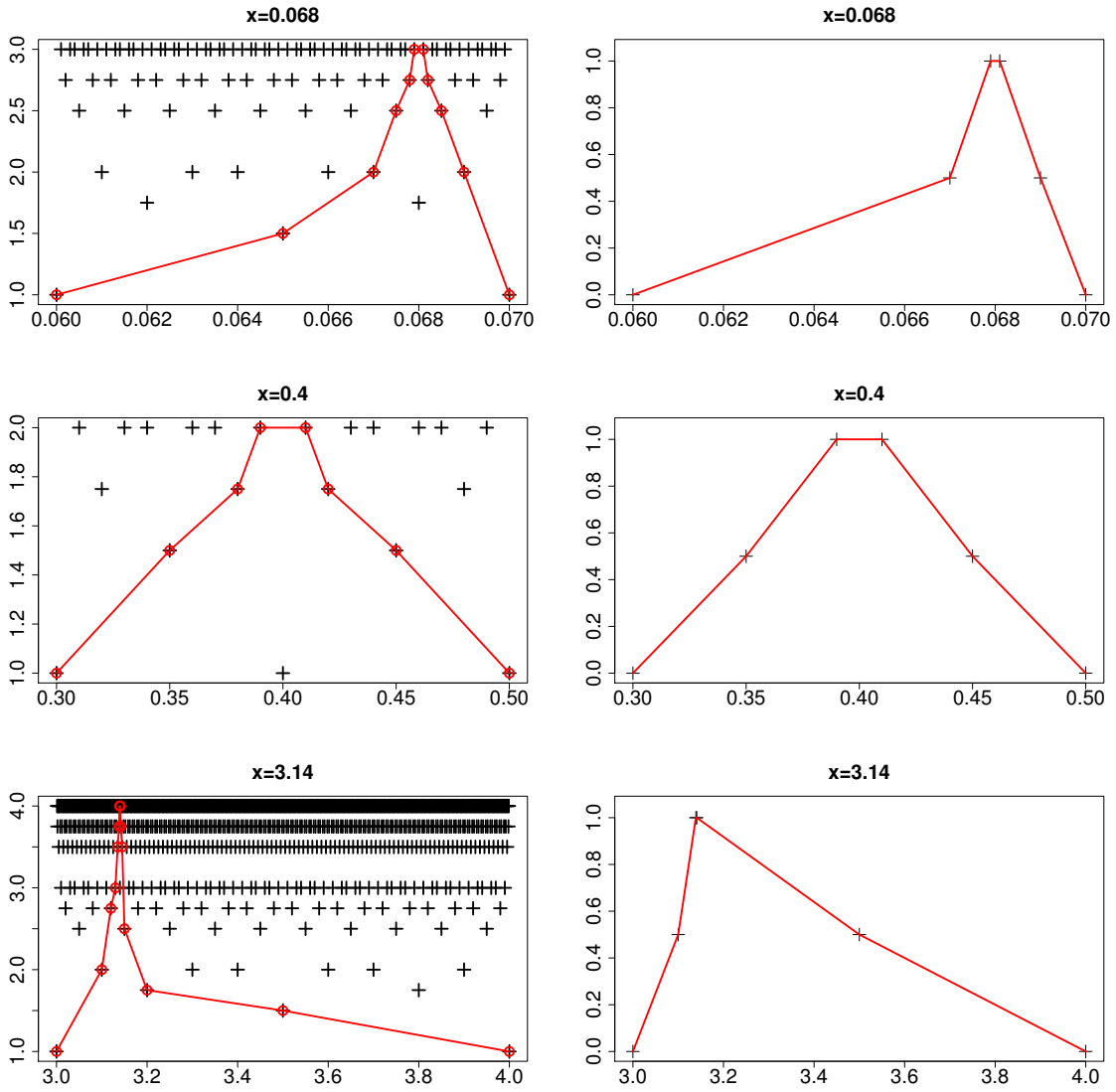


FIGURE 6.3 – Interprétation d'ENA impliquant des nombres décimaux : "*environ 0,068*" (haut), "*environ 0,4*" (centre) et "*environ 3,14*" (bas). A gauche : fronts de Pareto. A droite : fonctions d'appartenance générées par le modèle d'interprétation floue des ENA fRKM.

On peut observer que les résultats obtenus correspondent à ce qui peut être intuitivement attendu pour chacun des trois exemples. Toutefois, une étude empirique ainsi qu'une comparaison expérimentale avec les modèles existants seraient à conduire pour examiner systématiquement la plausibilité cognitive des résultats renvoyés par cette méthode.

ENA	Intervalle estimé
environ 0,068	$[0,067, 0,069]$
environ 0,4	$[0,35, 0,45]$
environ 3,14	$[3,1, 3,5]$

TABLEAU 6.1 – Intervalles estimés par le modèle des rangs RKM (voir section 5.2.2 p. 68), correspondant à trois ENA impliquant des nombres décimaux.

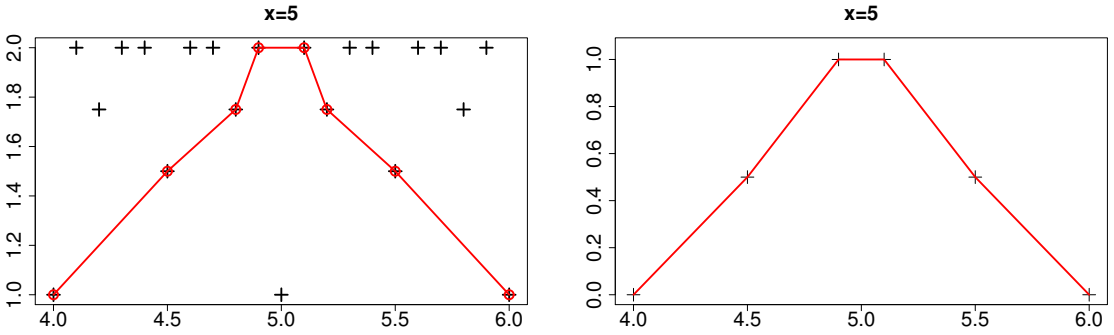


FIGURE 6.4 – Interprétation d'ENA impliquant un entier dont la granularité est 1 : "*environ 5*". A gauche : fronts de Pareto. A droite : fonction d'appartenance générée par le modèle d'interprétation floue des ENA fRKM.

6.2.3 Interprétation décimale d'ENA entières

La méthode que nous avons proposée pour interpréter des ENA impliquant des nombres décimaux permet d'obtenir des valeurs de bornes et d'α-coupes qui ne sont pas entières. Elle peut également être mise en œuvre pour affiner l'interprétation d'ENA dont la valeur de référence est entière et de granularité 1 (par ex., "*environ 5*"). En effet, les modèles décrits dans le chapitre précédent ne considèrent que les entiers naturels comme bornes d'intervalles et d'α-coupes. Dans ce cadre, l'interprétation d'ENA dont la granularité est 1 peut donner lieu à des résultats inattendus. Par exemple, dans le cas de "*environ 5*", seules deux valeurs sont incluses dans les fronts de Pareto : 4 et 6. Par conséquent, les modèles ne peuvent renvoyer qu'un seul intervalle, $[4, 6]$. Il semblerait donc raisonnable de considérer également les nombres décimaux comme valeurs possibles de bornes, par exemple $[4, 5, 5, 5]$ pour "*environ 5*".

Notons toutefois que le passage au système décimal peut ne pas être pertinent dans certains contextes, en particulier quand les valeurs de référence représentent un dénombrement. En effet, il nous parait absurde de considérer comme pertinent, par exemple, l'intervalle $[14, 5, 15, 5]$ pour l'expression "*environ 15 moutons*".

La figure 6.4 illustre l'application de la méthode d'interprétation décimale au cas de l'ENA "*environ 5*". Comme attendu, on peut observer que, au-delà de 4 et 6, des valeurs décimales sont incluses dans les fronts de Pareto (par ex., 4,8 et 5,5) et peuvent servir de bornes pour le support, la $0,5$-coupe et le noyau de la fonction d'appartenance générée par fRKM. L'intervalle estimé par le modèle RKM est également défini par des bornes décimales : $[4, 5, 5, 5]$.

6.3 Bilan

Dans ce chapitre, nous avons présenté la mise en œuvre concrète du modèle flou d'interprétation des ENA que nous avons élaboré au chapitre précédent ainsi que deux extensions des modèles d'interprétation des ENA, également présentés au chapitre précédent.

Nous avons implémenté le modèle flou d'interprétation des ENA dans le cadre d'un système de requêtes flexibles, ReqFlex (Smits et al., 2013). Nous avons ainsi montré l'intérêt concret du modèle que nous proposons en ce qu'il permet de se passer de l'étape de définition des prédicats flous, qui peut être fastidieuse pour l'utilisateur.

Nous avons ensuite proposé une extension des modèles précédemment décrits de manière à pouvoir interpréter des ENA dont la valeur de référence est un nombre décimal. Bien que les estimations produites grâce à cette extension paraissent intuitivement cohérentes, une étude empirique ainsi qu'une validation expérimentale vis-à-vis des modèles existants sont nécessaires pour en valider la plausibilité du point de vue cognitif.

Chapitre 7

Interprétation des ENA : contextualisation et interprétation implicite

Le cadre considéré dans les chapitres précédents porte sur les Expressions Numériques Approximatives non contextualisées, à la fois pour l'identification des dimensions pertinentes et la construction de modèles computationnels. Ce chapitre présente une seconde étude empirique que nous avons conduite pour compléter la première, selon deux axes. Le premier porte sur l'effet de la présence d'un contexte sémantique sur l'interprétation des ENA ; le second vise à examiner si l'interprétation implicite d'une ENA, pour laquelle nous proposons une tâche originale aux participants de l'étude, diffère de son interprétation explicite.

Nous détaillons dans la section 7.1 la problématique et les hypothèses motivant cette étude. Les méthodes mises en œuvre pour collecter, nettoyer, pré-traiter et analyser les données en vue de tester les hypothèses sont décrites dans la section 7.2. Les résultats des analyses sont présentés dans la section 7.3 et discutés dans la section 7.4.

7.1 Problématique et hypothèses

Dans cette section, nous présentons en premier lieu la problématique motivant la conduite de la seconde étude de l'interprétation des ENA. Les hypothèses que nous considérons pour traiter cette problématique sont décrites dans la seconde sous-section.

7.1.1 Problématique

L'étude empirique que nous présentons dans ce chapitre est motivée par deux objectifs originaux par rapport aux travaux présentés dans les chapitres précédents, et en particulier l'étude décrite au chapitre 4 : (i) examiner l'effet de la présence d'un contexte sémantique sur l'interprétation des ENA ; (ii) comparer l'interprétation explicite, telle que mise en œuvre dans la première étude, avec une interprétation implicite des ENA.

Le premier objectif porte sur la contextualisation des ENA. En effet, dans la première étude empirique, les ENA considérées, de la forme "*environ x*" ne sont pas contextualisées. Or, il se peut que la présence d'un contexte sémantique influence la manière dont l'être humain interprète une ENA (Lasersohn, 1999; Sadock, 1977; Williamson, 1994). Au delà de l'implication des dimensions arithmétiques, qui a été examinée dans le chapitre 4, nous nous intéressons dans ce chapitre à l'influence du contexte sémantique sur l'interprétation d'une ENA : à titre illustratif, il s'agit d'examiner si des différences peuvent exister, par exemple, entre "*environ 30km*" et "*environ 30 euros*".

Le second objectif porte sur la nature de l'interprétation elle-même. Dans la première étude, il a été demandé aux participants d'interpréter explicitement les ENA. Cette situation est particulière dans le sens où, dans une conversation normale, les locuteurs ne se forment que rarement une représentation consciente de la plage de valeurs que dénote une ENA (Cleeremans, 1997; Dienes & Perner, 1999). Il se peut donc que l'interprétation implicite d'une ENA diffère de son interprétation explicite.

Dans ce chapitre, nous considérons qu'une ENA est représentée explicitement lorsque les bornes de la plage de valeurs qu'elle dénote font l'objet d'une représentation explicite, qui peut être l'aboutissement d'un raisonnement. Autrement dit, nous considérons qu'une ENA est explicitement interprétée lorsque la décision quant à l'appartenance d'une valeur à l'imprécision d'une ENA est prise en fonction d'une borne clairement réfléchie et représentée. Par opposition, nous considérons qu'une ENA est représentée implicitement lorsque les bornes de la plage de valeurs qu'elle dénote ne font pas l'objet d'une représentation explicite, c'est-à-dire que la décision quant à l'appartenance d'une valeur à l'imprécision d'une ENA n'est prise qu'en fonction d'une intuition par rapport à la valeur de référence de l'ENA.

De plus, les données collectées lors de l'étude présentée dans ce chapitre nous permettent (i) d'examiner si les résultats obtenus dans la première étude sont reproductibles et (ii) de valider le principe général du compromis entre saillance cognitive et plage de valeurs dénotées ainsi que le modèle des rangs RKM [4] d'interprétation que nous avons proposés

4. Dans cette étude, nous ne considérons pas le modèle d'interprétation log-linéaire LLM et le modèle flou d'interprétation des ENA car la quantité de données collectées est trop faible pour que la régression linéaire de LLM et les fonctions d'appartenance élicitées à partir des données, auxquelles sont comparées

au chapitre 5 en nous basant sur un jeu de données indépendant.

7.1.2 Hypothèses

Pour traiter la problématique motivant la seconde étude de l'interprétation des ENA, cinq hypothèses sont formulées et détaillées, avec leurs hypothèses dérivées, ci-dessous. Formellement, nous notons l'intervalle correspondant à l'ENA "*environ* x" $I(x) = [x - \Delta^-(x), x + \Delta^+(x)]$.

H1 - L'interprétation des ENA serait dépendante du contexte dans lequel elle a lieu (Lasersohn, 1999; Sadock, 1977; Williamson, 1994). Nous nous attendons donc à ce que les intervalles correspondant à une ENA non contextualisée diffèrent des intervalles correspondant à une ENA contextualisée pour une même valeur de référence.

H2 - Demander aux participants de fournir explicitement les bornes de la plage de valeurs dénotées par une ENA s'inscrit dans la perspective épistémique du vague et fixe le principe de tolérance. En effet, dans ce cas, il est difficile d'examiner dans quelle mesure des valeurs proches des bornes fournies seraient également acceptées.

Au contraire, collecter implicitement les bornes de la plage de valeurs dénotées par une ENA, en proposant plusieurs valeurs possiblement dénotées au participant, permet de ne pas figer le processus à l'origine du principe de tolérance. Davantage de valeurs peuvent alors être considérées comme bornes de l'intervalle et rien ne permet d'affirmer qu'elles seront celles qu'aurait explicitement données le participant. Par conséquent, nous nous attendons à ce que les intervalles collectés explicitement diffèrent des intervalles collectés implicitement pour une même ENA.

H3 - La première étude de l'interprétation des ENA suggère que la magnitude, la granularité et le dernier chiffre significatif sont impliqués dans l'interprétation explicite des ENA. On devrait donc observer l'implication individuelle de chacune de ces trois dimensions aussi bien lors d'une interprétation explicite que lors d'une interprétation implicite. Nous proposons par conséquent de considérer trois hypothèses similaires aux hypothèses H1.1 à H1.3 de la première étude (voir section 4.2 p. 41) reproduites ici :

H3.1 - Magnitude : Nous prédisons qu'à granularité et dernier chiffre significatif constants, deux ENA dont la magnitude diffère sont interprétées de manière différente, aussi bien explicitement qu'implicitement, formellement :

$$\text{si } x > y \text{ et } Gran(x) = Gran(y) \text{ et } LSD(x) = LSD(y), \text{ alors } \Delta(x) \neq \Delta(y).$$

celles générées par le modèle flou d'interprétation, soient robustes.

H3.2 - Granularité : D'une manière similaire, nous prédisons qu'à dernier chiffre significatif constant, deux ENA dont la granularité diffère sont interprétées de façon différente, aussi bien explicitement qu'implicitement, formellement :

$$\text{si } Gran(x) > Gran(y) \text{ et } LSD(x) = LSD(y), \text{ alors } \Delta(x) \neq \Delta(y).$$

H3.3 - Dernier chiffre significatif : Enfin, nous prédisons qu'à granularité constante, deux ENA dont le dernier chiffre significatif diffère sont interprétées de manière différente, aussi bien explicitement qu'implicitement, formellement :

$$\text{si } LSD(x) > LSD(y) \text{ et } Gran(x) = Gran(y), \text{ alors } \Delta(x) > \Delta(y).$$

H4 - La première étude de l'interprétation des ENA a mis en évidence le fait que l'hypothèse de symétrie des intervalles n'est pas observée pour toutes les ENA. Nous prédisons donc que les intervalles collectés implicitement et explicitement ne sont pas toujours symétriques. Formellement, $\Delta^{-}(x) \neq \Delta^{+}(x)$.

H5 - Dans le chapitre 5, nous avons validé le principe général ainsi que les modèles d'interprétation à intervalles des ENA que nous proposons en se basant sur les données collectées lors de la première étude empirique de l'interprétation des ENA (voir section 5.2 p. 66). Nous nous attendons donc à ce que d'une part les intervalles collectés lors de la présente étude soient capturés par le principe général des fronts de Pareto et, d'autre part, que le modèle des rangs RKM que nous avons proposé (voir section 5.2.2 p. 68) prédise correctement les médianes des bornes collectées.

7.2 Méthodes

Nous décrivons dans cette section la méthodologie que nous avons mise en œuvre pour collecter, pré-traiter, nettoyer et analyser les données nous permettant de tester les cinq hypothèses considérées dans cette étude.

7.2.1 Population

37 adultes se sont portés volontaires pour cette étude, 18 femmes et 19 hommes, dont l'âge est compris entre 18 et 37 ans ($M = 21,0\,;\, \sigma = 4,11$). Tous ont pour langue maternelle le français et ont été recrutés sur le campus de l'Université d'Angers.

ENA	x	$Gran(x)$	$LSD(x)$	$RelMag(x)$	$NSD(x)$
environ 30	30	10	3	30	1
environ 70	70	10	7	70	1
environ 100	100	100	1	100	1
environ 110	110	10	1	10	2
environ 150	150	10	5	50	2
environ 500	500	100	5	500	1
environ 800	800	100	8	800	1
environ 1000	1000	1000	1	1000	1
environ 1100	1100	100	1	100	2
environ 1500	1500	100	5	500	2
environ 4700	4700	100	7	700	2
environ 4730	4730	10	3	30	3
environ 8000	8000	1000	8	8000	1
environ 8150	8150	10	5	50	3
environ 30 euros en poche	30	10	3	30	1
hôtel à environ 30km	30	10	3	30	1
environ 500km séparent deux villes	500	100	5	500	1
terrain d'environ 800 hectares	800	100	8	800	1
ville d'environ 1000 habitants	1000	1000	1	1000	1

TABLEAU 7.1 – ENA utilisées dans la seconde étude empirique de l'interprétation des ENA et leurs dimensions arithmétiques : magnitude (x), granularité ($Gran$), dernier chiffre significatif (LSD), magnitude relative ($RelMag$) et nombre de chiffres significatifs (NSD), telles que définies dans la section 4.1 p. 38.

7.2.2 Matériel

19 ENA, listées dans le tableau 7.1, sont considérées dans cette étude. 14 d'entre elles ne sont pas sémantiquement contextualisées. Leur valeur de référence est choisie pour répondre à l'hypothèse H3, portant sur l'implication des dimensions arithmétiques des ENA dans leur interprétation. Les 5 autres ENA sont sémantiquement contextualisées pour tester l'hypothèse H1 portant sur un effet du contexte. A chacune d'entre elles correspond une ENA de même valeur de référence mais non contextualisée.

Outre les items portant sur les 19 ENA, le questionnaire comporte également une question visant à contrôler la variabilité inter-individuelle quant à l'usage quotidien des mathématiques. La consigne correspondante est : "*Manipulez-vous très régulièrement des*

nombres dans le cadre de votre travail ou de vos études ?". Le participant doit répondre à cette question par *oui* ou par *non*.

7.2.3 Procédure

Trois tâches sont proposées l'une après l'autre à chaque participant : (i) une tâche d'entraînement pour le familiariser avec le matériel informatique, (ii) une tâche informatisée de collecte implicite des intervalles correspondant aux 19 ENA considérées et enfin (iii) une tâche de collecte explicite des intervalles correspondant aux mêmes 19 ENA.

Alors que dans la première étude, les deux bornes d'un intervalle collecté sont demandées comme une seule réponse (voir section 4.3.3 p. 46), dans les deux tâches de collecte de cette étude, les bornes supérieure et inférieure des intervalles sont données par un même participant de manière séparée et non consécutive. Autrement dit, la collecte d'un intervalle repose sur deux items qui ne se suivent pas dans le questionnaire et deux essais non consécutifs dans la tâche de collecte implicite.

L'ensemble des épreuves dure environ 30 minutes. Les prochaines sous-sections décrivent tour à tour les trois tâches exécutées par les participants.

7.2.3.1 Tâche d'entraînement

La première tâche exécutée par un participant a pour but de le familiariser avec le matériel informatique, les touches du clavier à utiliser ainsi que de l'entraîner à répondre le plus rapidement possible en vue de la tâche de collecte implicite des intervalles. La tâche consiste à juger de la parité de nombres affichés à l'écran. Ces nombres sont différents des valeurs de référence des ENA considérées dans cette étude afin de ne pas induire d'interférence avec les tâches de collecte d'intervalles.

En début de tâche, le participant voit apparaître la consigne suivante : " *Vous allez voir apparaître des nombres. Pour chacun d'entre eux, vous devez dire s'il est pair ou non : Appuyez sur la touche S s'il est pair. Appuyez sur la touche L sinon. Essayez de répondre le plus rapidement et le plus correctement possible. Appuyez sur une touche lorsque vous êtes prêt...*" Le participant appuie sur une touche quelconque du clavier lorsqu'il se sent prêt et la tâche proprement dite commence.

35 nombres sont présentés au participant, dans un ordre aléatoire. Pour chacun d'entre eux, la procédure, illustrée par la figure 7.1, est la suivante. D'abord l'écran reste noir durant 500 millisecondes. Ensuite un masque apparaît pendant 500 millisecondes. Enfin, le nombre apparaît au centre de l'écran et le participant doit juger le plus rapidement possible s'il est pair ou non. S'il est pair, il doit appuyer sur la touche S. S'il est impair, il doit appuyer sur la touche L.

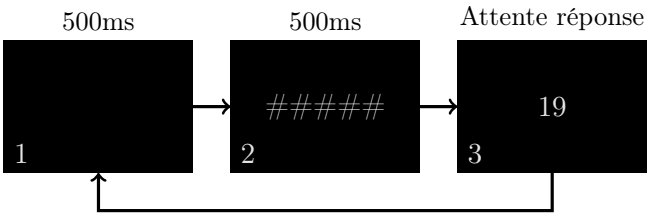


FIGURE 7.1 – Illustration de la tâche d'entraînement. Pour chacun des nombres pour lesquels le participant doit réaliser un jugement de parité : 1. l'écran reste noir pendant 500ms, 2. un masque apparaît durant 500ms. 3. le nombre proposé au participant est affiché jusqu'à l'appui de la touche correspondant à la réponse : *S* pour *pair* et *L* pour *impair*.

7.2.3.2 Tâche de collecte implicite des intervalles

Pour collecter implicitement les intervalles, nous proposons une tâche originale, dérivée du modèle de la recherche par partition binaire modifiée (Tyrrell & Owens, 1988) développée dans le cadre d'études psychophysiques pour déterminer les seuils d'intensité de détection de stimuli.

Le principe général permettant d'évaluer la valeur d'une borne de l'intervalle correspondant à une ENA est le suivant. Plusieurs valeurs sont proposées au participant et celui-ci doit, pour chacune d'entre elles, dire si elle correspond à l'ENA considérée. Les valeurs proposées sont générées en temps réel et dépendent des réponses précédentes du participant. Ces réponses permettent d'obtenir un intervalle dans lequel se trouve la valeur de la borne recherchée. La valeur exacte n'étant pas connue, une estimation de cette valeur de borne est réalisée, en prenant le centre de l'intervalle renvoyé par l'algorithme.

Nous décrivons dans un premier temps la tâche telle qu'elle est présentée au participant. Ensuite, nous décrivons la méthode utilisée pour générer en temps réel les valeurs proposées au participant durant l'exécution de la tâche.

Description de la tâche En début de tâche, le participant voit apparaître la consigne suivante : "*Vous allez voir apparaître une phrase contenant une expression numérique approximative suivie de plusieurs nombres. Pour chaque nombre, vous devrez dire si oui ou non il peut être approximé par l'expression numérique contenue dans la phrase. Appuyez sur S pour oui. Appuyez sur L pour non. Essayez de répondre le plus rapidement possible, il n'y a pas de bonne ou de mauvaise réponse. Prêt(e) ?*" Le participant peut alors demander des précisions à l'expérimentateur. Lorsqu'il est prêt, le participant appuie sur une touche quelconque et la tâche proprement dite commence.

Chaque borne de chacune des 19 ENA est collectée séparément et l'ordre de leur présentation est tiré aléatoirement pour chaque participant. Par conséquent, la procédure

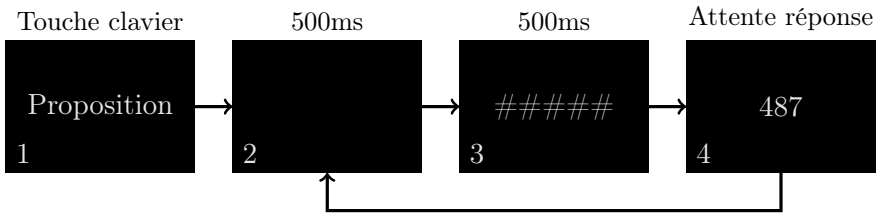


FIGURE 7.2 – Illustration de la tâche de collecte implicite des intervalles. (1) Une proposition contenant l'ENA considérée apparaît d'abord à l'écran. Une fois que le participant a appuyé sur une touche quelconque, l'algorithme de recherche de la valeur de la borne produit des nombres. (2) Pour chacun de ces nombres, l'écran reste noir pendant 500ms. (3) Un masque apparaît ensuite durant 500ms. (4) Le nombre proposé au participant est affiché jusqu'à l'appui de la touche correspondant à la réponse : *S* pour *oui* et *L* pour *non*.

suivante, illustrée par la figure 7.2 est répétée 38 fois pour chaque participant.

Lorsque le participant est prêt à commencer, une proposition contenant l'une des ENA est présentée à l'écran : "*Les nombres suivants vous paraissent-ils correspondre à environ x ?*", où x est la valeur de référence, lorsque l'ENA n'est pas contextualisée. Lorsque celle-ci est contextualisée, la proposition contenant l'ENA qui apparaît à l'écran prend la forme des propositions présentée dans le tableau 7.2. Le participant appuie ensuite sur une touche quelconque du clavier pour lancer la série de nombres proposés pour l'ENA considérée.

Avant la présentation de chaque nombre, l'écran reste noir durant 500 millisecondes puis un masque apparaît (voir figure 7.2) pendant 500 millisecondes. Le nombre proposé est ensuite affiché. Le participant doit alors répondre le plus rapidement possible si oui, en appuyant sur la touche S du clavier, ou non, en appuyant sur la touche L du clavier, le nombre correspond à l'ENA.

Des nombres sont proposés au participant jusqu'à ce que l'intervalle dans lequel se trouve la valeur de borne recherchée soit suffisamment étroit, comme décrit dans le paragraphe suivant.

Algorithme de génération des valeurs proposées La longueur de la liste des nombres présentés ainsi que leurs valeurs sont générées en temps réel et dépendent des réponses précédentes du participant, selon le principe d'une recherche par partition binaire. Pour ce faire, nous reprenons l'algorithme proposé et validé par Tyrrell et Owens (1988), dont une illustration du déroulement est donnée en annexe D.1 p. 195. Cet algorithme permet de déterminer l'intervalle dans lequel se trouve la valeur de la borne recherchée et comporte six règles. Comme exemple, nous proposons de considérer la recherche de la borne inférieure correspondant à "*environ 400*".

ENA	Proposition
environ 30 euros en poche	"*Il vous reste environ 30 euros en poche. Les sommes suivantes vous paraissent-elles rendre cette proposition vraie ?*"
hôtel à environ 30km	"*Vous cherchez un hôtel à environ 30km de Paris. Les distances suivantes vous paraissent-elles satisfaisantes comme réponses à cette recherche ?*"
environ 500km séparent deux villes	"*Environ 500km séparent deux villes. Les distances suivantes vous paraissent-elles rendre cette proposition vraie ?*"
terrain d'environ 800 hectares	"*Vous recherchez la liste des terrains d'environ 800 hectares. Les surfaces suivantes vous paraissent-elles satisfaisantes comme réponses à cette recherche ?*"
ville d'environ 1000 habitants	"*Vous recherchez la liste des villes d'environ 1000 habitants. Les populations suivantes vous paraissent-elles satisfaisantes comme réponses à cette recherche ?*"

TABLEAU 7.2 – Consignes affichées à l'écran pour collecter implicitement les intervalles correspondant aux 5 ENA contextualisées.

1. La plage de valeurs possibles pour la cible, c'est-à-dire la borne recherchée, est définie par deux bornes. A chacune des deux bornes est associée une pile de trois éléments. Le haut de la pile représente la valeur actuelle de la borne alors que les éléments plus bas dans la pile représentent ses précédentes valeurs.

Initialement, tous les éléments d'une pile représentent une extrémité des valeurs possibles alors que tous les éléments de l'autre pile représentent l'autre extrémité. Ces extrémités sont dépendantes du domaine auquel s'applique l'algorithme. Dans le cas des ENA, nous proposons de les représenter par l'intervalle $I_E(x)$ et de les calculer comme [5] :

$$I_E(x) = \left[x - \max\left(5, \frac{3 \cdot Gran(x)}{2}\right), x - 1\right] \tag{7.1}$$

5. Le choix d'un intervalle initial dont la taille est $3 \cdot Gran(x)/2$ a été fait indépendamment des résultats de la première étude empirique de l'interprétation. En effet, d'un point de vue chronologique, l'étude que nous présentons dans ce chapitre a été menée avant celle présentée au chapitre 4. Comme détaillé dans la présentation et la discussion des résultats (voir sections 7.3 et 7.4), cette taille s'est a posteriori révélée problématique pour deux ENA, "*environ 4730*" et "*environ 8150*", puisque les résultats de l'étude présentée au chapitre 4 montrent que les participants auraient pu avoir tendance à accepter des bornes d'intervalles au-delà de celles qui sont proposées. Ce choix n'affecte par conséquent qu'une partie très restreinte des résultats obtenus, ceux portant sur les intervalles implicites collectés pour ces deux ENA. Une nouvelle étude serait à conduire pour corriger ce biais vis-à-vis de ces ENA.

dans le cas d'une recherche de borne inférieure et comme :

$$I_E(x) = \left[x + 1, x + \max\left(5, \frac{3 \cdot Gran(x)}{2}\right)\right] \tag{7.2}$$

dans le cas d'une recherche de borne supérieure. Par exemple, dans le cas d'une recherche de la borne inférieure pour "*environ 400*", $I_E(400) = [250, 399]$. A chaque présentation d'un nombre, l'une des deux piles est mise à jour.

2. Le nombre proposé correspond à la moyenne des valeurs du haut des deux piles. Dans le cas de "*environ 400*", le premier entier proposé est $(250 + 399)/2 = 325$.

3. Lorsque le participant donne une réponse *oui* ou *non*, l'une des piles est mise à jour. Si le participant pense que le nombre ne correspond pas à l'ENA, ce nombre est placé en haut de la pile des valeurs les plus éloignées. Les valeurs déjà présentes dans la pile descendent d'un cran et occupent les deuxième et troisième places de la pile. La mise à jour des piles permet de scinder en deux la plage de recherche de la cible après chaque réponse tout en conservant les valeurs précédentes. Dans l'exemple de "*environ 400*", si le participant considère que 325 n'est pas dénoté par "*environ 400*", l'intervalle de travail est recalculé comme $[325, 399]$. S'il considère que 325 est dénoté, l'intervalle est $[250, 325]$. Dans ce dernier cas, l'entier proposé suivant est $(250 + 325)/2 = 287$.

4. Lorsque deux réponses consécutives sont identiques (i.e., *oui-oui* ou *non-non*), un test est implémenté pour confirmer la validité de la borne opposée. Par exemple, après deux réponses *non*, le nombre présenté au test suivant correspond à la valeur du haut de la pile des valeurs proches. Dans le cas de "*environ 400*", si le participant considère que 325 et 287 sont dénotés, l'entier proposé suivant est 250.

Tester la borne opposée permet de valider que la cible se trouve toujours dans l'intervalle de travail. Si ce n'est pas le cas, la borne opposée est considérée comme invalide et est mise à jour selon la règle 5 :

5. Si la cible se trouve au-delà d'une borne, la valeur de cette borne est mise à jour à sa valeur précédente. Pour ce faire, le haut de la pile correspondante est supprimé et les deux valeurs du dessous sont remontées d'un cran. La valeur la plus en bas de cette pile est réinitialisée selon la règle 1. Cet événement est appelé *régression*. Dans l'exemple de "*environ 400*", si le participant considère que 250 est également dénoté, la borne inférieure de l'intervalle de travail est réinitialisée, aboutissant à $[250, 287]$. Cet exemple est particulier puisque la borne ne change pas car 250 est déjà la valeur initiale. Toutefois, comme illustré dans le déroulement donné en exemple en annexe D.1 p. 195, la règle 5 est pertinente dans le cas où l'intervalle de travail courant diffère de celui qui a été initialisé.

6. Cette procédure se poursuit tant que deux critères d'arrêt ne sont pas remplis. Le premier critère est que deux *revirements* soient observés. Un *revirement* est défini comme deux réponses consécutives différentes (*oui-non* ou *non-oui*). Le second critère est que

la distance entre les deux bornes actuelles soit inférieure ou égale à 1% de la taille de l'intervalle $I_E(x)$ ou bien inférieure ou égale à 1.

Une fois ces deux critères atteints, la cible se calcule comme la moyenne des deux bornes actuelles de l'intervalle de travail.

7.2.3.3 Tâche de collecte explicite des intervalles

La dernière tâche prend la forme d'un questionnaire qui comporte 38 questions, correspondant chacune à une borne de l'une des 19 ENA considérées.

Pour que la collecte des intervalles soit similaire à celle mise en œuvre dans la tâche implicite, les bornes supérieures et inférieures des intervalles correspondant aux ENA sont collectées séparément. On peut noter que cette procédure est différente de celle mise en œuvre dans la première étude (voir section 4.3.3 p. 46), où les participants étaient invités à donner en une même réponse les deux bornes de l'intervalle correspondant à une ENA.

Pour les ENA non contextualisées, la consigne est : "*Selon vous, quelle est la valeur MINIMALE/MAXIMALE correspondant à environ x ?*", où x est la valeur de référence. Le participant doit donner explicitement, dans la case prévue à cet effet, la valeur de la borne, inférieure quand la consigne comporte le terme *MINIMALE*, et supérieure quand la consigne comporte le terme *MAXIMALE*. Le tableau 7.3 présente les consignes relatives aux 5 ENA contextualisées.

Une fois que le participant a répondu aux 38 items portant sur les ENA, il est invité à répondre à une question dont le but est de contrôler la variabilité inter-individuelle quant à l'usage des mathématiques.

7.2.4 Pré-traitement et nettoyage des données

L'intervalle implicite donné par le participant p pour l'ENA "*environ x*" est noté $II_p(x) = [II_p^-(x), II_p^+(x)]$. L'intervalle explicite correspondant est noté quant à lui $IE_p(x) = [IE_p^-(x), IE_p^+(x)]$.

Comme dans la première étude (voir section 4.3.6 p. 47), les intervalles sont utilisés pour calculer deux distances absolues $\Delta I_p^b(x)$ et $\Delta E_p^b(x)$ entre la valeur de référence x de l'ENA et les bornes $II_p^b(x)$ et $IE_p^b(x)$ pour $b \in \{-, +\}$ des intervalles implicites et explicites respectivement : $\Delta I_p^b(x) = |II_p^b(x) - x|$ et $\Delta E_p^b(x) = |IE_p^b(x) - x|$.

Les données sont ensuite nettoyées pour exclure les valeurs aberrantes ou extrêmes. La procédure de nettoyage est celle détaillée dans l'étude du chapitre 4 de l'interprétation des ENA (voir section 4.3.5 p. 46).

Sur les 1444 intervalles collectés, 1336 (92,5%) sont conservés pour les analyses. Un participant a été exclu de l'étude.

ENA	Proposition
environ 30 euros en poche	"*Il vous reste environ 30 euros en poche. Quelle est, selon vous, la somme MINIMALE/MAXIMALE qui rend cette proposition vraie ?*"
hôtel à environ 30km	"*Vous cherchez un hôtel a environ 30km de Paris. Quelle est, selon vous, la distance MINIMALE/MAXIMALE qui vous parait satisfaisante pour cette recherche ?*"
environ 500km séparent deux villes	"*Environ 500 km séparent deux villes. Quelle est, selon vous, la distance MINIMALE/MAXIMALE qui rend cette proposition vraie ?*"
terrain d'environ 800 hectares	"*Vous recherchez la liste des terrains d'environ 800ha. Quelle est, selon vous, la surface MINIMALE/MAXIMALE qui vous parait satisfaisante pour cette recherche ?*"
ville d'environ 1000 habitants	"*Vous recherchez la liste des villes d'environ 1000 habitants. Quelle est, selon vous, la population MINIMALE/MAXIMALE qui vous parait satisfaisante pour cette recherche ?*"

TABLEAU 7.3 – Consignes du questionnaire correspondant aux 5 ENA contextualisées.

7.2.5 Analyses statistiques

Cinq types d'analyses sont réalisés, successivement décrits dans les paragraphes suivants, selon l'hypothèse considérée. Elles permettent respectivement de comparer (i) deux ENA de même valeur de référence ; (ii) deux ENA de valeurs de référence différentes ; (iii) les bornes supérieure et inférieure d'une même ENA ; (iv) les bornes recueillies à celles prédites par les modèles d'interprétation proposés ; et (v) les bornes données par deux groupes de participants.

Le seuil de significativité pour tous les tests statistiques est fixé à $p = 0,01$.

Comparaisons entre deux ENA de même valeur de référence Le premier type d'analyse consiste à comparer deux distributions de bornes correspondant à deux ENA dont la valeur de référence est égale. Ces analyses nous permettent de tester les deux premières hypothèses : (i) H1, qui prédit une différence entre ENA sémantiquement contextualisées et ENA non contextualisées ; et (ii) H2, qui prédit une différence entre intervalles implicitement et explicitement collectés.

Pour chaque comparaison, les bornes inférieures et supérieures sont analysées séparément. Nous comparons la distribution des distances entre les bornes inférieures, ΔP_p^-, ou supérieures, ΔP_p^+, des intervalles et la valeur de référence des deux ENA, par exemple,

"*environ 30*" vs. "*environ 30 euros*". Pour faciliter la lecture, nous notons abusivement "*environ* x_1" et "*environ* x_2" les deux ENA comparées bien que leur valeur de référence soit égale.

Dans la mesure où les distributions observées ne sont pas normales, les comparaisons entre deux ENA sont réalisées en utilisant le test des rangs signés de Wilcoxon, conçu pour des mesures répétées.

En complément des tests de Wilcoxon, nous proposons d'évaluer l'égalité entre les deux distributions de distances des bornes considérées aux valeurs de références des ENA. Ce score d'égalité permet d'apprécier dans quelle mesure les valeurs des deux bornes considérées sont réellement proches ou non. Nous considérons que deux distances de bornes Δ_1 et Δ_2 à leur valeur de référence sont égales si leur différence absolue n'excède pas 10% de la plus petite des deux, ou si leur différence absolue est inférieure ou égale à 1, formellement :

$$Eq'(\Delta_1, \Delta_2) = \begin{cases} 1 & \text{si } \dfrac{|\Delta_1 - \Delta_2|}{\min(\Delta_1, \Delta_2)} \leq 0,1 \text{ ou } |\Delta_1 - \Delta_2| \leq 1 \\ 0 & \text{sinon} \end{cases} \tag{7.3}$$

Ce critère d'égalité est proche de celui défini dans l'équation 4.2 (voir section 4.3.6, p. 49). Il s'en différencie dans la mesure où dans cette étude, nous considérons aussi comme égales deux distances dont la différence absolue est inférieure ou égale à 1 car les valeurs des bornes implicites peuvent être décimales.

Le score d'égalité $Eq'(x_1, x_2, b)$ entre deux distributions de bornes $b \in \{-, +\}$ aux ENA "*environ* x_1" et "*environ* x_2" respectivement, correspond à l'agrégation des valeurs obtenues participant par participant, par la moyenne, formellement :

$$Eq'(x_1, x_2, b) = \frac{1}{|\mathcal{P}(x_1, x_2)|} \sum_{p \in \mathcal{P}(x_1, x_2)} Eq'(\Delta P_p^b(x_1), \Delta P_p^b(x_2)) \tag{7.4}$$

où $\mathcal{P}(x_1, x_2)$ est l'ensemble des participants pour lesquels les intervalles correspondant à "*environ* x_1" et "*environ* x_2" ne sont pas considérés comme aberrants.

Comparaisons entre deux ENA de valeurs de référence différentes Le deuxième type d'analyses consiste à comparer deux distributions de bornes correspondant à deux ENA dont les valeurs de référence diffèrent, par exemple "*environ 30*" vs. "*environ 70*", en traitant séparément les cas des bornes inférieures et supérieures. Nous proposons pour ce faire de réaliser des tests de Wilcoxon pour mesures répétées.

Ces analyses sont utilisées pour tester les trois hypothèses dérivées de l'hypothèse H3 prédisant un effet de la magnitude, de la granularité et du dernier chiffre significatif sur l'interprétation des ENA.

Toutes les comparaisons réalisées pour tester les hypothèse H3.1 à H3.3 impliquent uniquement des ENA non contextualisées. En effet, d'une part le nombre d'ENA sémantiquement contextualisées n'est pas assez conséquent pour réaliser des analyses robustes. D'autre part, ces hypothèses reposent sur un isolement des facteurs liés à la valeur de référence d'une ENA. Or, utiliser des ENA contextualisées résulterait en une interaction entre ces facteurs et ceux liés au contexte.

Comparaisons entre bornes supérieures et inférieures d'une même ENA Le troisième type d'analyses est mis en œuvre pour tester l'hypothèse H4, prédisant que les intervalles collectés ne sont pas nécessairement symétriques.

Comme dans le chapitre 4, nous proposons d'évaluer l'égalité stricte entre les distances des bornes inférieures à la valeur de référence x et les distances des bornes supérieures à la valeur de référence en comptant le nombre de celles-ci qui sont égales. Nous reprenons le critère d'égalité entre deux distances à la valeur de référence défini dans l'équation 7.3 ci-dessus.

Le score de symétrie d'une ENA "*environ* x", noté $Sym'(x)$, qui correspond au taux de bornes collectées qui sont égales, est formellement défini comme :

$$Sym'(x) = \frac{1}{|\mathcal{P}(x)|} \sum_{p \in \mathcal{P}(x)} Eq'(\Delta P_p^-(x), \Delta P_p^+(x)) \tag{7.5}$$

où $\mathcal{P}(x)$ est l'ensemble des participants pour lesquels l'intervalle correspondant à "*environ* x" n'est pas considéré comme aberrant.

Comparaisons des bornes recueillies avec les bornes prédites par les modèles d'interprétation proposés L'hypothèse H5 prédit que le principe général du compromis entre saillance cognitive et plage de valeurs dénotées que nous avons proposé (voir section 5.1 p. 62) rend compte des intervalles collectés et que le modèle des rangs RKM que nous avons élaboré (voir section 5.2.2 p. 68) estime correctement les médianes des bornes observées.

Pour ces analyses, nous ne considérons que les intervalles explicitement collectés. En effet, l'une des hypothèses sur lesquelles les modèles d'interprétation que nous proposons sont fondés est que les participants tendent à préférer des nombres ronds. Or, la méthode de collecte des intervalles implicites peut renvoyer des valeurs réelles comme bornes des intervalles (voir section 7.2.3). Par conséquent, les estimations des modèles que nous proposons, qui sont rondes, peuvent être considérées comme inexactes même si très proches des bornes observées.

Pour tester H5, nous comptons d'abord le nombre de valeurs de bornes collectées qui sont comprises dans les fronts de Pareto (voir section 5.1 p. 62). Formellement, pour l'ENA

"*environ x*", le score de capture des réponses $PFC(x)$ (*Pareto Frontier Capture*) est calculé comme :

$$PFC(x) = \frac{1}{2 \cdot |\mathcal{P}(x)|} \sum_{b \in \{-,+\}} |\{p \in \mathcal{P}(x) | IE_p^b(x) \in P^-(x) \cup P^+(x)\}| \tag{7.6}$$

où $\mathcal{P}(x)$ est l'ensemble des participants p pour lesquels l'intervalle explicite IE_p^b, pour $b \in \{-,+\}$, correspondant à "*environ x*" n'est pas considéré comme aberrant, et $P^-(x)$ et $P^+(x)$ les fronts de Pareto correspondant respectivement à la borne inférieure et à la borne supérieure de l'ENA "*environ x*".

Pour tester dans quelle mesure le modèle des rangs RKM que nous proposons (voir section 5.2.2 p. 68) estime correctement les médianes observées des bornes explicitement collectées, nous utilisons le critère de précision MA défini dans la section 5.2.3 p. 69.

Comparaisons entre deux groupes de participants Le dernier item du questionnaire de collecte explicite des intervalles est dédié au contrôle de l'usage quotidien des mathématiques. L'objectif est d'examiner si un usage régulier des mathématiques affecte l'interprétation des ENA.

L'effet de cette variable est testé en utilisant le test de la somme des rangs de Wilcoxon. Nous comparons les bornes données par deux groupes de participants, constitués selon qu'ils font un usage régulier ou non des mathématiques. Chaque borne de chaque ENA est traitée séparément.

7.3 Résultats

Dans cette section, nous présentons les résultats obtenus en analysant les données collectées. La première sous-section est dédiée à l'examen de l'effet d'un usage régulier des mathématiques sur les intervalles fournis par les participants. Les sous-sections suivantes décrivent les résultats portant successivement sur chacune des cinq hypothèses considérées dans cette étude.

Le tableau D.2 p. 197 de l'annexe D présente les moyennes et médianes des valeurs données par les participants comme bornes des intervalles explicites et implicites correspondant aux 19 ENA de l'étude.

7.3.1 Différence entre les intervalles selon l'usage des mathématiques

Sur les 37 participants inclus dans les analyses, 21 ont rapporté utiliser régulièrement les mathématiques au travail ou dans leurs études et 16 ont rapporté ne pas les utiliser régulièrement.

Aucune différence significative entre les deux groupes n'est observée, quelle que soit la borne de l'ENA considérée et la méthode de recueil employée (W allant de 96 à 225 ; $p = N.S.$).

Ces résultats indiquent donc qu'un usage régulier des mathématiques n'a pas d'effet sur l'interprétation des ENA, implicite comme explicite. Par conséquent, les analyses qui suivent portent sur l'ensemble des intervalles collectés auprès de tous les participants, sans distinction quant à l'usage régulier des mathématiques.

7.3.2 Effet du contexte (H1)

Pour tester l'effet du contexte, nous comparons cinq couples d'ENA de même valeur de référence, où l'une des ENA est sémantiquement contextualisée alors que l'autre ne l'est pas.

Le tableau 7.4 présente les résultats de ces cinq comparaisons. Sur l'ensemble de celles-ci, 49,2% des intervalles correspondant aux ENA sémantiquement contextualisées sont égaux à ceux correspondant aux ENA non contextualisées. Six différences significatives sont observées sur les 20 tests réalisés (voir tableau 7.4).

A un niveau d'analyse plus fin, on observe une différence entre les deux intervalles implicitement et explicitement collectés : 61,1% des intervalles explicites sont égaux lorsqu'on compare ENA contextualisées et non contextualisées, alors qu'une seule différence significative sur 10 comparaisons est observée entre distributions de bornes ("*environ 1000 habitants*" / "*environ 1000*", borne inférieure). Concernant les intervalles collectés implicitement, le score d'égalité est de 38,9% alors que 5 différences significatives sont observées sur les 10 tests de Wilcoxon réalisés (voir tableau 7.4).

Ces résultats nous invitent à nuancer l'hypothèse H1 selon laquelle l'interprétation d'une ENA sémantiquement contextualisée diffère de celle d'une ENA non contextualisée. En effet, le contexte sémantique semble avoir un effet plus important lors d'une interprétation implicite que lors d'une interprétation explicite.

7.3.3 Interprétations implicite et explicite (H2)

La seconde hypothèse prédit que les interprétations implicite et explicite diffèrent. Pour tester cette hypothèse, nous comparons les intervalles implicites aux intervalles explicites.

Le tableau 7.5 présente les résultats obtenus pour chacune des 19 ENA. Le score d'égalité pour l'ensemble des intervalles est de 32,7% et 11 tests de Wilcoxon sur 38 révèlent des différences significatives.

Une analyse plus détaillée permet de regrouper les ENA en trois catégories. La première concerne les ENA sémantiquement contextualisées, pour lesquelles le score d'égalité

Comparaison	Méthode	Borne	Test de Wilcoxon	Score d'égalité
environ 30 euros / environ 30	Explicite	Inférieure	$V = 1,5$	80,8%
		Supérieure	$V = 12$	72,0%
	Implicite	Inférieure	$V = 158$	61,1%
		Supérieure	$V = 70^*$	41,2%
environ 30km / environ 30	Explicite	Inférieure	$V = 64$	65,6%
		Supérieure	$V = 105,5$	63,6%
	Implicite	Inférieure	$V = 360,5^*$	50,0%
		Supérieure	$V = 352,5^*$	45,7%
environ 500km / environ 500	Explicite	Inférieure	$V = 65$	61,3%
		Supérieure	$V = 42,5$	64,5%
	Implicite	Inférieure	$V = 328$	28,6%
		Supérieure	$V = 378,5^*$	34,3%
environ 800 hectares / environ 800	Explicite	Inférieure	$V = 55$	67,7%
		Supérieure	$V = 90,5$	52,9%
	Implicite	Inférieure	$V = 277$	41,7%
		Supérieure	$V = 268$	31,4%
environ 1000 habitants / environ 1000	Explicite	Inférieure	$V = 61,5$	60,7%
		Supérieure	$V = 228^*$	24,1%
	Implicite	Inférieure	$V = 349$	28,6%
		Supérieure	$V = 402,5^*$	25,7%

TABLEAU 7.4 – Effet du contexte sémantique : résultats des comparaisons entre ENA sémantiquement contextualisées et non contextualisées, de même valeur de référence.
* : différence significative observée à $p < 0,01$.

est 39,4%, soit un score légèrement plus élevé que dans le cas d'ENA non contextualisées. D'autre part, les tests de Wilcoxon montrent des différences significatives pour 6 comparaisons de distributions de bornes sur 10.

La seconde catégorie concerne les deux ENA non contextualisées pour lesquelles l'ordre de grandeur de la magnitude relative est très inférieur à celui de la magnitude : 4730 ($RelMag(4730) = 30 << 4730$) et 8150 ($RelMag(8150) = 50 << 8150$). Pour ces ENA, un score d'égalité de 14,6%, soit très inférieur à la moyenne est observé (voir tableau 7.5). De plus, des différences significatives apparaissent pour chacune des bornes de ces ENA. La forte différence entre interprétations implicite et explicite pourrait être attribuée à un biais méthodologique lié à la manière dont sont générées les tailles des intervalles de départ

	Borne supérieure		Borne inférieure	
ENA	Test de Wilcoxon	Score d'égalité	Test de Wilcoxon	Score d'égalité
environ 30	$V = 188$	52,8%	$V = 213$	48,5%
environ 70	$V = 215$	36,4%	$V = 250,5$	39,4%
environ 100	$V = 156,5$	39,4%	$V = 269,5$	32,4%
environ 110	$V = 335,5$	47,1%	$V = 361,5$	32,4%
environ 150	$V = 333$	27,3%	$V = 435^*$	24,2%
environ 500	$V = 119$	39,4%	$V = 346,5$	31,3%
environ 800	$V = 280,5$	26,5%	$V = 284$	38,2%
environ 1000	$V = 161,5$	24,1%	$V = 132$	16,7%
environ 1100	$V = 211$	35,5%	$V = 279$	29,0%
environ 1500	$V = 331,5$	37,1%	$V = 269$	29,4%
environ 4700	$V = 227$	28,1%	$V = 315$	18,2%
environ 4730	$V = 389,5^*$	22,6%	$V = 385,5^*$	12,9%
environ 8000	$V = 307,5$	28,6%	$V = 248$	20,0%
environ 8150	$V = 516^*$	8,8%	$V = 554^*$	14,7%
environ 30 euros en poche	$V = 54^*$	62,7%	$V = 196,5$	53,9%
hôtel à environ 30km	$V = 124,5^*$	42,4%	$V = 112^*$	42,4%
environ 500km séparent deux villes	$V = 53^*$	43,3%	$V = 212$	39,4%
terrain d'environ 800 hectares	$V = 163,5$	41,2%	$V = 227,5$	29,4%
ville d'environ 1000 habitants	$V = 111^*$	18,8%	$V = 105^*$	18,8%

TABLEAU 7.5 – Différence entre interprétations implicite et explicite pour des ENA de même valeur de référence.

* : différence significative observée à $p < 0,01$.

dans l'algorithme de recherche par dichotomie. En effet, ceux-ci sont de petite taille. Il se peut donc que les participants aient été limités dans leurs possibilités de réponses alors que ce n'est pas le cas dans le questionnaire, dans lequel ils sont libres de fixer la valeur qui leur convient.

Enfin, la troisième catégorie concerne les autres ENA non contextualisées, dont l'ordre

de grandeur de la magnitude relative est comparable à celui de la magnitude. Leur score d'égalité est de 33,0% et une seule différence significative est observée ("*environ 150*", borne supérieure).

L'hétérogénéité observée sur l'ensemble des comparaisons réalisées sur les 19 ENA entre intervalles implicitement et explicitement collectés nous invite à nuancer l'hypothèse H2. Il apparaît globalement que l'interprétation implicite ne diffère pas de l'interprétation explicite. Cette absence de différence semble toutefois plus marquée pour les ENA sémantiquement contextualisées que pour les ENA non contextualisées, pour lesquelles une seule différence significative apparaît, bien que les scores d'égalité soient faibles.

7.3.4 Effet de la magnitude, de la granularité et du dernier chiffre significatif (H3)

L'hypothèse H3 consiste à généraliser les résultats de la première étude empirique de l'interprétation des ENA (voir chapitre 4) quant à l'implication de la magnitude, de la granularité et du dernier chiffre significatif dans celle-ci.

7.3.4.1 Effet de la magnitude (H3.1)

Pour tester l'effet de la magnitude sur les intervalles collectés, nous comparons 4 couples d'ENA. Dans chaque couple, la magnitude de la valeur de référence diffère alors que sa granularité et son dernier chiffre significatif sont constants, par exemple "*environ 8150*" vs. "*environ 150*". Nous traitons séparément les intervalles implicitement et explicitement collectés. De même, les bornes inférieures et supérieures sont analysées indépendamment.

Le tableau 7.6 présente les résultats de ces comparaisons. Seule l'une d'entre elles donne lieu à une différence non significative entre les deux distributions de bornes : "*environ 500*" vs. "*environ 1500*" pour la borne inférieure des intervalles implicites.

Ces résultats nous invitent donc à accepter l'hypothèse 3.1, et à confirmer les résultats de la première étude de l'interprétation des ENA (voir section 4.4 p. 50) : la magnitude d'une ENA a un effet sur son interprétation, que celle-ci soit implicite ou explicite.

7.3.4.2 Effet de la granularité (H3.2)

Pour tester l'hypothèse d'un effet de la granularité, deux couples d'ENA sont comparés. Pour chaque couple, le dernier chiffre significatif est identique alors que la granularité varie. Comme pour le test du rôle de la magnitude dans l'interprétation des ENA, les intervalles implicites et explicites ainsi que les bornes inférieures et supérieures sont traités séparément.

		Tests de Wilcoxon	
Comparaison	Méthode	Borne inférieure	Borne supérieure
30 vs. 4730	Explicite	$V = 1^*$	$V = 0^*$
	Implicite	$V = 19,5^*$	$V = 41^*$
100 vs. 1100	Explicite	$V = 7^*$	$V = 3^*$
	Implicite	$V = 16^*$	$V = 15^*$
150 vs. 8150	Explicite	$V = 0^*$	$V = 7,5^*$
	Implicite	$V = 82^*$	$V = 40^*$
500 vs. 1500	Explicite	$V = 22^*$	$V = 25^*$
	Implicite	$V = 112,5$	$V = 45,5^*$

TABLEAU 7.6 – Effet de la magnitude : résultats des comparaisons entre ENA de même granularité et dernier chiffre significatif mais de magnitude différente.
* : différence significative observée à $p < 0,01$.

		Tests de Wilcoxon	
Comparaison	Méthode	Borne inférieure	Borne supérieure
100 vs. 1000	Explicite	$V = 0^*$	$V = 12^*$
	Implicite	$V = 13^*$	$V = 13^*$
800 vs. 8000	Explicite	$V = 6,5^*$	$V = 12^*$
	Implicite	$V = 15^*$	$V = 24^*$

TABLEAU 7.7 – Effet de la granularité : résultats des comparaisons entre ENA de même dernier chiffre significatif mais de granularité différente.
* : différence significative observée à $p < 0,01$.

Le tableau 7.7 présente les résultats de ces comparaisons. Pour chacune d'entre elles, les tests de Wilcoxon montrent une différence significative. Ces résultats appuient donc l'hypothèse H3.2. En effet, la granularité d'une ENA a effet sur son interprétation, aussi bien implicite qu'explicite.

7.3.4.3 Effet du dernier chiffre significatif (H3.3)

L'hypothèse H3.3 prédit un effet du dernier chiffre significatif sur l'interprétation des ENA. Pour la tester, nous comparons trois couples d'ENA. Dans chaque couple, la granularité de la valeur de référence est la même et le dernier chiffre significatif varie. Comme pour les hypothèses H3.1 et H3.2, les intervalles implicites et explicites sont analysés séparément, tout comme les bornes inférieures et supérieures.

		Tests de Wilcoxon	
Comparaison	Méthode	Borne inférieure	Borne supérieure
30 vs. 70	Explicite	$V = 2^*$	$V = 35,5$
	Implicite	$V = 79^*$	$V = 216,5$
100 vs. 500	Explicite	$V = 11^*$	$V = 10,5^*$
	Implicite	$V = 22^*$	$V = 52,5^*$
500 vs. 800	Explicite	$V = 13,5$	$V = 51$
	Implicite	$V = 177,5$	$V = 76,5^*$

TABLEAU 7.8 – Effet du dernier chiffre significatif : résultats des comparaisons entre ENA de même granularité mais dont le dernier chiffre significatif diffère.
* : différence significative observée à $p < 0,01$.

Le tableau 7.8 présente les résultats des comparaisons. Les tests de Wilcoxon montrent 7 différences significatives sur les 12 comparaisons réalisées. Comme lors de la première étude empirique de l'interprétation des ENA (voir section 4.4 p. 50), on observe un effet de palier. En effet, alors que les bornes de "*environ 100*" et de "*environ 500*" diffèrent significativement, aucune différence n'est observée entre cette dernière et "*environ 800*" (voir tableau 7.8).

Si nous retrouvons des résultats similaires à ceux obtenus lors de la première étude, le nombre limité de comparaisons réalisables à partir des données collectées dans la présente étude ne permet pas d'effectuer de généralisation quant à l'hypothèse H3.3.

7.3.5 Symétrie des intervalles (H4)

La quatrième hypothèse prédit que toutes les ENA ne donnent pas lieu à des intervalles symétriques. En nous basant sur le critère d'égalité défini dans la section 7.2.5 p. 108, nous testons dans quelle mesure cette hypothèse est valable pour les intervalles implicitement et explicitement collectés.

Le tableau 7.9 présente les résultats des comparaisons, pour chacune des 19 ENA. Le score de symétrie de l'ensemble des intervalles explicites est 54,3%, plus élevé que dans le cas implicite, 36,2%. Ces scores sont plus faibles que ceux observés lors de la première étude de l'interprétation des ENA (voir chapitre 4), qui est de 78,7%.

D'autre part, on ne retrouve pas les mêmes particularités que dans la première étude. En effet, lors de celle-ci, nous avons observé que les ENA dont l'ordre de grandeur de la magnitude relative est très inférieur à celui de la magnitude (par ex., "*environ 8150*") donnent lieu à des scores de symétrie nettement moins élevés. Dans la présente étude, au contraire, les scores de symétrie de ces ENA ne sont pas particulièrement moins élevés

	Scores de symétrie	
ENA	Explicite	Implicite
environ 30	76,5%	68,6%
environ 70	58,8%	61,1%
environ 100	42,9%	30,3%
environ 110	62,9%	47,2%
environ 150	60,6%	22,2%
environ 500	58,5%	28,6%
environ 800	42,9%	36,1%
environ 1000	35,5%	8,8%
environ 1100	48,4%	22,2%
environ 1500	37,1%	30,6%
environ 4700	54,6%	27,8%
environ 4730	45,5%	44,1%
environ 8000	41,7%	30,6%
environ 8150	60,0%	47,2%
environ 30 euros en poche	81,5%	48,6%
hôtel à environ 30km	66,7%	47,2%
environ 500km séparent deux villes	71,9%	34,3%
terrain d'environ 800 hectares	60,0%	37,1%
ville d'environ 1000 habitants	30,3%	13,9%

TABLEAU 7.9 – Symétrie des intervalles : résultats des comparaisons entre bornes inférieures et supérieures, pour les intervalles explicites et implicites.

que ceux des autres ENA. Par exemple, le score des intervalles explicites correspondant à "*environ 8150*" est de 60% alors que le score de "*environ 1500*" est de 37,1%.

La méthode de calcul des intervalles implicites peut ici aussi expliquer les faibles scores de symétrie de ceux-ci. En effet, la valeur de la borne recherchée est calculée comme le centre d'un intervalle devant la contenir. La connaissance de cette valeur n'est pas exacte et peut donc conduire, si l'intervalle renvoyé par l'algorithme pour l'une des deux bornes est trop large, à des estimations moins fiables et entraîner ainsi une asymétrie de l'intervalle correspondant à l'ENA.

L'ensemble de ces résultats nous invite donc à valider l'hypothèse H4 : la tendance à la symétrie apparaît comme plus faible que dans la première étude.

ENA	Bornes contenues dans les fronts de Pareto
environ 30	67,6%
environ 70	67,6%
environ 100	63,5%
environ 110	60,8%
environ 150	55,4%
environ 500	70,3%
environ 800	70,3%
environ 1000	59,5%
environ 1100	67,6%
environ 1500	70,3%
environ 4700	65,9%
environ 4730	67,6%
environ 8000	70,3%
environ 8150	63,5%
environ 30 euros en poche	48,7%
hôtel à environ 30km	73,0%
environ 500km séparent deux villes	68,9%
terrain d'environ 800 hectares	75,7%
ville d'environ 1000 habitants	67,6%

TABLEAU 7.10 – Pourcentages de bornes explicites données par les participants et contenues dans les fronts de Pareto (voir section 5.1 p. 62).

7.3.6 Performances du principe général d'interprétation des ENA et du modèle RKM (H5)

L'hypothèse H5 prédit que le principe général du compromis entre saillance cognitive et plage de valeurs dénotées que nous avons proposé (voir section 5.1 p. 62) rend compte des intervalles explicitement collectés et que le modèle des rangs RKM que nous avons élaboré (voir section 5.2.2 p. 68) estime correctement les médianes des bornes explicites observées.

Le tableau 7.10 présente le pourcentage de bornes explicites données par les participants qui sont capturées par les fronts de Pareto pour chacune des 19 ENA. Ce pourcentage est de 65,9% pour l'ensemble des bornes données par les participants. Il est similaire pour les ENA sémantiquement contextualisées (66,8%) et pour les ENA non contextualisées (65,6%). Ce score est toutefois inférieur à celui obtenu sur les données de la première

étude : 84,4% (voir section 4.4 p. 50).

En outre, le taux de bonne prédiction de médiane (MA) obtenu par le modèle des rangs RKM que nous proposons est de 39,5%, plus faible que celui obtenu avec le corpus issu de la première étude de l'interprétation des ENA.

Ces résultats nous invitent à l'hypothèse H5. En effet, le principe du compromis entre saillance cognitive et plage de valeurs dénotées ainsi que le modèle des rangs RKM que nous proposons rendent compte des intervalles collectés, quoique dans une moindre mesure que lors de leur validation expérimentale décrite au chapitre 5, réalisée à partir des données collectées dans la première étude empirique de l'interprétation des ENA.

7.4 Discussion

La première étude empirique de l'interprétation des ENA, décrite au chapitre 4, que nous avons menée se focalisait sur les dimensions arithmétiques impliquées dans l'interprétation d'une ENA. L'étude présentée dans ce chapitre avait pour but de compléter la première en considérant, outre la reproduction des résultats obtenus précédemment et une seconde validation des modèles d'interprétation que nous proposons, deux objectifs originaux : (i) examiner l'effet de la présence d'un contexte sémantique sur l'interprétation des ENA ; et (ii) comparer l'interprétation implicite et explicite de ces expressions. Pour répondre au deuxième objectif et combler les lacunes méthodologiques de la littérature, nous avons proposé une tâche originale, dérivée des méthodes utilisées en psychophysique pour collecter implicitement les intervalles.

Dans les paragraphes suivants, nous discutons d'abord des biais induits par la tâche de collecte implicite avant d'aborder les résultats obtenus pour chacune des cinq hypothèses considérées dans cette étude.

Tâche de collecte implicite des intervalles La tâche de collecte implicite des intervalles que nous avons conçue est dérivée de celles utilisées en psychophysique pour déterminer des seuils de détection de stimuli (Tyrrell & Owens, 1988). Elle consiste à proposer plusieurs valeurs au participant, qui doit dire si elles sont dénotées ou non par une ENA, pour déterminer la valeur d'une borne d'un intervalle. Les valeurs proposées sont générées en temps réel, en fonction des réponses précédentes, et selon le principe d'une recherche par partition binaire (voir section 7.2.3.2 p. 103). Ce type de tâche n'a, à notre connaissance, pas été utilisé à ce jour dans le cadre des expressions numériques.

Après sa mise en œuvre, trois observations principales peuvent être faites. Premièrement, l'algorithme de recherche renvoie un intervalle dans lequel se trouve la borne recherchée, nous avons considéré le centre de cet intervalle comme valeur de borne. Aussi,

la valeur exacte de la borne n'est pas connue et la marge d'erreur peut être variable : par exemple, pour le participant 35, l'intervalle renvoyé par l'algorithme pour la borne inférieure de l'ENA "*environ 500*" est $[458, 459]$ alors qu'il est de $[905, 912]$ pour la borne inférieure de "*environ 1000*", résultant en une valeur de borne de $908,5$. Dans le second cas, la marge d'erreur peut biaiser les comparaisons entre bornes implicites et explicites ou entre les bornes implicites. Par conséquent, les résultats impliquant de telles comparaisons sont à prendre avec prudence.

Deuxièmement, un biais peut apparaître en lien avec la répétition des valeurs proposées pour une même ENA. En effet, il se peut qu'après quelques valeurs proposées, le participant forme une représentation explicite de la borne recherchée pour automatiser ses réponses par la suite (Schneider & Chein, 2003) : il se fixe explicitement une limite en dessous de laquelle il répondra toujours oui et toujours non au dessus, ou vice-versa. Aussi, au delà de quelques nombres proposés, la valeur estimée de la borne relèverait davantage de la représentation explicite qu'implicite. Ce biais potentiel nous invite donc à prendre avec prudence les comparaisons entre intervalles implicites et explicites, qui peut être en réalité des comparaisons entre deux intervalles explicites collectés différemment.

Troisièmement, l'intervalle de recherche défini lors de l'initialisation de l'algorithme (voir section 7.2.3.2 p. 103) s'est révélé trop étroit pour deux ENA : "*environ 4730*" et "*environ 8150*". Les résultats de la première étude de l'interprétation des ENA montrent qu'il se peut que certains participants aient souhaité accepter des valeurs au-delà des bornes de cet intervalle. Par conséquent, les résultats impliquant les intervalles implicites correspondant à ces deux ENA doivent être considérés avec prudence.

Le troisième biais est aisément surmontable, en élargissant l'intervalle de recherche, il nous paraît en revanche difficile de surmonter simultanément les deux autres biais. En effet, pour affiner l'intervalle renvoyé par l'algorithme, il serait nécessaire de proposer davantage de valeurs aux participants. Ceci aurait pour effet d'accentuer le biais de répétition conduisant à une représentation explicite. Inversement, pour réduire ce biais, il serait nécessaire de limiter le nombre de valeurs proposées, ce qui conduirait en revanche à augmenter la marge d'erreur liée à l'intervalle renvoyé par l'algorithme.

Aussi, il nous parait intéressant d'explorer d'autres types de tâches pour collecter implicitement les intervalles correspondant à des ENA. Une piste serait d'utiliser une application telle que ReqFlex, présentée au chapitre précédent, qui présente trois avantages. Le premier est qu'elle permet d'obtenir les données qu'un utilisateur considère comme satisfaisantes vis-à-vis d'une requête exprimée sous la forme d'une ENA : par exemple, l'ensemble des voitures dont le kilométrage est compris entre 140 000 et 160 000km pour "*environ 150 000km*". Le second avantage est que la collecte des données est écologique et totalement transparente à l'utilisateur puisqu'il n'est pas mis en condition expérimentale.

Le troisième avantage, enfin, est qu'il est simultanément possible d'étudier l'effet de plusieurs contextes, à la fois sémantiques et pragmatiques : sémantiques en faisant varier la base de données considérée (par ex., voitures vs. salariés) ; pragmatiques en considérant différents rôles de l'utilisateur lors des requêtes (par ex., acheteur vs. vendeur). Toutefois, un point mérite une attention particulière : la distribution des données contenues dans la base doit être prise en compte. En effet, dans l'exemple précédent, s'il n'y pas de véhicule avec un kilométrage compris entre 130 000 et 148 000km, on pourrait être amené à conclure que l'utilisateur considère 148 000km comme la borne inférieure de l'intervalle alors qu'il aurait sélectionné d'autres voitures, au kilométrage moins élevé, si elles avaient été présentes dans la base.

Effet du contexte sur les intervalles L'analyse détaillée que nous avons réalisée pour examiner l'effet du contexte sur les intervalles collectés met en évidence qu'il n'est pas toujours observé, et semble dépendre plus de la méthode de recueil que du contexte sémantique lui même. A la lumière de la littérature (Lasersohn, 1999; Sadock, 1977; Solt, 2014), l'absence d'effet clair du contexte peut paraître surprenant. En effet, outre le domaine des imprécisions, les études portant sur l'incertitude (Clark, 1990; Moxey & Sanford, 2000) montrent que la présence d'un contexte conduit souvent à une implication plus importante de l'individu vis-à-vis de l'événement incertain proposé dans la tâche. On observe ainsi que le contexte des événements incertains décrits peut encourager un raisonnement basé sur une opinion personnelle quant à la signification de l'évènement lié à une thématique (par ex., l'argent, la distance) plutôt que sur la signification des expressions d'incertitude elles-mêmes. Selon ces résultats, nous aurions dû observer une différence entre les aux ENA contextualisées non contextualisées.

Dans le cadre des études sur le raisonnement sous incertitude (Wallsten & Budescu, 1995), il est remarqué que si le contexte n'est pas défini avec une précision suffisante, les participants peuvent ajouter leur propre interprétation (par ex., traiter "*environ 30 euros*" du point de vue d'un vendeur ou d'un acheteur). De plus, un biais de raisonnement reconnu est que les individus ont une tendance générale à n'être pas très bons lors de la production d'un jugement précis, et qu'ils ont souvent tendance à résoudre l'ambiguïté en penchant vers leurs désirs ou attentes (Gilovich et al., 2002). Ces deux facteurs peuvent conduire à une grande variabilité dans leurs réponses et ainsi à une apparente absence d'effet du contexte.

Nos analyses montrent que lorsqu'on compare deux ENA de même valeur de référence, l'une étant contextualisée et l'autre non, les intervalles collectés de manière explicite ont davantage tendance à être égaux que les intervalles implicitement collectés. Compte tenu des biais liés à la méthodologie de recueil implicite des intervalles, il serait hasardeux de

tirer des conclusions tranchées. On peut toutefois noter que ces résultats sont en accord avec ce qui est décrit dans la littérature : par exemple, selon le principe qui sous-tend la cognition sociale implicite (Greenwald & Banaji, 1995), l'expérience passée et son contexte influenceraient automatiquement les évaluations des individus et leurs prises de décisions sans qu'ils en aient vraiment conscience.

De plus, la différence observée sur l'effet du contexte entre interprétation implicite et explicite des ENA peut être expliquée par le fait que la présence d'un contexte sémantique rende la tâche plus concrète et moins formelle pour les participants (Anderson et al., 1996). En effet, interpréter une ENA non contextualisée pourrait davantage s'apparenter à un problème abstrait à résoudre, faisant par conséquent l'objet d'un raisonnement explicite. Au contraire, la contextualisation des ENA pourrait rendre leur interprétation plus intuitive dans la mesure où, si le participant est familier avec le contexte proposé, cette interprétation renvoie à des expériences passées (Carraher et al., 1985).

Ainsi malgré les analyses détaillées que nous avons réalisées pour examiner un éventuel effet du contexte sémantique sur la production d'intervalles et à la lumière de la littérature, il nous semble nécessaire à ce stade de notre étude de considérer avec beaucoup de prudence nos résultats tendant à nier un rôle du contexte en imputant cette absence d'effet à un biais induits par le dispositif expérimental générant les bornes implicites.

Interprétation implicite vs. explicite La seconde question que nous avons abordée a été celle d'examiner dans quelle mesure l'interprétation implicite des ENA diffère de leur interprétation explicite. Nous nous focalisons sur les tests statistiques et non sur les scores d'égalités : ces derniers sont plus sensibles aux biais méthodologiques discutés ci-dessus que les premiers. Les résultats que nous avons obtenus suggèrent que l'interprétation implicite des ENA diffère peu de leur interprétation explicite.

Deux explications peuvent rendre compte de cette observation. On peut d'abord considérer que les résultats obtenus sont le reflet d'une équivalence entre les interprétations implicite et explicite des ENA. Cette explication apparaît toutefois peu vraisemblable : dans de nombreux domaines du raisonnent humain, il est montré que les traitements implicites et explicites différent (Frensch & Rünger, 2003; Nosek, 2007). La seconde explication consiste à considérer que les participants construisent, comme décrit ci-dessus, une représentation explicite même lors de la tâche implicite, nous conduisant ainsi à réaliser des comparaisons entre deux interprétations explicites collectées différemment.

De ce point de vue, nos résultats ne refléteraient donc pas une équivalence entre interprétations implicite et explicite des ENA en terme de processus sous-jacents mais témoigneraient peut-être d'une difficulté à produire chez les participants une interprétation strictement implicite.

Implication de la magnitude, de la granularité et du dernier chiffre significatif dans l'interprétation des ENA Conformément à ce que nous attendions, nous avons également pu reproduire et généraliser les résultats obtenus lors de la première étude concernant l'implication de la magnitude, de la granularité et du dernier chiffre significatif dans l'interprétation des ENA (voir section 4.4 p. 50). En effet, nous avons observé qu'elles sont impliquées non seulement dans l'interprétation explicite des ENA mais également dans leur interprétation implicite.

Ces observations appuient d'autre part notre hypothèse d'un compromis entre les deux sous-systèmes cognitifs responsables de la représentation et du traitement des nombres et quantités, le système approximatif des nombres et le système exact des nombres (voir section 2.2 p. 17), dans la représentation des ENA. En effet, la magnitude renvoie à l'encodage analogique dans le système approximatif des nombres alors que la granularité et le dernier chiffre significatif de la valeur de référence renvoient au traitement opéré par le système exact des nombres en tant que l'interprétation d'une ENA peut être envisagée comme un problème formel.

Toutefois, l'implication du dernier chiffre significatif est moins claire que celle de la magnitude et de la granularité. En effet, des différences significatives n'apparaissent pas dans toutes les comparaisons le mettant en jeu, ce qui est cohérent avec l'effet de palier précédemment observé (voir section 4.4 p. 50). Le faible nombre de comparaisons possibles impliquant cette dimension à partir du matériel proposé aux participants peut rendre compte de cette absence d'effet systématique.

Symétrie des intervalles Nous avons également examiné dans quelle mesure l'hypothèse de symétrie des intervalles est valable sur les nouvelles données collectées, qu'elles soient implicites ou explicites. Les résultats montrent que le taux de symétrie est plus faible que celui observé lors de la première étude. Cette observation peut s'expliquer par la différence de méthodologie de recueil des intervalles. En effet, lors de la première étude, il était demandé aux participants de donner simultanément les deux bornes d'un intervalle. Le participant devait alors former une seule représentation de la plage de valeurs dénotées (voir section 4.3 p. 44). Au contraire, dans l'étude que nous présentons dans ce chapitre, les participants devaient ne donner qu'une seule valeur de borne à la fois. Chaque intervalle est donc construit à partir de réponses à deux questions qui ne sont pas consécutives dans chacune des deux tâches de collecte. Par conséquent, les participants devaient former deux représentations, à deux moments différents d'une tâche, de la même plage de valeurs. L'intervalle final, formé par deux bornes dont chacune renvoie à une représentation différente, peut donc être asymétrique bien que les deux représentations prises individuellement soit symétriques.

Evaluation des modèles d'interprétation La dernière hypothèse examinée se place du point de vue de la modélisation computationnelle. Nous avons mis en évidence la pertinence du principe général du compromis entre saillance cognitive et plage de valeurs dénotées (voir section 5.1 p. 62), et du modèle des rangs RKM (voir section 5.2.2 p. 68) pour rendre compte des données collectées. En effet, il apparaît que 66% des valeurs de bornes données par les participants sont incluses dans les fronts de Pareto. Ce score, quoique inférieur à celui obtenu sur les données précédemment collectées, apporte une validation supplémentaire au principe général.

En revanche, le taux de bonnes prédictions du modèle RKM est plus faible que celui observé sur les données de la première étude. Cette faiblesse peut s'expliquer par le nombre plus limité de participants à l'étude. En effet, un participant ayant été exclu des analyses, au mieux 37 réponses sont disponibles pour calculer la médiane des distributions des bornes, rendant celles-ci peu robustes. Dans la présente étude, on peut observer que les médianes correspondent fréquemment à des valeurs de bornes peu données par les participants. Un plus grand nombre de participants peut rendre plus robuste l'estimation des médianes dans la population, offrant ainsi la possibilité au modèle d'interprétation d'obtenir de meilleures performances.

7.5 Bilan

Dans ce chapitre, nous avons présenté une étude empirique originale de l'interprétation des ENA, complémentaire à la première décrite au chapitre 4, dans laquelle nous nous sommes focalisé sur la contextualisation sémantique des ENA ainsi que sur leur interprétation implicite, pour laquelle nous avons exploré une méthodologie originale de collecte des données. Nous avons montré que l'interprétation des ENA contextualisées diffère peu de celle d'ENA non contextualisées. D'autre part, nos analyses montrent que l'interprétation implicite est semblable à l'interprétation explicite des ENA.

L'étude que nous avons présentée dans ce chapitre est aussi une reproduction et un prolongement de la première étude, présentée dans le chapitre 4. Elle nous a permis de vérifier la robustesse des résultats quant (i) aux dimensions arithmétiques de la valeur de référence des ENA impliquées dans leur interprétation ; (ii) à l'hypothèse de symétrie des intervalles représentant l'imprécision dénotée par les ENA ; et (iii) aux performances du modèle computationnel d'interprétation des ENA RKM que nous avons proposés dans le chapitre 5.

Chapitre 8

Combinaison des ENA : calcul arithmétique

Dans ce chapitre, nous nous intéressons à un traitement différent lié aux Expressions Numériques Approximatives : au-delà de leur interprétation, nous examinons la problématique de leur composition, qui vise à étudier la manière dont elles sont combinées arithmétiquement par l'être humain : la question est par exemple, d'estimer la surface d'une pièce dont les côtés mesurent "*environ 5*" et "*environ 10*" mètres.

La tâche de combinaison d'ENA par des calculs arithmétiques peut être interprétée comme une question de propagation des imprécisions à partir d'opérandes symboliques, c'est-à-dire reposant sur le langage, dans le raisonnement. Nous présentons une étude empirique qui constitue une première approche de la compréhension de cette propagation. Plus spécifiquement, notre objectif est double : (i) évaluer dans quelle mesure la propagation des imprécisions lors d'additions et de multiplications d'ENA correspond aux formalismes mathématiques de la représentation des imprécisions, l'arithmétique des intervalles et l'arithmétique floue, et (ii) identifier les heuristiques utilisées pour produire l'estimation de l'imprécision résultant du calcul.

Du point de vue méthodologique, cette problématique ouvre la question de la qualification et de la quantification de la propagation des imprécisions. Nous l'abordons en proposant deux méthodes originales d'analyse des données, correspondant respectivement à aux perspectives épistémique et sémantique du vague. La première consiste à examiner la propagation des imprécisions en se basant sur une représentation par intervalles des opérandes et des résultats des calculs. La seconde consiste à examiner cette propagation en se basant sur une représentation par sous-ensembles flous des opérandes et résultats.

Ce chapitre est structuré de la manière suivante. Nous introduisons dans la section 8.1 les définitions formelles et notations utilisées. La problématique et les hypothèses motivant

Intervalle	Description	Exemple
$I_C^p(z)$	Résultats des **C**alculs imprécis	"*environ 20*" · "*environ 30*"
$I_P^p(z)$	Approximations des résultats exacts z données par le **P**articipant p	"*environ 600*"
$I_P^p(x)$ et $I_P^p(y)$	Approximations des opérandes x et y données par le **P**articipant p	"*environ 20*" et "*environ 30*"
$I_A^p(z)$	Calculé par application de l'**A**rithmétique des intervalles à $I_P^p(x)$ et $I_P^p(y)$	

TABLEAU 8.1 – Notations et descriptions des quatre intervalles considérés, illustrés pour le calcul "*environ 20*" · "*environ 30*" dans la troisième colonne.

l'étude empirique sont détaillées dans la section 8.2. Nous décrivons dans la section 8.3 la méthodologie mise au point pour collecter les données et les pré-traiter. La section 8.4 détaille les deux méthodes originales d'analyses des données que nous proposons pour tester les hypothèses. Les résultats de l'examen épistémique de la propagation des imprécisions sont présentés dans la section 8.5. La section 8.6 expose les résultats de l'examen sémantique. Enfin, nous discutons de l'ensemble des résultats obtenus dans la section 8.7.

Plusieurs parties de ce chapitre ont fait l'objet d'une publication (Lefort et al., 2017a).

8.1 Définitions et notations

Cette étude se focalise sur l'addition et la multiplication d'ENA. Les calculs imprécis considérés sont de la forme "*environ x*" $\odot$ "*environ y*", où "*environ x*" et "*environ y*" sont les opérandes symboliques imprécises, avec x et $y \in \mathbb{N}^*$. Le symbole $\odot$ correspond à l'opération arithmétique, addition ou multiplication. Nous utilisons le terme *résultat exact* pour signifier z, le résultat du calcul exact, sans imprécision, de $x \odot y$.

Dans cette étude, deux types de représentation formelle des ENA sont utilisés : l'approche par intervalles (voir section 3.1.1 p. 23), correspondant à la perspective épistémique du vague et l'approche par sous-ensembles flous (voir section 3.1.2 p. 24), correspondant à la perspective sémantique du vague. Ces deux approches ne sont pas indépendantes dans la mesure où les sous-ensembles flous sont construits à partir des intervalles (voir section 3.1.2 p. 24).

Cinq types d'intervalles et de sous-ensembles flous, respectivement présentés dans les tableaux 8.1 et 8.2 sont considérés. Nous notons $I_P^p(x)$ et $I_P^p(y)$ les intervalles donnés par le **P**articipant p qui correspondent aux ENA utilisées comme opérandes imprécises, c'est-à-dire à "*environ x*" et "*environ y*". Les sous-ensembles flous représentant ces opérandes

Fonction d'appartenance	Description
μC_z	Elicitée à partir des résultats des **C**alculs imprécis $I_C^p(z)$
μP_z	Elicitée à partir des approximations des résultats exacts données par les **P**articipants $I_P^p(z)$
μP_x et μP_y	Elicitées à partir des approximations des opérandes données par les **P**articipants $I_P^p(x)$ et $I_P^p(y)$
μA_z	Calculée par application de l'**A**rithmétique floue à μP_x et μP_y
μM_z	Estimée par le **M**odèle flou d'interprétation des ENA (voir section 5.3)

TABLEAU 8.2 – Notations et descriptions des fonctions d'appartenance des quatre types de sous-ensembles flous considérés.

imprécises sont notés respectivement μP_x et μP_y.

$I_P^p(z)$ est l'intervalle correspondant à l'approximation du résultat exact z, c'est-à-dire l'intervalle correspondant à "*environ z*", donnés par le **P**articipant p. μP_z est le sous-ensemble flou représentant cette approximation.

L'intervalle $I_C^p(z)$ correspond au résultat fourni par le participant p au **C**alcul imprécis, "*environ x*" $\odot$ "*environ y*". μC_z est le sous-ensemble flou représentant les résultats au calcul.

$I_A^p(z)$ est l'intervalle calculé en appliquant l'**A**rithmétique des intervalles (Moore, 1966) aux intervalles $I_P^p(x)$ et $I_P^p(y)$ correspondant aux opérandes imprécises. De même, μA_z est le sous-ensemble flou obtenu par application de l'**A**rithmétique floue aux opérandes μP_x et μP_y.

Enfin, l'examen de la propagation des imprécisions dans la perspective sémantique du vague tient également compte des intervalles flous produits par le **M**odèle flou d'interprétation des ENA fRKM (voir section 5.3 p. 80). Ces intervalles flous sont notés ici μM_z.

8.2 Problématique et hypothèses

A notre connaissance, il n'existe pas, à ce jour, d'étude publiée portant sur la propagation des imprécisions dans le calcul imprécis réalisé par l'être humain. Cette problématique soulève pourtant la question des rôles joués par les systèmes approximatif et exact des nombres (voir section 2.2 p. 17) dans le calcul imprécis. En effet, d'un côté, le calcul arithmétique implique le système exact et, de l'autre, l'estimation des imprécisions implique le système approximatif (Dehaene, 2003).

Selon nous, deux hypothèses antagonistes peuvent être considérées pour rendre compte des traitements cognitifs opérés lors d'un calcul imprécis. Dans la première, le système exact des nombres prend en charge à la fois le calcul exact, c'est-à-dire $x \odot y$, par exemple $20 \cdot 30$ pour "*environ 20*" · "*environ 30*", et la propagation des imprécisions. Dans ce cas, le système approximatif des nombres ne joue aucun rôle spécifique et la propagation des imprécisions dans le calcul est conforme à ce qui serait attendu selon les formalismes mathématiques.

Dans la seconde hypothèse, le système exact des nombres prend en charge le calcul exact, par exemple $20 \cdot 30$ pour "*environ 20*" · "*environ 30*". L'estimation de l'imprécision résultant du calcul est quant à elle, comme pour l'interprétation des ENA, le produit d'un compromis entre les deux systèmes cognitifs. Autrement dit, l'imprécision finale est équivalente à celle estimée pour l'approximation du résultat exact, par exemple "*environ 600*" pour "*environ 20*" · "*environ 30*". Dans ce cas, l'imprécision résultant du calcul devrait être indépendante des opérandes et de l'opération arithmétique considérées et ne devrait pas correspondre à ce qui est attendu selon les formalismes mathématiques.

Nous avons mis en évidence dans les études empiriques de l'interprétation des ENA (voir chapitres 4 et 7) que les deux sous-systèmes responsables de la cognition des nombres sont tous les deux impliqués dans le traitement des ENA. Nous prenons par conséquent le parti de considérer a priori la seconde hypothèse, à savoir que lors d'une combinaison arithmétique d'opérandes imprécises, les individus réalisent dans un premier temps le calcul exact avant d'en produire une approximation. Par exemple, dans le cas de "*environ 150*" + "*environ 8000*", ils calculent d'abord $150+8000 = 8150$, sans tenir compte des imprécisions de "*environ 150*" et "*environ 8000*". Ensuite, ils estiment l'imprécision autour de 8150, comme ils le feraient pour "*environ 8150*".

Pour tester cette hypothèse principale, quatre propositions peuvent être formulées. La version formelle des hypothèses est donnée pour les intervalles. Elle s'applique identiquement aux sous-ensembles flous.

H1 - Résultat du calcul semblable à l'approximation du résultat exact : si l'imprécision assignée au résultat dépend uniquement du résultat exact, nous devrions observer que ce résultat est égal à l'approximation du résultat exact.

Formellement, cette hypothèse s'écrit : $I_C^p(z) = I_P^p(z)$.

H2 - Calcul non conforme aux formalismes mathématiques : si l'imprécision des opérandes n'est pas prise en compte, nous devrions observer que le résultat du calcul imprécis ne correspond pas à celui obtenu par application de l'arithmétique des intervalles ou de l'arithmétique floue.

Formellement, cette hypothèse s'écrit : $I_C^p(z) \neq I_A^p(z)$.

H3 - Insensibilité au calcul : si les participants ne tiennent pas compte de l'imprécision assignée aux opérandes, mais réalisent le calcul exact avant de produire une approximation du résultat, nous devrions observer la même imprécision finale dans les cas où deux opérations différentes donnent le même résultat exact, qu'elles varient selon les opérandes ou l'opérateur (addition ou multiplication). Plus spécifiquement, nous distinguons deux cas :

H3.1 - *Insensibilité à la granularité* : comme nous l'avons montré dans les deux études empiriques de l'interprétation des ENA, l'imprécision assignée à une ENA dépend, entre autres, de sa granularité (voir section 4.4 p. 50 et section 7.3 p. 111). Plus précisément, plus la granularité est importante, plus l'imprécision est grande.

Par conséquent, si les participants tiennent compte de l'imprécision des opérandes, les calculs dont les opérandes présentent une granularité plus importante (par ex., "*environ 400*" + "*environ 600*", où $Gran(400) = Gran(600) = 100$) devraient aboutir à une imprécision plus grande par rapport aux calculs dont les opérandes sont caractérisées par une granularité plus faible (par ex., "*environ 440*" + "*environ 560*", où $Gran(440) = Gran(560) = 10$).

Au contraire, si l'imprécision assignée aux opérandes n'est pas prise en compte, ces différents calculs devraient aboutir à la même imprécision.

Formellement, cette hypothèse s'écrit : $I_C^p(x_1+y_1) = I_C^p(x_2+y_2)$, avec $x_1+y_1 = x_2+y_2$, $Gran(x_1) \neq Gran(x_2)$ et $Gran(y_1) \neq Gran(y_2)$ [6].

H3.2 - *Insensibilité à l'opération arithmétique* : comme pour l'hypothèse H3.1, si les participants tiennent compte de l'imprécision assignée aux opérandes, celle assignée au résultat devrait être différente lorsqu'on considère deux calculs dont à la fois les opérandes et l'opérateur diffèrent mais dont le résultat exact est le même (par ex., "*environ 20*" · "*environ 50*" vs. "*environ 400*" + "*environ 600*").

Au contraire, si les imprécisions ne sont pas prises en compte, l'imprécision assignée au résultat devrait être la même, quelle que soit l'opération considérée.

Formellement, cette hypothèse s'écrit : $I_C^p(x_1 \cdot y_1) = I_C^p(x_2 + y_2)$, avec $x_1 \cdot y_1 = x_2 + y_2$.

6. Pour cette hypothèse, nous ne considérons que le cas des additions. En effet, il existe davantage de combinaisons possibles de granularité pour aboutir au même résultat dans le cas des additions que dans celui des multiplications. Par conséquent, tester cette hypothèse à partir de multiplications conduirait à considérer une plus grande quantité d'items dans le questionnaire que nous avons conçu pour collecter les données (voir section 8.3 p. 132).

H4 - Sensibilité aux paradoxes sorites : Les paradoxes sorites expriment une insensibilité à des petits changements conduisant à l'impossibilité d'un grand changement (voir section 2.1.1 p. 8). Ces paradoxes, caractéristiques des expressions vagues, peuvent être appliqués aux calculs imprécis. En effet, l'imprécision assignée à une ENA dont la granularité est importante (par ex., "*environ 6000*") ne devrait être que peu affectée lorsqu'on lui ajoute une ENA dont l'ordre de grandeur de la granularité est beaucoup plus faible (par ex., "*environ 10*") car l'imprécision de cette dernière se trouve noyée dans celle de la première. Ainsi, l'imprécision assignée à "*environ 6000*" + "*environ 10*" devrait être très proche de celle assignée à "*environ 6000*".

Au contraire, si les imprécisions ne sont pas prises en compte dans le calcul imprécis, l'imprécision assignée à "*environ 6000*" + "*environ 10*" devrait être proche de celle assignée à "*environ 6010*" et, par conséquent, nettement inférieure à celle assignée à "*environ 6000*".

Formellement, cette hypothèse s'écrit : $|I_C^p(z)| << |I_P^p(x)|$, où x est l'opérande dont la granularité est élevée et z est le résultat exact du calcul.

8.3 Méthodes

Nous décrivons dans cette section la méthodologie que nous avons mise en œuvre pour collecter et nettoyer les données relatives à l'étude du calcul imprécis chez l'être humain de manière à tester les quatre hypothèses considérées.

8.3.1 Population

146 participants se sont portés volontaires pour participer à cette étude : 102 femmes et 44 hommes, dont l'âge est compris entre 20 et 74 ans ($M = 38,6; \sigma = 14,2$). Ils ont été recrutés par une annonce diffusée sur une liste de diffusion et tous ont pour langue maternelle le français.

8.3.2 Matériel

Un questionnaire en ligne a été conçu pour collecter de manière explicite les données. Il se compose de trois parties. La première permet de collecter les résultats de 23 calculs imprécis, listés dans le tableau 8.3, 5 multiplications et 18 additions. Ils ont été sélectionnés pour minimiser le niveau nécessaire en calcul mental et pour éviter tout effet de taille du problème (Zbrodoff & Logan, 2005). Pour être réalisées, les additions ne nécessitent pas de retenue et leurs opérandes ne possèdent pas plus de deux chiffres significatifs. Les opérandes des multiplications ne possèdent quant à elles qu'un seul chiffre significatif.

De plus, les opérandes ont été sélectionnées de manière à pouvoir tester les hypothèses

Calcul			Résultat exact
environ 20	·	environ 30	600
environ 20	·	environ 400	8000
environ 20	·	environ 50	1000
environ 30	·	environ 50	1500
environ 40	·	environ 50	2000
environ 20	+	environ 30	50
environ 20	+	environ 80	100
environ 30	+	environ 4700	4730
environ 40	+	environ 110	150
environ 40	+	environ 400	440
environ 50	+	environ 100	150
environ 100	+	environ 500	600
environ 150	+	environ 8000	8150
environ 200	+	environ 400	600
environ 200	+	environ 800	1000
environ 400	+	environ 600	1000
environ 400	+	environ 1100	1500
environ 440	+	environ 560	1000
environ 500	+	environ 1000	1500
environ 500	+	environ 1500	2000
environ 2000	+	environ 6000	8000
environ 10	+	environ 6000	6010*
environ 20	+	environ 8000	8020*

TABLEAU 8.3 – Calculs proposés dans le questionnaire.
* : ces deux items sont dédiés à l'hypothèse H4 portant sur la sensibilité aux paradoxes sorites. Ce sont ceux pour lesquels l'approximation du résultat exact n'est pas demandée.

considérées. En effet, plusieurs calculs aboutissent au même résultat exact bien que leur opération soit différente (par ex., "*environ 20*" · "*environ 50*" et "*environ 400*" + "*environ 600*") ou que leurs opérandes diffèrent en granularité (par ex., "*environ 440*" + "*environ 560*" et "*environ 400*" + "*environ 600*").

La consigne est : "*selon vous, entre quelles valeurs MINIMALE et MAXIMALE se trouve le résultat de environ x ⊙ environ y ?*", où "*environ x*" et "*environ y*" sont les opérandes, avec $x, y \in \mathbb{N}^*$, et $\odot$ l'opération arithmétique considérée. Cette partie du questionnaire permet de collecter les intervalles $I_C^p(z)$.

La seconde partie du questionnaire est conçue pour collecter les intervalles correspondant aux approximations des 24 opérandes imprécises et résultats exacts de 21 des 23 calculs. Elle permet de collecter les intervalles $I_P^p(x)$, $I_P^p(y)$ et $I_P^p(z)$. Par exemple, "*environ 200*", "*environ 400*" et "*environ 600*" pour le calcul "*environ 200*" + "*environ 400*". Les approximations des résultats exacts des deux calculs utilisés pour tester l'hypothèse H4 concernant les paradoxes sorites, "*environ 6010*" et "*environ 8020*", ne sont pas demandées aux participants pour être certain que ceux-ci ne sont pas biaisés par celles-ci lorsqu'ils réalisent les calculs. En effet, il se peut qu'en donnant d'abord l'approximation du résultat exact, par exemple "*environ 6010*", les participants soient induits à donner les mêmes intervalles comme résultats aux calculs, par exemple "*environ 10*" + "*environ 6000*".

La consigne de cette partie est "*selon vous, entre quelles valeurs MINIMALE et MAXIMALE se trouve environ x ?*", où x est la valeur de référence de l'opérande ou du résultat exact considéré.

Tous les items du questionnaire, qu'ils portent sur les résultats des calculs ou sur les approximations des opérandes et résultats exacts, ne sont pas sémantiquement contextualisés.

Enfin, la troisième partie du questionnaire comporte deux questions dont le but est de contrôler, comme lors de l'étude de l'interprétation des ENA (voir section 4.3.2 p. 44), la variabilité interindividuelle quant :

— à l'usage des mathématiques, de manière à contrôler si un usage quotidien des mathématiques au travail ou dans les études affecte la combinaison des ENA.
— au niveau subjectif en calcul mental : comme information complémentaire à la précédente question, le participant estime, sur une échelle de type Likert graduée de 1 à 5, son niveau en calcul mental.

8.3.3 Procédure

Pour chacune des trois parties du questionnaire, l'ordre des questions est aléatoirement tiré pour chaque participant. De même, l'ordre dans lequel un participant remplit la partie calculs et la partie approximations est aléatoirement tiré. La troisième partie, comprenant les questions sur l'usage des mathématiques et le niveau subjectif en calcul mental, est toujours la dernière à être remplie.

Dans les deux premières parties, calculs et approximations, les participants doivent, comme dans la première étude de l'interprétation des ENA (voir section 4.3.3 p. 46), donner les valeurs des bornes inférieure et supérieure d'un intervalle comme une seule réponse, en entrant ces valeurs dans l'espace dédié.

8.3.4 Nettoyage des données

Pour exclure les intervalles extrêmes ou aberrants des données, celles-ci ont été nettoyées selon une procédure en quatre étapes. Les trois premières étapes sont identiques à celles décrites dans les deux études empiriques de l'interprétation des ENA (voir sections 4.3.5 p. 46 et 7.2.4 p. 107). Les étapes 1 à 3 sont exécutées pour $I_P^p(x)$, $I_P^p(y)$, $I_P^p(z)$ et $I_C^p(z)$.

La quatrième étape consiste à exclure tout calcul pour lequel au moins un des quatre intervalles collectés est manquant, extrême ou aberrant.

Sur les 3358 calculs présents dans le corpus original, 2516 (74,9%) sont utilisés dans les analyses. 28 participants ont été exclus.

8.4 Analyses des données

Dans la mesure où, à notre connaissance, aucune étude portant sur le calcul imprécis à partir d'opérandes symboliques n'a été publiée à ce jour, nous manquons d'une méthodologie d'analyse des données. Pour combler cette lacune méthodologique de la littérature, nous proposons deux types d'analyses originales, successivement détaillés dans cette section, pour tirer partie des données collectées. Le premier s'inscrit dans la perspective épistémique du vague et est basé sur une représentation par intervalles des ENA. Le second type est basé sur une représentation par sous-ensembles flous des ENA et s'inscrit donc dans la perspective sémantique du vague.

8.4.1 Analyse dans la perspective épistémique du vague : intervalles

Dans la perspective épistémique du vague, où les imprécisions sont représentées par des intervalles, toute l'analyse des données s'écrit comme des comparaisons d'intervalles. Cette section décrit les analyses réalisées dans cette perspective en vue de tester les quatre hypothèses considérées. En premier lieu, nous présentons brièvement l'arithmétique des intervalles (Moore, 1966), utilisée pour calculer $I_A^p(z)$. Dans un deuxième temps, nous détaillons les tests sur lesquels reposent les comparaisons des intervalles. Ensuite, nous montrons comment ces analyses s'appliquent aux hypothèses.

8.4.1.1 Arithmétique des intervalles

L'arithmétique des intervalles a été formalisée par Moore (1966). Nous en rappelons ici les résultats utiles pour les intervalles que nous considérons, qui ont la particularité d'avoir des bornes strictement positives, ce qui les place dans un cas plus simple que le cas général.

En notant deux intervalles $I_1 = [s, t]$ et $I_2 = [u, v]$ avec $s, t, u, v > 0$, on a (Moore, 1966) :

$$\begin{aligned} I_1 + I_2 &= [s + u, t + v] \\ \text{et } I_1 \cdot I_2 &= [s \cdot u, t \cdot v] \end{aligned}$$

Ces opérations permettent, dans notre cas, de définir $I_A^p(z)$ à partir de $I_P^p(x)$ et $I_P^p(y)$, c'est-à-dire ce que donne ce modèle formel de propagation des imprécisions à partir de la représentation qu'ont les participants de l'imprécision des opérandes.

Il faut souligner la différence essentielle de propagation des imprécisions observée entre les deux opérations. Si on note $I_1 = [x - \Delta_1^-, x + \Delta_1^+]$ et $I_2 = [y - \Delta_2^-, y + \Delta_2^+]$, les équations ci-dessus donnent :

$$\begin{aligned} I_1 + I_2 &= [x + y- \quad (\Delta_1^- + \Delta_2^-), \qquad\qquad x + y+ \quad (\Delta_1^+ + \Delta_2^+)] \\ \text{et } I_1 \cdot I_2 &= [x \cdot y- \quad (x \cdot \Delta_2^- + y \cdot \Delta_1^- - \Delta_1^- \cdot \Delta_2^-), \quad x \cdot y+ \quad (x \cdot \Delta_2^+ + y \cdot \Delta_1^+ + \Delta_1^+ \cdot \Delta_2^+)] \end{aligned} \tag{8.1}$$

Aussi, pour les additions, les imprécisions se propagent linéairement. Pour les multiplications, la propagation est quadratique du fait des produits $\Delta_1^- \cdot \Delta_2^-$ et $\Delta_1^+ \cdot \Delta_2^+$. De plus, elle est influencée par les valeurs de référence x et y, alors que ce n'est pas le cas pour l'addition.

Au regard de notre hypothèse générale, selon laquelle les participants ne tiennent pas compte des imprécisions liées aux opérandes pour estimer l'imprécision résultant du calcul, la différence dans la propagation des imprécisions entre multiplications et additions devrait donner lieu à un peu plus grand écart entre ce qui est attendu selon l'arithmétique des intervalles et les réponses fournies par les participants pour les multiplications que pour les additions.

8.4.1.2 Comparaisons d'intervalles

Dans cette section, nous présentons les analyses réalisées pour comparer les intervalles entre eux, ainsi que les distributions de leurs bornes entre elles. Dans un premier temps, nous introduisons les représentations utilisées. Les comparaisons numériques entre distributions de bornes d'intervalles sont ensuite détaillées. Enfin, nous décrivons les comparaisons topologiques réalisées pour caractériser les relations entre intervalles. Le seuil de significativité pour tous les tests statistiques est fixé à $p = 0,01$.

Représentation des intervalles Comme pour les analyses réalisées dans les études de l'interprétation des ENA (voir chapitres 4 et 7), nous transformons les intervalles collectés de manière à obtenir deux distances absolues, Δ^- et Δ^+, entre la valeur de référence de l'ENA considérée et les bornes de l'intervalle correspondant. En notant $I(x) = [x^-, x^+]$

l'intervalle considéré, $\Delta^-(x) = x - x^-$ et $\Delta^+(x) = x^+ - x$. Par exemple, $\Delta^-_C(z) = z - z^-_C$ et $\Delta^+_C(z) = z^+_C - z$.

Comparaisons numériques Ce paragraphe présente les analyses réalisées pour comparer deux distributions d'intervalles, par exemple $I^p_C(z)$ et $I^p_P(z)$ fournis par tous les participants p pour un calcul dont le résultat exact z est fixé. Dans chacune des analyses, les bornes inférieures et supérieures sont traitées séparément.

Dans ce qui suit, nous notons $\mathcal{I}_1$ et $\mathcal{I}_2$ deux distributions d'intervalles, indexées par les participants : $\mathcal{I}(p)$ désigne l'intervalle fourni par le participant p, il peut par exemple s'agir de $I^p_C(z)$. $\Delta^-(p)$ désigne la distance de la borne inférieure de cet intervalle à la valeur de référence du résultat exact z. De même, $\Delta^+(p)$ désigne la distance de la borne supérieure de cet intervalle à la valeur de référence du résultat exact z. Par exemple, $\Delta^-(p) = z - z^-$ et $\Delta^+(p) = z^+ - z$ si $\mathcal{I}(p)$ est $I^p_C(z) = [z - z^-, z + z^+]$. Ces deux distributions contiennent le même nombre d'intervalles, correspondant au nombre de participants dont les intervalles considérés ne sont pas exclus.

Nous proposons trois critères de comparaison numérique. Le premier évalue la proximité des deux distributions : elle est calculée comme la proportion de bornes suffisamment proches et elle repose d'abord sur la notion d'égalité entre deux distances à la valeur de référence que nous avons définie dans l'équation 4.2 (voir section 4.3.6, p. 49), rappelée ici pour faciliter la lecture :

$$Eq(\Delta_1, \Delta_2) = \begin{cases} 1 & \text{si } \dfrac{|\Delta_1 - \Delta_2|}{\min(\Delta_1, \Delta_2)} \leq 0,1 \\ 0 & \text{sinon} \end{cases}$$

Aussi deux bornes sont considérées comme égales si la valeur absolue de la différence entre leurs distances à la valeur de référence z ne dépasse pas 10% de la plus petite de ces deux distances [7]. La proximité entre les distributions de bornes $b \in \{-, +\}$ des intervalles $\mathcal{I}_1$ et $\mathcal{I}_2$, que nous nommons par abus de langage *score d'égalité* entre deux distributions, est alors formellement définie comme :

$$Eq(\mathcal{I}_1, \mathcal{I}_2, b) = \frac{1}{|\mathcal{I}_1|} \sum_p Eq(\Delta^b_1(p), \Delta^b_2(p)) \tag{8.2}$$

Le deuxième critère que nous utilisons permet d'évaluer dans quelle mesure les différences observées, s'il y en a, tendent vers l'inclusion de l'un des deux intervalles dans l'autre (par ex., si $I^p_P(z)$ tend à être inclus dans $I^p_C(z)$). Pour ce faire, nous examinons si des tests

7. Nous avons également réalisé l'ensemble des analyses sans autoriser une erreur relative de 10%. Dans la mesure où les scores obtenus ne diffèrent pas de plus de 4% de ceux obtenus avec une erreur relative autorisée, nous ne reportons dans les résultats que ces derniers.

statistiques montrent des différences significatives entre les Δ des deux intervalles. En effet, une différence significative indique que les bornes de l'un des deux intervalles tendent systématiquement à être plus éloignées ou plus proches de la valeur de référence que les bornes de l'autre intervalle. Les distributions des Δ n'étant pas normales, nous utilisons le test des rangs signés de Wilcoxon, conçu pour les mesures répétées.

Le troisième critère, enfin, est complémentaire au précédent et consiste à calculer la médiane de la distribution des différences entre les bornes $b \in \{-, +\}$ des deux intervalles comparés $\mathcal{I}_1$ et $\mathcal{I}_2$. Celle-ci est calculé comme :

$$MD(\mathcal{I}_1, \mathcal{I}_2, b) = \text{median}_p(\Delta_2^b(p) - \Delta_1^b(p)) \tag{8.3}$$

Comparaisons topologiques Pour examiner plus précisément la manière dont les participants réalisent les calculs imprécis, nous proposons d'une façon originale d'utiliser une autre famille de critères, basée sur les relations topologiques entre intervalles. Ces analyses permettent d'examiner, par exemple, si les participants compensent les résultats des calculs imprécis pour tendre vers les intervalles $I_A^p(z)$.

Pour ce faire, nous nous appuyons sur l'algèbre des relations topologiques entre intervalles proposée par Allen (1983), qui caractérise la position relative de deux intervalles en distinguant treize cas possibles, détaillés dans le tableau E.1 p. 200 de l'annexe E. Nous proposons d'en utiliser une version simplifiée à neuf cas. En effet, la valeur du résultat exact z étant nécessairement incluse dans les intervalles, quatre relations ne peuvent pas apparaître dans nos données : *meets*, *meets inverse*, *before* et *after*.

De plus, nous proposons de simplifier les 9 relations restantes, en réduisant l'expressivité, pour restreindre le nombre de combinaisons possibles de relations entre les trois intervalles considérés dans nos analyses, $I_C^p(z)$, $I_P^p(z)$ et $I_A^p(z)$. En effet, prendre en compte les 9 relations aboutirait à $9^3 = 729$ combinaisons possibles. Nous proposons les simplifications suivantes. Les relations *starts* et *finishes* ne sont pas distinguées de la relation *during* car dans les trois cas, un intervalle est inclus dans l'autre. De même, *starts inverse* et *finishes inverse* ne sont pas distinguées de la relation *during inverse*. De plus, *overlaps* et *overlaps inverse* sont regroupées dans la relation *overlaps*. Nous utilisons, abusivement, les noms *during*, *during inverse* et *overlaps* pour chacun de ces regroupements respectivement.

Par conséquent, quatre relations, illustrées dans le tableau 8.4, sont considérées dans cette étude : *during* (**d**), *during inverse* (**di**), *overlaps* (**o**) et *equals* (**eq**). Formellement, en notant $I_1 = [s, t]$ et $I_2 = [u, v]$ les deux intervalles à comparer, celle-ci, notée $RA(I_1, I_2)$

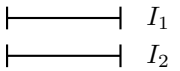
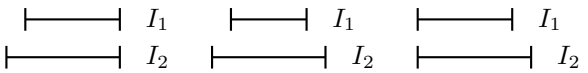
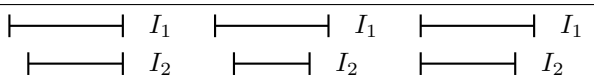
Relation	Illustration
I_1 **eq** I_2 (*equals*)	
I_1 **d** I_2 (*during*)	
I_1 **di** I_2 (*during inverse*)	
I_1 **o** I_2 (*overlaps*)	

TABLEAU 8.4 – Relations utilisées pour caractériser la relation topologique entre deux intervalles.

est calculée comme :

$$RA(I_1, I_2) = \begin{cases} \textbf{eq} & \text{si } s = u \wedge t = v \\ \textbf{d} & \text{si } (s \geq u \wedge t < v) \vee (s > u \wedge t \leq v) \\ \textbf{di} & \text{si } (s \leq u \wedge t > v) \vee (s < u \wedge t \geq v) \\ \textbf{o} & \text{sinon} \end{cases} \tag{8.4}$$

Comme pour le score d'égalité (voir équation 8.2), deux bornes sont considérées comme égales quand leur différence absolue n'excède pas un dixième de la plus petite des deux.

8.4.1.3 Analyses des hypothèses H1 à H3

Pour tester l'hypothèse H1, prédisant que le résultat du calcul imprécis est égal à l'approximation du résultat exact, nous comparons les intervalles $I_C^p(z)$ correspondant aux résultats des calculs imprécis (par ex., "*environ 20*" · "*environ 30*"), collectés dans la première partie du questionnaire, aux $I_P^p(z)$, correspondant aux approximations du résultat exact (par ex., "*environ 600*"), collectés dans la seconde partie du questionnaire.

Pour tester l'hypothèse H2, prédisant que le résultat du calcul imprécis est différent du résultat attendu selon l'arithmétique des intervalles, les intervalles $I_C^p(z)$ sont comparés aux intervalles $I_A^p(z)$, obtenus par application de l'arithmétique des intervalles aux intervalles correspondant aux approximations des deux opérandes du calcul (par ex., "*environ 20*" et "*environ 30*"), collectés dans la seconde partie du questionnaire.

Enfin, pour tester l'hypothèse H3.1, prédisant une insensibilité à la granularité des opérandes, et l'hypothèse H3.2, prédisant une insensibilité à l'opération, nous comparons les intervalles $I_C^p(z)$ correspondant à deux calculs différents dont les résultats exacts z sont égaux.

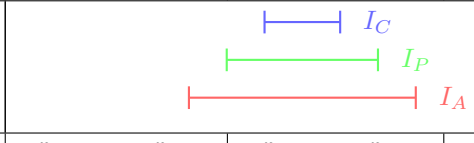
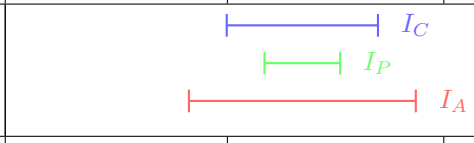
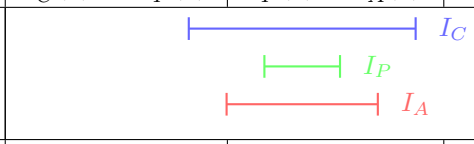
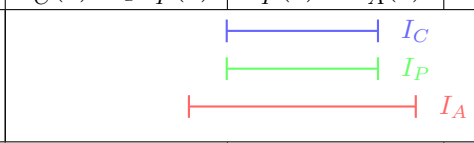
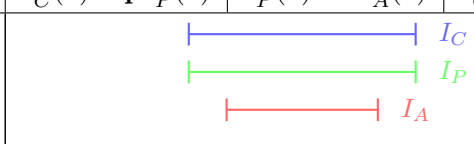
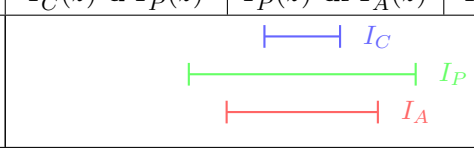
Combinaison	Illustration		
C1	$I_C^p(z)$ **d** $I_P^p(z)$	$I_P^p(z)$ **d** $I_A^p(z)$	$I_C^p(z)$ **d** $I_A^p(z)$
			
C2	$I_C^p(z)$ **di** $I_P^p(z)$	$I_P^p(z)$ **d** $I_A^p(z)$	$I_C^p(z)$ **d** $I_A^p(z)$
			
C3	$I_C^p(z)$ **di** $I_P^p(z)$	$I_P^p(z)$ **d** $I_A^p(z)$	$I_C^p(z)$ **di** $I_A^p(z)$
			
C4	$I_C^p(z)$ **eq** $I_P^p(z)$	$I_P^p(z)$ **d** $I_A^p(z)$	$I_C^p(z)$ **d** $I_A^p(z)$
			
C5	$I_C^p(z)$ **eq** $I_P^p(z)$	$I_P^p(z)$ **di** $I_A^p(z)$	$I_C^p(z)$ **di** $I_A^p(z)$
			
C6	$I_C^p(z)$ **d** $I_P^p(z)$	$I_P^p(z)$ **di** $I_A^p(z)$	$I_C^p(z)$ **d** $I_A^p(z)$
			

TABLEAU 8.5 – Combinaisons de relations topologiques entre les trois intervalles $I_C^p(z)$, $I_P^p(z)$ et $I_A^p(z)$ qui apparaissent avec une fréquence relative supérieure à 5% dans les données collectées.

Ces trois hypothèses sont testées indépendamment les unes des autres en comparant numériquement les intervalles selon la méthodologie proposée ci-dessus.

Dans un second temps, nous examinons les combinaisons de relations topologiques entre les trois intervalles considérés : $I_C^p(z)$, $I_P^p(z)$ et $I_A^p(z)$. Pour ce faire, nous nous basons sur les 4 cas de relation entre deux intervalles présentés ci-dessus. Parmi les $4^3 = 64$ combinaisons possibles de relations entre les trois intervalles, seules 6 apparaissent avec une fréquence relative de plus de 5% dans les données collectées. Nous nous focalisons donc sur ces 6 combinaisons, notées C1 à C6, qui sont illustrées dans le tableau 8.5.

Par exemple, dans C4 et C5, l'imprécision assignée au résultat du calcul imprécis est égale à celle assignée à l'approximation du résultat exact. Dans le premier cas, C4, ces imprécisions sont moins importantes que celle attendue selon le calcul des intervalles ; elles sont supérieures à cette dernière dans le cas de C5. C4 et C5 sont donc les combinaisons qui correspondent à notre hypothèse générale, qui prédit que, lors d'un calcul imprécis, les participants ne tiennent pas compte de l'imprécision des opérandes mais réalisent une approximation du résultat exact.

8.4.1.4 Analyses de l'hypothèse H4

Pour tester l'hypothèse H4, prédisant une sensibilité aux paradoxes sorites, nous comparons l'imprécision assignée au résultat du calcul imprécis, représentée par $\Delta_C(z)$, à celle assignée à l'opérande la plus grande, $\Delta_P(x)$. Pour comparer ces imprécisions, nous réalisons deux tests des rangs signés de Wilcoxon pour mesures répétées, le premier testant la différence entre $\Delta_C^-(z)$ et $\Delta_P^-(x)$ et le second testant la différence entre $\Delta_C^+(z)$ et $\Delta_P^+(x)$.

Aucune comparaison n'implique l'approximation du résultat exact, c'est-à-dire les intervalles $I_P^p(z)$, car ceux-ci ne sont volontairement pas demandés dans le questionnaire, comme discuté dans la section 8.3.2 p. 132.

8.4.1.5 Analyses de la variabilité inter-individuelle

Les derniers items du questionnaire permettent d'évaluer dans quelle mesure les participants utilisent régulièrement les mathématiques ainsi que leur niveau subjectif en calcul mental. Nous pouvons ainsi examiner si ces variables inter-individuelles sont impliquées dans la réalisation des calculs imprécis.

L'effet d'un usage régulier des mathématiques est testé en utilisant le test des rangs sommés de Wilcoxon en distinguant deux groupes : les participants rapportant un usage régulier des mathématiques et les participants rapportant ne pas les utiliser régulièrement. Une comparaison entre les deux groupes est réalisée pour chaque borne de chaque intervalle correspondant à chaque calcul imprécis

De même, nous testons l'effet du niveau subjectif en calcul mental en utilisant des tests des rangs sommés de Wilcoxon. Pour ce faire, dans un premier temps, deux groupes sont créés, selon le niveau en calcul mental rapporté. Les participants rapportant un score compris entre 1 et 3 inclus forment le groupe *faible niveau* tandis que ceux rapportant un niveau de 4 ou 5 composent le groupe *haut niveau*.

Enfin, nous examinons si l'ordre dans lequel les participants remplissent les deux parties du questionnaire a un effet sur les intervalles donnés comme résultats aux calculs imprécis. Pour ce faire, deux groupes sont créés, le premier remplissant d'abord la partie dédiée aux

calculs et le second, celle dédiée aux approximations des opérandes et résultats exacts. Comme pour le niveau en calcul mental, des tests des rangs sommés de Wilcoxon sont réalisés, pour chaque borne de chaque calcul.

8.4.2 Analyse dans la perspective sémantique du vague : sous-ensembles flous

La seconde approche d'analyse des données que nous proposons s'inscrit dans la perspective sémantique du vague et repose sur la comparaison de sous-ensembles flous construits à partir des intervalles collectés. Dans la première partie de cette section, nous décrivons les deux manières dont ces constructions sont réalisées : par élicitation et par application de l'arithmétique floue. Dans un second temps, nous définissons le critère quantitatif de comparaison utilisé.

8.4.2.1 Construction des sous-ensembles flous

Nous détaillons dans cette section les méthode mises en œuvre pour construire les quatre fonctions d'appartenance considérées dans cette étude (voir tableau 8.2 p. 129). La première partie décrit la méthode employée pour éliciter les fonctions d'appartenance μP et μC_z à partir des intervalles collectés. L'addition et le produit flous qui définissent μA_z sont présentés dans la seconde partie. La construction de μM_z est réalisée selon le modèle flou d'interprétation des ENA fRKM, détaillé dans la section 5.3 p. 80.

Calcul de μP et μC : élicitation à partir d'intervalles Parmi les interprétations classiques des fonctions d'appartenance (Bilgiç & Türkşen, 2000), la perspective *random set* apparaît comme la plus pertinente vis-à-vis de la manière dont les données sont collectées dans cette étude (voir section 3.1.2 p. 24).

Par conséquent, dans notre étude, le degré d'appartenance d'une valeur candidate x à la plage de valeurs est défini comme la fréquence relative cumulée de participants donnant un intervalle contenant x. Ainsi, si la moitié des participants incluent 595 dans "*environ 20*" · "*environ 30*", le degré d'appartenance de 595 est 0,5.

La figure 8.1 illustre les fonctions d'appartenance élicitées pour le calcul "*environ 200*" + "*environ 400*" : μP_{200} et μP_{400}, correspondant aux opérandes, et μC_{600}, correspondant au résultat du calcul imprécis.

Calcul de μA : addition et multiplication floues L'arithmétique floue est un domaine de la théorie des sous-ensembles flous qui étend le calcul pour les nombres imprécis (Hanss, 2005), de la même façon que l'arithmétique des intervalles pour des plages de valeurs.

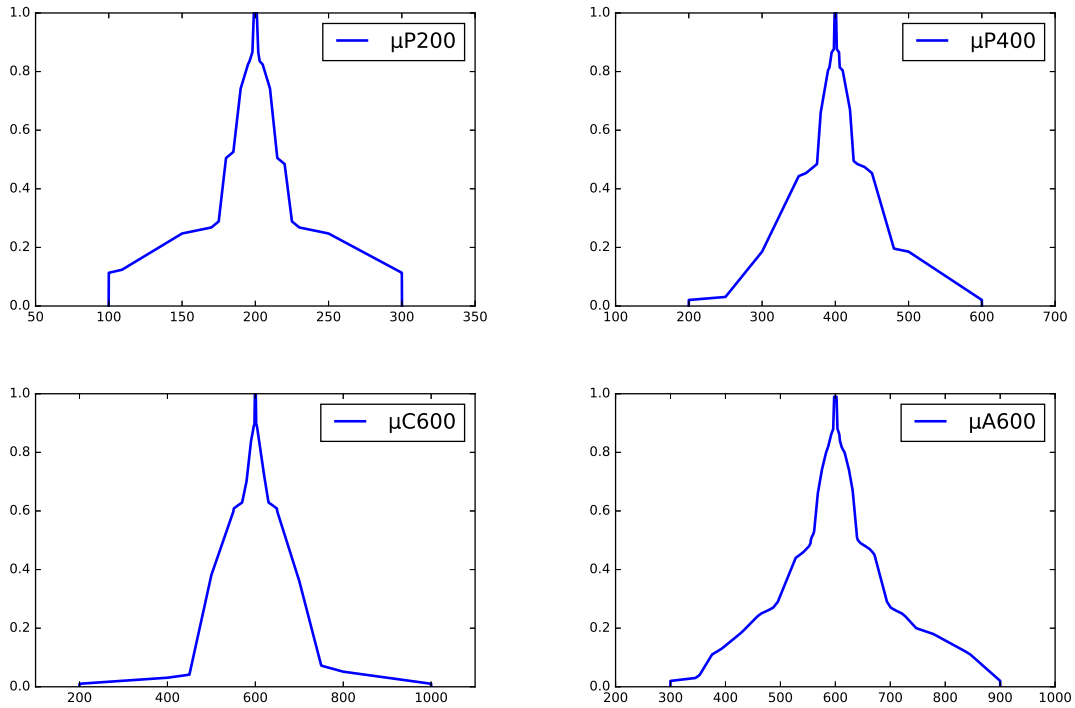


FIGURE 8.1 – Fonctions d'appartenance construites pour le calcul imprécis "*environ 200*" + "*environ 400*" : les fonctions d'appartenance élicitées à partir des intervalles collectés pour les opérandes μP_{200} (en haut, à gauche) et μP_{400} (en haut, à droite) et le résultat du calcul imprécis, μC_{600} (en bas, à gauche) ; et μA_{600} (en bas, à droite), obtenu par application de l'addition floue à μP_{200} et μP_{400}.

Deux approches principales peuvent être considérées pour réaliser des opérations arithmétiques sur des sous-ensembles flous. La première est basée sur le principe d'extension de Zadeh (Yager, 1986), la seconde repose sur l'utilisation de la représentation en $\alpha-$coupes des fonctions d'appartenance (Kauffman & Gupta, 1991). C'est cette dernière que nous utilisons pour calculer les sous-ensembles flous μA_z. Le résultat est obtenu en trois étapes.

Dans un premier temps, les fonctions d'appartenance μP_x et μP_y, correspondant aux opérandes "*environ x*" et "*environ y*" respectivement, sont décomposées en 100 α-coupes $[\mu P_x]_\alpha = [s,t]_\alpha$ et $[\mu P_y]_\alpha = [u,v]_\alpha$ (en omettant α en indice de s, t, u et v pour alléger la notation).

Ensuite, $[\mu P_x]_\alpha$ et $[\mu P_y]_\alpha$ sont combinées pour obtenir $[\mu A_z]_\alpha$, les α-coupes correspondant au sous-ensemble flou μA_z. Cette combinaison est réalisée selon le calcul des intervalles, présenté dans la section 8.4.1.1 p. 135. Aussi, dans le cas des additions, $[\mu A_z]_\alpha = [s+u, t+v]$ et dans le cas de multiplications, $[\mu A_z]_\alpha = [s \cdot u, t \cdot v]$.

Comme pour l'arithmétique des intervalles, la propagation des imprécisions est qua-

dratique dans le cas des multiplications floues et linéaire dans le cas des additions floues.

Enfin, la fonction d'appartenance μA_z est obtenue en interpolant linéairement les bornes de ses α-coupes, $[\mu A_z]_\alpha$. La figure 8.1 (en bas, à droite) illustre le sous-ensemble flou obtenu par application de l'addition floue aux opérandes μP_{200} et μP_{400}.

8.4.2.2 Etude comparative : méthodes

Cette section décrit les comparaisons expérimentales réalisées pour tester les hypothèses H1 et H2 de cette étude [8], qui prédisent respectivement que les participants donnent comme résultat du calcul imprécis l'approximation du résultat exact et que les résultats des calculs ne sont pas conformes à ce qui serait attendu selon l'arithmétique floue (voir section 8.2 p. 129).

Comparaisons réalisées Pour tester l'hypothèse H1, nous comparons μC_z à μP_z ainsi qu'à μM_z.

Si les participants réalisent d'abord le calcul exact avant de produire une approximation, μC_z devrait être proche de μP_z et de μM_z. On peut de plus noter que la comparaison de μC_z avec μM_z permet également d'évaluer dans quelle mesure le modèle fRKM permet d'estimer le résultat d'un calcul imprécis, fournissant ainsi une étude expérimentale complémentaire de celle proposée dans le chapitre 5, p. 83.

Pour tester l'hypothèse H2, selon laquelle les résultats des calculs imprécis diffèrent de ceux attendus selon les formalismes mathématiques, ici l'arithmétique floue, nous comparons μC_z et μA_z.

Nous proposons de réaliser deux types de comparaisons : la première, qualitative et visuelle, porte sur la globalité des deux fonctions d'appartenance considérées, la seconde quantifie la distance entre les deux fonctions d'appartenance. Le paragraphe suivant décrit le critère de qualité α_{Cmp} que nous proposons pour cette évaluation quantitative.

Comparaison numérique Pour comparer deux sous-ensembles flous quelconques, plusieurs mesures ont été proposées dans la littérature, comme les mesures de comparaison et de ressemblance (Rifqi et al., 2000). Les sous-ensembles flous que nous comparons de ce travail présentent la particularité d'être au moins partiellement superposés. En effet, leurs noyaux incluent toujours le résultat exact du calcul, comme illustré dans la figure 8.1. Par conséquent, utiliser les mesures classiques de comparaison amènerait à conclure que

8. Nous nous focalisons dans l'analyse par sous-ensembles flous aux deux premières hypothèses, sans traiter l'hypothèse H3 pour ne pas alourdir la présentation des résultats. Ce choix ne remet pas en cause les conclusions auxquelles nous arrivons. En effet, les deux premières hypothèses, impliquant les comparaisons des fonctions d'appartenance μC_z, μP_z et μA_z sont les plus informatives vis-à-vis notre hypothèse générale.

ceux-ci sont proches et ne rendrait pas compte des différences plus subtiles qui nous intéressent. Nous proposons donc une méthode originale d'évaluation de la proximité de deux sous-ensemble flous.

Plutôt que de produire un unique score tenant compte de l'ensemble de chacune des deux fonctions d'appartenance, nous proposons d'évaluer plusieurs α-coupes. En effet, la différence entre deux fonctions d'appartenance n'est pas nécessairement constante sur l'ensemble des α-coupes. Chaque α-coupe reçoit par conséquent son propre score et, à la fin, un graphique montrant le score de comparaison en fonction d'α est produit.

Ce score est calculé comme la somme normalisée des différences absolues entre les bornes des deux α-coupes comparées :

$$\alpha_{Cmp}(\mu 1_z, \mu 2_z, \alpha) = \frac{|u_1 - u_2| + |v_1 - v_2|}{RelMag(z)} \tag{8.5}$$

où $\mu 1_z$ et $\mu 2_z$ sont les deux fonctions d'appartenance comparées et $[\mu 1_z]_\alpha = [u_1, v_1]$, $[\mu 2_z]_\alpha = [u_2, v_2]$ leurs α-coupes respectives.

La normalisation est requise car le score n'est pas comparable d'un calcul à l'autre : une différence de 10 unités doit être considérée comme plus importante pour le calcul "*environ 20*" · "*environ 30*" que pour le calcul "*environ 500*" + "*environ 1500*", par exemple. Par conséquent, le score doit être normalisé par la valeur de référence du résultat exact z.

Toutefois, considérer la valeur de z elle-même peut conduire à un autre biais. En effet, la même différence doit être considérée comme plus importante pour 4730, par exemple, que pour 4700. Nous proposons donc d'utiliser la magnitude relative $RelMag(z)$ de la valeur de référence comme facteur de normalisation.

Un score α_{Cmp} élevé indique une grande différence. La comparaison de deux sous-ensembles flous est réalisée sur la base de 100 α-coupes ($\alpha \in \{0,01,\ 0,02, \ldots,\ 1\}$). Un score global, obtenu en moyennant les scores des 100 α-coupes est également proposé comme mesure synthétique.

8.5 Résultats des analyses par intervalles

Nous présentons dans cette section les résultats de l'analyse reposant sur les intervalles collectés, décrite dans la section précédente. Dans un premier temps, nous examinons l'effet de l'ordre de passation et du niveau en calcul mental. Ensuite, les résultats portant sur chacune des quatre hypothèses (voir section 8.2 p. 129) sont tour à tour détaillés. Enfin, après ces analyses basées sur des critères numériques, les résultats de l'analyse topologique sont décrits.

A titre d'illustration, le tableau 8.6 présente la médiane des bornes inférieure et supérieure des intervalles fournis par les participants comme résultats aux 23 calculs imprécis

Calcul			Résultat exact	Intervalle médian
environ 20	·	environ 30	600	[550, 650]
environ 20	·	environ 400	8000	[7500, 8500]
environ 20	·	environ 50	1000	[900, 1100]
environ 30	·	environ 50	1500	[1400, 1600]
environ 40	·	environ 50	2000	[1900, 2100]
environ 20	+	environ 30	50	[45, 55]
environ 20	+	environ 80	100	[90, 110]
environ 30	+	environ 4700	4730	[4700, 4750]
environ 40	+	environ 110	150	[140, 160]
environ 40	+	environ 400	440	[420, 458]
environ 50	+	environ 100	150	[140, 160]
environ 100	+	environ 500	600	[550, 650]
environ 150	+	environ 8000	8150	[8087, 5, 8200]
environ 200	+	environ 400	600	[550, 650]
environ 200	+	environ 800	1000	[900, 1100]
environ 400	+	environ 600	1000	[950, 1100]
environ 400	+	environ 1100	1500	[1400, 1600]
environ 440	+	environ 560	1000	[950, 1050]
environ 500	+	environ 1000	1500	[1400, 1600]
environ 500	+	environ 1500	2000	[1900, 2100]
environ 2000	+	environ 6000	8000	[7500, 8500]
environ 10	+	environ 6000	6010	[6000, 6020]
environ 20	+	environ 8000	8020	[8000, 8040]

TABLEAU 8.6 – Intervalles formés par les médianes des bornes fournies par les participants comme résultats aux calculs imprécis.

du questionnaire. Elles sont données comme indications des imprécisions.

8.5.1 Variabilité inter-individuelle : usage des mathématiques, niveau en calcul mental et ordre de passation

Parmi les 118 participants inclus dans les analyses, 70 ont d'abord rempli la partie calculs du questionnaire et 48 ont d'abord rempli la partie approximations. En comparant les intervalles donnés comme résultats des calculs imprécis, une seule différence significative est observée, pour la borne supérieure de "*environ 500*" + "*environ 1500*" ($W = 1338,5; p < 0,01$). Aucune autre différence significative n'est observée (W compris

entre 743 et 1766 ; $p = N.S.$).

Sur les 118 participants, 56 ont rapporté utiliser régulièrement les mathématiques dans leur travail ou dans leurs études alors que 62 ont rapporté ne pas les utiliser régulièrement. Les tests statistiques ne montrent aucune différence significative entre les deux groupes pour aucune borne (W compris entre 606 et 1393,5 ; $p = N.S.$).

Concernant le niveau subjectif en calcul mental, 68 participants sont inclus dans le groupe *faible niveau* et 50 dans le groupe *haut niveau*. Comme pour l'usage régulier des mathématiques, aucune différence significative n'est observée entre les deux groupes, quelle que soit la borne considérée (W compris entre 754,5 et 1467 ; $p = N.S.$).

Par conséquent, pas plus l'ordre dans lequel est rempli le questionnaire qu'un usage régulier des mathématiques ou que le niveau subjectif en calcul mental ne semblent avoir d'effet sur les résultats des calculs imprécis. Les analyses qui suivent portent donc sur l'ensemble des intervalles collectés auprès de tous les participants, sans distinction de niveau en calcul mental, d'usage régulier des mathématiques ou d'ordre de passation.

8.5.2 Résultat du calcul semblable à l'approximation du résultat exact (H1)

Pour tester la première hypothèse, prédisant que les participants ne prennent pas en compte les imprécisions assignées aux opérandes durant le calcul imprécis, nous comparons les intervalles donnés comme résultats à ces calculs et les intervalles fournis par les 118 participants comme approximations des résultats exacts, c'est-à-dire $I_C^p(z)$ et $I_P^p(z)$ (par ex., "*environ 20*" $\cdot$ "*environ 30*" vs. "*environ 600*").

Le tableau 8.7 présente les résultats de ces comparaisons. Trois observations peuvent être faites. Premièrement, les participants ne semblent pas donner strictement les mêmes valeurs de bornes aux intervalles $I_C^p(z)$ et $I_P^p(z)$. En effet, le score d'égalité varie entre 22,6%, pour "*environ 150*" + "*environ 8000*", et 46,3%, pour "*environ 20*" + "*environ 80*", il vaut 34,6% en moyenne.

Deuxièmement, les tests de Wilcoxon montrent des différences significatives pour 7 des 21 calculs considérés. Ces différences significatives sont particulièrement observées dans le cas des multiplications (4 sur 5). Ces résultats indiquent que les distributions des bornes tendent à être similaires pour les additions mais pas pour les multiplications, ce qui est cohérent avec les scores d'égalité observés, qui sont plus faibles pour les multiplications que pour les additions.

Enfin, la médiane des différences entre les valeurs des bornes ne diffère de zéro que pour deux calculs : "*environ 20*" $\cdot$ "*environ 30*" (bornes inférieure et supérieure) et "*environ 200*" + "*environ 400*" (borne supérieure seulement).

	Bornes inférieures			Bornes supérieures		
Calcul	Test de Wilcoxon	Eq	MD	Test de Wilcoxon	Eq	MD
$20 \cdot 30$	$V = 1756,5^*$	26,6%	-7	$V = 1962,5^*$	23,4%	-10
$20 \cdot 400$	$V = 1357,5^*$	32,2%	0	$V = 1202,5^*$	28,8%	0
$20 \cdot 50$	$V = 1337^*$	34,5%	0	$V = 1257,5^*$	31,0%	0
$30 \cdot 50$	$V = 1298,5^*$	34,5%	0	$V = 1201^*$	31,2%	0
$40 \cdot 50$	$V = 847$	31,8%	0	$V = 878,5$	27,0%	0
$20 + 30$	$V = 1123^*$	40,1%	0	$V = 1248,5^*$	38,3%	0
$20 + 80$	$V = 913,5$	46,3%	0	$V = 869,5$	40,0%	0
$30 + 4700$	$V = 1269,5$	30,1%	0	$V = 1345$	22,7%	0
$40 + 110$	$V = 876$	41,6%	0	$V = 898$	39,0%	0
$40 + 400$	$V = 1498^*$	35,3%	0	$V = 1401,5^*$	27,4%	0
$50 + 100$	$V = 1049,5$	37,0%	0	$V = 1116$	32,3%	0
$100 + 500$	$V = 1278,5^*$	34,2%	0	$V = 1331,5$	28,4%	0
$150 + 8000$	$V = 1922,5$	32,1%	0	$V = 1730,5$	22,6%	0
$200 + 400$	$V = 1466,5^*$	31,4%	0	$V = 1693^*$	26,8%	-1
$200 + 800$	$V = 788$	42,1%	0	$V = 720$	38,2%	0
$400 + 600$	$V = 986,5$	41,1%	0	$V = 899,5$	38,9%	0
$400 + 1100$	$V = 983,5$	43,2%	0	$V = 762$	39,6%	0
$440 + 560$	$V = 515,5$	43,8%	0	$V = 449,5$	40,9%	0
$500 + 1000$	$V = 957,5$	45,5%	0	$V = 1016$	43,4%	0
$500 + 1500$	$V = 954,5$	36,6%	0	$V = 768$	21,2%	0
$2000 + 6000$	$V = 1496$	31,3%	0	$V = 1358,5$	29,2%	0

TABLEAU 8.7 – Comparaisons entre $I^p_C(z)$ et $I^p_P(z)$ (H1). Pour chaque borne, la première colonne donne les résultats aux tests de Wilcoxon, la seconde la valeur du score d'égalité défini dans l'équation 8.2 p. 137 et la troisième la médiane des différence (voir équation 8.3 p. 138).

* : différence significative observée à $p < 0,01$.

L'ensemble de ces résultats nous invitent à nuancer la première hypothèse. En effet, les participants ne semblent pas donner strictement les mêmes intervalles comme résultats aux calculs imprécis et comme approximations des résultats exacts. Toutefois, comme le montre la médiane des différences entre les valeurs des bornes, il ne semble pas y avoir de tendance claire concernant une inclusion systématique des $I^p_C(z)$ dans les $I^p_P(z)$ ou vice-versa.

	Bornes inférieures			Bornes supérieures		
Calcul	Test de Wilcoxon	Eq	MD	Test de Wilcoxon	Eq	MD
$20 \cdot 30$	$V = 316^*$	7,5%	75,5	$V = 219,5^*$	7,5%	90
$20 \cdot 400$	$V = 447,5^*$	4,6%	700	$V = 270,5^*$	2,3%	1040
$20 \cdot 50$	$V = 501,5^*$	8,9%	90	$V = 727,5^*$	6,0%	110
$30 \cdot 50$	$V = 563^*$	8,6%	136	$V = 429^*$	6,9%	144
$40 \cdot 50$	$V = 536^*$	12,8%	112,5	$V = 331,5^*$	5,4%	165
$20 + 30$	$V = 1418$	16,0%	1	$V = 1314,5$	12,8%	1
$20 + 80$	$V = 2263^*$	13,2%	0	$V = 2437^*$	9,5%	0
$30 + 4700$	$V = 1075,5^*$	5,2%	25	$V = 763^*$	2,1%	35
$40 + 110$	$V = 2232$	16,8%	0	$V = 2021,5$	13,7%	0
$40 + 400$	$V = 1218,5^*$	7,4%	5	$V = 1121,5^*$	4,2%	9
$50 + 100$	$V = 1933$	10,4%	1	$V = 1924$	8,3%	1
$100 + 500$	$V = 2074$	11,6%	1	$V = 1893,5$	7,4%	1
$150 + 8000$	$V = 845^*$	3,8%	109,5	$V = 400^*$	2,8%	172,5
$200 + 400$	$V = 1750,5$	10,8%	1	$V = 1716,5$	8,3%	0
$200 + 800$	$V = 1675,5$	11,2%	0	$V = 1973,5$	9,0%	-3
$400 + 600$	$V = 1960$	13,2%	0	$V = 1796$	12,6%	0
$400 + 1100$	$V = 2530,5$	9,9%	0	$V = 1986,5$	7,3%	1
$440 + 560$	$V = 2270^*$	12,5%	-10	$V = 2597^*$	8,0%	-20
$500 + 1000$	$V = 1662$	12,6%	0	$V = 1696$	10,1%	0
$500 + 1500$	$V = 1845$	12,4%	1	$V = 1882$	10,8%	1
$2000 + 6000$	$V = 2280,5$	12,0%	0	$V = 2005,5$	7,3%	0

TABLEAU 8.8 – Comparaisons entre $I_C^p(z)$ et $I_A^p(z)$ (H2). Pour chaque borne, la première colonne donne les résultats aux tests de Wilcoxon, la seconde la valeur du score d'égalité défini dans l'équation 8.2 p. 137 et la troisième la médiane des différence (voir équation 8.3 p. 138).

* : différence significative observée à $p < 0,01$.

8.5.3 Calculs non conformes à l'arithmétique des intervalles (H2)

Notre seconde hypothèse prédit que les calculs imprécis réalisés par les participants ne correspondent pas à ce qui est attendu selon l'arithmétique des intervalles.

Pour tester cette hypothèse, nous comparons les intervalles $I_C^p(z)$ et $I_A^p(z)$. Le tableau 8.8 présente les résultats de ces comparaisons.

Le score d'égalité entre $I_C^p(z)$ et $I_A^p(z)$ varie entre 2,1%, pour "*environ 30*" + "*environ 4700*", et 16,8%, pour "*environ 40*" + "*environ 110*". En moyenne, sur l'ensemble des calculs, ce score est de 9,1%, indiquant que la manière dont les participants traitent le calcul imprécis ne correspond pas à l'arithmétique des intervalles.

Trois cas peuvent être distingués. Le premier comprend les multiplications, pour lesquelles le score d'égalité est inférieur à la moyenne alors que les distributions des bornes diffèrent significativement pour chaque calcul. De plus la médiane des différences entre les valeurs des bornes montre que les participants tendent à sous-estimer systématiquement l'imprécision du résultat. Cette observation est cohérente avec le fait que l'imprécision est propagée quadratiquement dans la multiplication des intervalles, ce qui tend à biaiser davantage les participants vers une sous-estimation de l'imprécision résultante que dans le cas des additions (voir section 8.4.1.1 p. 135).

Le second cas est composé des additions pour lesquelles l'ordre de grandeur de la granularité du résultat exact est comparable à celui de la granularité des opérandes (par ex., "*environ 200*" + "*environ 800*" ou "*environ 500*" + "*environ 1500*"). Dans ce cas, le score d'égalité est plus élevé que la moyenne et les tests de Wilcoxon ne montrent pas de différence significative entre les distributions de bornes. De plus, la médiane des différences entre ces valeurs est basse. La propagation de l'imprécision étant linéaire dans le cas des additions, il se peut que l'imprécision assignée aux résultats coïncide avec celle attendue selon l'arithmétique des intervalles.

Enfin, le dernier cas porte sur les additions pour lesquelles l'ordre de grandeur de la granularité du résultat exact diffère significativement de celui de la granularité de l'une ou des deux opérandes (par ex., "*environ 150*" + "*environ 8000*" ou "*environ 440*" + "*environ 560*"). Ces calculs donnent lieu au même profil que les multiplications. En effet, on observe des différences significatives dans la distributions des bornes, des scores d'égalité plus bas que la moyenne et des différences de médianes élevées.

Ces résultats appuient notre seconde hypothèse, prédisant que le calcul imprécis réalisé par les participants ne correspond pas à l'arithmétique des intervalles. Plus précisément, les participants tendent à sous-estimer l'imprécision du résultat, en particulier dans le cas des multiplications et des additions pour lesquelles l'ordre de grandeur de la granularité du résultat exact diffère significativement de celui de la granularité des opérandes.

8.5.4 Insensibilité au calcul (H3)

La troisième hypothèse prédit que l'imprécision assignée au résultat d'un calcul imprécis est indépendante des opérandes et de l'opération arithmétique considérées. En effet, si l'imprécision n'est pas prise en compte, les participants doivent donner le même intervalle

	Bornes inférieures			Bornes supérieures		
Comparaison	Test de Wilcoxon	Eq	MD	Test de Wilcoxon	Eq	MD
$50 + 100$ vs. $40 + 110$	$V = 260$	65,5%	0	$V = 230,5$	62,1%	0
$200 + 800$ vs. $440 + 560$	$V = 248,5$	63,2%	0	$V = 273,5$	59,2%	0
$400 + 600$ vs. $440 + 560$	$V = 327$	66,7%	0	$V = 291$	65,4%	0
$400 + 1100$ vs. $500 + 1000$	$V = 322$	60,4%	0	$V = 259$	57,1%	0

TABLEAU 8.9 – Comparaisons entre $I_C^p(z)$ de deux calculs dont le résultat exact est le même et dont la granularité des opérandes diffère (H3.1). Pour chaque borne, la première colonne donne les résultats aux tests de Wilcoxon, la seconde la valeur du score d'égalité défini dans l'équation 8.2 p. 137 et la troisième la médiane des différence (voir équation 8.3 p. 138).
* : différence significative observée à $p < 0,01$.

pour un même résultat exact, quel que soit le calcul y menant.

Pour tester cette hypothèse, nous comparons les résultats de différents calculs dont le résultat exact est le même. Deux cas sont en particulier considérés : (i) lorsque la granularité des opérandes diffère d'un calcul à l'autre et (ii) lorsque l'opération arithmétique est différente.

8.5.4.1 Insensibilité à la granularité (H3.1)

Le premier cas consiste à comparer différents calculs imprécis dont la granularité des opérandes est différente (par ex., "*environ 440*" + "*environ 560*", où $Gran(440) = Gran(560) = 10$ vs. "*environ 400*" + "*environ 600*", où $Gran(400) = Gran(600) = 100$). Quatre comparaisons, dont les résultats sont présentés dans le tableau 8.9, sont considérées.

Le score d'égalité varie entre 57,1% et 66,7%. En moyenne, il est de 62,5%. De plus, les tests de Wilcoxon ne montrent pas de différence significative et la médiane des différences entre valeurs de bornes est égale à zéro dans chaque cas.

Ces résultats nous invitent donc à accepter l'hypothèse H3.1. En effet, les participants tendent à donner les mêmes intervalles comme résultats des deux calculs imprécis de chaque comparaison, ce qui indique que la granularité des opérandes ne semble pas influencer l'imprécision assignée au résultat.

	Bornes inférieures			Bornes supérieures		
Comparaison	Test de Wilcoxon	Eq	MD	Test de Wilcoxon	Eq	MD
$20 \cdot 30$ vs. $200 + 400$	$V = 391$	56,0%	0	$V = 521,5$	50,0%	0
$20 \cdot 50$ vs. $200 + 800$	$V = 532^*$	52,1%	0	$V = 624^*$	46,6%	0
$20 \cdot 50$ vs. $440 + 560$	$V = 629^*$	47,1%	0	$V = 611^*$	42,9%	0
$20 \cdot 400$ vs. $2000 + 6000$	$V = 534^*$	44,8%	0	$V = 561$	40,3%	0
$30 \cdot 50$ vs. $400 + 1100$	$V = 589^*$	54,4%	0	$V = 568^*$	49,4%	0
$30 \cdot 50$ vs. $500 + 1000$	$V = 532,5^*$	56,2%	0	$V = 575^*$	54,3%	0
$40 \cdot 50$ vs. $500 + 1500$	$V = 577,5^*$	44,1%	0	$V = 640^*$	38,2%	0

TABLEAU 8.10 – Comparaisons entre $I_C^p(z)$ de deux calculs dont le résultat exact est le même et dont l'opération diffère (H3.2). Pour chaque borne, la première colonne donne les résultats aux tests de Wilcoxon, la seconde la valeur du score d'égalité défini dans l'équation 8.2 p. 137 et la troisième la médiane des différence (voir équation 8.3 p. 138).
* : différence significative observée à $p < 0,01$.

8.5.4.2 Insensibilité à l'opération (H3.2)

Le second cas de la troisième hypothèse est testé en comparant les résultats de calculs imprécis dont les résultats exacts sont égaux mais qui impliquent des opérations arithmétiques différentes (par ex., "*environ 40*" $\cdot$ "*environ 50*" vs. "*environ 500*" $+$ "*environ 1500*"). Sept comparaisons, dont les résultats sont présentés dans le tableau 8.10, sont considérées.

Le score d'égalité varie entre 38,2% et 56,2%. En moyenne, il est de 48,3%, plus bas que dans le cas de l'insensibilité à la granularité. Les tests de Wilcoxon montrent des différences significatives entre les distributions des bornes pour quatre comparaisons sur les sept réalisées. Toutefois, la médiane des différences entre bornes est égale à zéro dans chaque cas.

Ces résultats appuient l'hypothèse H3.2. En effet, bien que les scores d'égalité soient plus bas que dans le cas de l'insensibilité à la granularité, ils demeurent élevés et la médiane des

	Bornes inférieures			Bornes supérieures		
Comparaison	Médiane des Δ_P	Médiane des Δ_C	Test de Wilcoxon	Médiane des Δ_P	Médiane des Δ_C	Test de Wilcoxon
$10 + 6000$	410	10	$V = 16022^*$	225	10	$V = 16663^*$
$20 + 8000$	400	20	$V = 16022^*$	450	20	$V = 16663^*$

TABLEAU 8.11 – Sensibilité aux paradoxes sorites (H4). Comparaisons entre l'imprécision assignée au résultat, $\Delta_C(z)$, et celle assignée à la grande opérande des calculs, $\Delta_P(y)$.
* : différence significative observée à $p < 0,01$.

différences entre bornes ne montre aucune tendance claire. Le fait que les tests de Wilcoxon s'avèrent significatifs peut être expliqué par une forte variabilité dans les valeurs de bornes lorsque celles-ci diffèrent entre les deux calculs d'une même comparaison.

8.5.5 Sensibilité aux paradoxes sorites (H4)

La quatrième hypothèse prédit que les participants présentent une sensibilité aux paradoxes sorites lors du calcul imprécis : lorsqu'ils additionnent deux opérandes dont l'une est caractérisée par une granularité d'ordre de grandeur beaucoup plus élevée que la seconde (par ex., "*environ 10*" + "*environ 6000*"), ils ne tiennent pas compte de l'imprécision assignée à la grande opérande.

Pour tester cette hypothèse, nous comparons les imprécisions assignées aux résultats de deux additions, "*environ 10*" + "*environ 6000*" et "*environ 20*" + "*environ 8000*", aux imprécisions assignées aux grandes opérandes impliquées, "*environ 6000*" et "*environ 8000*", respectivement. Le tableau 8.11 présente les résultats de ces comparaisons.

Ceux-ci montrent que l'imprécision assignée au résultat du calcul imprécis est nettement inférieure à celle assignée à la grande opérande. Par conséquent, l'imprécision de cette dernière ne semble pas être prise en compte durant le calcul imprécis, validant l'hypothèse H4.

8.5.6 Examen des relations topologiques entre les trois intervalles

Afin de mieux caractériser la manière dont les participants traitent les calculs imprécis, nous analysons topologiquement les relations entre les trois types d'intervalles, $I^p_C(z)$, $I^p_P(z)$ et $I^p_A(z)$.

Comme indiqué dans la section 8.4 p. 135, six combinaisons émergent avec une fréquence relative supérieure à 5%. Le tableau 8.12 présente la fréquence relative ainsi que le nombre d'occurrences observées de chacune d'entre elles, en distinguant les additions

	Fréquence relative	
Combinaison	Multiplications	Additions
C1	19,1% ($N = 97$)	8,6% ($N = 131$)
C2	29,9% ($N = 152$)	7,9% ($N = 121$)
C3	11,8% ($N = 60$)	19,1% ($N = 291$)
C4	26,9% ($N = 137$)	21,0% ($N = 321$)
C5	1,0% ($N = 5$)	8,7% ($N = 132$)
C6	0,8% ($N = 4$)	7,3% ($N = 112$)
Total	89,5% ($N = 455/509$)	72,6% ($N = 1108/1526$)

TABLEAU 8.12 – Fréquence relative et, entre parenthèses, nombre d'occurrences des six combinaisons les plus fréquentes de relations topologiques entre les intervalles $I_C^p(z)$, $I_P^p(z)$ et $I_A^p(z)$. Voir tableau 8.5 p. 140 pour la définition illustrée de ces combinaisons.

des multiplications. On peut noter que ces six combinaisons représentent 89,5% de celles observées dans le cas des multiplications et 72,6% dans le cas des additions. Les 58 autres combinaisons ne sont pas considérées comme significatives car elles sont rares.

Trois observations peuvent être faites. D'abord, C1 et C2 apparaissent plus fréquemment dans le cas des multiplications que dans le cas des additions. Dans ces combinaisons, à la fois $I_C^p(z)$ et $I_P^p(z)$ sont inclus dans $I_A^p(z)$, indiquant que l'imprécision attendue selon l'arithmétique des intervalles est plus importante que celle assignée par les participants. Dans le cas de C1, l'imprécision assignée à $I_C^p(z)$ est plus faible que celle assignée à $I_P^p(z)$ alors que C2 représente le cas inverse. Le fait que ces combinaisons soient plus fréquentes dans le cas des multiplications est cohérent avec l'arithmétique des intervalles. En effet, lors de la multiplication d'intervalles, l'imprécision est quadratiquement propagée. Par conséquent, les participants peuvent plus facilement sous-estimer l'imprécision du résultat.

Deuxièmement, C4 est la combinaison la plus fréquente dans les additions et la seconde plus fréquente dans les multiplications. Dans ce cas, $I_C^p(z)$ et $I_P^p(z)$ sont égaux tandis qu'ils sont tous deux inclus dans $I_A^p(z)$. Le fait que C4 soit la plus fréquente tend à supporter notre hypothèse principale, à savoir que l'imprécision des opérandes n'est pas prise en compte durant le calcul imprécis et que le résultat de celui-ci correspond à l'approximation du résultat exact.

Troisièmement, C3 apparaît plus fréquemment dans les additions que dans les multiplications. Dans ce cas, les participants tendent à surestimer l'imprécision résultant du calcul, comparée à celle attendue selon l'arithmétique des intervalles. Ceci peut être dû aux additions pour lesquelles la granularité des opérandes est plus faible que celle du résultat exact (par ex., "*environ 440*" + "*environ 560*", où $Gran(440) = Gran(560) = 10$ et

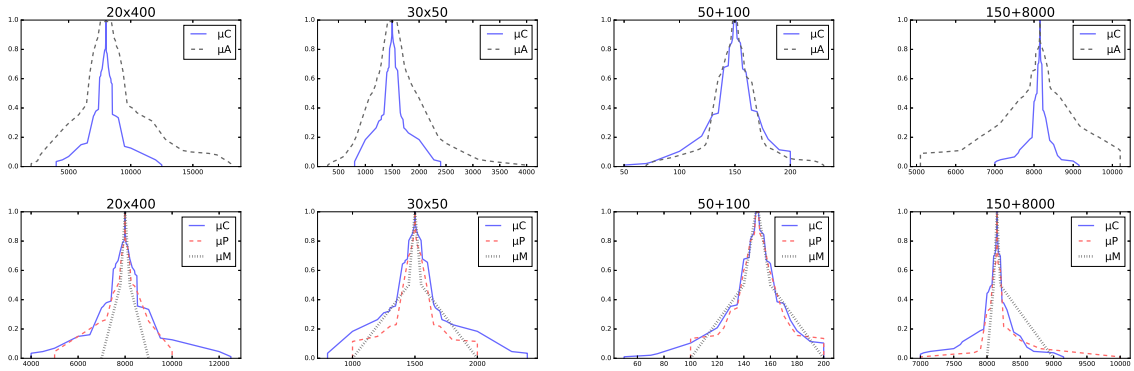


FIGURE 8.2 – Sous-ensembles flous correspondant à quatre calculs. Première ligne : μC_z (ligne pleine) et μA_z (en tirets) ; deuxième ligne : μC_z (ligne pleine), μP_z (en tirets) et μM_z (en pointillés).

$Gran(1000) = 1000$). En effet, dans ce cas, la propagation linéaire des imprécisions selon l'arithmétique des intervalles implique une faible imprécision liée au résultat. La granularité élevée des résultats correspondant peut par conséquent biaiser les participants vers une forte imprécision autour du résultat, biais qui tend à supporter notre hypothèse principale. Cette explication est également valable pour les combinaisons C5 et C6, qui n'apparaissent presque pas dans le cas des multiplications.

Enfin, on peut noter que l'imprécision assignée au résultat est très fréquemment inférieure à celle attendue selon l'arithmétique des intervalles dans les multiplications. En effet, en considérant à la fois C1, C2 et C4, qui représentent ce cas, leur fréquence d'apparition est de 75,9% alors que toutes les combinaisons prises ensemble représentent 89,5% des observations. Si l'on considère les additions, la distribution des combinaisons n'est pas aussi claire que dans le cas des multiplications. Les deux combinaisons les plus fréquentes sont C3 et C4 et représentent 40,1% de l'ensemble des observations. Comme indiqué plus haut, la fréquence élevée de C3 peut être due aux calculs pour lesquels la granularité des opérandes est nettement plus faible que celle du résultat exact.

8.6 Résultats des analyses par sous-ensembles flous

Cette section présente les résultats de l'analyse du calcul imprécis avec la représentation des ENA par sous-ensembles flous. Ceux portant sur les comparaisons qualitatives des fonctions d'appartenance sont détaillés dans un premier temps. En second lieu sont présentés les résultats des analyses quantitatives.

La figure 8.2 illustre les quatre types de sous-ensembles flous considérés : μC_z, μA_z, μP_z et μM_z. Ils correspondent à quatre calculs imprécis : "*environ 20*" · "*environ 400*", "*envi-*

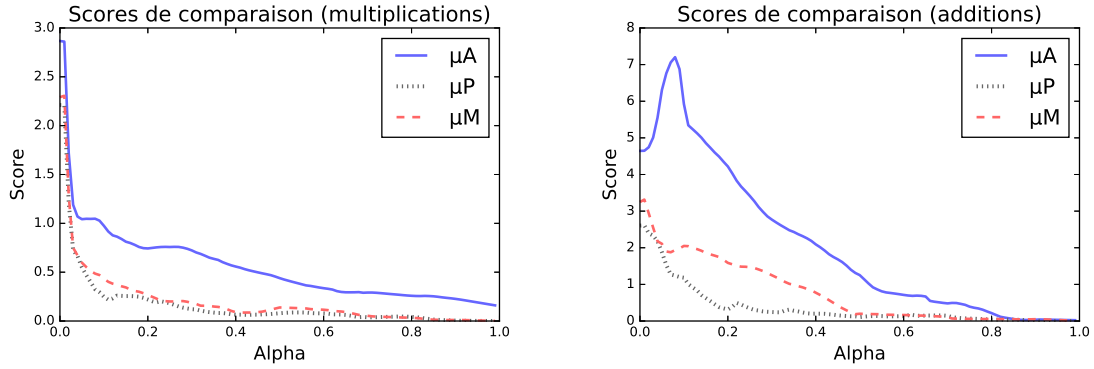


FIGURE 8.3 – Moyennes sur l'ensemble des multiplications (à gauche) et des additions (à droite) des scores de comparaison α_{Cmp} (voir équation 8.5 p. 145) entre μA_z (ligne pleine), μP_z (en pointillés) et μM_z (tirets) et μC_z, en fonction d'α.

ron 30" · "*environ 50*", "*environ 50*" + "*environ 100*" et "*environ 150*" + "*environ 8000*".

Les figures E.1, p. 201 et E.2, p. 202 de l'annexe E illustrent ces fonctions d'appartenance pour les 21 calculs considérés.

8.6.1 Comparaisons qualitatives entre μC_z, μP_z et μM_z (H1)

Pour tester la première hypothèse, selon laquelle les participants donnent un résultat aux calculs imprécis correspondant à l'approximation du résultat exact, nous comparons μC_z à μP_z et μM_z.

On observe, sur la seconde ligne de la figure 8.2 que μC_z tend à être très proche de μP_z, et ce de manière indépendante des opérandes menant au résultat. Par conséquent, il semble que les participants ne tiennent pas compte des imprécisions qu'ils assignent aux opérandes.

Ceci est particulièrement visible dans le cas d'additions comme "*environ 30*" + "*environ 4700*" (voir figures E.1, p. 201 et E.2, p. 202 de l'annexe E) et "*environ 150*" + "*environ 8000*" (voir colonne de droite de la figure 8.2). L'imprécision résultant du calcul (4730 et 8150) est nettement plus faible que celle assignée à la grande opérande (4700 et 8000). L'imprécision de cette dernière ne semble donc pas prise en compte dans le calcul.

Enfin, la second ligne de la figure 8.2 montre également que les sous-ensembles flous générés par le modèle d'interprétation floue des ENA, μM_z, sont proches de ceux élicités à partir des résultats aux calculs imprécis, μC_z.

8.6.2 Comparaisons qualitatives entre μC_z et μA_z (H2)

Pour tester la seconde hypothèse, selon laquelle les résultats des calculs imprécis réalisés par les participants ne sont pas conformes à ce qui est attendu selon l'arithmétique floue,

nous comparons μC_z à μA_z.

L'évaluation visuelle des fonctions d'appartenance montre que les sous-ensembles flous calculés d'après l'arithmétique floue, μA_z, tendent à être plus larges et à surestimer le degré d'appartenance des valeurs, comparé aux fonctions d'appartenance élicitées à partir des calculs des participants, μC_z.

Les résultats obtenus sont similaires à ceux observés lors de l'analyse portant sur les intervalles, décrits dans la section précédente. En effet, nous pouvons distinguer les trois mêmes cas. Premièrement, une imprécision plus importante est particulièrement observée dans le cas des multiplications. Cette observation est cohérente avec le fait que l'imprécision est propagée quadratiquement lors de multiplications alors qu'elle est propagée linéairement dans le cas d'additions.

Deuxièmement, μA_z est proche de μC_z dans le cas d'additions pour lesquelles la granularité de chacune des deux opérandes est comparable. Ce cas est illustré par le calcul "*environ 50*" + "*environ 100*" (voir troisième colonne de la figure 8.2), où $Gran(50) = 10$ et $Gran(100) = 100$.

Les additions pour lesquelles la granularité de l'une des deux opérandes est nettement plus élevée que la granularité de la seconde compose le second cas. Il est illustré par le calcul "*environ 150*" + "*environ 8000*" (voir dernière colonne de la figure 8.2), où $Gran(150) = 10$ et $Gran(8000) = 1000$. Dans ce cas, on observe que l'arithmétique floue produit des imprécisions plus importantes que celles données par les participants, plus importantes encore que dans le cas des multiplications.

Enfin, la surestimation de l'imprécision croît significativement lorsqu'on considère les α-coupes pour lesquelles α est petit, c'est-à-dire lorsqu'on s'approche du support. Cette observation est également cohérente avec la manière dont sont propagées les imprécisions dans l'arithmétique floue. En effet, l'imprécision assignée aux opérandes est plus importante près du support que près du noyau du fait que les intervalles les plus larges donnés par les participants constituent les α-coupes les plus proches du support. Par conséquent, l'imprécision résultante de l'application de l'arithmétique floue est d'autant plus importante que α tend vers 0, résultant en une plus grande surestimation de l'imprécision.

Nous pouvons donc conclure de ces observations que l'arithmétique floue diffère de la manière dont l'être humain propage les imprécisions lors de multiplications et d'additions imprécises.

Sous-ensembles flous comparés à μC_z	Scores de comparaison α_{Cmp}	
	Multiplications	Additions
μA_z	$0,564$	$2,057$
μP_z	$0,170$	$0,378$
μM_z	$0,209$	$0,771$

TABLEAU 8.13 – Scores de comparaison α_{Cmp} moyens obtenus dans le cas des multiplications et des additions.

8.6.3 Comparaisons quantitatives de μC_z avec μA_z, μP_z et μM_z (H1 et H2)

La figure 8.3 illustre les scores de comparaison α_{Cmp} moyens sur l'ensemble des multiplications et des additions, en fonction de α. Le tableau 8.13 présente ces scores moyennés sur α.

Les résultats sont cohérents avec l'analyse qualitative visuelle. En effet, les scores montrent que μP_z et μM_z sont plus proches de μC_z que μA_z.

Ils montrent également que les scores tendent à être plus bas dans le cas des multiplications, particulièrement en ce qui concerne μA_z. Cette observation peut être expliquée par le fait que, contrairement aux additions, les multiplications proposées aux participants ne comprennent pas de calculs dont les opérandes diffèrent grandement en granularité (par ex., "*environ 150*" · "*environ 8000*"). En effet, comme le montrent les analyses visuelles, les additions dont les opérandes sont très différentes en terme de granularité donnent lieu à une surestimation significative de l'imprécision lorsqu'on applique l'arithmétique floue.

De plus, cette observation appuie notre hypothèse principale prédisant que les participants ne prennent pas en compte les imprécisions assignées aux opérandes pour réaliser les calculs imprécis mais produisent une approximation du résultat exact. En effet, si les imprécisions étaient prises en compte, les scores de comparaison devraient être plus bas pour μA_z que pour μP_z et μM_z. Au contraire, on observe que les sous-ensembles flous obtenus par application de l'arithmétique floue, μA_z, sont les plus éloignés de μC_z. De plus, à la fois les sous-ensembles flous générés par le modèle fRKM, μM_z, et ceux élicités à partir des approximations des résultats exacts, μP_z sont proches de μC_z. Ces sous-ensembles flous rendent donc mieux compte des résultats aux calculs imprécis observés que ceux obtenus par application de l'arithmétique floue.

L'ensemble des résultats des analyses qualitatives et quantitatives nous invite à accepter les hypothèses H1 et H2. En effet, μC_z est plus proche de μP_z et μM_z que de μA_z, indiquant

que les participants tendent à donner l'approximation du résultat exact comme résultat du calcul imprécis. De plus, ce dernier n'est pas conforme à ce qui est attendu selon l'arithmétique floue.

8.7 Discussion

Cette étude se veut être une première approche de la compréhension de la manière dont sont propagées les imprécisions lors de tâches arithmétiques imprécises. La littérature portant sur l'arithmétique approximative (voir section 3.3 p. 34) montre que les adultes, les enfants et les animaux ne disposant pas de langage verbal peuvent additionner et soustraire deux tableaux de points (Barth et al. 2006; 2008 ; Gilmore et al., 2010; Knops et al., 2009). Toutefois, ces études ne permettent pas d'évaluer le lien entre l'imprécision des opérandes et celle du résultat. De plus, les opérandes utilisées dans ces travaux sont non-symboliques, c'est-à-dire ne sont pas exprimées via le langage. Or, comme pour les autres domaines du raisonnement humain, des biais peuvent apparaître pendant la combinaison de données numériques imprécises.

Les résultats que nous avons détaillés dans les sections précédentes vont dans le sens de notre hypothèse principale, à savoir que les participants ne tiennent pas compte de l'imprécision assignée aux opérandes lors des calculs imprécis, mais réalisent les calculs exacts avant d'en approximer le résultat. Par exemple, pour "*environ 20*" · "*environ 30*", ils calculent d'abord $20 \cdot 30 = 600$ puis estiment "*environ 600*".

Nous avons mis en lumière cinq éléments appuyant cette hypothèse. Tout d'abord, nous avons montré que les participants ont tendance à assigner la même imprécision à l'approximation du résultat exact qu'au résultat du calcul imprécis, que celle-ci soit représentée par des intervalles ou par des sous-ensembles flous.

Ensuite, nous avons observé que la combinaison des imprécisions assignées aux opérandes n'est pas conforme à l'arithmétique des intervalles ni à l'arithmétique floue. Ces différences entre résultats fournis et résultats attendus sont particulièrement visibles dans le cas des multiplications. Cette observation est cohérente avec l'arithmétique des intervalles et l'arithmétique floue. En effet, dans le cas des multiplications, les imprécisions sont propagées quadratiquement alors qu'elles le sont linéairement dans le cas des additions. Les multiplications imprécises ont par conséquent plus de chance de biaiser les participants si ceux-ci ne tiennent pas compte des imprécisions assignées aux opérandes.

Dans un troisième temps, nous avons montré que l'imprécision résultant d'un calcul imprécis ne dépend ni des opérandes ni de l'opération arithmétique considérées. Les résultats montrent qu'en effet, dans les deux-tiers des comparaisons, les intervalles ne varient pas d'un calcul à l'autre. Cette tendance apparaît toutefois plus importante dans le pre-

mier cas, lorsque la granularité des opérandes diffère, que dans le deuxième cas, lorsqu'une addition et une multiplication sont comparées.

Quatrièmement, nous avons mis en perspective les résultats des trois premières hypothèses, en proposant d'examiner la combinaison des relations topologiques entre les trois intervalles considérés dans cette étude. Nous montrons que la combinaison la plus fréquente est celle qui correspond à notre hypothèse principale : le résultat du calcul imprécis correspond à l'approximation du résultat exact et tous deux sont sous-estimés par rapport à l'arithmétique des intervalles. Concernant les multiplications, nous avons observé que la combinaison la plus fréquente est celle pour laquelle l'imprécision résultant du calcul est plus importante que celle assignée à l'approximation du résultat, alors que toutes deux sont sous-estimées par rapport à l'arithmétique des intervalles. D'autre part, cette combinaison apparaît nettement moins fréquemment dans le cas des additions. Par conséquent, il semble, au moins pour les multiplications, que les participants tendent à ajuster le résultat vers davantage d'imprécision.

Cinquièmement, par l'intermédiaire des paradoxes sorites, nous avons mis clairement en évidence que l'imprécision des opérandes n'est pas prise en compte durant le calcul imprécis. Toutefois, la méthodologie mise en œuvre ne nous permet pas d'évaluer dans quelle mesure l'imprécision du résultat correspond à celle assignée à l'approximation du résultat exact ou à celle assignée à la petite opérande, ces deux dernières n'étant pas demandées dans le questionnaire.

Enfin, grâce à l'analyse dans la perspective sémantique du vague, nous avons montré que les sous-ensembles flous générés par le modèle flou d'interprétation des ENA fRKM que nous avons proposé (voir section 5.3 p. 80) correspondent bien à ceux élicités à partir des intervalles donnés comme résultats des calculs imprécis. Ainsi, notre modèle flou d'interprétation permet également d'estimer le résultat d'un calcul imprécis là où l'arithmétique floue ne semble pas pertinente. D'autre part, ces résultats confortent les résultats de la validation expérimentale de ce modèle, présentés dans la section 5.3.3 p. 83.

L'ensemble des résultats que nous avons obtenus nous invite à proposer le traitement suivant pour rendre compte du calcul imprécis à partir d'opérandes symboliques chez l'être humain. Les individus, plutôt que de combiner les imprécisions dans le calcul imprécis, réalisent le calcul à partir des opérandes exactes, obtiennent ainsi le résultat exact, procèdent à son approximation et, selon les cas, réalisent un ajustement. Par exemple, en considérant le calcul "*environ 30*" + "*environ 4700*", ces trois étapes sont : (i) $30 + 4700 = 4730$; (ii) "*environ 4730*"$= [4720, 4740]$ et (iii) "*environ 30*" + "*environ 4700*"$= [4710, 4750]$.

Du point de vue cognitif, cette hypothèse implique que, comme pour l'interprétation des ENA, les deux sous-systèmes responsables de la cognition des nombres, les systèmes approximatif et exact des nombres sont impliqués dans le calcul imprécis. D'une part, le

système exact prend en charge le calcul exact et, éventuellement l'ajustement de l'imprécision finale. D'autre part, le système approximatif est responsable de la première estimation de l'imprécision résultant du calcul imprécis, à partir du résultat exact.

8.8 Bilan

Dans ce chapitre, nous avons présenté l'étude empirique que nous avons conduite en vue d'étudier le calcul imprécis symbolique tel que réalisé par l'être humain.

Nous avons collecté des données relatives à 5 multiplications et 18 additions impliquant des opérandes imprécises sous la forme d'ENA.

Nous avons proposé deux types d'analyses originales pour traiter les données collectées. Nous avons ainsi montré que les participants tendent à ne pas tenir compte de l'imprécision assignée aux opérandes durant le calcul imprécis mais à donner comme réponse l'approximation du résultat exact.

Chapitre 9

Conclusion et perspectives

> Si vous saviez, lorsque vous commencez à écrire un livre, ce que vous allez dire à la fin, croyez-vous que vous auriez le courage de l'écrire ?
>
> Michel FOUCAULT
> *Dits et écrits*

Nous présentons dans cette conclusion les différentes contributions de cette thèse ainsi que les perspectives qu'elle ouvre, dans les domaines théorique, méthodologique et applicatif.

Contributions

Nos contributions s'articulent autour de deux problématiques concernant les Expressions Numériques Approximatives, respectivement leur interprétation, aussi bien d'un point de vue cognitif que computationnel, et leurs combinaisons dans le calcul imprécis.

Interprétation des ENA

Nos contributions à l'étude de l'interprétation des ENA portent sur les deux aspects traités dans cette thèse : un aspect cognitif, c'est-à-dire la manière dont l'être humain se représente et interprète les ENA, et un aspect computationnel, avec nos propositions de modèles générant des plages de valeurs cohérentes avec l'interprétation humaine.

Aspects cognitifs de l'interprétation des ENA

La première contribution de cette thèse pour l'interprétation des ENA est la définition et la formalisation, à partir de l'état de l'art, des dimensions, à la fois arithmétiques et cognitive, caractérisant la valeur de référence d'une ENA. Les premières comprennent la

magnitude, la granularité et le dernier chiffre significatif. La complexité d'un nombre est la dimension cognitive que nous avons proposée comme contribution originale qui permet d'évaluer la saillance cognitive d'une valeur, c'est-à-dire le fait que certains nombres sont plus facilement évoqués que d'autres par les êtres humains.

Nous avons conduit une première étude empirique, au cours de laquelle nous avons collecté les intervalles correspondant à des ENA non contextualisées. Nous avons ainsi mis en question les modèles existants dans la littérature et montré que les dimensions arithmétiques que nous avons formalisées sont impliquées dans l'interprétation des ENA, à la fois séparément et en les combinant. Par cette même étude, nous avons également montré que l'hypothèse de symétrie, sur laquelle reposent les modèles de la littérature, n'est pas toujours valable et dépend des dimensions arithmétiques de l'ENA considérée.

Dans cette première étude, les ENA présentées aux participants étaient dénuées de contexte alors que celui-ci peut jouer un rôle dans leur interprétation. De plus, les intervalles correspondant aux ENA étaient collectés en interrogeant explicitement les participants sur ceux-ci. Dans la continuité de ces premiers résultats, nous avons examiné l'effet du contexte sur l'interprétation et nous avons étudié si une méthode de recueil implicite des intervalles pouvait donner lieu à une interprétation différente des ENA. Pour ce faire, nous avons conduit une seconde étude empirique de l'interprétation des ENA, pour laquelle nous avons proposé une tâche originale de collecte implicite des intervalles.

Bien que les résultats que nous avons obtenus doivent être pris avec prudence du fait de biais méthodologiques liés à la tâche de collecte implicite, nous avons pu observer que, contrairement à ce qui était attendu d'après la littérature, l'interprétation d'ENA sémantiquement contextualisées diffère peu de l'interprétation d'ENA non contextualisées. De même, nous n'avons observé une faible différence entre interprétations explicite et implicite. Enfin, cette seconde étude nous a permis de reproduire et généraliser les résultats obtenus lors de la première quant à l'implication des dimensions arithmétiques que nous avons proposées dans l'interprétation des ENA.

Finalement, d'un point de vue strictement cognitif, nous proposons de rendre compte de l'implication de la magnitude, de la granularité et du dernier chiffre significatif de la valeur de référence d'une ENA dans son interprétation, par le fait que les deux sous-systèmes cognitifs responsables de la cognition numérique chez l'être humain seraient sollicités : le système approximatif des nombres et le système exact. En effet, la magnitude est la dimensions qui est encodée dans le système approximatif des nombres. La granularité et le dernier chiffre significatif seraient liés aux traitements réalisés par le système exact des nombres en tant que l'interprétation d'une ENA représente un problème formel. Il se pourrait donc que l'interprétation, qu'elle soit implicite ou explicite, d'une ENA repose sur un compromis réalisé entre ces deux sous-systèmes, compromis observé par l'implication

des trois dimensions arithmétiques.

Modélisation computationnelle de l'interprétation des ENA

Le deuxième aspect de nos contributions à l'interprétation des ENA porte sur sa modélisation computationnelle.

Nous avons proposé un principe général fondé sur l'hypothèse que pour l'interprétation d'une ENA, les individus tendent à réaliser un compromis optimal entre la saillance cognitive, formalisée par la mesure de complexité que nous avons proposée, et la distance des bornes de la plage de valeurs dénotées à la valeur de référence de l'ENA.

Nous avons conforté l'hypothèse sous-jacente à ce principe en montrant qu'il permet de rendre compte de la plupart des intervalles recueillis auprès des participants lors des deux études empiriques que nous avons conduites.

Nous avons instancié ce principe dans deux types de modèles d'interprétation des ENA, correspondant respectivement à deux représentations des ENA : par intervalles et par sous-ensembles flous. Le premier type, comprenant les modèles log-linéaire (LLM) et des rangs (RKM), s'inscrit dans la perspective épistémique du vague et consiste à renvoyer un intervalle représentant la plage de valeurs dénotées. Les deux modèles LLM et RKM que nous avons élaborés tiennent compte des trois dimensions arithmétiques dont nous avons montré l'implication dans l'interprétation des ENA. Nous avons validé expérimentalement ces deux modèles en les comparant aux trois modèles d'interprétation identifiés dans la littérature : ils offrent de meilleures performances que ces derniers aux quatre critères de qualité utilisés.

Le second type de modèle que nous avons proposé, fRKM, s'inscrit dans la perspective sémantique du vague : il repose sur le formalisme de la logique floue et construit un intervalle flou représentant l'imprécision associée à une ENA. A l'aide d'une validation expérimentale, nous avons montré que les intervalles flous ainsi générés correspondent bien aux sous-ensembles flous élicités à partir des intervalles collectés lors de l'étude empirique de l'interprétation des ENA.

Nous avons ensuite proposé une méthode pour étendre les modèles LLM, RKM et fRKM de manière à interpréter des ENA impliquant des nombres décimaux. Les résultats illustratifs que nous avons obtenus montrent l'intérêt et la validité formelle de cette méthode. Cependant, une étude empirique serait à mener pour en valider la plausibilité du point de vue cognitif.

Enfin, nous avons implémenté le modèle flou d'interprétation des ENA fRKM dans une application de requêtes flexibles dans des bases de données. Nous avons ainsi concrètement mis en œuvre nos contributions et montré l'intérêt de notre modèle flou d'interprétation

pour se passer de l'étape de définition d'un sous-ensemble flou, qui peut être répétitive et fastidieuse pour l'utilisateur.

Combinaisons des ENA dans le calcul

La seconde problématique que nous avons traitée porte sur la combinaison des ENA dans le calcul chez l'être humain, en particulier dans le cas de multiplications et d'additions. Les contributions à cette problématique se focalisent sur les aspects cognitif et méthodologique.

Nous avons conduit une étude empirique pour collecter les résultats de multiplications et d'additions imprécises. D'un point de vue méthodologique, nous avons proposé deux types d'analyses originales pour les deux perspectives, épistémique et sémantique, du vague, soit à la fois pour les intervalles et pour les sous-ensembles flous.

Nous avons montré que les résultats des calculs tendent à correspondre aux approximations des résultats exacts. De plus, nous avons mis en évidence que les résultats des calculs imprécis ne dépendent pas des opérandes ou des opérations arithmétiques et ne correspondent pas à ce qui est attendu selon l'arithmétique des intervalles ou l'arithmétique floue. Nous avons également montré que les participants présentent une sensibilité aux paradoxes sorites.

L'hypothèse explicative que nous proposons pour rendre compte de l'ensemble de ces résultats est que les participants réalisent le calcul exact, produisent une approximation du résultat exact et ajustent ensuite celle-ci. Du point de vue cognitif, le système exact des nombres prendrait en charge le traitement du calcul exact et la phase d'ajustement de l'imprécision. Le système approximatif des nombres serait, quant à lui, responsable de l'estimation de l'imprécision avant ajustement.

Perspectives

Les différentes contributions de cette thèse ouvrent de multiples perspectives à la fois cognitives et computationnelles. Elles s'inscrivent dans le cadre d'interactions homme-machine basées sur le langage naturel, pour faciliter et fluidifier à la fois l'interaction et l'extraction intelligente d'informations. L'objectif principal est de proposer un ajustement cognitif en s'adaptant au mode de raisonnement de l'utilisateur.

Les perspectives s'articulent selon trois axes, successivement discutés dans les sections suivantes : la contextualisation des modèles d'interprétation, dans ses aspects cognitif, computationnel et lié aux données ; l'interprétation des expressions temporelles ; et la production d'approximations.

Adaptation aux utilisateurs : contextualisation des ENA

Dans cette thèse, nous avons étudié et mis en évidence les dimensions de l'interprétation des ENA qui ne dépendent ni du contexte, ni de l'individu considérés, et nous avons proposé des modèles génériques qui peuvent être instanciés dans différents contextes. Les travaux futurs pourront traiter la question essentielle du contexte, qui peut avoir un effet sur l'interprétation (Lasersohn, 1999; Sadock, 1977; Williamson, 1994), dans ses aspects à la fois cognitif et computationnel.

Aspects cognitifs de la contextualisation des ENA D'un point de vue cognitif, il nous semble intéressant de reproduire et d'approfondir la seconde étude empirique de l'interprétation des ENA que nous avons conduite (voir chapitre 7). Pour ce faire, il nous semble nécessaire de considérer une autre tâche que celle que nous avons proposée pour collecter implicitement les intervalles. Une piste qui nous semble pertinente est de s'appuyer sur une application concrète de requêtes flexibles, telle que ReqFlex (Smits et al., 2013), qui peut présenter trois avantages. Le premier est qu'elle permettrait de collecter, de manière transparente pour l'utilisateur, à la fois les requêtes, sous la forme d'ENA, et les réponses qui lui semblent satisfaisantes. Le deuxième avantage est qu'une telle application permet de contrôler le contexte sémantique via le choix de la base de données considérée, par exemple, une base de voitures d'occasions, ou de salariés. Enfin, le troisième avantage tient à la nature écologique de la tâche : celle-ci serait réalisée au plus près des conditions réelles d'utilisation et non dans un cadre expérimental, qui peut biaiser les participants (Brewer, 2000).

Au delà du contexte sémantique, il nous semble intéressant de prendre en compte et d'étudier le contexte pragmatique de l'interprétation d'une ENA. Dans le cadre des incertitudes, il a été montré que ce dernier, et plus particulièrement le but et l'intention d'un participant, peut avoir un effet sur l'interprétation d'une expression incertaine (Teigen & Brun, 1999). Nous pouvons donc penser que les buts et intentions des participants peuvent avoir également un effet sur l'interprétation d'expressions imprécises et plus particulièrement des ENA. Les travaux futurs devront traiter cette problématique en contrôlant le but des participants en les plaçant dans des rôles différents. Par exemple, dans le cas de la vente de voiture, le rôle d'acheteur devrait susciter le but d'obtenir un véhicule le moins cher possible, alors que le rôle de vendeur devrait susciter le but de vendre un véhicule le plus cher possible.

Aspects computationnels de la contextualisation des ENA D'un point de vue computationnel, deux approches peuvent être considérées pour une adaptation au contexte sémantique. La première consiste à instancier les modèles que nous avons proposés dans

chaque contexte spécifique. Il faut alors reproduire la méthode que nous avons mise en œuvre dans nos travaux : étudier l'interprétation réalisée par des individus dans le contexte considéré et modifier les modèles pour en rendre compte. Cette méthode peut s'avérer rapidement limitée du fait de sa fastidiosité.

La seconde approche, plus flexible, consiste à permettre aux modèles d'interprétation d'apprendre le contexte sémantique. Ainsi, le même modèle pourrait être instancié dans chaque contexte spécifique sans qu'il soit nécessaire de le modifier. Le modèle log-linéaire LLM que nous avons proposé et détaillé dans le chapitre 5, est particulièrement adapté à l'apprentissage du contexte. En effet, il repose sur une régression linéaire qui peut être mise à jour si nécessaire à l'aide de nouvelles données, celles-ci prenant la forme de l'association entre la requête d'une part, sous la forme d'ENA, et les réponses sélectionnées par l'utilisateur d'autre part.

Plusieurs choix sont à réaliser pour mettre au point la méthode d'apprentissage du contexte, en particulier sur deux points. Le premier point concerne la manière de construire les intervalles à partir desquels LLM peut réaliser l'apprentissage, par exemple en considérant les minimum et maximum des valeurs comprises dans les réponses sélectionnées, avec ou sans marge de tolérance. Le second point concerne le type d'apprentissage à considérer : il peut être effectué en temps réel, en mettant à jour les paramètres de la régression à chaque nouvelle donnée, ou consister à recalculer ces paramètres une fois qu'une certaine quantité de nouvelles données est rendue disponible au système. Les travaux futurs devront envisager et évaluer à la fois les pistes évoquées ici et les méthodes issues de l'état de l'art.

Adaptation aux données : contextualisation des modèles

Dans le cadre applicatif des bases de données, un autre type de contextualisation que celui décrit ci-dessus nous semble intéressant à explorer : l'adaptation automatique aux données. Celles-ci peuvent être pertinentes pour guider les modèles d'interprétation de manière à ce que les résultats qu'ils renvoient s'ajustent non seulement à l'utilisateur et au contexte sémantique, mais également à la distribution des données contenues dans la base. En effet, les modèles que nous avons proposés sont basé sur l'hypothèse que la distribution des données est uniforme. Or, en condition réelles, il se peut que ce ne soit pas le cas : la requête "*environ 100*", par exemple, peut n'avoir aucune réponse, ou les données peuvent être denses entre 80 et 100 et moins denses après 100, ce qui ne correspond pas exactement à la plage de valeurs renvoyée par le modèle. Cette problématique est en lien avec celles des réponses vides et pléthoriques (Gaasterland et al., 1992; Bosc et al., 2008; Bosc et al., 2012). La première correspond à l'absence de réponse à une requête, la seconde à une quantité trop importante.

Les travaux futurs devront donc explorer les méthodes permettant d'adapter les bornes renvoyées par les modèles que nous avons proposés à ces informations. Pour ce faire, deux transformations peuvent être en particulier considérées et combinées : la mise à l'échelle, pour élargir ou rétrécir la plage de valeurs dénotées, par exemple $[95, 105]$ ou $[80, 120]$ plutôt que $[90, 110]$ pour "*environ 100*", et la translation, pour décaler la plage de valeurs, par exemple $[85, 105]$ plutôt que $[90, 110]$.

Interprétation des expressions temporelles imprécises

Les modèles d'interprétation que nous avons proposés requièrent que les valeurs de référence des ENA soient exprimées dans le système d'échelles décimal. Or, certaines expressions numériques impliquent d'autres systèmes d'échelles, en particulier les expressions temporelles, par exemple "*environ 2 heures*" ou "*environ 15 minutes*". Outre les aspects contextuels, qui entrent dans le cadre des perspectives exposées ci-dessus, il nous semble pertinent d'envisager d'étendre les modèles que nous avons proposés pour tenir compte de tels systèmes d'échelles.

Pour parvenir à cet objectif, il nous semble nécessaire d'étendre la notion de granularité que nous avons formalisée dans le chapitre 4. En effet, la définition de la granularité sur laquelle reposent les modèles d'interprétation correspond au système d'échelles décimal, pour lequel les coefficients de passage d'un niveau de granularité à l'autre sont constants et valent 10. Dans le cas d'expressions temporelles, ces coefficients entre unités ne sont pas constants (par ex., 1s = 1000ms ; 1min = 60s ; 1 jour = 24h). De plus, au sein de chaque unité, différents niveaux de granularité peuvent être envisagés. Par exemple, pour les minutes, les niveaux 15 et 30, correspondant aux quarts d'heure et aux demi-heures, jouent un rôle particulier.

De même, la mesure de complexité devrait être révisée à la lumière du système d'échelles temporel. En effet, la mesure de complexité telle que nous l'avons définie repose sur l'observation que certains nombres sont cognitivement plus saillants que d'autres. Or, une même valeur numérique dans le système décimal peut être plus ou moins complexe dans un autre système d'échelles (Krifka, 2007). Par exemple, alors que 15 est plus complexe que 10 dans le système décimal, c'est l'inverse qui se produit au niveau des minutes dans le système temporel puisque 15 minutes correspond au quart d'heure, qui est à un niveau de granularité supérieur à celui des dizaines de minutes.

La complexité d'une même valeur peut de plus varier entre deux unités au sein d'un même système. En effet, alors que 15 appartient à un niveau de granularité pertinent pour les minutes, ce n'est pas le cas pour les heures. Au contraire, 12 appartient à un niveau de granularité pertinent pour les heures, car correspondant à une demi-journée, alors que ce

n'est pas le cas pour les minutes.

La composition des unités dans les expressions temporelles soulève un autre problème quant à la définition de la complexité. En effet, par exemple 1h1min et 1h1s sont, d'un point de vue verbal aussi complexes l'un que l'autre. En revanche, la seconde expression est beaucoup plus précise que la première. Nous pouvons donc nous interroger quant à la pertinence à ne considérer que les valeurs numériques incluses dans l'expression pour calculer sa complexité. Une piste pour surmonter cette difficulté serait, à la suite de Krifka (2007), de distinguer deux types de complexité. Le premier est celui que nous avons mis en œuvre dans notre travail et est lié à la représentation de l'expression considérée dans le système d'échelles sélectionné. Dans cette perspective, "*24 heures*" est plus saillante et moins complexe que "*23 heures*" car 24 heures correspondent à la durée d'une journée. Le second type de complexité serait davantage lié à la plus ou moins importante simplicité de l'expression du point de vue verbal, en tenant compte, par exemple, du nombre de syllabes qu'elle comporte (Krifka, 2007). Dans cette perspective, par exemple "*23 heures*" et "*24 heures*" sont équivalents en terme de complexité puisque les deux expressions comportent chacune quatre syllabes. Autrement dit, prononcer "*23 heures*" n'est ni plus long ni plus difficile que prononcer "*24 heures*".

Enfin, une quatrième problématique est soulevée par les expressions temporelles, du fait qu'une même durée peut être exprimée de différentes manières. Ainsi, 2h et 120 minutes sont strictement équivalents du point de vue quantitatif. En revanche, d'un point de vue pragmatique, nous pouvons considérer comme différent de dire "*le film dure environ 2 heures*" et "*le film dure environ 120 minutes*". En effet, nous pouvons nous attendre à ce que l'imprécision associée à la première soit plus importante que celle associée à la seconde.

On voit donc les multiples difficultés qui sont soulevées par une généralisation des notions de granularité et de complexité sur lesquelles reposent les modèles d'interprétation des ENA que nous avons proposés. Les travaux futurs pourront s'attacher à surmonter ces difficultés en formalisant ces deux notions tout en tenant compte des contraintes évoquées ici.

Composition des ENA dans le calcul

Nous avons montré, lors de l'étude empirique du calcul imprécis, que les être humains ne tiennent pas compte de l'imprécision assignée aux opérandes pour estimer l'imprécision du résultat. Cette étude ouvre plusieurs perspectives d'un point de vue cognitif.

D'abord, il nous semble intéressant d'étudier l'effet de la présence d'un contexte sémantique sur le calcul imprécis. En effet, il a été montré (Carraher et al., 1985) que le fait d'instancier un calcul dans un cadre concret améliore les performances des individus.

On peut donc s'attendre à ce qu'un contexte sémantique familier des participants, par exemple l'estimation d'une somme de coûts approchés, conduise les participants à donner des résultats aux calculs imprécis plus proches de ceux attendus selon l'arithmétique des intervalles et l'arithmétique floue.

Ensuite, explorer le cas de calculs imprécis impliquant plus de deux opérandes nous semble intéressant, par exemple pour estimer la durée totale d'un trajet comportant plusieurs segments dont les durées sont elles-mêmes imprécises, ou l'heure de fin estimée d'une après-midi comprenant plusieurs réunions d'"*environ 1 heure*". En effet, il se pourrait que les individus mettent en place des stratégies de compensation, du type : "*il y a des chances pour qu'une réunion soit plus courte et qu'une autre dure plus longtemps*". Dans ce cas, il devrait être observé une nette absence de relation entre les imprécisions de départ et l'imprécision résultante.

Enfin, plus généralement, il nous paraît pertinent d'examiner plus en détails les stratégies de compensation mises en œuvre par les participants lors de calculs imprécis. En effet, l'analyse topologique que nous avons réalisée met en évidence que les participants ne donnent pas exactement le même intervalle comme résultat d'un calcul imprécis et comme approximation du résultat exact correspondant, en particulier dans le cas des multiplications. Une analyse basée sur les distances entres les bornes de ces deux intervalles permettrait de mieux caractériser les compensations réalisées par les participants et éventuellement de faire émerger des profils de participants. Une telle analyse, pour être robuste, nécessite toutefois une grande quantité de données, à la fois en terme de participants et de calculs imprécis proposés.

Production d'Expressions Numériques Approximatives

Enfin, la situation considérée dans cette thèse était celle d'une communication allant de l'utilisateur vers le système. De nombreuses problématiques actuelles concernent le résumé et la présentation de données, dans une situation de communication allant du système vers l'utilisateur. Deux cas peuvent être considérés en lien avec les Expressions Numériques Approximatives : la production d'approximations à partir de valeurs numériques précises et la description d'une distribution de données numériques. Dans ce cadre, et à plus long terme, les principes que nous avons proposés peuvent être mis en œuvre.

En effet, cette problématique recoupe celles qui concernent l'interprétation des Expressions Numériques Approximatives que nous avons traitées dans le cadre de cette thèse et celles que nous avons exposées comme perspectives : le compromis entre saillance cognitive et distance à la valeur de référence, le contexte sémantique, les systèmes d'échelles et le contexte des données.

D'un point de vue pragmatique, produire une approximation consiste à simplifier la valeur numérique précise (Bautista et al. 2013), par exemple "*environ 3,1*" pour 3,141592. La notion de complexité d'un nombre que nous proposons peut dans ce cadre être utilisée pour évaluer dans quelle mesure les valeurs candidates pour être l'approximation répondent à l'exigence de simplicité. De plus, la valeur choisie pour être l'approximation doit communiquer l'information de manière pertinente et donc ne pas être trop éloignée de la valeur précise et ne pas couvrir une plage de valeur trop large. Aussi, le principe du compromis entre saillance cognitive et distance à la valeur de référence, que nous avons proposé, peut être mis en œuvre pour sélectionner les valeurs approchées qui réalisent le meilleur compromis entre ces deux contraintes.

Le choix d'une approximation peut également dépendre du contexte sémantique, et plus particulièrement du domaine considéré (Moxey & Sanford, 1997). En effet, dans le cas de la température corporelle, par exemple, le domaine est compris chez l'être humain entre 36 ° et 43 ° . Dans le cas d'un haut fourneau, le domaine comprend plusieurs milliers de degrés. Aussi, 37,9 ° peut être une bonne approximation de 37,89 ° dans le premier cas, alors que 40 ° peut être acceptable dans le second.

En lien avec la problématique du contexte se pose celle du système d'échelles. En effet, dans le cas d'expressions temporelles, une valeur précise peut être approchée par plusieurs expressions dénotant une quantité identique. Par exemple, à 112 minutes peut correspondre, outre "*environ 110 minutes*", "*environ 120 minutes*" et "*environ 2 heures*".

Enfin, le résumé d'une distribution de données pose le problème de la couverture, c'est-à-dire la taille et la place de la plage de valeurs dénotées, d'une approximation. Dans nos travaux, nous avons proposé des modèles qui estiment les bornes de cette plage. En situation réelle, il est toutefois possible qu'un groupe dense de données soit observé auquel aucune ENA ne correspond, par exemple, un groupe compris entre 70 et 100, qui ne correspond ni à "*environ 80*", *90* ou *100*. Dans ce cas, il faut pouvoir produire une description qui désigne ce groupe et non pas un autre et, par conséquent, envisager la description de manière approchée non pas de valeurs précises mais d'intervalles.

Bibliographie

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*, 832–843.

Alxatib, S., & Pelletier, F. J. (2011). The psychology of vagueness : Borderline cases and contradictions. *Mind & Language*, *26*, 287–326.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational researcher*, *25*, 5–11.

Barth, H., Beckmann, L., & Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children : Abstract addition prior to instruction. *Developmental Psychology*, *44*, 1466–1477.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*, 199–222.

Bautista, S., Hervás, R., Gervás, P., Power, R., & Williams, S. (2013). A system for the simplification of numerical expressions at different levels of understandability. *Natural Language Processing for Improving Textual Accessibility (NLP4ITA 2013)*, 10–19.

Bennett, B. (2010). Spatial vagueness. In R. Jeansoulin, O. Papini, H. Prade and S. Schockaert (Eds.), *Methods for handling imperfect spatial information*, vol. 256 of *Studies in Fuzziness and Soft Computing*, 15–47. Springer Berlin Heidelberg.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*, 545.

Bilgiç, T., & Türkşen, I. B. (2000). Measurement of membership functions : Theoretical and empirical work. In *Fundamentals of fuzzy sets*, vol. 7 of *The Handbooks of Fuzzy Sets Series*, 195–227. Boston : Springer.

Blakemore, D. (1992). *Understanding utterances. an introduction to pragmatics*. Oxford : Blackwell.

Bonini, N., Osherson, D., Viale, R., & Williamson, T. (1999). On the psychology of vague predicates. *Mind and Language*, *14*, 377–393.

Bosc, P., Hadjali, A., & Pivert, O. (2008). Empty versus overabundant answers to flexible relational queries. *Fuzzy Sets and Systems*, *159*, 1450–1467.

Bosc, P., Hadjali, A., Pivert, O., & Smits, G. (2012). An approach based on predicate correlation to the reduction of plethoric answer sets. In F. Guillet, G. Ritschard and D. A. Zighed (Eds.), *Advances in knowledge discovery and management*, 213–233. Berlin : Springer.

Bosc, P., & Pivert, O. (1992). Some approaches for relational databases flexible querying. *Journal of Intelligent Information Systems*, *1*, 323–354.

Bosc, P., & Pivert, O. (1993). On the evaluation of simple fuzzy relational queries : Principles and measures. In R. Lowen and M. Roubens (Eds.), *Fuzzy logic : State of the art*, 355–364. Dordrecht : Springer Netherlands.

Bosc, P., & Pivert, O. (1995). Sqlf : a relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems*, *3*, 1–17.

Bosc, P., & Pivert, O. (2000). Sqlf query functionality on top of a regular relational database management system. In O. Pons, M. A. Vila and J. Kacprzyk (Eds.), *Knowledge management in fuzzy databases*, 171–190. Heidelberg : Physica-Verlag HD.

Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis and C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology*, 3–16. New York : Cambridge University Press.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, *3*, 21–29.

Chameau, J.-L., & Santamarina, J. C. (1987). Membership functions i : Comparing methods of measurement. *International Journal of Approximate Reasoning*, *1*, 287–301.

Channell, J. (1980). More on approximations : a reply to wachtel. *Journal of Pragmatics*, *4*, 461–476.

Channell, J. (1994). *Vague language*. Oxford : Oxford University Press.

Clark, D. A. (1990). Verbal uncertainty expressions : A critical review of two decades of research. *Current Psychology*, *9*, 203–235.

Cleeremans, A. (1997). Principles for implicit learning. In *How implicit is implicit learning ?*, 195–234. Oxford : Oxford University Press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all ? rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Coventry, K. R., Cangelosi, A., Newstead, S. N., & Bugmann, D. (2010). Talking about quantities in space : Vague quantifiers, context and similarity. *Language and Cognition*, *2*, 221–241.

Degroot, M. (1970). *Optimal statistical decisions.* New York : McGraw-Hill.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law : a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*, 145–147.

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, *1156*, 232–259.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*, 1–29.

Delboni, T. M., Borges, K. A., Laender, A. H., & Davis, C. A. (2007). Semantic expansion of geographic web queries based on natural language positioning expressions. *Transactions in GIS*, *11*, 377–397.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735–808.

Douven, I., Decock, L., Dietz, R., & Egré, P. (2013). Vagueness : A conceptual spaces approach. *Journal of Philosophical Logic*, *42*, 137–160.

Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems : Theory and applications.* New York : Academic Press.

Dubois, D., & Prade, H. (1988). *Possibility theory.* New York : Plenum.

Dubois, D., & Prade, H. (1989). Fuzzy sets, probability and measurement. *European Journal of Operational Research*, *40*, 135–154.

Dummett, M. (1975). Wang's paradox. *Synthese*, *30*, 301–324.

Egré, P. (2013). Vagueness : why do we believe in tolerance ? *Journal of Philosophical Logic*, *44*, 663–679.

Egré, P., & Klinedist, N. (2011). Vagueness and language use. In P. Egré and N. Klinedist (Eds.), *Vagueness and language use*, 1–21. Palgrave Macmillan.

Ferson, S., O'Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C. C., Sentz, K., & Finkel, A. M. (2015). Natural language of uncertainty : numeric hedge words. *International Journal of Approximate Reasoning*, *57*, 19–39.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*, 265–300.

Fine, T. (1973). *Theories of probability.* New York : Academic Press.

Fisher, P. (2000). Sorites paradox and vague geographies. *Fuzzy Sets and Systems*, *113*, 7–18.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current directions in psychological science*, *12*, 13–18.

Gaasterland, T., Godfrey, P., & Minker, J. (1992). Relaxation as a platform for cooperative answering. *Journal of Intelligent Information Systems*, *1*, 293–321.

Gelman, R., & Butterworth, B. (2005). Number and language : how are they related ? *Trends in Cognitive Sciences*, *9*, 6–10.

Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science*, *306*, 441–443.

Gideon, L. (2012). *Handbook of survey methodology for the social sciences.* New York : Springer.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*, 394–406.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases : The psychology of intuitive judgment.* Cambridge : Cambridge University Press.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition : attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics*, vol. 3, 41–58. New York : Academic Press.

Grice, H. P. (1991). *Studies in the way of words.* Harvard : Harvard University Press.

Groza, V. S. (1968). *A survey of mathematics : Elementary concepts and their historical development.* New York : Holt, Rinehart and Winston.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense" : The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668.

Hanss, M. (2005). *Applied fuzzy arithmetic.* Berlin : Springer.

Hersh, H. M., & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology : General*, *105*, 254–276.

Hisdal, E. (1985). Reconciliation of the yes-no versus grade of membership in human thinking. In M. M. Gupta, A. Kandel, W. Bandler and J. B. Kiszka (Eds.), *Approximate reasoning in expert systems*, 33–46. Amsterdam : North-Holland.

Holyoak, K. J., & Walker, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior*, *15*, 287–299.

Hunink, M. M., Weinstein, M. C., Wittenberg, E., Drummond, M. F., Pliskin, J. S., Wong, J. B., & Glasziou, P. P. (2014). *Decision making in health and medicine : integrating evidence and values*. Cambridge : Cambridge University Press.

Jansen, C., & Pollmann, M. (2001). On round numbers : Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics*, *8*, 187–201.

Jucker, A. H., Smith, S. W., & Lüdge, T. (2003). Interactive aspects of vagueness in conversation. *Journal of Pragmatics*, *35*, 1737–1769.

Kahneman, D. (2003a). Maps of bounded rationality : Psychology for behavioral economics. *The American Economic Review*, *93*, 1449–1475.

Kahneman, D. (2003b). A perspective on judgment and choice : mapping bounded rationality. *American Psychologist*, *58*, 697–720.

Kamp, H. (1975). Two theories of adjectives. In E. Keenan (Ed.), *Formal semantics of natural language*, 123–155. Cambridge : Cambridge University Press.

Kauffman, A., & Gupta, M. M. (1991). *Introduction to fuzzy arithmetic, theory and application*. New York : Van Nostrand Reinhold.

Keefe, R. (2000). *Theories of vagueness*. Cambridge : Cambridge University Press.

Kennedy, C. (2007). Vagueness and grammar : The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1–45.

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation : Evidence from the operational momentum effect. *Attention, Perception, & Psychophysics*, *71*, 803–821.

Krantz, D., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement*, vol. 1. San Diego : Academic Press.

Krifka, M. (2007). Approximate interpretations of number words : A case for strategic communication. In G. Bouma, I. Krämer and Z. Joost (Eds.), *Cognitive foundations of interpretation*, 111–126. Amsterdam : Koninklijke Nederlandse Akademie van Wetenschapen.

Lakoff, G. (1973). Hedges : A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, *2*, 458–508.

Lakoff, G. (1987). *Women, fire, and dangerous things : What categories reveal about the mind*. Chicago : The University of Chicago Press.

Lasersohn, P. (1999). Pragmatic halos. *Language*, *75*, 522–551.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C., & Detyniecki, M. (2016a). How much is "about" ? fuzzy interpretation of approximate numerical expressions. *Proc. of the 16th Int Conf on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'16)* (pp. 226–237). Eindhoven, Netherlands : Springer.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C., & Detyniecki, M. (2016b). Interprétation floue des expressions numériques approximatives. *Proc. of Rencontres Francophones sur la Logique Floue et ses Applications (LFA'16)*. La Rochelle, France.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C., & Detyniecki, M. (2017a). How arithmetically fuzzy are we ? an empirical comparison of human imprecise calculation and fuzzy arithmetic. *FUZZ-IEEE - 2017 IEEE Int Conf on Fuzzy Systems*. Naples, Italy.

Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C., & Detyniecki, M. (2017b). Interpretation of approximate numerical expressions : Computational model and empirical study. *International Journal of Approximate Reasoning*, *82*, 193–209.

Lefort, S., Zibetti, E., Lesot, M.-J., Detyniecki, M., & Tijus, C. (2017c). Dimensions for automatic interpretation of approximate numerical expressions : An empirical study. *Proc. of the 22nd Int Conf on Intelligent User Interfaces* (pp. 107–117). Limassol, Cyprus : ACM.

Lewis, D. (1986). *On the plurality of worlds*. Oxford : Blackwell.

Loch, C. H., DeMeyer, A., & Pich, M. (2011). *Managing the unknown : A new approach to managing high uncertainty and risks in projects*. Hoboken : John Wiley & Sons.

Mendel, J. M. (2007a). Advances in type-2 fuzzy sets and systems. *Information sciences*, *177*, 84–110.

Mendel, J. M. (2007b). Computing with words and its relationships with fuzzistics. *Information Sciences*, *177*, 988–1006.

Moore, R. E. (1966). *Interval analysis*, vol. 4. Englewood Cliffs : Prentice-Hall.

Moxey, L. M., & Sanford, A. J. (1997). Choosing the right quantifier. Usage in the context of communication. In T. Givon (Ed.), *Conversation : Cognitive, communicative, and social perspectives*, 207–231. Amsterdam : John Benjamins.

Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities : A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, *14*, 237–255.

Moyse, G., & Lesot, M.-J. (2016). Linguistic summaries of locally periodic time series. *Fuzzy Sets and Systems*, *285*, 94–117.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science*, *16*, 65–69.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*, 499–503.

Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning*, *22*, 78–98.

Prince, E., Bosk, C., & Frader, J. (1982). On hedging in physician-physician discourse. In di J. Pietro (Ed.), *Linguistics and the professions*, 83–97. Norwood : Ablex.

Rifqi, M., Berger, V., & Bouchon-Meunier, B. (2000). Discrimination power of measures of comparison. *Fuzzy Sets and Systems*, *110*, 189–196.

Rosch, E., & Mervis, C. B. (1975). Family resemblances : Studies in the internal structure of categories. *Cognitive psychology*, *7*, 573–605.

Sadock, J. M. (1977). Truth and approximations. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*.

Sainsbury, R. M. (1989). What is a vague object ? *Analysis*, *49*, 99–103.

Samson, S., Reneke, J. A., & Wiecek, M. M. (2009). A review of different perspectives on uncertainty and risk and an alternative modeling paradigm. *Reliability Engineering & System Safety*, *94*, 558–567.

Sauerland, U., & Stateva, P. (2007). Scalar vs. epistemic vagueness : Evidence from approximators. *Proceedings of SALT*, 228–245.

Sauerland, U., & Stateva, P. (2011). Two types of vagueness. In P. Egré and N. Klinedist (Eds.), *Vagueness and language use*, 121–145. Palgrave Macmillan.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing : behavior, theory, and biological mechanisms. *Cognitive Science*, *27*, 525–559.

Serchuk, P., Hargreaves, I., & Zach, R. (2011). Vagueness, logic and use : Four experimental studies on vagueness. *Mind & Language*, *26*, 540–573.

Shapiro, S. (2006). *Vagueness in context*. Oxford : Oxford University Press.

Smets, P. (1991). Varieties of ignorance and the need for well-founded theories. *Information Sciences*, *57*, 135–144.

Smets, P. (1997). Imperfect information : Imprecision and uncertainty. In A. Motro and P. Smets (Eds.), *Uncertainty management in information systems*, 225–254. Boston : Springer.

Smith, N. J. J. (2008). *Vagueness and degrees of truth.* Oxford : Oxford University Press.

Smits, G., Pivert, O., & Girault, T. (2012). Postgresqlf : Un système d'interrogation floue. *Journées Bases de Données Avancées (BDA'12) - Session démo.* Clermont-Ferrand, France.

Smits, G., Pivert, O., & Girault, T. (2013). Reqflex : Fuzzy queries for everyone. *Proceedings of the VLDB Endowment*, *6*, 1206–1209.

Smits, G., Pivert, O., & Lesot, M.-J. (2014). A vocabulary revision method based on modality splitting. *Proc. of 14th Int Conf on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'14)* (pp. 140–149). Springer.

Solt, S. (2014). An alternative theory of imprecision. *Semantics and Linguistic Theory*, *24*, 514–533.

Sorensen, R. (2001a). *Blindspots.* Oxford : Oxford University Press.

Sorensen, R. (2001b). *Vagueness and contradiction.* Oxford : Oxford University Press.

Straszecka, E. (2006). Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences*, *176*, 3026–3059.

Stubbs, M. (1986). 'a matter of prolonged field work' : notes towards a modal grammar of english. *Applied Linguistics*, *7*, 1–25.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*, 27–38.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions : Effects on decisions, predictions, and probabilistic reasoning. *Organizational behavior and human decision processes*, *80*, 155–190.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty : Heuristics and biases. *Science*, *185*, 1124–1131.

Tyrrell, R., & Owens, D. (1988). A rapid technique to assess the resting states of the eyes and other threshold phenomena : The Modified Binary Search (MOBS). *Behavior Research Methods, Instruments, & Computers*, *20*, 137–141.

Van Fraassen, B. C. (1966). Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy*, *63*, 481–495.

Varzi, A. C. (2007). Supervaluationism and its logics. *Mind*, *116*, 633–676.

Wachtel, T. (1980). Pragmatic approximations. *Journal of Pragmatics*, *4*, 201–211.

Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing : General principles and empirical evidence. *The Knowledge Engineering Review*, *10*, 43–62.

Walsh, V. (2003). A theory of magnitude : common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488.

Wierzbicka, A. (1986). Precision in vagueness : The semantics of english 'approximatives'. *Journal of Pragmatics*, *10*, 597–613.

Williamson, T. (1994). Vagueness. In R. Asher and J. Simpson (Eds.), *The encyclopedia of language and linguistics*, 4869–4871. Oxford : Pergamon Press.

Winer, B., Brown, D., & Michels, M. (1991). *Statistical principles in experimental design.* Boston : McGraw-Hill.

Wright, C. (1976). Language mastery and the sorites paradox. In G. Evans and J. McDowell (Eds.), *Truth and meaning : Essays in semantics*, 223–247. Clarendon Press.

Yager, R. R. (1986). A characterization of the extension principle. *Fuzzy Sets and Systems*, *18*, 205–217.

Yule, G. (1996). *Pragmatics.* Oxford : Oxford University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, *8*, 199–249.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*, 3–28.

Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with Applications*, *9*, 149–184.

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds : The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition*, 331–345. New York : Psychology Press.

Zhang, Q. (1998). Fuzziness-vagueness-generality-ambiguity. *Journal of Pragmatics*, *29*, 13–31.

Annexes

Annexe A

Quelques éléments de la théorie des sous-ensembles flous

Dans la théorie classique des ensembles, un objet appartient ou n'appartient pas à un ensemble : son degré d'appartenance est 0 ou 1. Ainsi, dans le cas des Expressions Numériques Approximatives, par exemple "*environ 100*", une valeur, par exemple 95, est dénotée ou n'est pas dénotée par celle-ci.

Principe général Dans le cadre de la théorie des sous-ensembles flous, introduite par Zadeh en 1965, un objet peut appartenir plus ou moins à un ensemble : son degré d'appartenance prend une valeur réelle dans l'intervalle $[0, 1]$. Appliquée aux ENA, une valeur peut donc être plus ou moins dénotée par celle-ci. Par exemple, 95 peut être dénoté à 0,6 ou 0,7 par "*environ 100*".

Formellement, soit une variable v (par ex., une valeur possiblement dénotée par une ENA) et un univers de référence U (les réels $\mathbb{R}$ dans le cas des ENA). Un sous-ensemble flou A, par exemple le sous-ensemble des valeurs dénotées par l'ENA "*environ x*", est défini par une fonction d'appartenance μ_A qui décrit le degré avec lequel la valeur v appartient à (est dénotée par, dans le cas des ENA) A (Zadeh, 1975). La figure A.1 illustre les différences entre la théorie classique, où le degré d'appartenance d'une valeur v à une ENA "*environ x*" ne peut être que 0 ou 1, et la théorie floue, dans laquelle le degré d'appartenance de v à "*environ x*" prend une valeur réelle dans l'intervalle $[0, 1]$.

Caractérisation de sous-ensembles flous Soit une fonction d'appartenance μ_A décrivant le sous-ensemble flou A défini sur l'univers X, représentant la dénotation d'une ENA "*environ x*". Une α-*coupe* de A est le sous-ensemble classique des valeurs ayant un degré

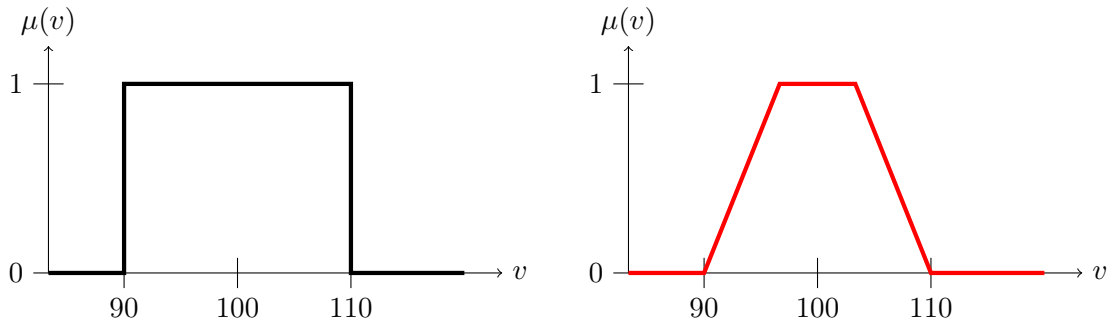


FIGURE A.1 – Comparaison entre ensembles classiques (à gauche) et sous-ensembles flous (à droite), illustrée pour l'ENA "*environ 100*".

d'appartenance supérieur ou égal à α, formellement :

$$\alpha\text{-coupe}(A) = \{v \in \mathbb{X} \mid \mu_A(v) \geq \alpha\}.$$

Deux α-coupes particulières peuvent être définies. Le *noyau* correspond à la 1-coupe, formellement :

$$noy(A) = \{v \in \mathbb{X} \mid \mu_A(v) = 1\}.$$

Le *support* est défini comme l'ensemble des valeurs v dont le degré d'appartenance est non nul, formellement :

$$supp(A) = \{v \in \mathbb{X} \mid \mu_A(v) > 0\}.$$

Annexe B

Validation des dimensions d'interprétation des ENA : résultats détaillés

Dans cette annexe, nous reportons les résultats détaillés de l'étude empirique de l'interprétation des ENA présentée au chapitre 4.

Le tableau B.1 donne les intervalles formés par la moyenne et la médiane des bornes collectées auprès des 146 participants. Les tableaux B.2 à B.7 présentent les résultats obtenus aux tests post-hoc réalisés pour tester l'effet de la magnitude, de la granularité et du dernier chiffre significatif sur l'interprétation des ENA. Enfin, le tableau B.8 donne les taux de symétrie observés pour chacune des 24 ENA considérées dans cette étude.

x	Intervalle moyen	Intervalle médian	Nombre de valeurs différentes par borne (inférieure, supérieure)
20	$[16,2, 24,5]$	$[18, 22]$	9, 10
30	$[24,5, 35,5]$	$[25, 35]$	11, 12
40	$[34,0, 44,8]$	$[35, 45]$	12, 10
50	$[40,9, 59,1]$	$[45, 55]$	14, 12
80	$[71,0, 87,6]$	$[75, 85]$	12, 10
100	$[86,7, 113,4]$	$[90, 110]$	14, 13
110	$[102,3, 123,0]$	$[100, 117]$	16, 13
150	$[131,5, 167,3]$	$[140, 160]$	17, 13
200	$[171,8, 232,9]$	$[185, 210]$	14, 15
400	$[357,8, 439,1]$	$[380, 420]$	16, 15
440	$[419,4, 457,4]$	$[430, 450]$	14, 14
500	$[458,6, 541,1]$	$[470, 530]$	16, 17
560	$[540,6, 577,2]$	$[550, 570]$	14, 13
600	$[552,7, 649,0]$	$[570, 630]$	16, 17
800	$[733,1, 858,5]$	$[750, 850]$	16, 17
1000	$[878,0, 1131,4]$	$[950, 1050]$	20, 18
1100	$[1038,4, 1168,0]$	$[1050, 1150]$	16, 14
1500	$[1351,9, 1643,7]$	$[1400, 1600]$	16, 17
2000	$[1774,3, 2266,6]$	$[1900, 2100]$	19, 20
4700	$[4543,7, 4826,9]$	$[4600, 4800]$	17, 17
4730	$[4632,1, 4794,2]$	$[4712, 4742]$	17, 20
6000	$[5574,8, 6589,8]$	$[5725, 6200]$	19, 19
8000	$[7423,8, 8514,8]$	$[7800, 8200]$	20, 16
8150	$[8021,4, 8301,7]$	$[8100, 8200]$	21, 22

TABLEAU B.1 – Description des intervalles collectés, correspondant aux 24 ENA proposées dans l'étude empirique : intervalles moyens, intervalles médians et nombre de valeurs différentes données par les participants comme bornes des intervalles.

	50	150
150	*	-
8150	*	*

TABLEAU B.2 – Effet de la magnitude : résultats des tests post-hoc pour la comparaison du triplet 50/150/8150. * : $p < 0,01$.

	20	200
200	*	-
2000	*	*

TABLEAU B.3 – Effet de la granularité : résultats des tests post-hoc pour la comparaison du triplet 20/200/2000. * : $p < 0,01$.

	80	800
800	*	-
8000	*	*

TABLEAU B.4 – Effet de la granularité : résultats des tests post-hoc pour la comparaison du triplet 80/800/8000. * : $p < 0,01$.

	20	30	40	50
30	*	-	-	-
40	*	*N.S.*	-	-
50	*	*	*	-
80	*	*	*	*N.S.*

TABLEAU B.5 – Effet du dernier chiffre significatif : résultats des tests post-hoc pour les dizaines. * : $p < 0,01$.

	100	200	400	500	600
200	*	-	-	-	-
400	*	*	-	-	-
500	*	*	*N.S.*	-	-
600	*	*	*	*N.S.*	-
800	*	*	*	*N.S.*	*N.S.*

TABLEAU B.6 – Effet du dernier chiffre significatif : résultats des tests post-hoc pour les centaines. * : $p < 0,01$.

	1000	2000	6000
2000	*	-	-
6000	*	*	-
8000	*	*	*N.S.*

TABLEAU B.7 – Effet du dernier chiffre significatif : résultats des tests post-hoc pour les milliers. * : $p < 0,01$.

Valeur de référence x	20	30	40	50	80	100	110	150
$Sym(x)$ (%)	88,1	82,7	79,7	86,8	75,9	87,8	77,8	82,4
Valeur de référence x	200	400	440	500	560	600	800	1000
$Sym(x)$ (%)	85,0	84,6	61,5	88,1	67,7	85,8	78,6	83,3
Valeur de référence x	1100	1500	2000	4700	4730	6000	8000	8150
$Sym(x)$ (%)	82,0	89,6	82,0	65,7	50,7	80,3	78,4	64,2

TABLEAU B.8 – Scores de symétrie observés pour chacune des 24 ENA proposées dans l'étude empirique.

Annexe C

Modèles d'interprétation des ENA : résultats détaillés

Dans cette annexe, nous présentons les résultats détaillés de l'étude expérimentale des modèles d'interprétation que nous proposons dans le chapitre 5.

Le tableau C.1 présente, pour chacune des 24 ENA, les taux de bornes des intervalles fournis par les participants de l'étude empirique qui sont capturées par les fronts de Pareto.

Nous détaillons ensuite les résultats obtenus par les cinq modèles d'interprétation des ENA comparés sur la base des quatre critères de qualité utilisés. Plus précisément, nous distinguons le cas des ENA dont la valeur de référence ne possède qu'un seul chiffre significatif des autres. Le tableau C.2 présente les résultats détaillés au benchmark Participant alors que le tableau C.3 présente les résultats détaillés relatifs au benchmark ENA.

Enfin, nous présentons les résultats de l'évaluation des fonctions d'appartenance générées par le modèle flou d'interprétation des ENA fRKM. Le tableau C.4 présente les scores obtenus par le modèle pour chaque ENA alors que la figure C.1 illustre à la fois les fonctions d'appartenance générées par le modèle et celles obtenues par élicitation à partir des intervalles collectés.

Valeur de référence	20	30	40	50	80	100	110	150
Taux (%)	88,1	82,0	85,7	77,9	74,1	84,4	86,7	78,7
Valeur de référence	200	400	440	500	560	600	800	1000
Taux (%)	90,2	88,5	83,5	83,3	87,6	86,6	78,6	83,7
Valeur de référence	1100	1500	2000	4700	4730	6000	8000	8150
Taux (%)	91,0	87,0	88,0	89,6	83,6	86,4	80,2	81,3

TABLEAU C.1 – Taux, en pourcentage, de bornes données par les participants aux intervalles correspondant aux 24 ENA de l'étude empirique qui sont capturées par les fronts de Pareto.

Scores au benchmark Participant - Détails				
$NSD = 1$	Prédiction de borne	Distance relative	Prédiction de médiane	Erreur de médiane
Modèle	PA (%)	RD	MA (%)	$MErr$
RM	10,2 (1,4)	0,10 (0,02)	22,2 (9,0)	0,63 (0,14)
SBM	21,5 (3,1)	0,15 (0,01)	34,5 (15,4)	0,70 (0,18)
REGM	7,2 (1,8)	**0,10 (0,02)**	18,6 (7,5)	0,64 (0,19)
LLM	**22,9 (3,1)**	**0,09 (0,02)**	53,6 (11,9)	0,36 (0,13)
RKM	**22,8 (3,1)**	**0,09 (0,02)**	**59,1 (11,9)**	**0,30 (0,13)**
$NSD > 1$	Prédiction de borne	Distance relative	Prédiction de médiane	Erreur de médiane
Modèle	PA (%)	RD	MA (%)	$MErr$
RM	6,7 (1,2)	1,91 (0,11)	11,9 (7,8)	0,97 (0,10)
SBM	16,2 (3,3)	**0,89 (0,22)**	17,3 (14,8)	0,86 (0,22)
REGM	9,2 (3,0)	**0,87 (0,21)**	22,4 (11,1)	0,72 (0,19)
LLM	**21,5 (2,9)**	**0,86 (0,20)**	54,9 (11,3)	**0,40 (0,14)**
RKM	21,0 (2,8)	**0,87 (0,20)**	**57,1 (10,3)**	0,43 (0,13)

TABLEAU C.2 – Moyennes et, entre parenthèses, écart-types des scores obtenus au benchmark Participant par chaque modèle aux quatre critères de qualité. Ces paramètres sont calculés sur deux groupes d'ENA : celles dont la valeur de référence ne possède qu'un seul chiffre significatif (en haut, $NSD = 1$) et celles en possédant plusieurs (en bas, $NSD > 1$). Les scores en gras sont statistiquement les meilleurs, selon les tests ANOVA et post-hoc. Plusieurs scores en gras indiquent une absence de différence significative lors des analyses post-hoc.

Scores au benchmark ENA - Détails				
$NSD = 1$	Prédiction de borne	Distance relative	Prédiction de médiane	Erreur de médiane
Modèle	PA (%)	RD	MA (%)	$MErr$
RM	10,0 (4,2)	**0,10 (0,01)**	32,0 (19,1)	0,54 (0,25)
SBM	21,2 (4,4)	0,15 (0,05)	33,4 (21,4)	0,72 (0,27)
REGM	7,7 (5,2)	**0,10 (0,06)**	17,2 (17,7)	0,59 (0,14)
LLM	**22,6 (4,3)**	**0,09 (0,01)**	64,1 (19,3)	0,22 (0,15)
RKM	**22,7 (4,4)**	**0,09 (0,01)**	**69,4 (17,8)**	**0,17 (0,14)**
$NSD > 1$	Prédiction de borne	Distance relative	Prédiction de médiane	Erreur de médiane
Modèle	PA (%)	RD	MA (%)	$MErr$
RM	6,6 (5,3)	2,03 (1,84)	11,8 (19,8)	0,94 (0,25)
SBM	16,3 (2,5)	**0,92 (0,64)**	11,8 (19,8)	0,91 (0,24)
REGM	8,9 (7,8)	**1,06 (2,20)**	12,8 (21,4)	0,74 (0,25)
LLM	**20,9 (5,6)**	**0,91 (0,62)**	45,7 (30,4)	**0,45 (0,29)**
RKM	**21,1 (6,1)**	**0,90 (0,60)**	**55,2 (30,3)**	**0,41 (0,38)**

TABLEAU C.3 – Moyennes et, entre parenthèses, écart-types des scores obtenus au benchmark ENA par chaque modèle aux quatre critères de qualité. Ces paramètres sont calculés sur deux groupes d'ENA : celles dont la valeur de référence ne possède qu'un seul chiffre significatif (en haut, $NSD = 1$) et celles en possédant plusieurs (en bas, $NSD > 1$). Les scores en gras sont statistiquement les meilleurs, selon les tests ANOVA et post-hoc. Plusieurs scores en gras indiquent une absence de différence significative lors des analyses post-hoc.

Valeur de référence x	20	30	40	50	80	100	110	150
$MFQ(x)$	0,198	0,342	0,185	0,244	0,242	0,517	0,614	0,204
Valeur de référence x	200	400	440	500	560	600	800	1000
$MFQ(x)$	0,266	0,216	0,199	0,250	0,295	0,312	0,283	0,601
Valeur de référence x	1100	1500	2000	4700	4730	6000	8000	8150
$MFQ(x)$	0,951	0,249	0,357	0,378	0,776	0,279	0,298	0,467

TABLEAU C.4 – Scores d'adéquation MFQ entre fonctions d'appartenance générées par le modèle flou d'interprétation des ENA fRKM (voir section 5.3 p. 80) et fonctions d'appartenance élicitées à partir des intervalles collectés, pour chacune des 24 ENA.

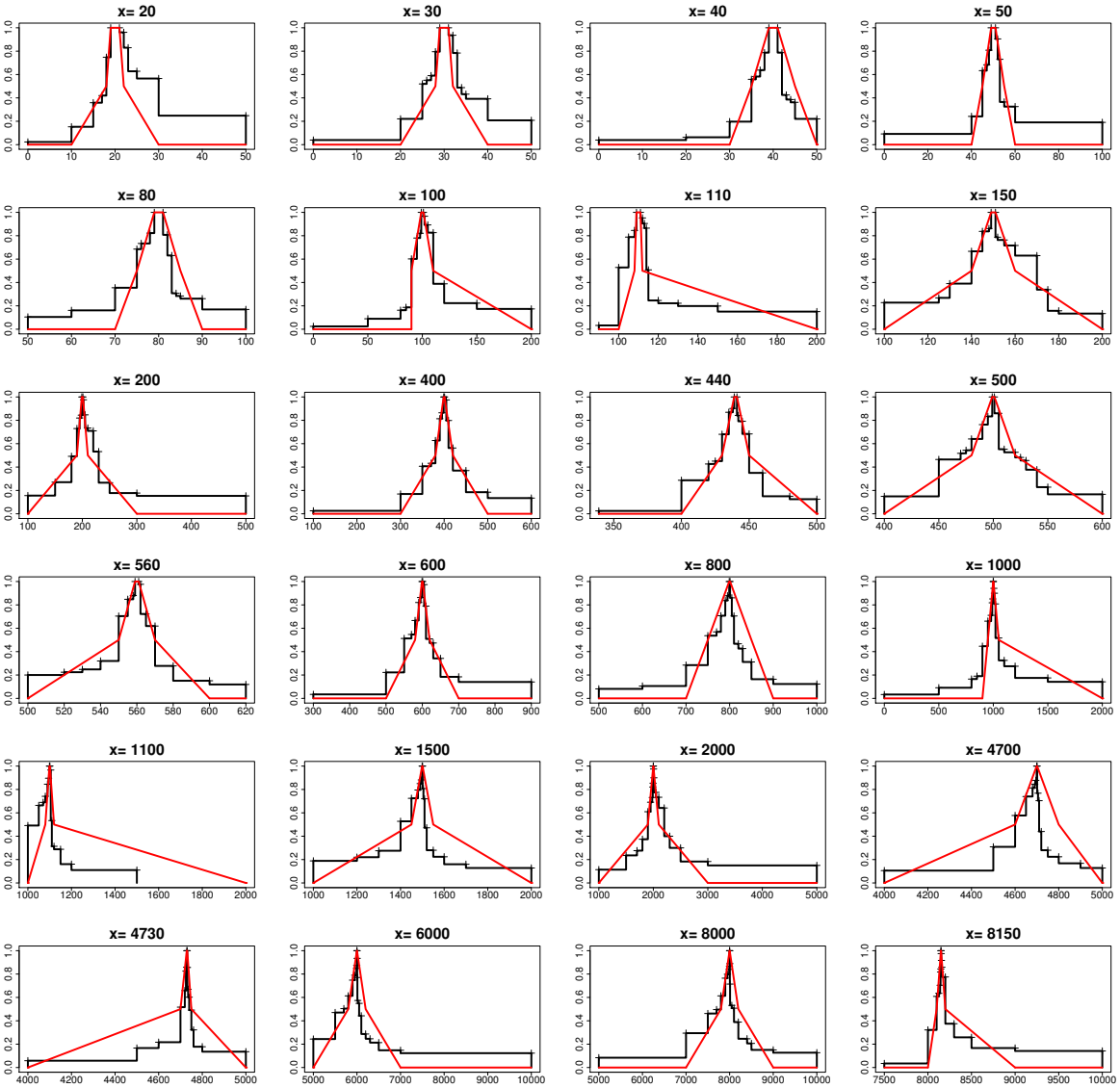


FIGURE C.1 – Fonctions d'appartenance générées par le modèle flou d'interprétation des ENA fRKM (en rouge) et élicitées (en noir) des 24 ENA considérées dans ce travail.

Annexe D

Contextualisation et interprétation implicite des ENA : résultats détaillés

Dans cette annexe, nous illustrons dans un premier temps le déroulement de l'algorithme de recherche par dichotomie, mis en œuvre pour collecter implicitement les intervalles correspondant aux 19 ENA de la seconde étude empirique de l'interprétation des ENA décrite au chapitre 7. Dans un second temps, les statistiques descriptives des intervalles implicites et explicites collectés sont présentées.

D.1 Algorithme de recherche par dichotomie

Le tableau D.1 présente l'exemple d'un déroulement de recherche par dichotomie selon l'algorithme MOBS (Tyrrell & Owens, 1988) pour la borne inférieure de l'ENA "*environ 800*". L'intervalle de départ est $[650, 799]$. La deuxième colonne donne les nombres successivement proposés au participant. Ses réponses sont présentées dans la troisième colonne. Le numéro de la règle qui est appliquée pour générer le nombre proposé d'une ligne est donné dans la quatrième colonne. La dernière colonne, enfin, correspond à l'intervalle dans lequel se trouve la valeur de borne recherchée, mis à jour après que le participant ait répondu.

On peut remarquer que l'algorithme met à jour une borne de l'intervalle de recherche par nombre proposé. A l'étape 5 et à l'étape 8, le même nombre, 746, est affiché. Aux étapes 4 et 5, le participant a répondu oui, déclenchant ainsi la règle 4, selon laquelle après deux réponses successives identiques, le prochain nombre proposé sera l'une des bornes de l'intervalle de recherche. Le nombre suivant est 742, que le participant ne considère pas

dénoté par "*environ 800*". La valeur de la borne doit donc se trouver entre 742 et 746. A l'étape 7, le participant considère que 744 n'est pas dénoté non plus. Ces deux réponses identiques déclenchent à nouveau la règle 4 et le nombre suivant est 746, la borne supérieure de l'intervalle de recherche. Le participant considère cette fois que 746 n'est pas dénoté par "*environ 800*". Par conséquent, l'algorithme retrouve la précédente valeur de la borne supérieure, 751. La recherche par dichotomie se poursuit ensuite, sans incohérence de la part du participant, pour aboutir à l'intervalle $[749, 750]$. La valeur estimée de la borne inférieure de l'intervalle correspondant à "*environ 800*" est calculée comme le centre de cet intervalle, 749,5.

Etape	Nombre proposé	Réponse	Règle appliquée	Nouvel intervalle
1	724	Non	2	$[724, 799]$
2	761	Oui	2	$[724, 761]$
3	742	Non	2	$[742, 761]$
4	751	Oui	2	$[742, 751]$
5	746	Oui	2	$[742, 746]$
6	742	Non	4	$[744, 746]$
7	744	Non	2	$[744, 751]$
8	746	Non	5	$[744, 751]$
9	751	Oui	4	$[744, 751]$
10	747	Non	2	$[747, 751]$
11	749	Non	2	$[749, 751]$
12	751	Oui	4	$[749, 751]$
13	750	Oui	2	$[749, 750]$

TABLEAU D.1 – Exemple d'un déroulement de recherche par partition binaire selon l'algorithme MOBS (Tyrrell & Owens, 1988) de la borne inférieure de l'ENA "*environ 800*". L'intervalle de départ est $[650, 799]$.

D.2 Résultats détaillés

Le tableau B.1 donne les intervalles formés par la moyenne et la médiane des bornes collectées explicitement et implicitement auprès des 37 participants non experts.

ENA	Intervalle explicite moyen	Intervalle explicite médian	Intervalle implicite moyen	Intervalle implicite médian
environ 30	[25, 9, 34, 9]	[25, 35]	[25, 2, 35, 3]	[25, 5, 35, 5]
environ 70	[64, 6, 75, 6]	[65, 75]	[63, 7, 75, 7]	[64, 5, 75, 5]
environ 100	[90, 7, 115, 8]	[90, 110]	[87, 2, 114, 3]	[89, 5, 110, 5]
environ 110	[102, 0, 117, 5]	[104, 115]	[103, 3, 116, 5]	[104, 5, 115, 5]
environ 150	[136, 9, 161, 6]	[140, 160]	[140, 9, 157, 3]	[139, 5, 156, 5]
environ 500	[470, 1, 535, 5]	[475, 535]	[464, 8, 527, 4]	[469, 521, 5]
environ 800	[763, 4, 834, 6]	[775, 830]	[766, 9, 835, 8]	[772, 5, 832, 5
environ 1000	[944, 2, 1073, 3]	[950, 1050]	[925, 3, 1124, 2]	[932, 1037, 5]
environ 1100	[1062, 6, 1149, 8]	[1063, 1150]	[1060, 4, 1146, 0]	[1075, 5, 1148, 5]
environ 1500	[1436, 8, 1550, 6]	[1450, 1550]	[1451, 1, 1553, 6]	[1450, 1549, 5]
environ 4700	[4636, 1, 4760, 1]	[4650, 4750]	[4643, 0, 4751, 9]	[4650, 4739, 5]
environ 4730	[4708, 2, 4749, 1]	[4715, 4749]	[4720, 5, 4739, 6]	[4719, 5, 4740, 5]
environ 8000	[7716, 5, 8178, 2]	[7840, 8060]	[7758, 0, 8218, 0]	[7805, 8117]
environ 8150	[8115, 6, 8197, 0]	[8125, 8175]	[8139, 2, 8159, 7]	[8137, 5, 8161]
environ 30 euros en poche	[26, 3, 34, 0]	[27, 35]	[24, 7, 33, 9]	[25, 33, 5]
hôtel à environ 30km	[24, 9, 35, 6]	[25, 35]	[23, 0, 37, 2]	[24, 38]
environ 500km séparent deux villes	[468, 3, 537, 0]	[470, 550]	[452, 1, 538, 3]	[450, 543, 5]
terrain d'environ 800 hectares	[758, 1, 840, 0]	[750, 850]	[752, 5, 842, 3]	[750, 840, 5]
ville d'environ 1000 habitants	[933, 1, 1154, 0]	[950, 1100]	[880, 8, 1203, 8]	[904, 5, 1190, 5]

TABLEAU D.2 – Description des intervalles implicites et explicites collectés, correspondant aux 19 ENA proposées dans la seconde étude empirique, décrite au chapitre 7 : intervalles moyens, intervalles médians.

Annexe E

Composition des ENA : résultats détaillés

Dans cette annexe, nous rappelons en premier lieu les 13 relations originales proposées par Allen (1983) pour décrire les relations topologiques entre deux intervalles. Le tableau E.1 les illustre.

Dans un second temps, nous détaillons les résultats des analyses réalisées lors de l'étude empirique du calcul imprécis. La figure E.1 illustre les fonctions d'appartenance μC_z et μA_z alors que la figure E.2 illustre les fonctions μC_z, μP_z et μM_z pour 21 des 23 calculs proposés aux participants. Les deux autres calculs, utilisés pour tester l'hypothèse de sensibilité aux paradoxes sorites, ne sont pas représentés car l'approximation du résultat exact n'est pas demandée dans le questionnaire.

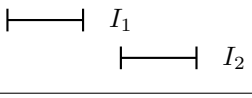
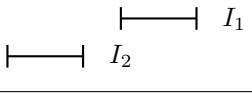
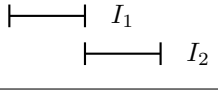
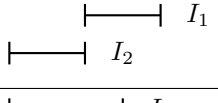
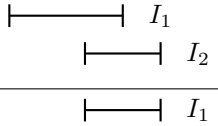
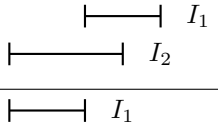
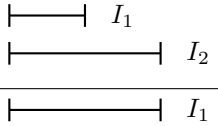
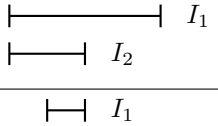
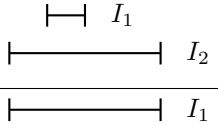
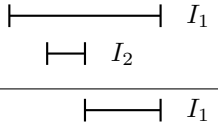
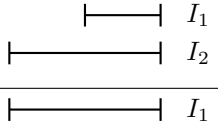
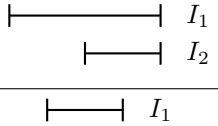
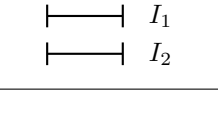
Relation	Illustration
I_1 **before** I_2	
I_1 **before inverse** I_2	
I_1 **meets** I_2	
I_1 **meets inverse** I_2	
I_1 **overlaps** I_2	
I_1 **overlaps inverse** I_2	
I_1 **starts** I_2	
I_1 **starts inverse** I_2	
I_1 **during** I_2	
I_1 **during inverse** I_2	
I_1 **finishes** I_2	
I_1 **finishes inverse** I_2	
I_1 **equals** I_2	

TABLEAU E.1 – Les 13 relations originales d'Allen (1983).

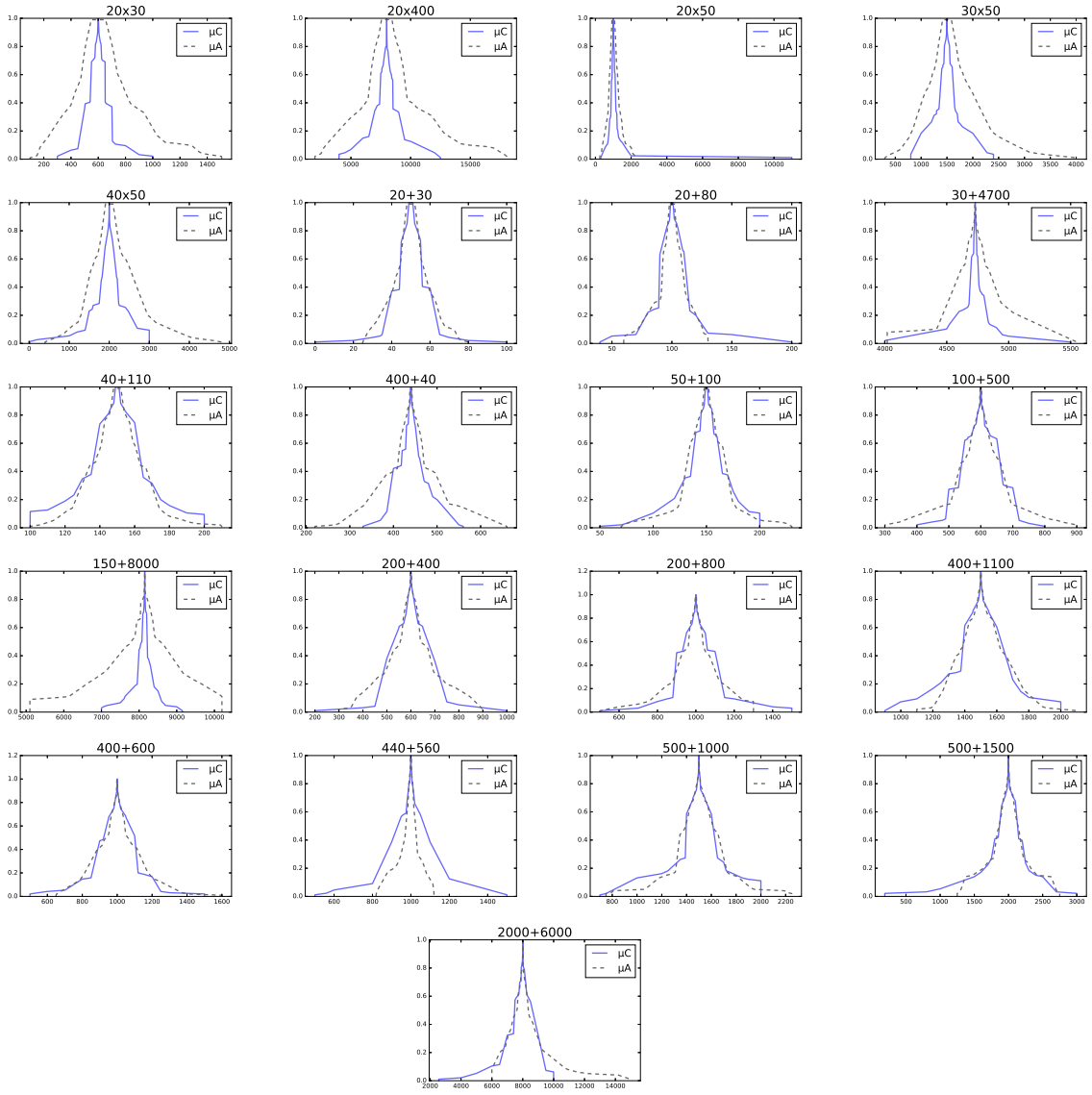


FIGURE E.1 – Comparaisons entre μC_z (ligne pleine) et μA_z (en tirets) pour 21 des 23 calculs proposés aux participants.

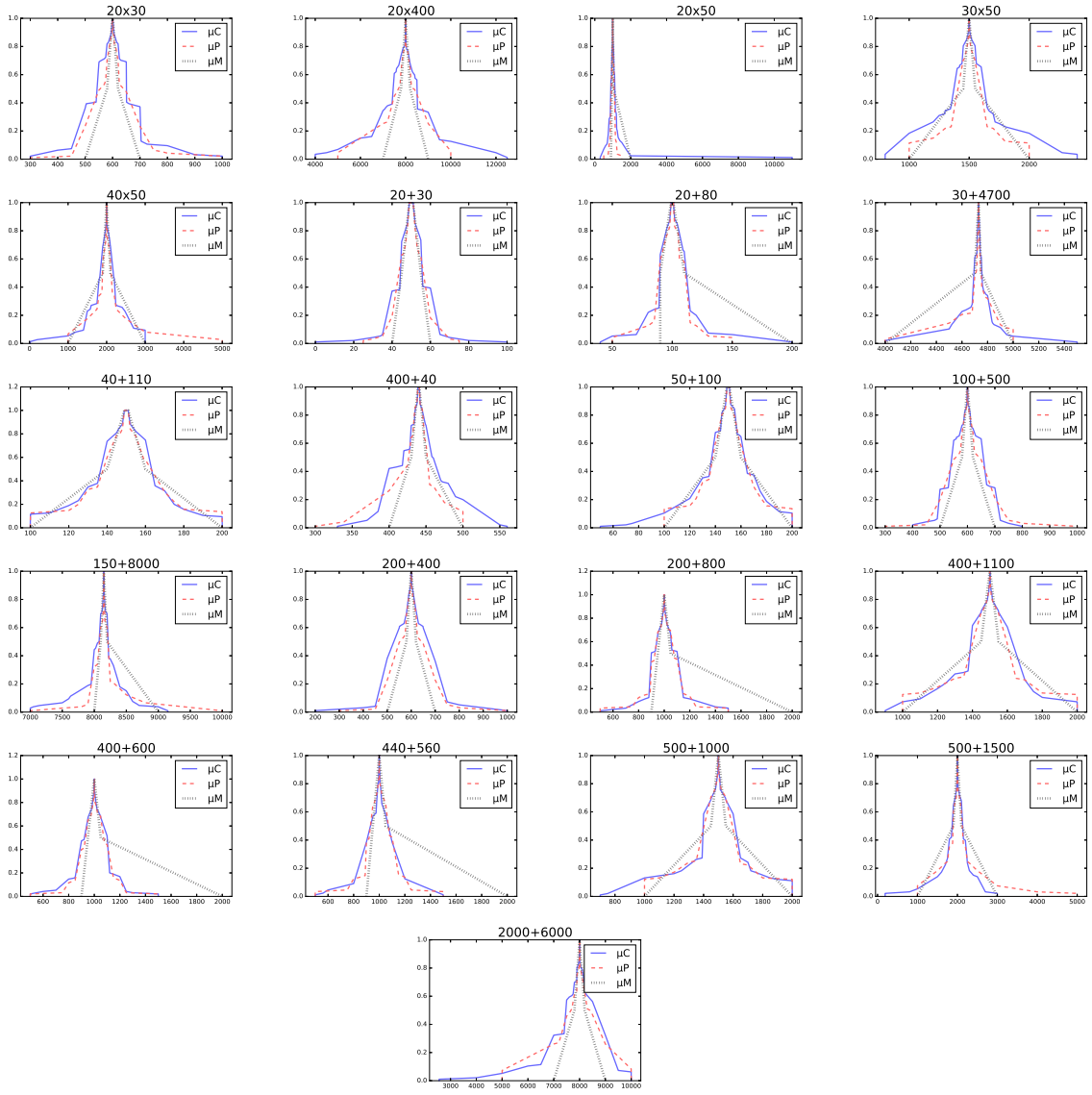


FIGURE E.2 – Comparaisons entre μC_z (ligne pleine), μP_z (en tirets) et μM_z (en pointillés) pour 21 des 23 calculs proposés aux participants.