

**La reconnaissance des mots des langues romanes  
médiévales. Étude sur le français, l'italien,  
l'espagnol et le portugais**

Lassi Luomanen

Mémoire de master

Programme de master : Expertise en langues, Département de français

Institut de langues et de traduction

Faculté des Lettres

Université de Turku

Mars 2023

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

UNIVERSITÉ DE TURKU

Institut de langues et de traduction / Faculté des Lettres

LUOMANEN LASSI : La reconnaissance des mots des langues romanes médiévales. Étude sur le français, l'italien, l'espagnol et le portugais.

Mémoire de master, 54 p. (13 p. d'annexes)

Département de français

Programme de master : Expertise en langues, Département de français

Mars 2023

-----

La présente recherche se focalise sur la reconnaissance des mots dans les langues médiévales (français, espagnol, italien et portugais) apparentées au français moderne, observée à travers la distance d'édition orthographique ainsi que la juxtaposition constituée par l'absence et la présence du contexte. Nous essaions de percevoir s'il y existe des modèles de traduction réguliers et si les résultats obtenus pourraient être appliqués à l'apprentissage des langues étrangères – surtout celles qui appartiennent à la même famille de langue. Afin de développer notre recherche, nous avons abordé sa problématique depuis de nombreuses perspectives : ont été analysées notamment les influences de la langue maternelle, du contexte et des compétences linguistiques en langues romanes. Nous estimions que la présence du contexte, la transparence du mot déterminé et la forte connaissance préexistante dans les langues romanes actuelles étaient tous des facteurs facilitant la reconnaissance de mots.

L'étude contient ainsi trois objectifs primaires liés aux influences des susnommés contexte et transparence (à travers la distance d'édition), mais aussi à celle des compétences linguistiques déjà acquises. A été formée, pour les analyser, une épreuve lexicale comportant 30 mots tirés de quatre épopées nationales sous leurs versions originales (datant des XII<sup>e</sup>-XVI<sup>e</sup> siècles). Cette épreuve est divisée en deux parties : sans et avec contexte. Nous avons eu 30 participants, dont une moitié comprend des locuteurs natifs du français et l'autre moitié des locuteurs natifs du finnois qui étudient le français à présent ou l'ont étudié dans le passé (tous au niveau universitaire). Il y a eu, par ailleurs, une distribution de tous les participants dans plusieurs groupes selon le nombre des langues romanes parlées ou connues : ces compétences sont très variées, mais nos délimitations nous ont permis seulement leur description.

Les résultats démontrent que la distance d'édition affecte la reconnaissance, mais que cette dernière est néanmoins contrôlée par la syntaxe, la morphologie et la lisibilité générale du contexte auquel le mot en question appartient. A été affirmé également que la présence du contexte élève les taux de reconnaissance. L'une des constatations moins attendues est l'uniformité apparente des points obtenus par les francophones et les finnophones, malgré quelques flottements inexplicables : les compétences individuelles jouent donc un rôle très important dans cette étude. La connaissance antérieure d'autres langues romanes a un effet appréciable sur la reconnaissance de mots, ce qui souligne l'importance de l'apprentissage des langues étrangères. En général, les taux d'identification chez les mots transparents étaient assez élevés dans les deux parties et les différences dans la reconnaissance se sont accentuées au fur et à mesure que l'opacité accroissait. Finalement, les mots tirés d'autres langues anciennes ont été reconnus bien plus facilement que ceux de l'ancien français, le fait expliqué en partie par notre choix de mots.

Mots-clés : plurilinguisme réceptif, intercompréhension, distance de Levenshtein, influence du contexte, compétences linguistiques, langues romanes médiévales

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Cadre théorique</b>	<b>3</b>
<b>2.1</b>	<b>La reconnaissance de mots</b>	<b>3</b>
2.1.1	La recherche sur la reconnaissance du vocabulaire	4
2.1.2	Le plurilinguisme traditionnel vs le plurilinguisme réceptif	5
2.1.3	Le rôle de la psycholinguistique dans la reconnaissance des mots	6
<b>2.2</b>	<b>Le décodage des éléments textuels inconnus</b>	<b>8</b>
2.2.1	La compréhension du texte	8
2.2.2	La transparence du lexique	10
<b>2.3</b>	<b>Les points communs entre les langues romanes</b>	<b>11</b>
2.3.1	L'intercompréhension dans les pays romanophones	12
2.3.2	La méthode EuroCom	14
2.3.3	L'effet de la graphie et l'intelligibilité mutuelle	15
<b>2.4</b>	<b>Les spécificités du français par rapport aux autres langues romanes</b>	<b>17</b>
<b>3</b>	<b>Corpus et méthode</b>	<b>18</b>
<b>3.1</b>	<b>Sélection du matériau de l'étude</b>	<b>18</b>
3.1.1	Choix des textes littéraires sources	19
3.1.2	Les mots retenus pour l'épreuve lexicale	19
3.1.2.1	Choix des mots	19
3.1.2.2	Distance de Levenshtein	21
3.1.3	Reformulation du matériau	22
<b>3.2</b>	<b>Déroulement de l'épreuve lexicale</b>	<b>25</b>
3.2.1	Les sujets-cibles de l'épreuve	25
3.2.2	Conception du questionnaire	26
3.2.3	Recueil des réponses	26
3.2.4	Les profils comparés des participants francophones	27
3.2.5	Les profils comparés des participants finnophones	28
<b>3.3</b>	<b>Méthode d'analyse</b>	<b>29</b>
<b>4</b>	<b>Résultats</b>	<b>31</b>
<b>4.1</b>	<b>Bilan</b>	<b>31</b>
<b>4.2</b>	<b>Résultats comparés</b>	<b>34</b>
4.2.1	Les francophones versus les finnophones	34

4.2.2	L'attribution des points donnés et les modèles de traduction	37
4.2.3	La reconnaissance selon les compétences linguistiques : l'espagnol, l'italien et le portugais anciens	41
4.2.4	La reconnaissance selon les compétences linguistiques : l'ancien français	44
4.2.5	Entrevue informelle	48
<b>5</b>	<b>Discussion et conclusion</b>	<b>51</b>
	<b>Références</b>	<b>55</b>
	<b>Annexes</b>	<b>i</b>
	<b>Annexe 1. Questionnaire préliminaire</b>	<b>i</b>
	<b>Annexe 2. Épreuve lexicale (avec réponses)</b>	<b>iii</b>
	<b>Annexe 3. Liste simplifiée des mots du matériau</b>	<b>iv</b>
	<b>Annexe 4. Les regroupements selon les compétences linguistiques</b>	<b>v</b>
	<b>Annexe 5. Résumé en finnois – Suomenkielinen tiivistelmä</b>	<b>vi</b>

# 1 Introduction

Dans ce mémoire, nous étudions la reconnaissance des mots dans quelques langues romanes médiévales. Il s'agit d'une étude qualitative, basée sur un corpus collecté à partir de textes anciens et sur une épreuve lexicale soumise à des locuteurs du français, de niveaux variés. Elle se situe principalement dans le domaine de la lexicologie, mais elle se rapproche également du domaine de la psycholinguistique (Murillo Puyal et al., 2008).

Cette recherche a été inspirée par l'apprentissage général des langues romanes et, surtout, par la constatation des changements historiques qui les rendirent indépendantes : en raison de ces changements, la compréhension des langues ayant la même origine devint plus laborieuse (Blanche-Benveniste, 2008 : 36). Par égard pour tout ce que la présente recherche peut apporter aux études linguistiques contemporaines, nous nous concentrerons sur l'aspect actuel de la reconnaissance du vocabulaire, en faisant appel au contraste entre la présence et l'absence du contexte autour d'un mot individuel. La parenté des langues dans la reconnaissance des mots a été examinée notamment par Bezooijen et Gooskens (2007), McCann et al. (2003), mais aussi Palokangas (2018) dans son mémoire de master. Son œuvre nous a été très utile de plusieurs manières, car elle nous a permis de trouver les concepts essentiels pour rédiger le cadre théorique, tels que le *plurilinguisme réceptif*, l'*intercompréhension* et le *transfert de connaissances*. Grâce à cette étude très approfondie, nous avons découvert les travaux des chercheurs des domaines susmentionnés, ceux qui nous ont guidé dans la création des moyens de mesure et de classification, non seulement pour la constitution du matériau, mais aussi pour son analyse.

Notre étude se focalise sur trois questions de recherche :

- 1) *Dans quelle mesure la présence du contexte aide-t-elle à identifier le sens du mot, selon la langue maternelle du participant ?*
- 2) *Comment les mots de l'ancien français sont reconnus par rapport aux mots venant d'autres anciennes langues romanes, selon leur distance orthographique ?*
- 3) *Comment les compétences linguistiques dans les langues romanes influent sur la reconnaissance des mots ?*

Nos hypothèses sont que le contexte aide les sujets d'expérience à mieux reconnaître les mots et que les mots qui sont graphiquement les plus proches de leur correspondant en français moderne sont

les mieux reconnus, au moins chez ceux qui ne connaissent que le français. Cette possibilité est soutenue par Murillo Puyal et al. (2008 : 359) qui prétendent que les compétences d'intercompréhension pourront être améliorées très sensiblement en intégrant des règles d'équivalence ou de transformation afin de faciliter le passage d'une langue à une autre (cf. 2.2.2). Cela étant dit, nous estimons également qu'une expertise linguistique étendue aide à comprendre les mots plus facilement et que les francophones natifs réussissent mieux dans cet exercice, ce dernier appuyé par Oller et Tullius (1973).

Afin de trouver des réponses à nos questions de recherche, nous avons créé une épreuve lexicale (v. 3.1.) consistant en 30 mots tirés des épopées nationales de quatre pays romanophones (France, Italie, Espagne, Portugal), chacune en version originale, illustrant le langage d'antan. Pour que le pouvoir de la *supposition* (Lutjeharms 1997 ; 2007) ne soit pas trop élevé, nous avons fait le choix de mots en nous appuyant sur toutes les parties du discours en même temps. Comme l'échelle de comparaison, nous avons employé la distance d'édition orthographique, connue également sous le nom « *distance de Levenshtein* » (v. 3.1.2.2). L'épreuve lexicale est constituée de deux parties dont la première expose les mots sans contexte et la deuxième les montre contextualisés. 30 personnes ont pris part à ce test : 15 locuteurs natifs du français et 15 locuteurs natifs du finnois. En plus de la répartition faite selon la langue maternelle des participants, nous les avons regroupés ultérieurement dans quatre groupes, déterminés par leurs compétences linguistiques dans les langues romanes modernes.

Les recherches de Doetjes (2010 : alors en préparation, mais déjà présentées par Zeevaert & ten Thije, 2007) révèlent que le choix de la méthode utilisée pour l'analyse des résultats d'une telle épreuve engendre des conséquences différentes sur le plan de la notation. Sans prendre en considération le niveau de difficulté, les réponses à choix multiple aboutissent à une moyenne élevée de scores puisque les sujets d'expérience peuvent se reposer sur les textes aussi bien que sur l'information donnée implicitement dans le questionnaire. Les questions ouvertes, en revanche, mènent à une moyenne abaissée de scores puisque l'information doit d'abord être ramassée dans les textes eux-mêmes.

Etant donné l'espace réduit dont nous disposons, nous avons dû délimiter notre recherche et cela se manifestera clairement tout au long du mémoire. Entre autres, pour notre épreuve, nous avons été obligé de renoncer, presque complètement, aux questions phonétiques, syntaxiques et didactiques, ainsi qu'à un corpus exhaustif des phénomènes traités.

Nous aborderons premièrement, dans le chapitre 2, les concepts généraux liés à l'intercompréhension, la perception des lettres et de la syntaxe, la transparence du lexique et la parenté des langues romanes. Il faut remarquer tout de suite que, dans cette étude, le terme *intercompréhension* ne renvoie pas à la notion de l'entendement mutuel de plusieurs parties, mais plutôt à celui en sens unique, à savoir la capacité des participants de reconnaître des groupes de lettres semblables ou non à la langue qu'ils connaissent bien, le français. Nous expliciterons cette notion par le moyen du *plurilinguisme réceptif*, introduit en 2.1.2.

Ensuite, dans le chapitre 3, nous verrons plus précisément comment le matériau a été compilé, selon quel classement il sera analysé et quelles difficultés nous avons rencontrées en le construisant. Les éléments discutés dans la section 3.1 nous permettront de réfléchir à comment une épreuve lexicale devrait être construite afin de pourvoir la communauté scientifique de résultats qui donneraient un aperçu aussi précis que possible sur la reconnaissance des mots apparentés. Nous poursuivrons par la présentation des méthodes d'analyse, la description de l'épreuve (3) et de ses résultats (4). Enfin, nous conclurons notre recherche par une discussion où nous parlerons brièvement de nouvelles possibilités d'étude.

## 2 Cadre théorique

Afin de répondre aux questions de recherche, nous devons introduire quelques notions appartenant à ce domaine de la linguistique. Nous tenons compte de la reconnaissance des mots et des structures syntaxiques face au plurilinguisme réceptif, la transparence du lexique, ainsi que les relations partagées par la famille des langues sous inspection, c'est-à-dire, les langues romanes. Pour illustrer l'importance de cette thématique, Zeevaert et ten Thije (2007) relèvent les notions suivantes : le plurilinguisme constitue une dimension sociale importante dans l'histoire des langues européennes, pour une compréhension plurilingue ne sont pas requises les compétences natives, et dernièrement, la position de l'anglais comme *lingua franca* n'est pas la seule solution pour la communication interlinguale en Europe. Être conscient de l'unité culturelle des Européens et le fait de valoriser sa propre langue maternelle donnent la place à la diversification linguistique, en ouvrant la voie vers les compétences linguistiques bien plus étendues (McCann et al., 2003 : 13-14).

### 2.1 La reconnaissance de mots

La compréhension écrite est influencée par de nombreux facteurs (tels que les connaissances de base ou les stratégies de lecture ; Laufer, 1996) mais encore plus évidemment par la connaissance du

vocabulaire, ce qui justifie notre travail et les méthodes que nous avons choisies d'utiliser. En effet, plusieurs chercheurs (Beck, Perfetti & McKeown, 1982 ; Kameenui, Carninen & Freschi, 1982 ; Stahl, 1983 – tous cités par Laufer, 1996) ont constaté qu'une meilleure compréhension écrite est rattachée à l'accroissement de la connaissance du vocabulaire. Dans ce qui suit, nous donnerons quelques prémisses théoriques (2.1.1) pour discuter le plurilinguisme réceptif (2.1.2) et le rôle de la psycholinguistique (2.1.3) ci-après.

### 2.1.1 La recherche sur la reconnaissance du vocabulaire

Une des études sur la reconnaissance du vocabulaire dans des langues étroitement apparentées est celle de Bezooijen et Gooskens (2007), dans laquelle elles comparent la compréhension des textes écrits en frison et en afrikaans, par des sujets néerlandophones. Selon ces chercheuses, une présence substantielle des *cognats* (cf. sa définition en 2.2.2), c'est-à-dire les mots dérivés de la même racine, rend les textes correspondants plus faciles à comprendre. Néanmoins, cette racine mutuellement partagée n'est pas toujours requise pour que la signification d'un mot soit reçue. En effet, il y existe des synonymes qui servent à saisir le sens d'un terme, comme l'expliquent les auteures :

For example, the Dutch word *samenleving* ('society') in the original newspaper article was translated in the Frisian text with *maatskippij*. These two words are non-cognates. Nevertheless, the Dutch reader can easily understand the Frisian word *maatskippij* because of the existence of the Dutch synonym *maatschappij*. (Par exemple, le mot néerlandais *samenleving* ('société') dans l'article original du journal a été traduit par *maatskippij* en frison. Ces deux mots ne sont pas de cognats. Toutefois, le lecteur néerlandais comprend facilement le mot frison *maatskippij* grâce à l'existence du synonyme *maatschappij* en néerlandais. Traduit par nous ; Bezooijen & Gooskens, 2007 : 255).

Un autre exemple, tiré de l'estonien et du finnois – deux langues appartenant à la branche fennique de la famille finno-ougrienne (Frick et al. 2018) –, serait le mot estonien « *uks* » ('porte') qui a un équivalent en finnois, « *uksi*<sup>1</sup> » ; ce mot est pourtant très peu usité de nos jours. Pour revenir à l'étude de Bezooijen et Gooskens, elles précisent que la déduction d'un mot non apparenté est impossible. Qui plus est, l'attitude envers la culture et/ou la langue en question influe sur la motivation de vouloir comprendre cette langue. (Bezooijen & Gooskens, 2007 : 250.)

---

<sup>1</sup> **Uksi** : (déclinaison: *uksen*, *ukseä*, ou rarement; *usta*, *uksessa* etc.) archaïque, soutenu : *porte*. (Le dictionnaire de l'institut des langues, 2021) © 2021 Institut des langues de la Finlande et Kielikone Oy.



Parfois – et surtout dans notre recherche – la notion de l'étymologie acquiert un rôle important. Après tout, ce sont les liens étymologiques qui nous ont orienté dans la création du matériau (v. 3.1.2.1). Bien que Bogaards (1994) ne discute pas autant ce domaine d'étude dans son ouvrage traitant l'apprentissage des langues étrangères, il remarque (p. 14) que les locuteurs non natifs d'une langue seraient généralement moins au courant des sources historiques des mots.

### 2.1.2 Le plurilinguisme traditionnel vs le plurilinguisme réceptif

Le plurilinguisme est souvent considéré comme une notion moderne, lié à la mondialisation et la mobilité de la main-d'œuvre – cela du moins dans les pays européens. Ce que l'on ignore, toutefois, c'est que le monolinguisme comme norme n'existe que dès la création des États-nations pendant le XVIII<sup>e</sup> siècle. En Europe, des dizaines de langues sont quand même présentes et cette réalité est désormais prise en compte par la Commission européenne, ce qui reflète également le slogan officiel de l'UE « *unité dans la diversité* ». (Zeevaert & ten Thije, 2007.) En octobre 2003, même le Président de la République Italienne – Carlo Azeglio Ciampi – a déclaré que « *les langues romanes constituent une partie de l'histoire commune pour les Européens et qu'une didactique des langues conduisant à leur intercompréhension soit divulguée* » (Stegmann, 2005 : 35 ; traduit par nous). De ceci nous pouvons déduire qu'il y existe de différentes approches au concept de plurilinguisme. Comme l'expliquent Zeevaert et ten Thije (2007 : 1) : « *receptive multilingualism refers to the language constellation in which interlocutors use their respective mother tongue while speaking to each other* » (Notre traduction : le plurilinguisme réceptif réfère à la constellation des langues dans laquelle les interlocuteurs utilisent leurs propres langues maternelles pour se parler).

Bien souvent, le plurilinguisme traditionnel – ici renvoyant à l'enseignement d'une langue étrangère – fait référence à l'apprentissage *compréhensif* de celle-ci, et le côté de l'apprentissage *partiel* (ou *réceptif*) – que cela vise uniquement à la lecture, à la compréhension ou à l'usage oral – n'est point considéré dans l'enseignement des langues vivantes. De fait, cette idée a épouvanté la plupart des professeurs de langue. (Blanche-Benveniste, 2008 : 34.) Il est vrai que le plurilinguisme réceptif ne constituait pas alors un domaine établi à l'intérieur des études sur le plurilinguisme. Cependant, les développements économiques et politiques (que l'on appelle effectivement « mondialisation ») signifient un accroissement marquant de la communication internationale, ce qui, à son tour, aboutit à des entraves communicatives ne pouvant être résolues par l'entremise des concepts traditionnels du plurilinguisme. (Zeevaert & ten Thije, 2007.)

### 2.1.3 Le rôle de la psycholinguistique dans la reconnaissance des mots

En ce qui concerne l'analyse du contexte autour des mots individuels, des remarques sont faites par Blanche-Benveniste et Valli (1997 : 36) sur les niveaux linguistiques des lecteurs (v. *infra*). Ils en parlent pour souligner l'utilité d'une grammaire contrastive mais nous tenons pour vrai que les mêmes règles s'appliquent à notre étude aussi. Les grands débutants utilisent toute leur énergie à déchiffrer le lexique et lorsqu'ils voient plus de mots, ils se sentent accablés par la tâche. S'ils sont un peu rassurés, les marquages grammaticaux leur permettent de mieux s'intéresser à la structure de la phrase. Ceux qui connaissent bien le comportement des langues, deviennent des observateurs et c'est à ce stade-là que la curiosité de la comparaison est née. En plus de cela, des connaissances antérieures sont requises à chaque niveau de l'interprétation de l'écrit (Lutjeharms, 2007 : 276). À condition que les formes soient interprétables, le transfert de ces connaissances se fait le plus tôt possible et il peut même être soutenu par la connaissance générale du contenu conjointement aux formes linguistiques reconnaissables (par exemple les cognats ou les anticipations pour l'ordre des mots). D'un autre côté, le transfert des connaissances inadaptées s'avère fallacieux, car une interprétation erronée peut susciter davantage d'erreurs lorsque le lecteur essaie de garder la cohérence (ibid.)

Bien que nous n'ayons pas étudié – spécifiquement – la reconnaissance des faux cognats dans notre épreuve (v. 3.1.3), nous voulons relever la série des erreurs d'interprétation que les mots insidieux sont capables de provoquer. Laufer (1996 : 26-27) emploie le terme *deceptive transparency*, qui est abrégé comme DT et que l'on peut traduire par « transparence trompeuse ». Voici le cheminement en anglais que nous expliquerons plus bas :

Lack of awareness of ignorance of DT words > misinterpretation of DT words > distortion of immediate context > using distorted context for further interpretation > distortion of larger context. (Laufer, 1989 : 16).

Brièvement, tout cela veut dire que l'interprétation erronée d'un faux cognat est capable de causer une réaction en chaîne : le contexte subissant des déformations sémantiques provoque la présence de faux indices qui aboutissent, à leur tour, à d'autres interprétations erronées, jusqu'à défigurer tout le texte. Dans notre étude, le mot *parfunz* ('profonds', qui paraît signifier *parfums* selon plusieurs répondants) pourrait être considéré comme un faux cognat, même s'il a été classifié par nous comme mot opaque (v. nos méthodes de classification en 3.1.2.2). Qui plus est, les cognats – ayant une représentation commune dans le lexique mental (v. *infra*) – profitent d'un processus automatique, tandis que les faux cognats ont besoin d'être traités consciemment afin d'éviter

l'activation du sens erroné de la langue maternelle (donc la langue première, que nous désignerons désormais par L1) (Zeevaert & ten Thije, 2007 : 16).

Lire un texte rédigé dans une langue qui n'est pas encore maîtrisée nécessite l'aide de ressources attentionnelles pour traiter l'information linguistique basée sur la morphologie. Toutefois, la mémoire de travail est restreinte en vue d'exécuter ces tâches attentionnelles et ainsi sont employées les stratégies d'*évitement* et de *supposition*, qui consistent à sauter certains passages dans le texte, afin de saisir quelque peu de son sens. Les lecteurs débutants en L1 exploitent ces stratégies faute d'automatisme dans la reconnaissance de mots et l'analyse syntaxique. (Lutjeharms, 1997.) De plus, lorsque le lecteur ne s'est pas familiarisé avec le contenu ou que les signaux permettant la supposition sont insuffisants, il se peut que des malentendus se produisent. La lecture courante implique, en revanche, que la forme linguistique (p. ex. l'orthographe) soit soumise au traitement automatique : ainsi, les ressources attentionnelles peuvent se diriger vers le niveau sémantique, c'est-à-dire la lecture de contenu. (Lutjeharms, 2007 : 269.)

Pour désigner la représentation des connaissances linguistiques dans le cerveau, Lutjeharms (2007 : 270-271) emploie le terme **lexique mental**. Il appartient à la mémoire à long terme, qui contient essentiellement tout ce que nous savons. Les mots qui sont rencontrés - sans qu'ils soient compris ou acquis - ne se transfèrent pas à la mémoire à long terme instantanément, mais à la mémoire épisodique qui fonctionne selon un lieu ou un moment spécifique. Ainsi, nous pouvons nous rappeler les conditions sous lesquelles nous avons rencontré ledit mot. Afin de le transmettre vers le lexique mental, il faudra le convertir en profitant d'un code linguistique fixé, tel que la traduction du sens du mot. Une fois entré dans le lexique mental, le mot peut être de nouveau accédé et analysé, ce qui rend possible sa reconnaissance (ibid.) Bogaards (1994 : 70) nous suggère qu'une personne, à l'aide d'une liste de mots, commence à les classer, et c'est ainsi que l'on peut mieux connaître la structure du lexique mental. Lorsque cette liste est suffisamment longue et composée de mots appartenant à une catégorie quelconque, les éléments ne sont pas tous mémorisés, mais plutôt regroupés ensemble, en bloc. Cette liste ne vise pas à examiner la connaissance totale du vocabulaire, mais plutôt les manières de classer l'information selon les relations sémantiques entre les mots.

Un autre point de vue sur cette thématique est offert par Lutjeharms (2007 : 269) en nous pourvoyant des détails sur l'importance de l'oculométrie. Le mouvement des yeux consiste en des sauts rapides entre de courtes périodes de fixation. Tandis que l'on lit un texte rédigé dans une langue apparentée – ayant un ordre de mots analogue à la L1 – les habitudes du mouvement des

yeux peuvent être transférées de la L1 vers la L2. En revanche, les compétences linguistiques basses chez le lecteur et un texte sémantiquement difficile à déchiffrer sont souvent au fond des *régressions*, ce qui signifie un avancement ralenti lors de la lecture. L'étude sur les points du regard en lecture menée par Oller et Tullius (1973 ; citée par Coirier et al., 1996 : 210) montre que les locuteurs natifs de l'anglais – ayant des durées de fixation plus courtes que les apprenants de l'anglais – réussissent le mieux dans l'expérience. Cela va sans dire que les conditions dans cette expérimentation ne sont point les mêmes que dans la nôtre, mais puisque l'épreuve lexicale a été établie sur la base du français, il serait intrinsèque pour les locuteurs natifs d'avoir davantage de succès. D'ailleurs, en observant le mouvement du regard chez les lecteurs de l'hébreu et de l'anglais, Deutsch et Rayner (1999 : 416 ; cité par Lutjeharms, 2007) ont constaté que leurs yeux se fixent d'abord à mi-chemin entre le début et le centre du mot.

## 2.2 Le décodage des éléments textuels inconnus

Pour approfondir ce qui a été dit dans le chapitre précédent (2.1), nous mettrons en relief l'importance du vocabulaire acquis lors du décodage du texte. Le premier sous-chapitre se concentrera sur cette importance dans la compréhension plus générale du texte, et le deuxième apportera des bases théoriques pour la réalisation de notre matériau, présentée dans la section 3.

### 2.2.1 La compréhension du texte

Bien que notre étude s'oriente vers la compréhension des mots individuels, nous tenons à souligner l'importance de la compréhension plus générale du texte, parce que cela explique en partie comment les mots étudiés sont reconnus lorsqu'ils sont entourés par du contexte. Comme le déclare Bogaards (1994 : 64), « *c'est en effet le contexte qui fournit les interprétations plausibles et détermine le sens des éléments* ». Cet avis il partage avec Rastier (cité par Bogaards, 1994 : 64), selon qui « *les relations contextuelles sont constitutives du sens* » (Rastier, 1991 : 154) et « *les mots hors contexte ne sont que des artefacts dont la linguistique n'a guère à s'occuper* » (p. 195).

Tout d'abord, nous voulons insister sur le **seuil lexical** (*threshold vocabulary*). C'est un terme employé par Laufer (1996 : 23) qui désigne le point critique au-delà duquel les moyens du niveau supérieur, autrement dit les stratégies efficaces de lecture, peuvent être mises en œuvre. En d'autres

mots, il s'agit de la connaissance d'au moins 3,000 familles de mots<sup>2</sup> pour assurer que la lecture soit courante, sans être empêchée par une connaissance insuffisante du vocabulaire. Laufer a examiné la position de ce seuil en 1992. Elle a remarqué que même les étudiants les plus avancés ne pouvaient bien lire dans la L2 (langue seconde, donc langue étrangère) s'ils ne dépassaient pas le seuil lexical. En tout, ceux dont les connaissances lexicales se trouvent au-dessous de ce seuil ont du mal à distinguer quelle information est importante, laquelle est périphérique et laquelle est inutile. Si le message n'est pas compris, l'information considérée « inutile » sera celle qui est mal comprise. Parfois, pourtant, l'omission peut être dépendante des facteurs culturels, liée aux propres attentes et connaissances du lecteur sur le sujet. Steffensen et Joag-Dev (1984, cité par Laufer, 1996) déclarent que les lecteurs peuvent compléter l'information qui « devrait » être présente, en négligeant l'information qui leur semble futile selon leur vision du monde.

Ensuite, Laufer (1996 : 31) mentionne plusieurs points qui influent sur la compréhension du texte donné. Il ne peut être tenu pour certain qu'il y aurait des indices à « découvrir » : soit le texte en contient ou non. De plus, ces indices doivent être compris par le lecteur. Plus il y existe de mots familiers, plus il est probable que les mots qui s'expliquent l'un l'autre soient familiers. Les deux derniers points démontrent que la nature humaine, elle aussi, joue un rôle. Certains indices peuvent être captieux et l'apprenant de la L2 ne voit aucune raison de réinterpréter le mot si ce dernier semble convenable dans le contexte : pendant l'épreuve lexicale, l'un de nos participants avait traduit *altre* ('autre') comme *korkeus* ('hauteur') dans la phrase « *L'un gist sur l'altre e envers e adenz* », le contexte acceptant grammaticalement les deux propositions. Ainsi, nous semble-t-il que le contexte lui est resté assez opaque (trad. par Bédier : *Ils gisent l'un sur l'autre, face au ciel, face contre terre*).

L'on a également tendance à faire l'interprétation de sa propre manière, notamment quand une opinion partielle se présente, au cas où le contexte est moins clair et si les deux vues diffèrent substantiellement (Laufer, 1996 : 30). Lorsque nous appliquons cette théorie à notre travail, nous constatons qu'elle a raison : le sens des mots inconnus est souvent deviné en s'appuyant sur ce qui convient le mieux aux propres interprétations du lecteur en essayant de déchiffrer le contexte (v. 4.2.5). Tout compte fait, les traductions erronées peuvent provenir de multiples sources différentes, comme le montrera notre entrevue informelle (v. 4.2.5).

---

<sup>2</sup> Une famille est composée de différentes expressions idiomatiques (*garder la tête froide*) ou formes du même lexème, telles que *venir, venu, vint, vienne*. Ceci est la définition de Nation (1983), citée par Laufer (1996).

## 2.2.2 La transparence du lexique

Selon une classification proposée par Bezooijen et Gooskens (2007 : 256) pour analyser la transparence, il y aurait quatre cas de figure : le sens peut être complètement transparent, semi-transparent, semi-opaque ou complètement opaque. Nous présentons ci-dessous cette classification en reprenant les exemples des auteures pour les langues néerlandaise, frisonne et afrikaans. Cette classification nous servira de base pour la classification des mots de notre matériau aussi. Pour ce qui vient après, le dictionnaire de Collins<sup>3</sup> définit le terme « cognat[e] » comme suit : « *related through the same source; derived from a common original form* » (Notre traduction : ‘apparenté par la même source ; dérivé d’une forme originale commune’). Ce terme connaît plusieurs acceptions, mais nous adoptons celle-ci. La classification ci-dessous, rédigée à l’origine en anglais, est traduite par nous.

- **Sens complètement transparent**

Quand deux cognats ont une forme identique, il n’y a aucun problème de reconnaissance. Exemple : afr. *uitbuiting* vs néerl. *uitbuiting* ‘exploitation’. Les cognats ayant une petite différence orthographique qui ne se manifeste pas au niveau de la prononciation sont de même inclus dans cette catégorie. Exemple : fris. *buro’s* vs néerl. *bureaus* ‘bureaux’.

- **Sens semi-transparent**

Deux cognats se ressemblent autant pour présumer que le lecteur reconnaîtrait leur rapport assez facilement. Dans la plupart des cas, il y a la différence d’une seule lettre. Exemple : afr. *sewentig* vs néerl. *zeventig* ‘soixante-dix’.

- **Sens semi-opaque**

Deux cognats ont si peu en commun qu’il sera difficile pour le lecteur néerlandophone de reconnaître leur rapport. Souvent, plusieurs lettres diffèrent. Exemple : fris. *jierren* vs néerl. *jaren* ‘ans’.

- **Sens complètement opaque**

La ressemblance de deux cognats est si minuscule que le lecteur néerlandophone ne pourrait voir aucun rapport entre les deux. Fort souvent, la majorité des lettres sont différentes. Exemple : afr. *hê* vs néerl. *hebben* ‘avoir’. (Bezooijen & Gooskens, 2007)

En analysant la transparence des mots, Bezooijen et Gooskens (2007) utilisent un système de points très pratique pour les études appliquées : elles attribuent respectivement 0, 1, 2 et 3 points selon que le sens est complètement transparent, semi-transparent, semi-opaque ou complètement opaque. Elles profitent de cette notation pour calculer la mesure de la distance entre les mots comparés.

---

<sup>3</sup> [Cognate definition and meaning | Collins English Dictionary \(collinsdictionary.com\)](https://www.collinsdictionary.com/)

Quant à notre étude, le système de notation se construit par la distance d'édition que nous présenterons plus tard, dans le chapitre 3.1.2.2.

Une échelle similaire à celle susmentionnée est proposée par Piccaluga et Harmegnies (2008 : 182-183). Ils font la distinction entre *transparence*, *semi-transparence* et *non-transparence*. Il faut remarquer toutefois que leur objectif était d'observer la distance *phonético*-sémantique des mots, tandis que nous nous intéressons à la distance *graphico*-sémantique. Pour des raisons concernant la complexité d'une étude phonétique (surtout dans un cadre historique), nous ne pouvons suivre le chemin pavé par Piccaluga et Harmegnies.

Souvent, les mots peuvent sembler différenciés à première vue mais à travers l'observation et l'analyse de leurs transformations ainsi que d'autres faits linguistiques, leurs équivalents sont saisis à un moindre coût. Ceci veut dire qu'il s'agit d'un processus psycholinguistique et psychocognitif. Nous estimons que cette théorie « procédurale » explique une meilleure reconnaissance de mots chez les personnes qui se sont familiarisées avec les changements spécifiques et les règles générales (des langues romanes), telles que le suffixe latin *-tatem*, donnant *-té* (fr.), *-tà* (it.), *-dad* (esp.), *-dade* (port.), ou encore les terminaisons verbales. (Murillo Puyal et al., 2008 : 360, 362.) La connaissance des changements systématiques sur le plan phonétique est aussi à envisager (McCann et al., 2003 : 53).

### **2.3 Les points communs entre les langues romanes**

Les langues romanes sont issues du latin vulgaire (leur développement est traité par Herman, 1967a et Posner, 1996, entre autres), par moyen de l'évolution phonétique, influencée par des facteurs internes (changements universels) ainsi que externes (contacts avec d'autres langues). À travers cette parenté historique, elles constituent une famille de langue. Nous avons consacré cette section pour le traitement des langues actuelles car cela nous donne la possibilité de mieux traiter la thématique. Nous avons adopté cet aspect de parenté car l'appartenance familiale des langues facilite la compréhension : Sundqvist (1986 ; cité par Ringbom, 1987 : 80) a montré que les écoliers suédophones de la ville de Pietarsaari ont mieux réussi dans une compréhension de texte en anglais que les finnophones, en dépit d'avoir eu moins de l'enseignement dans cette langue. Comme indiqué en 2.1.1, le finnois fait partie des langues finno-ougriennes, tandis que l'anglais et le suédois sont tous les deux des langues germaniques, qui est une branche sous la famille indo-européenne.

### 2.3.1 L'intercompréhension dans les pays romanophones

L'intercompréhension entre les langues romanes se base – en grande partie – sur le **vocabulaire pan-roman**, qui est presque entièrement composé des mots provenus du latin (McCann et al., 2003). Elle est facilitée davantage par la terminologie internationale dont la plupart est dérivée du latin et de ses langues filles, mais nous n'y entrerons pas plus en détail ici. Il est vrai que le latin vulgaire variait considérablement selon la position géographique : pourtant, ses dialectes étaient mutuellement intelligibles – de proche en proche (v. notamment le tableau 2b). L'unité de leurs pratiques linguistiques était garantie, pendant longtemps, par l'orthographe du latin écrit. (Blanche-Benveniste, 2008 : 36.) Vu l'inexistence des médias imprimés à l'époque, les conventions écrites n'empêchèrent pourtant pas les dialectes du latin de s'évoluer en langues (McCann et al., 2003 : 29). Les langues romanes modernes gardent dès lors un vocabulaire commun, composé des éléments fréquemment employés, rendant leur apprentissage plus économique : cet aspect nous a guidé également dans la création de notre corpus (chapitre 3).

La distinction entre les diverses langues d'une même famille se fait non seulement par les différences phonétiques, syntaxiques ou lexicales, mais aussi par l'entente mutuelle partagée par les locuteurs de ladite langue (Janson, 2012 : 122). Ainsi dit, quelques idiomes très anciens – tels que le chinois et le grec – persistent, parce que leurs locuteurs le pensent. De la même façon, la « mort » du latin est difficile à dater, car il cesse d'exister au moment où les locuteurs ne pensent plus le parler. Il faut dire que nous ne sommes pas tout à fait d'accord avec Janson en ce qui concerne l'identité linguistique comme grand tournant dans la distinction des langues très semblables – p. ex. le suédois et le norvégien – dont les divergences structurelles nous considérons plutôt dialectales. McCann et al. (2003 : 29) nous rappellent que la classification des dialectes et des langues est une question de volonté politique et culturelle, non pas toujours liée aux matières linguistiques. Ainsi, nous admettons que l'identité linguistique (en tant qu'une volonté politico-culturelle) joue depuis toujours un rôle important dans la société.

Pour ce qui est du commencement d'usage du français, Janson (2012 : 125-126) souligne l'importance du XI<sup>e</sup> siècle. Jusqu'à cette époque, l'infrastructure ainsi que le côté culturel européens avaient été assez minables, seul s'employait le latin pour les écritures, souvent proches de l'Eglise. Une fois améliorés l'économie et les moyens de communication, une classe sociale – la noblesse – apparut. Les membres de cette classe voulaient employer leur temps pour la lecture, mais vu qu'ils ne s'intéressaient pas forcément à la théologie ou aux vies des saints, était-il venu ainsi le moment opportun pour la littérature épique, les chansons de geste et l'époque des ménestrels. Grâce à ce



changement *métalinguistique*, la langue parlée avait obtenu une forme écrite ; et cela presque simultanément au nord (*français*) et au sud (*provençal*) de la France moderne. Cette tendance se propagea plus tard aux autres régions romanophones : d'abord en Italie, ensuite en Espagne et enfin au Portugal (*ibid.*)

Pottier (1997 : 75) emploie le concept « air de famille » en parlant des langues romanes. Sont constatées, selon lui, les ressemblances avec d'autres langues, en particulier lorsque l'on lit des journaux ou entend des informations à la radio. L'idée que la presse d'information est le plus facilement partageable est soutenue par Blanche-Benveniste et Valli (1997 : 34). Les rubriques sont assez standardisées, les thèmes traités sont semblables dans n'importe quel pays et, maintes fois, ils ne nécessitent pas beaucoup de connaissances antérieures. Même sans formation préalable, les apprenants d'une autre langue romane saisissent certains éléments des messages linguistiques et c'est ainsi que ces langues ne sont jamais complètement étrangères (Murillo Puyal et al., 2008 : 367). Au niveau de la conversation courante, cependant, ces ressemblances ne sont pas perçues aussi clairement (Pottier, 1997 : 75).

Blanche-Benveniste et Valli affirment toutefois (1997 : 33) que les langues romanes ont servi à l'intercommunication à travers les frontières des Etats et des régions, au temps du commerce maritime. Selon eux, il y a encore quelques professions voyageuses où l'on « jongle » avec plusieurs langues. Schmitt Jensen (1997 : 107) nous rappelle qu'aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles il y avait des écrivains français qui étaient capables de lire couramment l'espagnol, l'italien et le portugais grâce à leur éducation classique, c'est-à-dire, l'apprentissage du latin. Ce dernier s'est clairement décrié, et Schmitt Jensen constate que de nos jours, la majorité des francophones n'ont jamais étudié le latin. Pour notre étude, nous tiendrons compte des connaissances du latin auprès des participants et nous inviterons à faire l'épreuve également des étudiants l'ayant étudié.

Même si la parenté des langues romanes constitue un domaine d'études connu depuis longtemps, il n'est offert qu'au niveau universitaire, ni présenté qu'assez rarement à ceux voulant apprendre les langues modernes (Blanche-Benveniste, 2008 : 34-35). Les principes de la comparaison et la hiérarchie entre les différents domaines de langue ne sont pas exposés. L'on néglige le fait que les plus grandes différences se trouvent entre les manières de prononcer et qu'elles sont moindres au niveau du lexique – et encore moindres pour tout ce qui concerne la grammaire. Même la grande ressemblance entre les lexiques « savants » n'est pas prise en compte, ce qui justifie en partie notre choix de matériau (v. l'introduction au chapitre 3) (*ibid.*)

### 2.3.2 La méthode EuroCom

Il y existe une méthode, appelée EuroCom (mentionnée par Lutjeharms, 2007 : 269), qui a pour but de faciliter la lecture dans des langues apparentées. Cette méthode nous est présentée en détail par McCann et al. (2003) : elle est destinée à l'origine à l'intercompréhension des langues romanes. Son objectif primaire est d'augmenter des compétences partielles en plusieurs langues, au lieu de se concentrer sur la *maîtrise quasi-native* d'une langue (McCann et al., 2003 : 13-14 ; v. aussi 2.1.2).

EuroCom contient sept « tamis » pour trouver des similarités entre les langues apparentées – plus explicitement, des lentilles à travers lesquelles l'on peut observer les phénomènes langagiers. De ces tamis, nous considérerons plus en détail trois, qui sont les suivants : le vocabulaire partagé par la famille de langue, correspondances ou changements de sons systématiques et, dernièrement, structures syntaxiques communes pour la famille de langue. Maintenant, nous montrons quelques correspondances ayant lieu dans les langues romanes (Stegmann, 2005 : 37-38 ; la version originale se trouvant dans McCann et al., 2003 : 54).

Tableau 1. Correspondances systématiques des groupes consonantiques dans les langues romanes.

<i>Langue</i>	<i>Exemple</i>
<b>Français</b>	La <u>c</u> lef est dans une salle <u>p</u> leine de <u>f</u> lammes.
<b>Catalan</b>	La <u>cl</u> au està en una sala <u>pl</u> ena de <u>fl</u> ames.
<b>Italien</b>	La <u>ch</u> iave è in una sala <u>pi</u> ena di <u>fi</u> amme.
<b>Portugais</b>	A <u>ch</u> ave está numa sala <u>che</u> ia de <u>ch</u> amas.
<b>Espagnol</b>	La <u>ll</u> ave está en una sala <u>ll</u> ena de <u>ll</u> amas.

Bien que le tableau 1 ci-dessus soit destiné uniquement à montrer les différentes évolutions des groupes constitués par des lettres *cl*, *pl* et *fl* en latin, nous avons pu faire des rapprochements aux deux autres tamis, en y voyant quelques divergences aussi. D'abord, les traits syntaxiques sont partagés par toutes ces langues (genre et concordance, articles défini et indéfini, construction « plein + de + qqch. »), même si le verbe et la préposition pour exprimer la localisation varient selon la langue. Deuxièmement, au niveau lexical, l'on ne peut douter la parenté de ces langues parce que les mots sont clairement issus d'origine latine (*clef* = *clavis*, *plein* = *plenus*, *flamme* = *flamma*).

Dernièrement, nous voulons insister sur les changements qui correspondent. C'est bien ce trait que nous étudions chez les locuteurs de français, s'ils réussissent à reconnaître les mots apparentés à travers la voile historico-orthographique. Bien que certains mots puissent paraître méconnaissables, être conscient de ce systématisme derrière leur développement (McCann et al., 2003 : 53) permet de reconstruire sa traduction dans une autre langue. Cette expérience se base sur nos connaissances,

ayant suivi des cours des plusieurs langues romanes au cours des années académiques. Pour en donner un exemple, il faut savoir qu'en espagnol, tous ces groupes consonantiques donnent *ll* (prononcé [ʎ]) au début du mot (cf. tableau 1). En revanche, dans l'enseignement de la parenté des langues romanes, la régularité de ces correspondances n'est pas envisagée. (Blanche-Benveniste, 2008 : 35).

Ajoutons encore qu'EuRom4 – une expérience préparatoire où les jeunes locuteurs romanophones ont été placés devant des textes écrits et parlés dans les trois langues romanes autres que la leur – a révélé que les ressemblances de ladite langue étaient perçues immédiatement. Ces ressemblances ont abouti à l'établissement inconscient des correspondances, ce qui, à travers des généralisations, les a aidés à apprendre de nouvelles langues – tout en utilisant la connaissance de sa propre langue comme l'outil principal. (Simone, 1997 : 28.) Mout curieusement, Blanche-Benveniste (2008 : 35) a observé chez les adultes abordant l'EuRom4 qu'ils ne savaient rien des possibilités de comparaison, étant dépourvus de modèles de comparaison. Ils s'abstenaient de faire les rapprochements les plus évidents, car ils craignaient de suivre leur intuition à tort : il est donc probable que l'on leur avait enseigné que l'intuition est mauvaise conseillère (ibid.)

### 2.3.3 L'effet de la graphie et l'intelligibilité mutuelle

Pour ce travail, nous relevons l'importance des deux notions suivantes (Pottier, 1997 : 78) : le *contact avec la graphie* et le *contact visuel*. Elles semblent être très proches, mais opèrent une distinction importante. Le contact avec la graphie concerne l'acquisition d'habitudes de reconnaissance de séquences complexes, souvent dénotant des consonnes palatalisées (comme [ɲ], [ʎ], [ʝ] et [ʒ]). Les combinaisons de consonnes – telles que *ch*, *qu*, *sc*, *nh*, *lh* – renvoyant parfois à une seule unité phonétique dans une langue, marquent une prononciation différente dans une autre, et c'est ainsi qu'elles ne seront pas reconnues aussi facilement à l'écrit : *chiave* ('clé') en italien se prononce ['kja.ve], tandis que *chave* en portugais est ['ʃa.vi], donc la graphie *ch* correspond à deux phonèmes complètement différents, [k] et [ʃ] (cf. tableau 1). Pour aller plus loin, même le français connaît deux palatalisations différentes de [k] latine : celle qui a eu lieu plus tôt – devant [e] et [i] – se prononce aujourd'hui [s] (*centum* > *cent*), tandis que la prononciation [ʃ] se manifeste devant [a], palatalisée plus tard (*cambiare* > *changer*) (Andrieux-Reix, 1993 : 93). Malgré ces quelques écarts orthographiques, les langues romanes s'écrivent assez cohéremment, en règle générale. (Pottier, 1997 ; McCann et al., 2003 : 89.) Le français y fait exception, et nous en parlerons dans la section suivante (2.4). D'autre part, le contact avec la graphie s'intéresse aussi à la connaissance des règles de marquage de l'accent tonique (Pottier, 1997) ; notons toutefois que ce phénomène n'est pas très

important dans cette recherche, car les mots contenant ce type de marquage sont inexistant dans notre corpus.

Le contact visuel, en revanche, se focalise sur le lexique, et il est donc plus séyant du point de vue du corpus de cette étude. Nous en présenterons ci-dessous deux tableaux. Le tableau 2a montre deux mots (*lune*, *vert*) qui sont restés presque inchangés à travers les pays romans, et le tableau 2b souligne la notion d'*originalité*, devenue la force majeure derrière les obstacles de la reconnaissance des mots. Il s'agit très souvent de différents étymons latins que les langues gardèrent (lat. *parabolare* vs *fabulare*), mais quelquefois, des mots empruntés ont pénétré dans le lexique. Dans les deux tableaux, les langues sont organisées de façon géographique, allant de l'ouest à l'est.

Tableau 2a. Comparaison de désignations restées très proches. (Pottier, 1997 : 78)

Portugais	Espagnol	Catalan	Français	Italien	Roumain
lua	luna	lluna	lune	luna	lună
verde	verde	verd	vert	verde	verde

Tableau 2b. Comparaison de désignations qui se sont éloignées. (Pottier, 1997 : 78)

Portugais	Espagnol	Catalan	Français	Italien	Roumain
segunda-feira	lunes	dilluns	lundi	lunedì	luni
falar	hablar	parlar	parler	parlare	vorbi
irmã	hermana	germana	sœur	sorella	soră
madeira	madera	fusta	bois	legno	lemn
cabeça	cabeza	cap	tête	testa	cap

Il est bien visible que l'espagnol et le portugais ont conservé le même étymon dans la plupart des cas, le français et l'italien également. Le catalan et le roumain (qui ne sont pas présents dans notre matériau, v. chapitre 3.1.1) montrent plus de variation étymologique. Cette répartition met en évidence l'ancienne dialectisation du latin vulgaire, en démontrant des taux de ressemblance élevés entre les langues avoisinantes (cf. la remarque de Blanche-Benveniste en 2.3.1).

Tout cela signifie qu'il y a, dans une certaine mesure, de l'*intelligibilité mutuelle* parmi les locuteurs des langues romanes. Comme nous rappelle Posner (1996 : 87), ce sont les différences phonétiques qui causent davantage d'obstacles pour la compréhension que les différences lexicales. Ceci doit être pris en compte parce que les sujets d'expérience auxquels nous avons fait appel (v. 3.2.1) connaissent bien l'orthographe du français, qui n'est pourtant pas si dissemblable des autres idiomes issus du latin vulgaire, et qui les aide à reconnaître les mots bien plus souvent à l'écrit qu'à l'oral. Nous entamerons cette discussion dans le chapitre suivant.

## 2.4 Les spécificités du français par rapport aux autres langues romanes

Toutes les langues romanes que nous étudions dans cette étude présentent un aspect unificateur : les *emprunts lexicaux* au latin classique. Tandis que nous ne nous concentrons pas sur cette partie du lexique (nous en parlerons plus dans le chapitre 3.1.2, consacré au recueil du matériau), nous sommes toutefois d'avis qu'il faut relever ce fait, car cela explique pourquoi les langues romanes modernes – mais surtout le français – sont souvent considérées fort proches l'une de l'autre. Ce phénomène est bien illustré chez Posner (1996 : 80), subséquemment traduit par nous :

“Among loanwords, a large proportion are borrowed directly from Latin or the other Romance languages, so that the ‘Romance’ character of the vocabulary is overwhelmingly obvious.”  
(Notre traduction : Une vaste proportion de mots a été empruntée directement au latin ou aux autres langues romanes, et pour cette raison le caractère « roman » du vocabulaire est pesamment évident.)

L'auteure précise plus loin (Posner, 1996 : 151) que la plupart des emprunts au latin datent d'entre le XIII<sup>e</sup> et le XVII<sup>e</sup> siècle – après cela, l'on parle des emprunts internationaux qui se trouvent dans plusieurs groupes de langues du monde – et la raison principale derrière ces afflux était l'inadéquation du vocabulaire vernaculaire face à la création des discours plus élaborés.

Bien que le latin ait transmis à tous ses descendants une quantité considérable d'emprunts, c'est en français que la différence entre ce qui est hérité et ce qui est emprunté reste la plus évidente. Posner (1996) nous indique que le changement phonétique y a été très régulier et radical, et cela signifie, inversement, que le français s'est bien éloigné d'autres parlars romans. Cet avis est renforcé par Simone (1997 : 30) en disant que « *le français est ressenti comme plus distant de toutes [les langues romanes], comme s'il était lancé sur une trajectoire de fuite hors de la famille romane* ». Quoiqu'il en soit, les emprunts l'ont empêché de se diverger davantage vers la fin de la période médiévale, notamment sur le plan des conventions orthographiques, voir *infra*. (Posner, 1996 : 150-151).

En scrutant la Chanson de Roland – l'épopée nationale de la France, l'un des corpus utilisés – nous avons remarqué que le vocabulaire n'est pas très ample : il se concentre essentiellement sur un lexique belliqueux avec des structures grammaticales assez simples pour exprimer maintes choses. Qui plus est, l'orthographe de l'ancien français n'est pas loin de celle du français moderne : « *Mais lui **meisme** ne volt mettre en oubli* » se traduit en français moderne « *Mais il ne veut pas se mettre lui-même en oubli* », phrase tirée de notre matériau (trad. par Bédier). Puisqu'il s'agit d'un stade antérieur de la même langue, les formes actuelles suivent en principe les mêmes lignes d'évolution, et ne contiennent pas beaucoup de changements. Selon Herman (1967b : 51), la période de l'ancien

français commence par les **Serments de Strasbourg** en 842 et termine au début du XIV<sup>e</sup> siècle. Les serments sont également traités par Janson (2012 : 124) qui, en faisant remarquer l'écart notable entre le latin classique et l'ancien français, observe que plusieurs traits caractéristiques au français moderne existaient autrefois déjà : ceci voulant dire que la langue n'a pas évolué aussi profondément par la suite. Herman constate (1967b : 118) que l'orthographe du français cesse de suivre l'évolution de la prononciation (excepté des cas isolés) au XIII<sup>e</sup> siècle.

Tout compte fait, cette orthographe « vieillotte » du français facilite l'intercompréhension entre les langues romanes, puisqu'elle permet de retrouver les liens bien plus facilement que l'expression orale. Pour le bon fonctionnement de la méthode EuroCom (v. 2.3.2), il faut s'appuyer sur cette « solidarité écrite », offerte par l'orthographe française. (McCann et al., 2003 : 89, 47.)

### 3 Corpus et méthode

Nous nous intéressons surtout à la compréhension des mots semblables dans des langues apparentées, des mots hérités par la « voie populaire » et modifiés au fil des siècles. Cela étant dit, il s'avère nécessaire de nous appuyer sur des textes littéraires anciens, aussi dépourvus de mots d'emprunt (latinismes, hellénismes, anglicismes) que possible, pour évoquer combien d'obstacles le changement linguistique naturel peut poser à la compréhension d'une langue issue de la même origine. Nous présenterons les textes littéraires choisis ainsi que les mots que nous en avons extraits pour l'épreuve lexicale dans les sous-sections 3.1.1 et 3.1.2, respectivement. Dans ce travail, il n'est guère possible d'entrer dans le détail des changements linguistiques (c'est-à-dire les règles systématiques et les caractéristiques concomitantes du changement), étant donné que cela nous obligerait à approfondir notre étude dans le domaine de la phonétique. Ensuite, la sous-section 3.2 présentera les participants à cette étude ainsi que le recueil des réponses ; 3.3 exposera notre méthode d'analyse.

#### 3.1 Sélection du matériau de l'étude

Nous étudions la reconnaissance de mots à l'aide d'une épreuve qui se base sur un corpus de 30 mots extrait des épopées nationales des pays parlant une langue romane (v. *infra*). Dans cette sous-section, nous expliquerons les étapes parcourues pour aboutir aux 30 mots, à savoir le choix des textes littéraires sources et le choix des mots eux-mêmes. Avant de passer l'épreuve aux participants, nous avons fait un test pilote auprès d'une enseignante, locutrice native du français, pour tester comment l'épreuve fonctionne.

### 3.1.1 Choix des textes littéraires sources

Comme textes sources nous avons choisi quatre épopées nationales provenant des pays romans. Nous avons adopté le format original de chaque œuvre, tout en sachant que ces œuvres sont rédigées dans des formes de langue assez anciennes. Ce sont surtout l'accessibilité et le regroupement thématique de ces textes qui ont joué un rôle important dans la sélection du matériau. L'on aurait pu l'assembler d'autres sources convenables, mais pour donner à ce travail plus de cohérence, nous avons opté pour celles-ci.

Tableau 3. Les quatre épopées nationales utilisées pour la collecte du matériau.

<b>Pays</b>	<b>Titre</b>	<b>Composition</b>
<b>Espagne</b>	Le Poème du Cid	XIII <sup>e</sup> siècle (1200-1207)
<b>France</b>	La Chanson de Roland	XII <sup>e</sup> siècle (ou avant)
<b>Italie</b>	La Divine Comédie	XIV <sup>e</sup> siècle (1303-1321)
<b>Portugal</b>	Les Lusiades	XVI <sup>e</sup> siècle (paru en 1572)

Comme indiqué plus haut (v. 2.3.3), de ce recueil sont exclus le roumain et le catalan. La raison principale pour cette omission a été la délimitation du travail : la date très tardive où la littérature purement « nationale » fut rédigée dans ces deux langues (le XIX<sup>e</sup> siècle). De plus, le corpus est déjà suffisamment volumineux même sans ces langues : il fallait penser aux participants pour que le test ne devienne pas trop lourd à faire.

### 3.1.2 Les mots retenus pour l'épreuve lexicale

Afin d'établir une liste de mots adaptée pour les objectifs de notre étude (voir l'Annexe 3), nous avons exploré les textes littéraires sources sous plusieurs aspects, que nous présenterons dans cette sous-section. Une centaine de mots ont été parcourus avant que nous ayons délimité le matériau définitivement. Initialement, l'épreuve comportait 40 mots au lieu des 30 mots actuels (v. 3.1.3) ; nous avons réduit la liste en raison de la pesanteur générale de l'épreuve qui aurait pu embrouiller les sujets d'expérience par la perte de motivation.

#### 3.1.2.1 Choix des mots

Pour définir la sélection des mots, nous avons bénéficié de trois aspects primaires : l'étymologie, la partie du discours et le contexte environnant. Bien souvent, nous avons cherché des mots spécifiques en tenant compte des aspects susmentionnés, mais parfois, le choix de mots s'est fait « par hasard », en lisant des passages aléatoires.

- I. Étymologie – Tous les termes présents dans les groupes apparentés de notre matériau sont directement reliés à une traduction en français moderne, en passant par le latin vulgaire. Cet état de langue est bien sûr discutable (Herman, 1967a : 5), et dont la forme précise ne pourra jamais être ratifiée. Cependant, grâce aux connaissances acquises dans le domaine de la phonétique historique (surtout celle du français ; voir l’ouvrage de Herman (1967b), *Précis d’histoire de la langue française*), nous sommes convaincu que les formes correspondent. Dans le cas du français, il suffit d’indiquer tout simplement qu’il s’agit d’une étape antérieure de la langue et il n’est pas obligatoire de recourir au latin vulgaire.
- II. Partie du discours – Quatre catégories lexicales sont représentées dans notre matériau : noms communs, adjectifs, verbes et pronoms. Ceci nous permet de mieux examiner la perception des mots, en offrant de nombreux points de vue. Il manque certains groupes (adverbes, conjonctions) et il n’y a pas d’équilibre quantitatif total entre ceux qui existent. Une partie des mots dans notre matériau sont accordés. Les noms et les adjectifs s’accordent en nombre et en genre, les verbes sont encore plus variés : de différents temps et personnes, ainsi que des participes passés.
- III. Contexte environnant – Puisque la présence du contexte joue un rôle essentiel dans l’épreuve lexicale (Bogaards, 1994 ; v. aussi 3.2), il est primordial de sélectionner des phrases qui aident à reconnaître le terme demandé. Les contextes consistent souvent en des structures grammaticales qui se trouvent également en français moderne, mais il y a des exceptions, comme celle-ci en ancien français : « *De la citet l’une **meitet** est sue* » (littéralement : *une moitié de la cité est à lui*, trad. par Bédier). Il se peut donc que certaines tournures soient trop difficiles à déchiffrer et la traduction du mot reste impossible : le taux de reconnaissance pour *meitet* est étonnamment bas (7 %) par rapport à sa distance d’édition, et cela surtout dans la deuxième partie du test lexical (v. tableau 17b). Plusieurs contextes par nous choisis présentent des archaïsmes ou des milieux fictifs.

Nous tenons à souligner ici que les mots retenus de l’espagnol, l’italien et le portugais anciens présentent beaucoup plus de similarités au langage moderne par rapport à ceux de l’ancien français. Parfois, il est question des corpus un peu plus récents ou du temps très limité qui nous a été délégué pour la reformulation du matériau (v. 3.1.3), mais avant tout, il s’agit du fait que notre étude est fondée sur la perspective francophone d’observer les mots. Pourtant, la connaissance des autres langues romanes modernes s’est immiscée trop souvent dans la procédure de l’identification des



mots, et c'est pour ces motifs qu'il est d'une importance primordiale de séparer les corpus venant d'autres langues de celui de l'ancien français. Cela se réalise par la constitution de deux sous-chapitres détachés (4.2.3 et 4.2.4) pour traiter la même thématique. Nous parlerons davantage de l'influence des langues modernes dans la section consacrée aux commentaires reçus lors de l'entrevue informelle (4.2.5).

### 3.1.2.2 *Distance de Levenshtein*

Nous avons donné à tous les termes présents dans notre matériau la valeur obtenue de la distance d'édition (ou plus couramment dit : *distance de Levenshtein*), qui sert à calculer et proportionner la divergence des mots cognats. Dans sa version élémentaire, une unité de modification est ajoutée si une lettre doit être **additionnée**, **supprimée** ou **substituée** ; ou si deux lettres doivent **changer de place**. Ensuite, les unités de modifications sont mises en relation avec la longueur du mot et c'est ainsi que l'on obtiendra la valeur. (Borin, 2013 : 10.)

Nous avons choisi de suivre la version simple (le *coût* restant '1' pour les quatre actions), mais puisque nous sommes d'avis qu'une modification au début du mot aboutirait à une plus grande difficulté de compréhension, nous y attribuons '2'. Notre choix est appuyé par la constatation de Lutjeharms (2007 : 269), en faisant référence à l'étude oculométrique de Deutsch et Rayner (1999), v. 2.1.3. Décidément, après avoir effectué l'épreuve lexicale, l'un des participants à notre étude nous a confirmé qu'il essayait de traduire les mots habituellement en regardant la première lettre.

Lorsque les deux mots commencent par des lettres différentes – et pour garantir que la distance est calculée correctement – le premier point de l'alignement est lui aussi doublé. Autrement, le nombre du dividende risque d'être plus grand que celui du diviseur. En la modifiant, nous reprenons la proposition de Bezooijen et Gooskens (2007 : 257) pour clarifier ce moyen de calculer.

Tableau 4. Distance de Levenshtein pour le mot « voisines », exemple modifié.

Alignement <sup>1</sup>	1	2	3	4	5	6	7	8	9
Portugais	V		I	Z	I	N	H	A	S
Français	V	O	I	S	I	N		E	S
Coût <sup>2</sup>		1		1			1	1	

<sup>1</sup>Alignement (*Alignment*) = moyen de proportionner les mots

<sup>2</sup>Coût (*Cost*) = addition, suppression ou substitution d'une lettre

*Vizinhas* et *voisines* sont tous les deux issus du latin *vicinas*, mais avec des évolutions bien distinctes, propres aux régions où ces langues sont parlées. Les unités de modification (notées par 1

ou 2 en bas du tableau) sont additionnées (4), puis divisées par le nombre des alignements :  $4/9 = 0,44$ . La distance de Levenshtein entre ces deux mots est ainsi d'environ 44 % et *vizinhas* est pour cette raison considéré comme un mot semi-opaque pour les francophones. Un taux de 0 % voudrait dire que les mots sont complètement identiques et 100 % signifierait qu'aucune lettre n'est restée telle quelle.

En ce qui concerne les distances indiquées par le pourcentage, nous avons repris l'échelle de Bezooijen et Gooskens (2007 ; v. 2.2.2), en l'adaptant à nos besoins. Nous présenterons ci-dessous un tableau qui contient les quatre catégories, ainsi qu'un exemple représentatif de chacune. Pour plus d'informations sur la structuration du matériau pour l'épreuve, voir le sous-chapitre 3.1.3.

Tableau 5. Les catégories de transparence, avec des exemples et leurs distances des mots français.

%	Groupe	Mot du matériau	Distance (%)	Traduction en français
0-25	Transparent	<i>sofferto</i>	22	'souffert'
26-41	Semi-transparent	<i>tuneire</i>	38	'tonnerre'
42-57	Semi-opaque	<i>culcez</i>	50	'couché'
58-100	Opaque	<i>capra</i>	67	'chèvre'

Nous avons distribué tous les termes apparentés dans les catégories présentées en 2.2.2 (transparence – opacité), selon la distance de Levenshtein. Pour les mots *non apparentés*, nous avons créé une catégorie supplémentaire (abrégée comme NA) qui nous permet de voir si la reconnaissance d'une autre langue romane aide à la compréhension. Le mot portugais *amarelo* (jaune) ne voudrait rien dire à un francophone sans la connaissance d'une langue ibéro-romane, tandis que celui qui connaît l'espagnol (*amarillo*) le traduirait directement. *Amarelo* est étymologiquement lié au mot français *amer*, mais cette traduction serait erronée.<sup>4</sup>

Notons que depuis la perspective du français moderne, la grande majorité du vocabulaire de la Chanson de Roland consiste en des mots (semi-)transparents et non apparentés. Les mots opaques sont bien moins fréquents dans cette œuvre que dans les autres sources littéraires, le phénomène étant lié étroitement au « stade antérieur » de la langue que nous avons mentionné en 2.4.

### 3.1.3 Reformulation du matériau

L'épreuve a encouru des modifications radicales bientôt après l'achèvement du test pilote. Nous avons constaté que le vocabulaire choisi comprenait trop de mots opaques pour les participants, et que la présence proéminente d'autres langues romanes induirait le déplacement de l'objectif

<sup>4</sup> [amarelo - Infopédia \(infopedia.pt\)](http://infopedia.pt)

principal, qui était de mesurer la reconnaissance de mots étrangers, au lieu de ceux qui sont familiers : cette « familiarité » persiste pourtant dans la version renouvelée, chez certains mots sélectionnés (v. *supra* 3.1.2.1). C'est ainsi que nous avons reconstruit les cinq catégories. Par ailleurs, nous avons réduit le nombre de mots de 40 à 30, et ajouté la colonne « C'est quelle langue ? », sans donner aucun indice sur de quels parlars il s'agissait. En définitive, nous n'avons pas analysé cette colonne (v. chapitre 5).

Dans un premier temps, nous présentons le tableau 6 qui contient de l'information sur la répartition des langues dans le test. Nous employons AF pour désigner l'ancien français. Rappelons que notre première question de recherche s'intéresse à la reconnaissance des mots de l'AF, c'est donc à ce phénomène que nous avons donné le plus de relief. Pour ce faire, nous avons augmenté la proportion de lexèmes en AF par rapport à ceux d'autres idiomes. Il en résulte que les mots de l'AF constituent une moitié du matériau de la version finale, tandis que la proportion des quatre langues dans la première version était tout à fait commensurable.

Tableau 6. La répartition des mots selon la langue, une comparaison entre les deux versions.

Groupe ↓ Langue →	Version 1 (40 mots)			Version 2 (30 mots)		
	AF	Autres	Total	AF	Autres	Total
Transparent	2	6	8	2	3	5
Semi-transparent	2	6	8	4	3	7
Semi-opaque	2	6	8	4	3	7
Opaque	1	6	7	2	3	5
Non apparenté	3	6	9	3	3	6
<b>TOTAL</b>	<b>10</b>	<b>30</b>	<b>40</b>	<b>15</b>	<b>15</b>	<b>30</b>

Dans un second temps, la difficulté globale du test lexical a été amortie. Les limites des catégories ont été redéfinies : elles ne suivent plus la répartition proportionnée ('Version 1' dans la figure 1), mais sont encore régies par le pourcentage obtenu de la distance de Levenshtein (v. chapitre 3.1.2.2). Dans la première version, le mot *capra* (67 %) aurait été considéré comme semi-opaque, mais il est désormais opaque. La figure sous-jacente montre comment les limites ont été remues, pour le bien de la simplification de l'épreuve lexicale. Ce déplacement nous a contraint de remplacer la plupart des lexèmes choisis : en effet, 13 mots seulement sont restés inchangés.

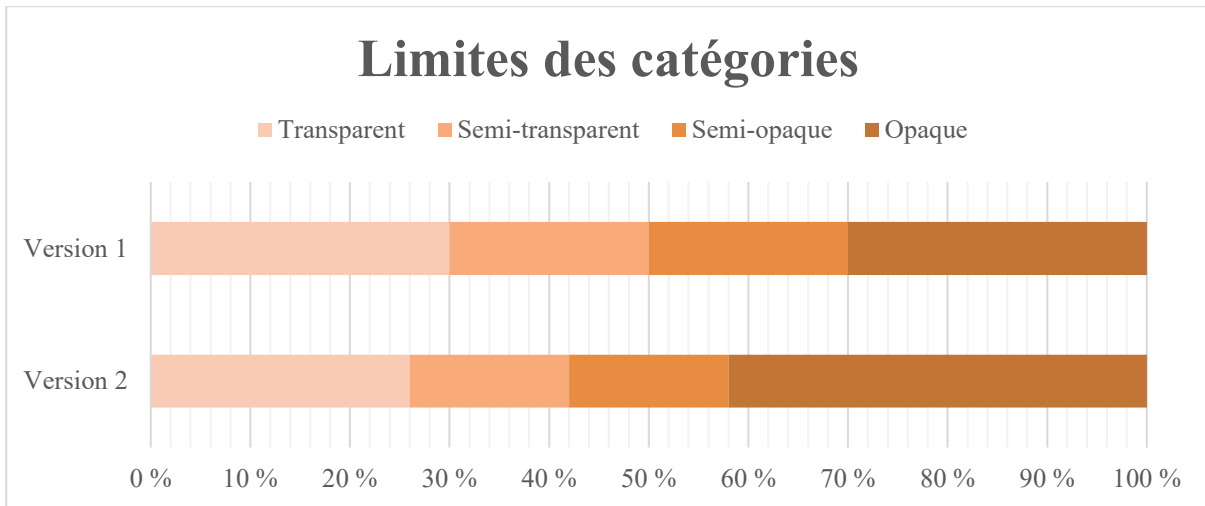


Figure 1. Les limites des catégories selon le taux d'opacité, une comparaison entre les deux versions.

La figure 2 est reliée à la même thématique (difficulté du test). Elle a pour but d'indiquer que l'accent est mis sur la compréhension des mots semi-transparents et semi-opaques dans la deuxième version. Les trois autres groupes apportent des renseignements sur la marche générale de l'épreuve. En dessus des piliers sont les pourcentages des mots distribués en chaque classe. En bas de la figure, en revanche, sont les noms des catégories de transparence (TP = mots transparents, STP = semi-transparents, SO = semi-opaques, O = opaques, NA = non apparentés).

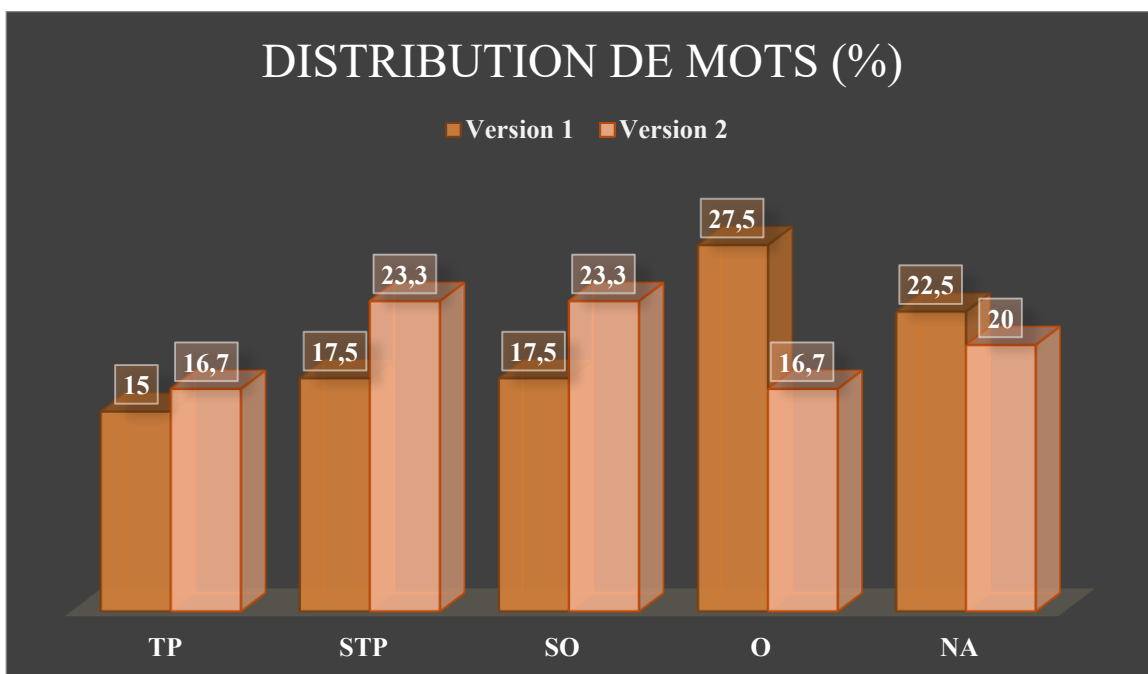


Figure 2. La comparaison de la distribution de mots explicitant la simplification du test.

Notons également que plusieurs expressions ne seraient plus utilisées de nos jours, et le choix des mots a été fait en tenant compte de la compréhension du mot lui-même. Nous aurions voulu inclure l'italien *conocchia* qui veut dire « quenouille » en français, mais il risque de ne pas être compris par les locuteurs non-spécialistes. Le même souci s'applique au cas du contexte : un ordre des mots atypique ou la présence trop élevée des termes inexplicables compliquerait l'interprétation en général (v. 2.1.3), comme nous le verrons plus tard, dans l'analyse des résultats. Trois graphiques montrent tous les termes de notre matériau : ceux accompagnés par les résultats et la distance de Levenshtein peuvent être visualisés en 4.1 et en 4.2.4, celui présentant les traductions équivalentes est disponible dans l'Annexe 3.

## **3.2 Déroulement de l'épreuve lexicale**

Pour mieux examiner la compréhension des langues apparentées, nous avons créé une épreuve lexicale, divisée en deux parties (une liste des mots seuls, entourés par du contexte dans la deuxième). L'épreuve était constituée des mots du matériau, qui devaient être traduits en français moderne ou en finnois. Afin d'accepter la participation, il a fallu donner quinze réponses sur trente, dans les deux parties du test. Nous présenterons les participants dans les sous-chapitres 3.2.4 et 3.2.5.

### **3.2.1 Les sujets-cibles de l'épreuve**

Le groupe-cible de notre recherche est les francophones natifs, en grande partie des étudiants, dont quelques-uns étaient en échange à Turku. D'autre part, des finnophones natifs connaissant le français ont été invités à participer à cet examen. L'objectif et les méthodes de la recherche n'ont été dévoilés à aucun sujet. Pour recueillir des participants, une invitation a été transmise sur plusieurs plateformes médiatiques (courriel, Facebook Messenger, WhatsApp), accompagnée par un lien menant à un questionnaire de base, que nous traiterons plus en détail dans le sous-chapitre suivant. Après y avoir répondu, les participants pouvaient suggérer des dates. Lorsque l'on est parvenu à une heure mutuellement admissible – que ce soit par Doodle (un outil de planification des réunions) ou par une autre plateforme médiatique – l'on a arrangé un rendez-vous sur le service de visioconférence appelé Zoom, dont l'hyperlien nous avons communiqué par courriel ou, moins fréquemment, par WhatsApp. Avec trois personnes, un rendez-vous en présentiel a été organisé.

### 3.2.2 Conception du questionnaire

En ce qui concerne les profils des participants, nous leur avons demandé de remplir un questionnaire (v. Annexe 1), qui avait été créé sur le site Webropol. À la première page, il leur était demandé de communiquer des informations de base (telles que le nom complet, l'année de naissance, le sexe, l'éducation) afin que nous puissions identifier les sujets d'expérience et observer la répartition des résultats selon de divers regroupements que nous expliquerons dans le chapitre 4 consacré à l'analyse des résultats. Malgré le fait d'avoir demandé le genre des participants, ce dernier ne sera aucunement pris en compte lors de l'analyse. La deuxième page du questionnaire était consacrée à l'enquête de la connaissance des langues romanes et du latin. Le niveau de la langue était mesuré sur une échelle de 0 à 5 – pour le latin, la durée des études était demandée. Il fallait savoir impérativement si la personne connaissait bien un autre idiome roman parce que cela entraînerait sans doute des conséquences quant à la compréhension de certains mots.

Le questionnaire a été testé par une professeure de français, avant d'être transmis aux participants par courriel, par Facebook Messenger ou par WhatsApp. D'abord, ils ont dû accepter le RGPD<sup>5</sup> (Règlement Général sur la Protection des Données, mis en place par l'Union Européenne en 2018) : ensuite, toute information a été traitée de manière anonyme. Le contenu du questionnaire était le même en finnois, avec une seule exception : le compte rendu sur le niveau du français (v. tableau 8).

### 3.2.3 Recueil des réponses

Le recueil des réponses s'est déroulé en deux parties. Dans la première partie, les participants devaient essayer de traduire en français moderne des mots isolés, sans aucun contexte. La deuxième partie contenait les mêmes mots dans le même ordre, cette fois faisant partie du contexte. Nous nous sommes assuré que l'ordre dans lequel les mots se trouvent sur les feuilles était complètement aléatoire. Les sujets d'expérience ont été invités à traduire chaque mot, même s'ils n'avaient aucune idée de ce qu'il pourrait signifier. Il leur a été rappelé que l'usage des dictionnaires et des traducteurs n'était pas autorisé. Qui plus est, dans la colonne à droite était demandé de quelle langue il s'agissait, selon eux. Les réponses dans cette colonne n'avaient aucun poids en ce qui concerne la notation : cette question avait pour but de savoir comment les répondants identifient ces langues.

---

<sup>5</sup> [EUR-Lex - 32016R0679 - EN - EUR-Lex \(europa.eu\)](https://eur-lex.europa.eu/eur-lex.do)

Peu s'en est fallu que nous eussions écarté ce « jeu de supposition » : enfin, il ne répondait pas à nos questions de recherche.

Notre objectif principal est de voir si les répondants réussissent à discerner des ressemblances avec le français moderne. Le matériau est constitué de mots assez fréquents, mais en s'assurant en même temps qu'ils ne sont pas trop évidents (comme *fiesta*, *crescendo*, etc.) Nous avons utilisé Microsoft Word pour l'élaboration de l'épreuve (v. Annexe 2). La durée maximale de l'examen était de 60 minutes, mais la plupart des participants l'ont terminé en moins de 45 minutes.

Après avoir collecté des réponses avec le test, nous avons préparé une entrevue informelle (v. 4.2.5) visant à expliquer les résultats plus en détail. Il ne s'agit pas d'un sondage systématique : les quelques commentaires que nous avons reçus permettent plutôt le raccordement de la base théorique avec les résultats. Nous avons demandé auprès de dix participants de nous spécifier les raisons pour lesquelles les sept mots choisis après l'analyse des résultats avaient été reconnus de manière inattendu. Quoi qu'il en soit, notre entrevue ne peut être justifiée strictement par les études antérieures – de fait, les corpus que nous avons recueillis contenaient les termes choisis par nous, auxquels seuls nos sujets d'expérience pouvaient conférer des exégèses. Pour avoir un aperçu aussi complet que possible sur la structuration d'une étude abordant l'intercompréhension, la présence d'une telle entrevue s'avère indispensable.

### 3.2.4 Les profils comparés des participants francophones

En ce qui concerne le questionnaire rédigé en français, nous avons reçu 32 réponses complètes par des francophones, à partir de la fin janvier 2022. Ensuite, nous avons contacté ceux qui montraient de l'intérêt et quant à l'épreuve elle-même, seize participations ont été recueillies entre le milieu février et le début avril, en y ajoutant une supplémentaire en décembre de la même année, ce qui a haussé leur nombre à dix-sept.

Nous avons rejeté deux participants : une femme et un homme d'âge mûr. Elle n'avait pas complété la deuxième partie du test de manière régulière ; quant à lui, il n'avait pas donné suffisamment de réponses sur la feuille, lors de la première partie. C'est pour cela que le nombre définitif des réponses données s'est fixé à quinze.

L'anonymisation des participants s'est faite en donnant à chacun deux lettres (FR ou FI) selon sa langue maternelle, puis en les mettant dans l'ordre numérique. Le tableau 7 met en comparaison les

réponses écrites sur la version française du questionnaire, sans prise en considération des personnes qui n'avaient pas participé à l'épreuve lexicale, ni de ces deux qui en avaient été rejetés.

Tableau 7. Les profils comparés des participants francophones.

	Sexe	Année de naissance	Profession	Domaine d'études ou de spécialisation	La plus forte langue romane (hormis FR)
FR1	H	1997	Ingénieur	Bâtiment, numérique	Espagnol
FR2	F	1998	Etudiante alternante	Littérature	Espagnol
FR3	H	1992	Entrepreneur	Gestion d'entreprise	Espagnol
FR4	F	1998	Etudiante	Gestion culturelle	Espagnol
FR5	H	1980	Enseignant-chercheur	Langue, littérature, civilisations françaises	Italien
FR6	F	1997	Enseignante FLE	Linguistique	Espagnol
FR7	H	1987	Enseignant	Arts plastiques, philosophie	Italien
FR8	H	2001	Etudiant	Droit	Espagnol
FR9	F	1990	Sans emploi	Sciences politiques, sciences sociales	S/O
FR10	F	1999	Etudiante	Droit, études européennes	Espagnol
FR11	F	1974	Infirmière	Santé	Espagnol
FR12	F	2000	Etudiante	Physique	Espagnol
FR13	H	2000	Etudiant	Sciences politiques, développement durable	Italien
FR14	F	1972	Enseignante FLE	Politiques locales, développement	Espagnol
FR15	H	1979	Enseignant FLE	Linguistique	Italien

**F = Femme, H = Homme, S/O = Sans objet (ici : la personne ne connaît que le français)**

Le groupe-cible francophone consiste en huit femmes et sept hommes, dont l'âge moyen est de 31 ans. Environ un tiers des participants sont des étudiants, un autre tiers des enseignants et quatre personnes n'appartiennent à aucune de ces deux catégories. FR5, FR6, FR14 et FR15 se sont spécialisés dans le domaine de l'enseignement du français. L'espagnol est en général la langue romane la plus forte parmi les participants francophones : seule FR9 disait de ne connaître aucune langue romane hormis le français. FR3 et FR9 ont une autre langue maternelle : le néerlandais et le finnois, respectivement.

### 3.2.5 Les profils comparés des participants finnophones

En octobre 2022, recommence le recueil des réponses sur le questionnaire rédigé en finnois : cette fois les personnes parlant le finnois comme langue maternelle étaient invités à le remplir. Vingt personnes ont participé à l'épreuve entre le début novembre et le milieu décembre.

De cette catégorie, cinq participants ont été rejetés : quatre jeunes adultes et une femme d'âge mûr. Excepté une jeune femme, chacun d'eux avait laissé trop de points vides sur la feuille, soit dans la première, soit dans la deuxième partie. La jeune femme, à son tour, avait déjà pris part à la version



d'essai (v. 3.1.3) : comme elle se souvenait encore des mots et connaissait les objectifs de l'épreuve, nous avons dû l'exclure, et ainsi la totalité des réponses données s'est fixée – ici aussi – à quinze.

Le tableau 8 met en comparaison les réponses écrites sur la version finlandaise du questionnaire, sans prise en considération des personnes qui n'avaient pas participé à l'épreuve lexicale, ni de ces quatre qui en avaient été rejetés.

Tableau 8. Les profils comparés des participants finnophones.

	Sexe	Niveau du français	Année de naissance	Profession	Domaine d'études ou de spécialisation	La plus forte langue romane (hormis FR)
FI1	F	4	1997	Etudiante	Linguistique	Espagnol
FI2	F	4	1997	Chercheuse	Linguistique	Espagnol
FI3	F	4	1966	Enseignante	Linguistique	Espagnol
FI4	F	4	1993	Sans emploi	Littérature comparée	Espagnol
FI5	F	3	1996	Enseignante	Linguistique	Espagnol
FI6	F	4	1996	Enseignante	Linguistique, pédagogie	Espagnol
FI7	F	3	1998	Etudiante	Linguistique	Italien
FI8	F	4	1995	Enseignante	Linguistique	Espagnol
FI9	H	2	1989	Expert en législation	Droit	Espagnol
FI10	F	3	2000	Etudiante	Linguistique	S/O
FI11	F	4	1996	Etudiante	Linguistique	Italien
FI12	F	4	1999	Etudiante	Linguistique	Espagnol
FI13	F	3	1998	Etudiante	Linguistique	Italien
FI14	H	4	1984	Etudiant diplômé	Linguistique	Espagnol
FI15	H	4	1990	Enseignant	Linguistique	Espagnol

**F = Femme, H = Homme, S/O = Sans objet (la personne ne connaît que le français)**

Le groupe de référence finnophone comporte douze femmes et trois hommes, dont l'âge moyen est de 29 ans. Presqu'une moitié des participants sont des étudiants, un tiers des enseignants et trois personnes n'appartiennent à aucune de ces deux catégories. Contrairement au groupe-cible, tous les participants finnophones se sont spécialisés – plus ou moins – dans le domaine de la linguistique, ce qui est assez compréhensible car il s'agit d'une langue étrangère. Parmi eux aussi, l'espagnol est bien souvent la langue romane la plus forte, après le français : pourtant, FI13 fait l'exception en connaissant l'italien mieux que le français. Seule FI10 disait de ne connaître aucun idiome roman hormis le français. FI10 et FI14 sont des locuteurs natifs en suédois aussi, FI13 en italien.

### 3.3 Méthode d'analyse

Pour analyser les résultats, nous emploierons une échelle à six étapes, établie par nous, sur la base de l'échelle disponible dans Palokangas (2018), composée de neuf étapes. Nous l'avons écartée parce qu'elle semblait bien trop détaillée pour nos besoins. Cependant, elle nous a aidé à élaborer la nôtre, que nous introduirons dans le tableau 9. Contrairement à ce que Palokangas a fait, nous

n'avons réalisé aucune entrevue approfondie avec les sujets d'expérience, mais plutôt une qui fait l'esquisse de quelques phénomènes proéminents.

La comparaison des résultats se fait par l'attribution de points, selon la réponse donnée. Dans les deux parties, il y avait 30 mots à traduire : la bonne réponse sur chaque ligne (en assignant cinq points) équivaut donc à un maximum de 150 points. Le sujet d'expérience devait traduire du moins quinze mots pour que sa participation fût acceptée (cf. 3.2). Le tableau sous-jacent explique plus en détail la méthode selon laquelle nous avons attribué les points. En guise de clarification, nous utilisons le terme « forme », qui renvoie à la morphologie : p. ex. les traductions peuvent être correctes, mais si les verbes ne sont pas conjugués, ou le pluriel d'un nom commun n'est pas pris en compte, cela se verra également dans la notation. Nous attribuons un point aux réponses qui n'ont rien à voir avec celle qui serait correcte, mais dont la forme révèle que sa fonction syntaxique dans le contexte a été identifiée.

Tableau 9. La description des six étapes comprises dans la notation des résultats.

Pts	Description	Exemple
5	La réponse est correcte par sa forme et par son sens. Un synonyme n'ayant aucune différence sémantique est accepté aussi. Il peut y avoir deux bonnes réponses dont l'une est pourtant mal accordée.	<b>Amarelllos = jaunes</b> (tout correct)
4	La réponse est correcte mais mal accordée. La réponse est correcte mais son sens diffère légèrement. La bonne réponse est accompagnée par une ou deux mauvaises réponses qui font étroitement partie de la même thématique.	<b>Amarelllos = jaune</b> (sans marqueur de pluriel)
3	La bonne réponse est accompagnée par une mauvaise réponse. Trois réponses, dont l'une est bonne et deux sont mauvaises. La réponse est thématiquement correcte, mais elle peut être mal accordée ou appartenir à une autre partie du discours.	<b>Amarelllos = blanches</b> (une autre couleur)
2	La réponse contient une référence sémantique ou étymologique claire. Trois réponses, dont l'une est sur la bonne voie. La bonne réponse a été éliminée ou remplacée par une mauvaise réponse.	<b>Amarelllos = brillantes</b> (une qualité visuelle des dents, pouvant être rapprochée à la couleur)
1	La réponse contient une référence sémantique ou étymologique faible, et peut appartenir à une autre partie du discours. La mauvaise réponse est sous la bonne forme verbale. La partie du discours est identifiée dans la deuxième partie du test (sauf pour les noms, v. <i>infra</i> ). ➤ Il peut y avoir une ou deux mauvaises réponses sans rapport.	<b>Amarelllos = pointues</b> (une formation spécifique des dents)
0	Une ou plusieurs réponses sont incorrectes.	<b>Amarelllos = amerrissage</b> (rien à voir)

Nous remarquons ici que les synonymes sont acceptés également. Prenons l'exemple de l'ancien français « *Paien d'Arabe des nefes se sunt eissut, puis sunt muntez es chevals e es muls* ». Il est vrai que le lien étymologique entre 'eissut' et 'sortis' est inexistant, et que dans le contexte donné, il

serait préférable de le traduire comme '*descendus*'. La compréhension du sens du mot montre fort clairement que le lien est fait (cf. la constatation de Bezooijen & Gooskens en 2.1.1). Obtenir un point pour un adjectif bien décliné ou un verbe bien conjugué est – selon nos constatations – une indication de la reconnaissance. Un substantif, toutefois, est souvent donné comme réponse primitive. Bien évidemment, nous n'attribuons jamais aucun point pour un adjectif ou un verbe erroné, qui est en outre mal accordé.

## 4 Résultats

Dans ce qui suit, nous essaierons de répondre à nos questions de recherche, sur la base des résultats obtenus de l'épreuve lexicale, dont la structure et le déroulement ont été éclaircis dans le chapitre 3. Pour ce faire, nous présenterons plusieurs tableaux qui seront expliqués et commentés en temps voulu. Comme mentionné en 3.2.2, nous ne considérerons pas le genre des participants, parce que l'échantillonnage de cette expérimentation est resté plutôt restreint, surtout en ce qui concerne la distribution entre les genres.

Tout d'abord, nous parlerons du bilan au sens général, en approfondissant la manière d'observer vers un niveau plus spécifique au fur et à mesure. Notre question sur l'influence du contexte (4.1) selon la langue maternelle aura sa réponse en 4.2.1, la distance orthographique et la comparaison des corpus seront traitées en 4.2.2, 4.2.3 et 4.2.4, puis le rôle des compétences linguistiques en 4.2.3 et 4.2.4. Nous conclurons ce chapitre par une entrevue informelle, où nous soulèverons des commentaires individuels reçus de la part des participants.

### 4.1 Bilan

Le tableau 10 ci-dessous montre les résultats complets selon les catégories de transparence. Le nombre des mots inclus dans chaque groupe est mis entre parenthèses : il nous permet de calculer le maximum des points obtenables à l'intérieur dudit groupe, en l'employant comme multiplicateur d'une somme totale (150) qu'un mot quelconque peut atteindre. Plus explicitement, si tous les participants avaient reçu le maximum de 5 points pour un mot donné, sa somme totale serait alors  $30 \times 5 = 150$  points. C'est pratiquement le cas de *esteiles* (étoiles) dans la deuxième partie du test, ayant eu 146/150 points (donc 97 % de tous les points distribuables). Réciproquement, la somme totale de 0 points voudrait dire que personne n'a reconnu ni même quelque peu de son sens, ce qui est p. ex. le cas du mot *velz* (vieux) dans la première partie du test. La colonne TR (%) signifie le taux de reconnaissance, qui est donc calculé en divisant les points complets par le maximum obtainable.

Enfin, la colonne C marque les points de pourcentage (p. %) indiquant le changement entre les parties. Il est calculé en divisant le résultat plus récent par celui antérieur (p. ex. *transparentes* :  $[(691/540)-1] \times 100 = 27,96\dots$ )

Tableau 10. Les résultats de l'épreuve selon la transparence.

Groupe	Maximum obtainable	Première partie		Deuxième partie		C (p. %)
		Points complets	TR (%)	Points complets	TR (%)	
Transparentes (5)	750	540	72	691	92	28
Semi-transparentes (7)	1050	530	50	785	75	50
Semi-opaques (7)	1050	342	33	503	48	45
Opagues (5)	750	173	23	262	35	52
Non apparentés (6)	900	160	18	393	44	144
<b>Total (30)</b>	<b>4500</b>	<b>1745</b>	-	<b>2634</b>	-	-
<b>Moyenne</b>	-	349	39	527	59	51

Il est évident que la présence du contexte aide le participant à mieux identifier le terme demandé (v. les commentaires de Rastier et de Bogaards en 2.2.1). Cette croissance se voit surtout dans la catégorie des non apparentés (144 points de pourcentage du changement). En scrutant plus minutieusement le pourcentage désignant le taux de reconnaissance (TR %), nous pouvons déduire que l'opacité influe également sur la reconnaissance du mot. Cependant, ce tableau n'évite aucunement toute la vérité. Il y a eu de considérables écartements entre la reconnaissance des mots spécifiques, auxquels nous reviendrons *infra*. L'une des caractéristiques les plus saillantes de cette observation générale est la meilleure reconnaissance des mots non apparentés par rapport aux mots opaques dans la deuxième partie du test. Nous sommes convaincu que ce phénomène s'explique par notre choix de mots et la complexité du contexte circonvoisin. Ces questions seront traitées plus tard, dans la partie consacrée à la discussion (chapitre 5).

Nous poursuivrons notre analyse à l'aide d'un classement plus détaillé des résultats, c'est-à-dire, un graphique mettant en parallèle à la fois tous les mots de l'épreuve, dont l'ordre est déterminé par la distance de Levenshtein (v. 3.1.2.2). Nous verrons combien de points chacun des termes a obtenu dans les deux parties, quel est exactement leur taux de reconnaissance et de quelle façon les résultats ont changé après la contextualisation. La formule pour calculer le changement (C) par points de pourcentage a été introduite *supra*, avant le tableau précédent (ici p. ex. *esteiles* :  $[(146/69)-1] \times 100 = 111,59\dots$ ) Cette fois, nous avons utilisé quatre nuances de vert et deux nuances de rouge afin de souligner les différences concernant le changement dans la reconnaissance. Les limites des catégories de transparence sont visibles également.

Tableau 11. Les points complets de l'épreuve lexicale.

Mot	Distance	Première partie		Deuxième partie		C (p. %)
		Points (150)	%	Points (150)	%	
Ventos	17	97	65	141	94	45
Altre	20	124	83	140	93	13
Viene	20	130	87	134	89	3
Sofferto	22	120	80	130	87	8
Esteiles	25	69	46	146	97	112
Arvores	29	60	40	111	74	85
Lontane	30	116	77	119	79	3
Meïsome	33	89	59	135	90	52
Piedra	33	45	30	60	40	33
Sustienget	36	99	66	89	59	-10
Curune	38	110	73	140	93	27
Tuneire	38	11	7	131	87	1090
Meitet	43	5	3	10	7	100
Vuldreie	44	96	64	140	93	46
Vizinhas	44	38	25	47	31	24
Culcez	50	24	16	53	35	121
Podriemos	56	118	79	141	94	19
Ascolte	56	46	31	73	49	59
Preium	57	15	10	39	26	160
Velz	60	0	0	5	3	S/O
Parfunz	63	4	3	27	18	575
Conhecem	64	68	45	83	55	22
Capra	67	35	23	50	33	43
Habedes	75	66	44	97	65	47
Amarellous	NA	68	45	117	78	72
Mugier	NA	45	30	129	86	187
Fiume	NA	25	17	70	47	180
Eissut	NA	10	7	44	29	340
Toluz	NA	0	0	28	19	S/O
Amsdous	NA	12	8	5	3	-58
Total (4500)	-	1745	-	2634	-	-
Moyenne	-	-	39	-	59	121

FR      IT      ES      PT

À première vue alors, nous pouvons tirer la même conclusion qu'avant : les termes transparents sont reconnus mieux que les termes opaques et les taux se sont élevés après la contextualisation.

Néanmoins, nous remarquons maints détails intéressants : même si l'identification des transparents et des semi-transparentes (ici, de *ventos* jusqu'à *meïsome*) est bien réussie lors de la deuxième partie, elle ne l'est pas toujours dans la première. L'on voit des taux assez bas pour *arvores* (40 %) et *esteiles* (46 %). D'autre part, dès la mise en contexte, nous observons une amélioration considérable chez les mots non apparentés (à l'exception de *amsdous*), le phénomène déjà indiqué *supra*.

L'énorme accroissement des points dans le cas de *tuneire* ('tonnerre') s'expliquerait par un contexte dénotant des conditions météorologiques (*Orez i ad de tuneire e de vent, pluies e gresilz desmesurement*) : sans lui, sa reconnaissance reste bien plus difficile, vu le changement important de 'u' en 'o' dans la première syllabe du mot (v. 4.2.5). De même, *parfunz* ('profonds') a été

beaucoup mieux identifié avec contexte, qui, pour les plus astucieux, a dévoilé sa vraie signification (...*cez roches plus haltes, cez vals parfunz...*) en tant qu'un adjectif spatial.

## 4.2 Résultats comparés

Afin de connaître les particularités des résultats, nous avons mis à part les sections contenant cette information plus détaillée. Dans les sous-chapitres qui suivent, nous insisterons d'abord sur la séparation des données selon la langue maternelle du participant. Ensuite, nous ferons des remarques sur l'attribution des points et mettrons en relief quelques modèles de traduction pertinents. Dernièrement, nous présenterons les données à travers une classification par des groupes, afin de savoir quel est le rôle des compétences linguistiques dans la reconnaissance des mots.

### 4.2.1 Les francophones versus les finnophones

Pour répondre à notre première question de recherche, – *Dans quelle mesure la présence du contexte aide-t-elle à identifier le sens du mot, selon la langue maternelle du participant* – nous comparons dans un premier temps les résultats individuels des deux groupes – les locuteurs natifs du français et ceux qui l'ont appris en l'étudiant – en y ajoutant des observations ci-après. Les tableaux 12a et 12b ci-dessous montrent les points obtenus tels quels dans les deux parties du test : 150 étant toujours le maximum. Est indiquée également la durée de chaque séance, mise en parenthèses au cas où la durée exacte manque de nos renseignements. Cette exactitude est possible grâce à la fonction d'enregistrement sur Zoom. Le point de pourcentage (Changement) servant à l'analyse de l'amélioration des résultats figure dans la colonne droite : sa formule est expliquée dans le chapitre précédent. Donnons-en encore un exemple aplanissant – FR1 :  $[(96/57)-1] \times 100 = 68,421\dots$ )

Tableau 12a. Les points complets reçus par les participants francophones.

Participant(e)	Première partie (mots seuls)			Deuxième partie (mots contextualisés)			Changement (p. %)
	Points (/150)	%	Temps (approx.)	Points (/150)	%	Temps (approx.)	
FR1	57	38	3 min	96	64	11 min	68
FR2	50	33	9 min	65	43	15 min	30
FR3	120	80	6 min	142	95	18 min	18
FR4	68	45	12 min	84	56	19 min	24
FR5	99	66	6 min	135	90	16 min	36
FR6	63	42	10 min	99	66	20 min	57
FR7	64	43	5 min	96	64	18 min	50
FR8	30	20	12 min	56	37	27 min	87

FR9	43	29	13 min	81	54	41 min	88
FR10	57	38	4 min	87	58	12 min	53
FR11	39	26	10 min	60	40	14 min	54
FR12	42	28	8 min	99	66	42 min	136
FR13	53	35	12 min	77	51	20 min	45
FR14	39	26	(14 min)	83	55	(20 min)	113
FR15	63	42	(13 min)	102	68	(18 min)	62
<b>Total (/2250)</b>	<b>887</b>	-	-	<b>1362</b>	-	-	-
<b>Moyenne</b>	<b>59</b>	<b>39</b>	<b>9 min</b>	<b>91</b>	<b>61</b>	<b>21 min</b>	<b>61</b>

Tableau 12b. Les points complets reçus par les participants finnophones.

Participant(e)	Première partie (mots seuls)			Deuxième partie (mots contextualisés)			Changement (p. %)
	Points (/150)	%	Temps (approx.)	Points (/150)	%	Temps (approx.)	
FI1	56	37	11 min	94	63	20 min	68
FI2	77	51	12 min	106	71	18 min	38
FI3	83	55	5 min	117	78	14 min	41
FI4	55	37	21 min	80	53	30 min	45
FI5	47	31	13 min	60	40	28 min	28
FI6	69	46	10 min	93	62	10 min	35
FI7	31	21	8 min	52	35	11 min	68
FI8	25	17	11 min	56	37	25 min	124
FI9	54	36	25 min	82	55	27 min	52
FI10	35	23	14 min	59	39	29 min	69
FI11	40	27	8 min	62	41	27 min	55
FI12	60	40	18 min	82	55	24 min	37
FI13	90	60	11 min	119	79	25 min	32
FI14	49	33	(6 min)	90	60	(20 min)	84
FI15	87	58	4 min	120	80	12 min	38
<b>Total (/2250)</b>	<b>858</b>	-	-	<b>1272</b>	-	-	-
<b>Moyenne</b>	<b>57</b>	<b>38</b>	<b>12 min</b>	<b>85</b>	<b>57</b>	<b>21 min</b>	<b>54</b>

Comme nous pouvons le constater dans ces tableaux, la reconnaissance de mots est bien plus élevée dans la partie comprenant le contexte par rapport à celle où les mots restaient seuls. Les taux crurent pour chaque participant, mais en tout, il n'y a eu pas de différence considérable entre les francophones et les finnophones – ces derniers ayant généralement moins de succès. Lorsqu'il n'y avait pas de contexte, le taux de succès se trouvait généralement entre 20 et 60 %, avec rares exceptions. Une fois les mots contextualisés, le taux de succès allait typiquement de 40 jusqu'à 80 %. FR12 a vu son résultat s'améliorer nettement (136 p. de pourcentage), en employant beaucoup plus de temps pendant la deuxième partie. L'on peut évoquer le fait que les finnophones ont utilisé un peu plus de temps à réfléchir aux sens que les francophones, mais ce n'est pas non plus de grande importance, car en général nous avons discuté un peu plus librement avec les premiers lors du déroulement de l'épreuve lexicale. Pour ce qui est de l'amélioration des résultats (v. la colonne droite des tableaux 12a et 12b), il suffit de dire que son comportement est – lui aussi, en premier lieu – relié à la performance individuelle, étant parfois déterminé par l'emploi plus prolongé du temps mais non pas toujours.

Or, regardons les résultats sous un angle différent – c’est-à-dire – la mise en parallèle des francophones et des finnophones selon les catégories de transparence, pour voir si les premiers ont eu plus de succès. Nous organiserons les données de la façon semblable à celle présentée dans la section du bilan (v. 4.1 ; tableau 10). En dessous, il y aura quatre tableaux dans lesquels sont exposées les catégories susmentionnées, la totalité des points obtenables pour chacune d’elles, tout comme la moyenne des résultats et leurs pourcentages. Nous ne nous intéressons pas à l’analyse exhaustive des différences entre chaque mot : c’est pour cela que nous démontrons ici un survol des tendances qui sourdent. D’abord, nous comparons la reconnaissance des mots de l’AF (tableaux 13a et 13b).

Tableau 13a. La comparaison des résultats selon la langue maternelle. AF sans contexte.

Matériau	Ancien français (sans contexte)						
	Groupe	/	FR	%	FI	%	Moyenne
<i>Transparents</i>	150	87	58	106	71	96,5	64
<i>Semi-transparents</i>	300	153	51	156	52	154,5	52
<i>Semi-opaques</i>	300	78	26	62	21	70,0	23
<i>Opagues</i>	150	4	3	0	0	2,0	1
<i>Non apparentés</i>	225	10	4	12	5	11,0	5
<b>Total</b>	<b>1125</b>	<b>332</b>	<b>30</b>	<b>336</b>	<b>30</b>	<b>334</b>	<b>30</b>

Tableau 13b. La comparaison des résultats selon la langue maternelle. AF contextualisé.

Matériau	Ancien français (avec contexte)						
	Groupe	/	FR	%	FI	%	Moyenne
<i>Transparents</i>	150	145	97	141	94	143,0	95
<i>Semi-transparents</i>	300	258	86	237	79	247,5	83
<i>Semi-opaques</i>	300	148	49	94	31	121,0	40
<i>Opagues</i>	150	16	11	16	11	16,0	11
<i>Non apparentés</i>	225	54	24	23	10	38,5	17
<b>Total</b>	<b>1125</b>	<b>621</b>	<b>55</b>	<b>511</b>	<b>45</b>	<b>566</b>	<b>50</b>

Puisque nos délimitations ne permettent pas d’entrer en détail des causes derrière la fluctuation, nous nous concentrerons à sa description. Il ressort que dans toutes les catégories (v. aussi les tableaux suivants 13c et 13d), les participants francophones ont mieux reconnu les mots semi-opaques, à l’encontre des finnophones : ceci se passe le plus distinctement lorsque les mots de l’AF sont contextualisés (tableau 13b). Dans cette même classification (l’AF contextualisé), la différence des taux de reconnaissance généraux est la plus évidente entre les deux groupes de participants (FR : 55 % vs FI : 45 %). Les semi-transparents et les non apparentés y présentent de clairs écarts aussi, mais d’ailleurs ces écarts sont bien moindres. Maintenant nous comparons la reconnaissance des autres langues romanes (13c et 13d).

Tableau 13c. La comparaison des résultats selon la langue maternelle. ES+IT+PT sans contexte.

Matériau	Ancien espagnol, italien et portugais (sans contexte)						
	Groupe	/	FR	%	FI	%	Moyenne
<i>Transparents</i>	225	177	79	170	76	173,5	77



<b>Semi-transparents</b>	225	108	<b>48</b>	113	<b>50</b>	110,5	<b>49</b>
<b>Semi-opaques</b>	225	109	<b>48</b>	93	<b>41</b>	101,0	<b>45</b>
<b>Opaques</b>	225	92	<b>41</b>	77	<b>34</b>	84,5	<b>38</b>
<b>Non apparentés</b>	225	69	<b>31</b>	69	<b>31</b>	69,0	<b>31</b>
<b>Total</b>	<b>1125</b>	<b>555</b>	<b>49</b>	<b>522</b>	<b>46</b>	<b>539</b>	<b>48</b>

Tableau 13d. La comparaison des résultats selon la langue maternelle. ES+IT+PT contextualisés.

Matériau	Ancien espagnol, italien et portugais (avec contexte)						
	Groupe	I	FR	%	FI	%	Moyenne
<b>Transparents</b>	225	206	<b>92</b>	199	<b>88</b>	202,5	<b>90</b>
<b>Semi-transparents</b>	225	146	<b>65</b>	144	<b>64</b>	145,0	<b>64</b>
<b>Semi-opaques</b>	225	131	<b>58</b>	130	<b>58</b>	130,5	<b>58</b>
<b>Opaques</b>	225	116	<b>52</b>	114	<b>51</b>	115,0	<b>51</b>
<b>Non apparentés</b>	225	142	<b>63</b>	174	<b>77</b>	158,0	<b>70</b>
<b>Total</b>	<b>1125</b>	<b>741</b>	<b>66</b>	<b>761</b>	<b>68</b>	<b>751</b>	<b>67</b>

Quand les autres anciennes langues se trouvent sans contexte (tableau 13c), les francophones reconnaissent mieux les mots avec plus d'opacité, même si cet écart disparaît lorsque l'on arrive aux mots non apparentés. Ailleurs, les différences sont assez minimales, mais nous observons une meilleure reconnaissance des mots chez les participants finnophones dans deux circonstances : les transparents de l'AF sans contexte (71 % vs 58 % ; tableau 13a) aussi bien que les non apparentés des autres anciennes langues contextualisées (77 % vs 63 % ; tableau 13d). De façon générale, nonobstant, les francophones remportent cette mise en comparaison.

La fluctuation subsiste, mais notre échantillonnage n'est pas suffisant pour dévoiler toute la vérité, et c'est ainsi que nous effaçons la séparation entre le groupe-cible et le groupe de référence : les deux seront dorénavant analysés en même temps.

#### 4.2.2 L'attribution des points donnés et les modèles de traduction

Pour commencer l'étude plus minutieuse sur la reconnaissance des mots individuels, nous nous concentrerons d'abord sur la distribution de tous les points donnés. De cette façon, il sera possible d'observer dans quelle mesure les mots étaient reconnus, quelle était la proportion de réponses fausses ou correctes et ainsi de suite. Cette analyse correspond en partie à notre deuxième question de recherche sur la comparaison de la reconnaissance entre les mots spécifiques. En 4.2.4, nous reprendrons la même question, mais nous l'aborderons d'une autre perspective.

Comme déjà expliqué dans la méthode d'analyse (3.3), chaque réponse est attribuée une valeur dans une échelle en six étapes (0-5). Les deux graphiques suivants affichent la quantité de ces valeurs chez tous les participants à la fois. Dans la colonne à droite est indiquée la moyenne des points, ce qui sert également à la mise en ordre des mots de l'épreuve. Elle est calculée en multipliant les points annoncés en haut du tableau (0-5) par le nombre des unités, puis en les additionnant et

divisant leur somme par le maximum des points (150) que tous les participants ensemble peuvent obtenir. Clarifions-le avec *podriemos* (v. tableau 14a ci-dessous) :  $2 \times 0 + 4 \times 1 + 6 \times 4 + 18 \times 5 = 118 / 150 = 0,787$ . Puisque – pour déchiffrer la moyenne de cette échelle – le maximum est cinq points, nous devons multiplier la somme par cinq :  $0,787 \times 5 = 3,9$ .

Le pourcentage des attributions est exposé tout en bas : il est calculé depuis les totalités des unités d'une colonne, en les divisant par le maximum (30 mots x 30 participants = 900). Ce pourcentage sert à proportionner chaque niveau de la notation et nous en discuterons après le tableau 14b. Dans cette section, par souci de clarté, nous ne traiterons pas l'amélioration globale des résultats liée à la présence du contexte.

Tableau 14a. L'attribution des points de l'épreuve lexicale. Première partie.

1 <sup>ère</sup> partie	0 points	1 point	2 points	3 points	4 points	5 points	Moyenne
viene	2				10	18	4,3
altre	4		1		3	22	4,1
sofferto	2			2	16	10	4,0
podriemos	2	4			6	18	3,9
lontane			1	5	21	3	3,9
curune	8					22	3,7
sustienget	6		1	3	12	8	3,3
ventos	8			3	7	12	3,2
vuldreie	7			1	17	5	3,2
meïsme	12				1	17	3,0
esteiles	15		1	1	1	12	2,3
amarellos	16				2	12	2,3
conhecem	13			1	15	1	2,3
habedes	2	14	3		9	2	2,2
arvores	17		1	1		11	2,0
mugier	21					9	1,5
pedra	21					9	1,5
ascolte	18		1		11		1,5
vizinhas	21			1	5	3	1,3
capra	23					7	1,2
fiume	25					5	0,8
culcez	24				6		0,8
preium	26			2	1	1	0,5
tuneire	27	1				2	0,4
amsdous	27		1			2	0,4
eissut	26	1		3			0,3
meitet	29					1	0,2
parfunz	29				1		0,1
toluz	30						0,0
velz	30						0,0
<b>Total (900)</b>	<b>491</b>	<b>20</b>	<b>10</b>	<b>23</b>	<b>144</b>	<b>212</b>	<b>-</b>
<b>% des réponses</b>	<b>55</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>16</b>	<b>24</b>	<b>1,9</b>

Parmi les nombreuses mauvaises réponses, nous relevons quelques modèles de traduction.

L'influence la plus proéminente vient de l'espagnol, peu importe de quel corpus il s'agit : *velz* >

*fois, esteiles > styles, amarellos > aimer* ou son dérivé sémantico-étymologique – même *toluz > lumière*. À la source de ces traductions seraient sans doute les mots espagnols *vez, estilos, amar, luz*. Le fait que *meitet* est traduit comme *rencontrer* s'explique vraisemblablement par l'anglais *to meet* (cf. 4.2.5), *arvores* comme *chères* par le finnois *arvo*, et *altre* comme *vieux* par l'allemand *alt*. Étant donnée la fréquence amoindrie des derniers exemples, nous pouvons présumer que les participants ont bien souvent essayé de traduire les mots en s'appuyant sur les langues romanes qu'ils connaissent déjà – c'est ce que nous voulions mais n'avions pas dit explicitement à nos participants. Voici encore des traductions erronées très typiques lors de la première partie, les exemples comprenant les formes conjuguées aussi : *meitet > mettre, fiume > fume, piedra > pied, preium > premier, toluz > tous, tuneire > tenir/tunnel* ainsi que le *parfum(s)* quasi omniprésent pour *parfunz*. Le traitement des autres aspects aura lieu après le tableau suivant.

Tableau 14b. L'attribution des points de l'épreuve lexicale. Deuxième partie.

2 <sup>ème</sup> partie	0 points	1 point	2 points	3 points	4 points	5 points	Moyenne
<b>esteiles</b>			1		1	28	<b>4,9</b>
<b>curune</b>	2					28	<b>4,7</b>
<b>altre</b>	2					28	<b>4,7</b>
<b>ventos</b>	1				4	25	<b>4,7</b>
<b>vuldreie</b>	1				5	24	<b>4,7</b>
<b>podriemos</b>				1	7	22	<b>4,7</b>
<b>meïsme</b>	3					27	<b>4,5</b>
<b>viene</b>					16	14	<b>4,5</b>
<b>tuneire</b>	2		2	1	1	24	<b>4,4</b>
<b>mugier</b>	1			8		21	<b>4,3</b>
<b>sofferto</b>	2			1	8	19	<b>4,3</b>
<b>lontane</b>			2	2	21	5	<b>4,0</b>
<b>amarellos</b>	2	3	1	3	2	19	<b>3,9</b>
<b>arvores</b>	4		5	2		19	<b>3,7</b>
<b>habedes</b>	3	7		1	8	11	<b>3,2</b>
<b>sustienget</b>	10			2	7	11	<b>3,0</b>
<b>conhecem</b>	8	3	2		9	8	<b>2,8</b>
<b>ascolte</b>	6	10	1	1	2	10	<b>2,4</b>
<b>fiume</b>	16					14	<b>2,3</b>
<b>piedra</b>	18					12	<b>2,0</b>
<b>culcez</b>	18				7	5	<b>1,8</b>
<b>capra</b>	20					10	<b>1,7</b>
<b>vizinhas</b>	19	2				9	<b>1,6</b>
<b>eissut</b>	13	8	1	1	4	3	<b>1,5</b>
<b>preium</b>	20	1		2	3	4	<b>1,3</b>
<b>parfunz</b>	19	7				4	<b>0,9</b>
<b>toluz</b>	19	3	5			3	<b>0,9</b>
<b>meitet</b>	28					2	<b>0,3</b>
<b>amsdous</b>	29					1	<b>0,2</b>
<b>velz</b>	25	5					<b>0,2</b>
<b>Total (900)</b>	<b>291</b>	<b>49</b>	<b>20</b>	<b>25</b>	<b>105</b>	<b>410</b>	<b>-</b>
<b>% des réponses</b>	<b>32</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>12</b>	<b>46</b>	<b>2,9</b>

À la lumière de ces valeurs, il est évident que la reconnaissance soit se réalise presque ou tout correctement (4 ou 5 pts), soit ne se réalise pas du tout (0 pts) – ceci se produisant la plupart du temps. Ce qui change considérablement entre les parties (cf. les colonnes à gauche et à droite des toutes dernières lignes des tableaux 14a et 14b), c’est le taux de non-reconnaissance (55 %) qui décroît radicalement (32 %) pour donner de la place à la croissance de la reconnaissance complète, allant de 24 % jusqu’à 46 %. Les autres points n’ont connu que de légères variations. Selon nos méthodes de notation (présentées bien plus en détail en 3.3), le participant obtient donc 1-3 points lorsqu’il identifie quelque peu du sens du mot et 4-5 points lorsqu’il reconnaît véritablement le mot demandé.

Lors de la première partie, tous les mots qui avaient cette distinction très nette entre la non-reconnaissance (0 pts) et la reconnaissance complète (5 pts) étaient des substantifs au singulier : *curune, mugier, piedra, capra, fiume, meitet*. Ceci n’est pourtant pas non plus le cas dans la deuxième partie, car des mots appartenant à d’autres parties du discours ont été bien identifiés aussi (notamment *altre* et *meïsme*). Plusieurs participants ont négligé la conjugaison des verbes, en laissant leur traduction sous la forme infinitive. Parfois – bien entendu – les quatre points reçus pour une traduction sont le résultat d’une déclinaison incorrecte d’un substantif ou d’un adjectif. L’exemple le plus saillant de ce phénomène est *lontane*, qui n’acceptait que la forme plurielle en féminin « lointaines ». Bien souvent, les mots sont mieux accordés lors de la deuxième partie (comme *ascolte : écoute > écoutes* ou *vuldreie : vouloir > voudrais*), mais non pas toujours. Dans le cas de « *¿No te viene en miente en Valençia lo del leon?* », la particule « *te* » avait confondu quelques personnes si profondément qu’ils avaient interprété le verbe suivant comme s’il s’agissait de la deuxième personne au singulier « *viens* » au lieu de celle correcte en troisième : *vient*. D’autres confusions se sont produites aussi, mais nous y revenons prochainement.

Les modèles de traduction se sont – eux aussi – remués dans la deuxième partie. Il n’y a plus autant de recours à d’autres langues : très souvent, le contexte permet de saisir la fonction du terme dans la phrase, même s’il n’a pas été traduit correctement. Ceci s’est passé principalement chez les termes venant d’autres langues que l’ancien français. Nous avons attribué 0-3 p. pour ces quelques traductions erronées figurant *infra*, dans le tableau 15.

Tableau 15. Traductions incorrectes démontrant la compréhension de la partie du discours.

Contexte	Traduction correcte	Réponses erronées démontrant la reconnaissance
[...] de verdes ervas e <b>arvores</b> floridas	‘arbres’	<i>Plantes carnivores</i> (FR2), <i>arbustes</i> (FR7), <i>mot lié à la nature</i> (FR9), <i>champs</i> (FI3), <i>niityt</i> (‘prés’ ; FI5), <i>niitty</i> (‘pré’ ; FI7)

Del agua feziste vino y de la <b>pedra</b> pan	'pierre'	<i>Pâte</i> (FR10), <i>miche</i> (FR7), <i>croute</i> (FR1), <i>farine</i> (FI4, FI14), <i>pain</i> (FI7)
Charles li magnes <b>velz</b> est e redotez [...]	'vieux'	<i>Vaillant</i> (FR1, FR6), <i>sage</i> (FI3), <i>courageux</i> (FI4)
[...] se tu m' <b>ascolte</b> , io ti dirò	'écoutes'	<i>Escortes</i> (FR2), <i>rencontres</i> (FR1), <i>demandes</i> (FI14), <i>consultes</i> (FR10), <i>assignes</i> (FI1), <i>m'aider</i> (FR11), <i>m'acoste</i> (FR12), <i>accueilles</i> (FR14), <i>moitit</i> ('reproches' ; FI4), <i>välitât</i> ('occupes' ; FI5)
[...] a boca negra, os dentes <b>amarellós</b>	'jaunes'	<i>Blanches</i> (FR1, FR11), <i>brillantes</i> (FR7), <i>joku väri</i> ('une couleur' ; FI11)
Por amor de mi <b>mugier</b> [...]	'femme'	<i>Mère</i> (FR1, FR2, FR8, FR10, FR11, FI10), <i>mari</i> (FR13), <i>fille</i> (FI7)

Comme déjà noté *supra*, dans le paragraphe précédant l'introduction du tableau 15, le contexte a parfois tordu le sens général de la phrase (et nous le traiterons plus amplement en 4.2.5). Les remplacements de *parfum* par *parfumé* pour *parfunz* ('profonds') et *fumer* par *fumée* pour *fume* ('fleuve') montrent bien l'opacité du mot cherché : cette dernière aboutit plus souvent encore à l'absence complète d'une traduction dans la deuxième partie par rapport à celle première. Plus explicitement, l'on voit moins de « parfums » dans la deuxième partie, et plutôt des cadres vides. L'exemple le plus intrusif de ce phénomène est la traduction de « *Cuntre le ciel amsdous* ('tous les deux') *ses mains ad juinz* [...] ». Sont offerts des termes suivants : *joindre*, *baisse*, *leva*, *lever*, *au-dessus*, *nostaa* ('lever'), *claire*, *synkkä* ('maussade'), *assombrit*, *bleu*. Ceux-ci manifestent que la fonction de *amsdous* est constamment reliée soit avec « le ciel », soit avec « le mouvement des mains » et même rapproché au sujet de la phrase – au lieu de reconnaître sa fonction correcte comme un déterminant.

#### 4.2.3 La reconnaissance selon les compétences linguistiques : l'espagnol, l'italien et le portugais anciens

Afin de répondre à notre deuxième question de recherche sur la reconnaissance spécifique des mots de l'AF, ainsi que la troisième question sur l'influence des compétences linguistiques des langues romanes dans cette expérimentation, nous avons établi une division selon le nombre des langues parlées par les participants (v. Annexe 4). Étant donné que les résultats ne montrent pas une grande variation entre les francophones et les finnophones, nous avons aboli cette distinction (v. *supra* 4.2.1). La nouvelle division des participants se réalise maintenant par la constitution de quatre groupes : les G1, G2, G3 et G4. Le premier est composé de sept personnes, qui ne disent connaître d'autres langues romanes que le français ou qui ne les connaissent guère. Le deuxième groupe consiste en neuf personnes qui connaissent une autre langue romane en plus du français. Cette langue est bien souvent l'espagnol, qui se voit aussi sur le plan de la reconnaissance des mots d'origine espagnole (et même portugaise). Le groupe numéro trois compte sept personnes, ayant appris deux autres langues romanes. Dans le quatrième groupe sont inclus ceux qui ont déclaré

d'avoir étudié ou expérimenté avec au moins trois langues romanes – parfois pas autant pour une compréhension courante de ces idiomes – mais nous les avons regroupés ensemble tout de même : le total étant sept ici également. La connaissance du latin ou de l'ancien français n'est pas prise en compte dans cette classification-ci : néanmoins, nous analyserons l'influence potentielle des études sur l'AF en 4.2.4.

Nous avons donc formé les quatre groupes, combiné les points à l'intérieur de chaque rassemblement, puis divisé la somme par le nombre maximum atteignable. Le maximum est 35 pour les G1, G3 et G4 comprenant sept personnes et 45 pour le G2 consistant en neuf personnes – chaque participant pouvant toujours recevoir cinq points pour un mot. Une fois les points divisés, nous avons obtenu les pourcentages qui sont indiqués sous les titres 'G1-G4'. En les additionnant et divisant la somme par quatre, il nous est resté la moyenne (M) de chaque mot, ce qui représente le taux de reconnaissance général parmi tous les sujets d'expérience. Toutes les valeurs du 16a pourront être mises en comparaison avec celles des tableaux 16b, 17a et 17b qui suivent. Les mots se trouvent dans l'ordre de transparence, régi par la distance de Levenshtein (la colonne D) et les catégories de transparence sont séparées par des lignes : rappelons-nous que les transparents ont une distance de 0-25, les semi-transparentes 26-41, les semi-opaques 42-57 et les opaques 58-100. Les mots n'ayant aucun lien étymologique sont classifiés comme non apparentés (NA). Bien que notre objectif d'étude principal soit le matériau de l'AF, nous le présentons plus tard. Ainsi, nous ne rompons pas la continuité du texte, établie par l'étude plus minutieuse sur les aspects de l'AF. D'abord, donc, nous affichons les résultats de la première partie du test lexical chez les termes venant d'autres anciennes langues romanes.

Tableau 16a. Les taux de reconnaissance des mots seuls (ES, IT, PT) selon un regroupement basé sur les compétences linguistiques dans les langues romanes.<sup>6</sup>

Les phrases complètes (1 <sup>ère</sup> partie : mots seuls)	D	G1	G2	G3	G4	M
Lá nas covas marítimas entrarão fugindo á tempestade e <b>ventos</b> duros	17	46	71	66	74	64
¿No te <b>viene</b> en mente en València lo del leon?	20	60	96	94	94	86
E come noi lo mal ch'avem <b>sofferto</b> perdoniamo a ciascuno	22	80	76	77	89	81
Estava a terra em montes revestida, de verdes ervas e <b>arvores</b> floridas	29	6	33	57	66	41
Quant' è che tu venisti a piè del monte per le <b>lontane</b> acque?	30	66	78	83	83	78
Del agua feziste vino y de la <b>pedra</b> pan	33	14	11	43	57	31
Vio ter a muitas villas suas <b>vizinhas</b> , que por armas tu, Sancho, humildes tinhas	44	0	24	37	40	25
Si no, en esta tierra angosta no <b>podriemos</b> vivir	56	40	87	86	100	78
O dolce padre mio, se tu m' <b>ascolte</b> , io ti dirò	56	11	9	46	63	32
Mais bravos e mais feros se <b>conhecem</b> pela fama nas obras e nos feitos	64	34	38	46	66	46
Quando 'l corno de la <b>capra</b> del ciel col sol si tocca	67	0	0	43	57	25
D'esto que <b>habedes</b> fecho vos no perderedes nada	75	31	40	69	37	44

<sup>6</sup> Par souci de clarté, nous n'incluons que les pourcentages dans les tableaux 16a-17b.

Cheios de terra e crespos os cabelos, a boca negra, os dentes <b>amarelos</b>	NA	14	64	83	14	44
Por amor de mi <b>mugier</b> y de mis fijas amas	NA	14	22	29	57	31
Vidi genti a la riva d'un gran <b>fiume</b>	NA	0	0	14	57	18
<b>Moyenne des groupes</b>	-	28	43	58	64	48

Nous voyons qu'ici, l'augmentation des compétences linguistiques dans les langues romanes entraîne de claires améliorations qui sont plutôt linéaires. Tenant compte de vastes aptitudes du G4, le plongeon le plus drastique lui advient pour *amarelos* : il n'est reconnu que par une seule personne. Notons aussi *habedes* ('avez') chez le même groupe, souvent traduit par '*habits*' ou '*habitez*'. Quant aux mots d'origine italienne (les moins transparents – *ascolte*, *capra* et *fiume*), leur reconnaissance paraît très laborieuse chez les G1 et G2, mais s'améliore chez le G3 et atteint finalement son faite chez le G4. Quelques participants n'ont pas été si précis avec la bonne déclinaison des termes, surtout s'ils n'ont pas étudié les langues romanes. Puisque les membres des groupes ne sont pas nombreux, les points d'un seul membre ont un effet irréfutable sur le système de notation. Il s'agit habituellement d'une reconnaissance individuelle au cas où le total ne surpasse pas les 15 points.

Le tableau ci-dessous fonctionne d'une manière parallèle au tableau précédent : la seule différence étant la contextualisation des mots, c'est-à-dire les résultats de la deuxième partie du test lexical.

Tableau 16b. Les taux de reconnaissance des mots contextualisés (ES, IT, PT) selon un regroupement basé sur les compétences linguistiques dans les langues romanes.

<b>Les phrases complètes (2<sup>ème</sup> partie : en contexte)</b>	<b>D</b>	<b>G1</b>	<b>G2</b>	<b>G3</b>	<b>G4</b>	<b>M</b>
Lá nas covas marítimas entrarão fugindo á tempestade e <b>ventos</b> duros	17	83	98	100	94	94
¿No te <b>viene</b> en mente en València lo del leon?	20	83	89	94	91	89
E come noi lo mal ch'avem <b>sofferto</b> perdoniamo a ciascuno	22	83	91	86	86	87
Estava a terra em montes revestida, de verdes ervas e <b>arvores</b> floridas	29	31	78	91	94	74
Quant' è che tu venisti a piè del monte per le <b>lontane</b> acque?	30	71	76	83	89	80
Del agua feziste vino y de la <b>pedra</b> pan	33	29	22	43	71	41
Vio ter a muitas villas suas <b>vizinhas</b> , que por armas tu, Sancho, humildes tinhas	44	14	27	43	43	32
Si no, en esta tierra angosta no <b>podriemos</b> vivir	56	83	98	94	100	94
O dolce padre mio, se tu m' <b>ascolte</b> , io ti dirò	56	23	27	77	74	50
Mais bravos e mais ferros se <b>conhecem</b> pela fama nas obras e nos feitos	64	26	56	63	77	56
Quando 'l corno de la <b>capra</b> del ciel col sol si tocca	67	14	0	57	71	36
D'esto que <b>habedes</b> fecho vos no perderedes nada	75	34	60	80	86	65
Cheios de terra e crespos os cabelos, a boca negra, os dentes <b>amarelos</b>	NA	49	78	94	91	78
Por amor de mi <b>mugier</b> y de mis fijas amas	NA	77	82	86	100	86
Vidi genti a la riva d'un gran <b>fiume</b>	NA	14	22	71	86	48
<b>Moyenne des groupes</b>	-	48	60	77	84	67

L'amélioration des résultats selon les compétences linguistiques dans les langues romanes est évidente dans cette classification aussi. Elle est moins perceptible parmi les mots transparents (où les valeurs se sont un peu morcelées) et plus accentuée parmi les autres catégories. Les réponses du

G3 et du G4 ne sont plus aucunement régies par la distance de Levenshtein : la reconnaissance se fait très couramment sans considérer ni le mot ni son contexte, avec des fluctuations habituelles, dont la plongée la plus profonde touche *vizinhas*. La grande majorité des termes contextualisés ont été donc traduits tout correctement chez les G3 et G4. Presque tous les participants appartenant au G2 avaient perçu la fonction grammaticale de *ascolte* ('écoutes'), sans aucune idée de sa vraie traduction (v. tableau 15). Les G1 et G2 avaient en général traduit *mugier* ('femme') par 'mère', 'mari' ou 'fille', indiquant donc une bonne compréhension du contexte.

#### 4.2.4 La reconnaissance selon les compétences linguistiques : l'ancien français

Ce sous-chapitre offre de la perspective pour l'étude des mots de l'espagnol, l'italien et le portugais anciens, que nous venons de traiter en 4.2.3. Bien que les distances de Levenshtein n'aillent pas à des niveaux analogues entre ces langues et l'AF, les termes par nous choisis fonctionnent de cette façon, donnant ainsi un terrain fécond pour la comparaison des résultats. Curieusement, les disparités existent bel et bien. La façon de calculer les taux de reconnaissance est identique à celle présentée en 4.2.3, et les valeurs du tableau suivant (17a) sont pour cette raison approchables avec celles du tableau 16a (mots seuls des autres anciennes langues).

Nous présentons d'abord les résultats comparables à ceux du sous-chapitre précédent, mais évoluons la thématique ensuite par examiner si les études de l'AF aident à identifier ces mots. Cet aspect – bien entendu – n'est pas applicable aux autres matériaux, car la connaissance des langues romanes actuelles devance (et même annule) l'influence potentielle des études sur le plan des stades historiques de ces langues, ce que nous n'avons même pas étudié dans ce travail.

Tableau 17a. Les taux de reconnaissance des mots seuls (AF) selon un regroupement basé sur les compétences linguistiques dans les langues romanes.

Les phrases complètes (1 <sup>ère</sup> partie : mots seuls)	D	G1	G2	G3	G4	M
L'un gist sur l' <b>altre</b> e envers e adenz	20	77	78	97	80	83
Clere est la lune e les <b>esteiles</b> flambient	25	17	44	66	57	46
Mais lui <b>meïsme</b> ne volt mettre en ubli	33	0	78	57	97	58
N'en avrai ja ki <b>sustienget</b> m'onur; suz ciel ne quid aveir ami un sul	36	46	69	69	80	66
Puis si li tolent sun sceptre e sa <b>curune</b>	38	71	56	86	86	75
Orez i ad de <b>tuneire</b> e de vent, pluies e gresilz desmesurement	38	3	0	14	14	8
De la citet l'une <b>meitet</b> est sue	43	0	0	0	14	4
Si grant doel ai que jo ne <b>vuldreie</b> estre!	44	49	73	60	71	63
Passet li jurz, la nuit est aserie: <b>culcez</b> s'est li reis en sa cambre voltice	50	0	9	11	46	17
Laisum le plait e si <b>preium</b> le rei que Guenelun cleimt quite ceste feiz	57	11	11	17	0	10
Carles li magnes <b>velz</b> est e redotez: recreanz ert de sa guerre mener	60	0	0	0	0	0
Passent cez puis e cez roches plus haltes, cez vals <b>parfunz</b> , cez destreiz anguisables	63	0	0	0	11	3
Païen d'Arabe des nefz se sunt <b>eissut</b> , puis sunt muntez es chevaux e es muls	NA	0	7	9	11	7



Li reis Marsilie est de guere vencud: vos li avez tuz ses castels <b>toluz</b>	NA	0	0	0	0	0
Cuntre le ciel <b>amsdous</b> ses mains ad juinz, si priet Deu que pareïs li duinst	NA	0	4	14	14	8
<b>Moyenne des groupes</b>	-	18	29	33	39	30

Tout contrairement à l'analyse des autres langues (v. tableau 16a et 16b), ici nous ne constatons aucune linéarité. Malgré tout cela, les groupes eux-mêmes sont souvent en concordance l'un avec l'autre. La connaissance d'une ou plusieurs langues romanes joue un certain rôle, bien qu'elle ne veuille pas toujours dire que la reconnaissance s'améliore selon le nombre des langues maîtrisées. Ce qui est très caractéristique à ces résultats, c'est l'identification mauvaise des mots opaques et non apparentés – un trait qui peut être généralisé. *Meïsme* reste inidentifiable pour le G1, tandis que *vuldreie* y est bien reconnu même s'il est classifié semi-opaque.

Le tableau ci-dessous fonctionne d'une manière parallèle au tableau précédent : la seule différence étant la contextualisation des mots, c'est-à-dire les résultats de la deuxième partie du test lexical.

Tableau 17b. Les taux de reconnaissance des mots contextualisés (AF) selon un regroupement basé sur les compétences linguistiques dans les langues romanes.

Les phrases complètes (2 <sup>ème</sup> partie : en contexte)	D	G1	G2	G3	G4	M
L'un gist sur l' <b>altre</b> e envers e adenz	20	100	89	100	86	94
Clere est la lune e les <b>esteiles</b> flambient	25	89	100	100	100	97
Mais lui <b>meïsme</b> ne volt mettre en ubli	33	86	89	86	100	90
N'en avrai ja ki <b>sustienget</b> m'onur; suz ciel ne quid aveir ami un sul	36	20	51	91	77	60
Puis si li tolent sun sceptre e sa <b>curune</b>	38	71	100	100	100	93
Orez i ad de <b>tuneire</b> e de vent, pluies e gresilz desmesurement	38	71	87	100	91	87
De la citet l'une <b>meitet</b> est sue	43	0	0	14	14	7
Si grant doel ai que jo ne <b>vuldreie</b> estre!	44	83	96	97	97	93
Passet li jurz, la nuit est aserie: <b>culcez</b> s'est li reis en sa cambre voltice	50	11	47	14	66	35
Laisum le plait e si <b>preium</b> le rei que Guenelun cleimt quite ceste feiz	57	9	20	46	31	27
Carles li magnes <b>velz</b> est e redotez: recreanz ert de sa guerre mener	60	0	4	9	0	3
Passent cez puis e cez roches plus haltes, cez vals <b>parfunz</b> , cez destreiz anguisables	63	11	22	17	20	18
Païen d'Arabe des nefz se sunt <b>eissut</b> , puis sunt muntez es chevaux e es muls	NA	3	38	34	40	29
Li reis Marsilie est de guere vencud: vos li avez tuz ses castels <b>toluz</b>	NA	9	4	34	31	20
Cuntre le ciel <b>amsdous</b> ses mains ad juinz, si priet Deu que pareïs li duinst	NA	0	0	0	14	4
<b>Moyenne des groupes</b>	-	38	50	56	58	50

La même tendance se voit également après la contextualisation des mots : les groupes montrent une forte reconnaissance des mots transparents et semi-transparentes face à une identification plus troublée au fur et à mesure que le taux d'opacité augmente. L'on remarque que la divergence moyenne entre le G3 et le G4 est presque inexistante (hormis le cas du semi-opaque *culcez*, notamment) et le groupe ayant la moindre réussite est fort souvent le G1. Cette fois, le rôle des compétences linguistiques paraît avoir accru quelque peu – cela surtout parmi les mots non apparentés – mais les grandes variations non-linéaires insinuent que la reconnaissance est encore

affectée par les compétences individuelles ; d'où la présomption que l'AF se perçoit en tant qu'une langue étrangère. En ce qui concerne les deux parties de l'épreuve, les mots de l'AF ne sont pas toujours identifiés le mieux chez ceux qui l'ont étudié : nous traiterons cette question dans les paragraphes qui suivent.

Ainsi, pour donner plus de profondeur sur la reconnaissance des mots de l'AF, nous avons sélectionné 14 participants, faisant partie des deux derniers groupes (le G3 et le G4) pour une comparaison de l'influence des études antérieures en AF. Comme indiqué dans l'Annexe 4, ceux qui connaissent plusieurs langues romanes ont étudié l'AF avec beaucoup plus de probabilité que les autres participants : le G3 est constitué des personnes connaissant deux autres parlers romans modernes en plus du français, le G4 à son tour compte ceux en connaissant au moins trois. Nous faisons cette distinction en vertu de ne pas confondre les résultats suivants par un taux de reconnaissance général réduit, provoqué par la quasi-inexistence des personnes ayant étudié l'AF chez les G1 et G2. Il est à noter que nous ne pourrions pas établir de clairs rapprochements à la réalité, car l'échantillonnage reste limité et la durée, la période et l'intensité des études de l'AF n'avaient même pas été demandées dans le questionnaire ; partant, elles ne peuvent être prises en compte.

Les deux tableaux sous-jacents mettent en parallèle ces deux regroupements (les instruits et les non instruits dans l'AF ; tous venant des G3 et G4), chacun comptant sept personnes (d'où  $7 \times 5 = 35$ ). Quatre finnophones (FI2, FI6, FI14, FI15) et trois francophones (FR5, FR7, FR15) forment le groupe des instruits, alors que le groupe des non instruits consiste en cinq finnophones (FI3, FI9, FI11, FI12, FI13) et deux francophones (FR3, FR6). Sont indiqués également la distance de Levenshtein (qui détermine toujours l'ordre des termes) et les pourcentages de ces résultats.

Tableau 18a. Les mots seuls de l'AF, comparés selon les études de ce stade de langue.

Sans contexte	Mots	Distance	Instruits (IN) dans l'AF		Non instruits (NI) dans l'AF	
			Total / 35	%	Total / 35	%
	Altre	20	34	97	28	80
	Esteiles	25	18	51	25	71
	Meïsme	33	24	69	30	86
	Sustienget	36	29	83	23	66
	Curune	38	25	71	35	100
	Tuneire	38	5	14	5	14
	Meitet	43	0	0	5	14
	Vuldreie	44	24	69	22	63
	Culcez	50	16	46	4	11
	Preium	57	3	9	3	9
	Velz	60	0	0	0	0
	Parfunz	63	0	0	4	11
	Eissut	NA	4	11	3	9
	Toluz	NA	0	0	0	0
	Amsdous	NA	0	0	10	29

<b>Total / 525</b>	-	<b>182</b>	<b>35</b>	<b>197</b>	<b>38</b>
--------------------	---	------------	-----------	------------	-----------

Tableau 18b. Les mots contextualisés de l'AF, comparés selon les études de ce stade de langue.

Avec contexte	Mots	Distance	Instruits (IN) dans l'AF		Non instruits (NI) dans l'AF	
			Total / 35	%	Total / 35	%
	Altre	20	35	100	30	86
	Esteiles	25	35	100	35	100
	Meïsme	33	35	100	30	86
	Sustienget	36	31	89	28	80
	Curune	38	35	100	35	100
	Tuneire	38	35	100	32	91
	Meitet	43	5	14	5	14
	Vuldreie	44	35	100	33	94
	Culcez	50	19	54	9	26
	Preium	57	8	23	19	54
	Velz	60	1	3	2	6
	Parfunz	63	8	23	5	14
	Eissut	NA	12	34	14	40
	Toluz	NA	15	43	8	23
	Amsdous	NA	0	0	5	14
	<b>Total / 525</b>	-	<b>309</b>	<b>59</b>	<b>290</b>	<b>55</b>

Bien que la totalité des points des deux parties (IN : 182 + 309 = **491** vs NI : 197 + 290 = **487**) ne montre aucune divergence, les différences intérieures existent bel et bien. Même sans l'omniscience de FR3, qui – rappelons-nous – avait obtenu 95 % de tous les points dans la deuxième partie – le groupe des non instruits a eu maintes réussites dans cette mise en parallèle. Dans la première partie, nous pouvons relever *esteiles*, *meïsme* et *curune*. Les défaites des participants instruits se voient de manière plutôt accentuée dans ces cas-là, mais une fois le contexte apparu, ces quelques défaillances se changent en victoires. Souvent, en ce qui concerne les mots penchés vers l'opacité, les instruits les ont mieux identifiés par rapport aux non instruits (notamment *culcez*, *parfunz* et *toluz*). Malgré tout cela, ces derniers ont beaucoup mieux reconnu *preium* dans la deuxième partie. Sous la lumière de cette distribution des points, nous faisons valoir le fait qu'un échantillonnage plus étendu et mieux structuré est requis : par ailleurs, il ne semble pas pertinent de tirer la conclusion que seules les études de l'AF apportent de l'assistance décisive chez l'interprétation de ses mots.

En fin de compte, nous exposons les résultats de tous les corpus ensemble. Non seulement suscitent-ils de l'intérêt pour une comparaison générale, mais ils attestent méticuleusement les divergences à l'encontre de la linéarité impliquée par la distance de Levenshtein (D). Dans la colonne indiquant la moyenne des taux (M), nous avons marqué en bleu clair les mots identifiés mieux que prévu et en bleu foncé ceux identifiés pis. Les valeurs du tableau désignent les taux de reconnaissance, étant identiques à ceux des tableaux 16a-17b.

Tableau 19. Les taux de reconnaissance complets de l'épreuve lexicale.

Ancien français	D	1 <sup>ère</sup>	2 <sup>ème</sup>	M	IT, ES, PT anciens	D	1 <sup>ère</sup>	2 <sup>ème</sup>	M
Altre	20	83	94	89	Ventos	17	64	94	79
Esteiles	25	46	97	72	Viene	20	86	89	88
Meisme	33	58	90	74	Sofferto	22	81	87	84
Sustienget	36	66	60	63	Arvoses	29	41	74	58
Curune	38	75	93	84	Lontane	30	78	80	79
Tuneire	38	8	87	48	<b>Piedra</b>	<b>33</b>	<b>31</b>	<b>41</b>	<b>36</b>
<b>Meitet</b>	<b>43</b>	<b>4</b>	<b>7</b>	<b>6</b>	<b>Vizinhas</b>	<b>44</b>	<b>25</b>	<b>32</b>	<b>29</b>
<b>Vuldreie</b>	<b>44</b>	<b>63</b>	<b>93</b>	<b>78</b>	<b>Podriemos</b>	<b>56</b>	<b>78</b>	<b>94</b>	<b>86</b>
Culcez	56	17	35	26	Ascolte	56	32	50	41
Preium	57	10	27	19	Conhecem	64	46	56	51
<b>Velz</b>	<b>60</b>	<b>0</b>	<b>3</b>	<b>2</b>	Capra	67	25	36	31
Parfunz	63	3	18	11	Habedes	75	44	65	55
Eissut	NA	7	29	18	Amarellous	NA	44	78	61
Toluz	NA	0	20	10	Mugier	NA	31	86	59
<b>Amsdous</b>	<b>NA</b>	<b>8</b>	<b>4</b>	<b>6</b>	Fiume	NA	18	48	33
<b>Total</b>	<b>-</b>	<b>30</b>	<b>50</b>	<b>40</b>	<b>Total</b>	<b>-</b>	<b>48</b>	<b>67</b>	<b>58</b>

Les plus grands écarts entre les deux matériaux s'expliquent donc par la reconnaissance assez élevée des mots semi-opaques, opaques et non apparentés en espagnol, en italien et en portugais. Dans la sous-section suivante alors, nous essaierons de trouver des justifications pour ces quelques phénomènes inattendus, colorés en bleu clair et en bleu foncé.

#### 4.2.5 Entrevue informelle

Nous avons vu que la distance de Levenshtein influence sur la reconnaissance de mots, le phénomène se manifestant le plus nettement chez les mots de l'AF (v. 4.2.4), et cela surtout dans un environnement qui néglige le contexte. Or, nous relevons le sort de sept mots qui se sont le plus distinctement égarés de cette échelle prédéfinie (v. tableau 19). Cinq d'entre eux n'ont pas été reconnus aussi bien que prévu (*meitet*, *piedra*, *velz*, *vizinhas*, *amsdous*) et deux d'entre eux ont été identifiés mieux que l'on n'aurait pu le supposer (*podriemos*, *vuldreie*). Ces derniers, regroupés ici comme des mots « faciles », ne sont pas plus nombreux que cela, parce qu'il ne s'agit pas de notre objectif principal. Qui plus est, leur amélioration est le plus souvent expliquée par une connaissance antérieure du mot, l'aspect déjà commenté en 3.1.2.1. La nécessité de cette entrevue est motivée par son importance générale pour ce genre de travail et par la complexité de cette perspective pour qu'un traitement cohérent puisse être établi dans le cadre théorique.

Fin janvier 2023, nous avons organisé une entrevue informelle avec dix participants sélectionnés au hasard : trois francophones (FR6, 7, 10) et sept finnophones (FI2, 4, 6, 7, 8, 12, 14). Nous leur avons posé plusieurs questions sur la morphologie du terme, les manières d'associer et la transparence du contexte – notons maintenant que seuls les mots contextualisés étaient visualisés :

nous fournissant ainsi davantage de possibilités pour une discussion élargie. Nous exposons plus loin dans cette sous-section d'autres observations intéressantes, faites par certains participants sitôt l'épreuve finie. Avant cela toutefois, dans le tableau 20 qui suit, nous mettons en relief les commentaires les plus pertinents derrière les taux de reconnaissance inattendus. Les finnophones ont donné leurs commentaires en finnois et nous les avons traduits en français – tout en condensant le contenu de chaque observation. De plus, nous avons souligné tout ce qui nous paraît essentiel au niveau de la reconnaissance des mots. Pour chaque mot, sont désignés également la catégorie de transparence (Cat.) et le taux de reconnaissance général (Re.) chez tous les participants (cf. tableau 19).

Tableau 20. Les commentaires donnés par les participants.

Mot	Trad.	Cat.	Re.	Commentaires
<b>Des mots difficiles à traduire</b>				
<b>Piedra</b>	'pierre'	STP	41	<b>FR6, FI12</b> : <u>N'appartient pas au vocabulaire essentiel, malgré le fait d'avoir appris l'espagnol</u> <b>FR6, FR7, FI7</b> : La lettre 'd' cause une perturbation <b>FI2, FI4</b> : L'idée du <i>pie</i> d dans la 1 <sup>ère</sup> partie qui a persistée <b>FR10, FI14</b> : Association sémantique avec la nourriture (grâce au contexte) (cf. Laufer, 1996)
<b>Vizinhas</b>	'voisines'	SO	32	<b>FR7, FR10, FI2, FI4, FI6</b> : Un contexte long qui aide peu, tout n'est pas compris <b>FR6, FR7, FI2, FI6</b> : L'orthographe perturbe ( <b>FR7</b> : <u>surtout 'h' qui coupe le mot en deux</u> et éloigne le mot de l'italien <i>vicine</i> ) (cf. Pottier, 1997) <b>FR10</b> : Recours à l'anglais <i>vicinity</i> , qui a pourtant aidé <b>FR6</b> : Connaissance du portugais seulement à l'oral <b>FR7, FI14</b> : <u>Influence des premières lettres</u> → <i>vision, visité</i> <b>FI4, FI7</b> : Appui sur la langue plus forte (le français), donnant parfois <u>trop de confiance en ses propres compétences</u>
<b>Meitet</b>	'moitié'	SO	7	<b>FR6, FI6, FI8, FI14</b> : <i>Moitié</i> ne vient pas à l'esprit <b>FR6, FI2, FI4, FI7, FI14</b> : <u>La lettre 'e' à la place de 'o' perturbe</u> <b>FR10</b> : Recours à l'anglais <i>to meet</i> <b>FR7, FI7</b> : Le contexte fait penser à la politique, à l'architecture <b>FI4, FI12</b> : L'ordre des mots ( <b>FI4</b> : n'est pas suivi par <i>de</i> ), un contexte obscur <b>FI6</b> : Le contexte ne délimite rien
<b>Velz</b>	'vieux'	O	3	<b>FI2, FI7, FI8, FI12</b> : Le contexte difficile n'a pas aidé <b>FI2, FI14</b> : Ne ressemble à rien ( <b>FI2</b> : 'lz' une rare combinaison de lettres) <b>FR6</b> : La forme verbale <i>est</i> qui suit → adjectif <b>FR6, FI4</b> : En voyant <i>Charlemagne</i> , sa position est venue à l'esprit <b>FR7, FI6</b> : L'ordre inhabituel de mots attire l'attention <b>FR10</b> : Le contexte fait penser à la politique <b>FI14</b> : Le mot est très court
<b>Amsdous</b>	'les deux'	NA	4	<b>FR7, FR10, FI2, FI8</b> : Ne ressemble à rien ( <b>FR10</b> : <u>Il est important de pouvoir associer le mot à quelque chose que l'on connaît déjà</u> ) <b>FI4, FI14</b> : Le début de la phrase ( <i>Cuntre le ciel</i> ) est trop évident, provoquant <u>une traduction sans réfléchir au sens du terme demandé</u> <b>FR7</b> : <u>Le début du mot (ams) ne permet pas de percevoir sa fin (dous)</u> <b>FI6</b> : Très éloigné du français moderne
<b>Des mots faciles à traduire</b>				
<b>Podriemos</b>	'pourrons'	SO	94	<b>Remarque générale</b> : Connaissance de l'espagnol actuel

			<b>FI2, FI4, FI6, FI7, FI8, FI12</b> : <u>Le contexte permet l'identification</u> (particule négatif <i>no</i> + verbe à l'infinitif <i>vivir</i> → la présence d'un verbe auxiliaire commençant par 'p' ; personne clarifiée par la désinence -os) (cf. Murillo Puyal et al., 2008) <b>FR7</b> : Le contexte montre la conséquence ( <i>si...no</i> ) <b>FR10</b> : <i>Poder</i> est parmi les verbes espagnols tôt appris	
<b>Vuldreie</b>	'voudrais'	SO	93	<b>FI2, FI4, FI6, FI7, FI12, FI14</b> : Le contexte permet l'identification (v. <i>supra</i> ) <b>FR6, FR10, FI6, FI7, FI14</b> : <u>La ressemblance au verbe <i>vouloir</i> par le « timbre » et la prononciation latine [u]</u> (noté également par <b>FI4</b> entre <i>sustienget</i> et sa traduction <i>soutienne</i> ) <b>FR7, FI2, FI4, FI7, FI14</b> : Le verbe à l'infinitif <i>estre</i> qui suit <b>FI6</b> : Un verbe auxiliaire commençant par 'v' <b>Mais FI8 perturbée par la non-reconnaissance de la langue !</b>

**STP = semi-transparent, SO = semi-opaque, O = opaque, NA = non apparenté**

Pour résumer le contenu de tous ces commentaires, la reconnaissance courante de mots est rendue possible par une certaine clarté du contexte, une structure syntaxique logique, l'association des notions et l'absence des lettres inhabituelles dans le mot. Il semblerait donc que la présence d'une seule lettre « étrange » puisse entraver la compréhension du mot entier. Qui plus est, la confiance excessive en ses propres compétences ne doit pas non plus être négligée (cf. les cas de *vizinhas* et *amsdous* chez FI4 et FI14). En revenant à l'étude oculométrique de Deutsch et Rayner (1999), il est fort évident que le début du mot reçoit le plus d'attention. C'est pour cette raison que nous estimons le changement de la deuxième lettre du mot d'être d'une importance majeure. Étant donné que la variation entre 'u' et 'ou' n'avait pas importuné autant nos sujets d'expérience, cet effet peut être laissé de côté. L'augmentation de la valeur de la deuxième lettre entraîne sans doute des conséquences concernant l'ordre et les classes de transparence, mais le travail déjà terminé – bien évidemment – nous n'entrerons pas dans leur examen.

Remarquons encore que les commentaires aléatoires que nous avons reçus pendant les séances permettent – eux aussi – de mieux comprendre la structuration de l'épreuve à partir d'une perspective extérieure. FI7 a révélé (après l'entrevue) que la mise en doute de l'existence réelle des mots (1<sup>ère</sup> partie) ainsi qu'être consciente de la présence d'une autre langue étrangère – incitant donc à appliquer une méthode de traduction différente (2<sup>ème</sup> partie) – avaient troublé son propre processus de la reconnaissance. Le même vaut pour une situation où la langue elle-même n'est pas reconnue (FI8) : en vue du participant, délimiter le nombre des langues possibles avant de commencer le test serait donc avisé. FR9 a avoué que les réponses données dans la première partie ont pu rester à l'esprit pendant la deuxième partie (cf. le tableau 20 : FI2 et FI4 pour *pedra*). FR9 avait donc traduit *fume* comme « *fumier* » sans contexte : une fois le contexte apparu, elle a remarqué qu'il s'agissait d'un sens complètement différent, mais la première traduction « hantait » son esprit. FR11 a constaté que la deuxième partie était la plus difficile, parce que le lien logique entre le mot

cherché et son contexte n'était pas toujours évident. Dans la première partie l'on pouvait tout simplement suivre sa propre intuition, malgré le fait que la réponse soit erronée. F11 a ajouté que certains mots s'étaient rendus beaucoup plus clairs avec contexte, tandis que d'autres avaient complètement perdus cette « clarté ». Ces deux dernières déclarations (F11, FR11) soutiennent la théorie de Laufer (1996) sur la transparence de faux cognats – ce qu'ici doit être appliquée plutôt au contexte – que l'interprétation réussie d'un mot nécessite la présence d'un contexte qui ne déforme pas le sens de celui-ci.

## 5 Discussion et conclusion

Nous venons d'étudier comment les mots des langues romanes médiévales sont reconnus par les locuteurs du français. Afin d'effectuer cette recherche qualitative, nous avons élaboré une épreuve lexicale qui était constituée de 30 mots recueillis dans les textes anciens. Nous estimions que la reconnaissance des mots est facilitée par le contexte, par la proximité graphique (c'est-à-dire la *transparence*) des mots face à leurs traductions en français moderne, par la langue française comme langue maternelle ainsi que par une expertise linguistique étendue dans les langues romanes. Puisque l'importance de connaître ses origines linguistico-culturelles est pertinente de nos jours encore (le fait constaté par plusieurs chercheurs : McCann et al., 2003 ; Zeevaert & ten Thije, 2007 ; Blanche-Benveniste, 2008 ; entre autres), nous avons analysé le sujet du point de vue plutôt pratique. Voici encore nos questions de recherche, auxquelles les réponses ont été obtenues grâce à l'épreuve lexicale :

- 1) *Dans quelle mesure la présence du contexte aide-t-elle à identifier le sens du mot, selon la langue maternelle du participant ?*
- 2) *Comment les mots de l'ancien français sont reconnus par rapport aux mots venant d'autres anciennes langues romanes, selon leur distance orthographique ?*
- 3) *Comment les compétences linguistiques dans les langues romanes influent sur la reconnaissance des mots ?*

Notre épreuve était constituée de deux parties, ce qui nous a permis d'analyser l'effet du contexte. Les participants ont été classifiés d'abord selon leur langue maternelle (le français ou le finnois) : c'est ainsi que nous avons pu établir une comparaison statistique au niveau des classes de transparence. Les corpus par nous collectés ont été séparés (l'AF vs les autres anciennes langues romanes) pour donner plus d'accent à notre objectif principal, c'est-à-dire, la reconnaissance des

mots de l'ancien français. Nous avons organisé tous les mots selon leur distance orthographique afin de répondre à notre hypothèse sur la meilleure identification des mots transparents. Ensuite, les participants ont été regroupés derechef, cette fois en quatre groupes selon leurs compétences linguistiques en langues romanes actuelles, à l'aide duquel il nous a été possible de repérer les différences liées à ces compétences. Puis, nous avons analysé brièvement l'influence potentielle des études de l'AF sur l'identification de ses mots, et enfin, nous avons comparé les commentaires des participants lors d'une entrevue informelle.

Concernant la discussion autour de l'analyse des résultats, nous la commençons par répondre à nos questions de recherche. Nous avons vu que le contexte améliore les résultats (Bogaards, 1994) – cela presque sans exception. Le changement des résultats (v. tableaux 10 & 11) est soumis à l'envie personnel du participant de comprendre ce qui est dit (Laufer, 1996), au temps qu'il est prêt à y consacrer (non pas toujours généralisable), à la totalité des points obtenus dans les deux parties, ainsi qu'aux méthodes de lecture et de perception, que nous n'avons pas observées plus en détail. Selon nos résultats, la réussite s'est déterminée souvent par les compétences individuelles et non pas tant par la langue maternelle du participant – une réorganisation du matériau et une classification bien plus systématique de données sont requises afin de conduire une étude plus approfondie sur l'effet que la langue maternelle puisse avoir sur la compréhension des mots étrangers. La remarque de Bogaards (1994 : v. 2.1.1.) sur ce que les locuteurs natifs d'une langue fussent plus au courant des changements historiques ne semble pas ressortir des résultats que nous avons obtenus.

Il se peut que le domaine d'expertise et la profession influent sur la capacité de réussir dans une telle expérimentation, mais notre échantillonnage n'était pas suffisant pour permettre de le savoir. Ce qui est clair, pourtant, c'est que les compétences étendues sur le plan des langues étrangères habilite la compréhension et l'identification des mots et des contextes, même historiques. Il ressort de l'entrevue que la reconnaissance d'un mot est quelquefois gênée par la connaissance existante d'une langue. Nous pouvons déduire, conséquemment, que la création d'une épreuve lexicale pour comparer les langues apparentées nécessite bien plus d'attention que nous avons pu y prêter. Si le matériau était constitué des textes modernes, il se pourrait que la compréhension se fasse plus couramment, ce qui n'est pas assez souvent le cas dans notre travail. Lors de la sélection du matériau, la négligence des compétences linguistiques acquises dans les langues romanes modernes a abouti au déséquilibre général chez l'identification des termes choisis, et encore, à la séparation des corpus.



Bien que nous craignions que le test soit trop exigeant pour nos sujets d'expérience, il s'est bien avéré qu'il était parfait pour la juxtaposition des versés et les non-connaisseurs. Maintes fois, la fonction du mot a été identifiée, même si le sens exact n'est pas venu à l'esprit. Eussions-nous fait le choix des mots du matériau encore plus méthodiquement, et eût leur nombre été plus élevé, nous aurions pu établir – éventuellement – la vraie influence exercée par la distance de Levenshtein. Si la complexité du contexte était régulée, une description plus précise sur l'influence de la distance orthographique des mots individuels serait atteinte : nous n'avions fait aucune attention à l'analyse des contextes historiques et poétiques dans le cadre théorique, même si les structures syntaxiques difficiles ont clairement troublé nos sujets d'expérience (v. 4.2.5). Il ne faut pas non plus oublier l'altération des résultats si nous avons inclus les participants rejetés ou si nos méthodes de notation avaient été recalculées et vérifiées par une autre personne. Nous aurions dû suivre le classement de Bezooijen et Gooskens (v. 2.2.2) encore plus minutieusement, et non pas augmenter la difficulté de l'épreuve autant. Par ailleurs, analyser le temps écoulé lors de l'épreuve serait sûrement lucratif dans des études à venir : nos résultats insinuent que les résultats s'amélioreraient considérablement par l'emploi élevé du temps – plus de preuve est cependant requise.

Pour répondre à nos hypothèses (les facteurs qui rendent la reconnaissance plus facile, v. l'introduction de ce chapitre), il est évident que le contexte et une expertise linguistique étendue aplanissent la compréhension de mots, comme prévu. La proximité orthographique n'est pourtant pas toujours le facteur facilitant : sans prendre en considération la connaissance de langues, un contexte et une syntaxe très transparents rendent l'identification du mot plus aisée, pourvu qu'il puisse être associé morphologiquement ou sémantiquement à sa traduction. En revanche, la reconnaissance devient bien plus difficile lorsque le contexte est sémantiquement complexe et surtout si les lettres du début du mot sont modifiées (Lutjeharms, 2007) : cela étant dit, nous aurions pu recueillir des données intéressantes en ce qui concerne le changement régulier de *f* latine en *h* de l'espagnol moderne (*facere* > *hacer*, 'faire'), en étudiant ce phénomène chez les non-connaisseurs de l'espagnol. Hélas, cet aspect ne se présentait pas encore pendant la période de la rédaction du Poème du Cid, attesté par notre phrase « *D'esto que habedes **fecho** [...]* » (mod. *hecho*, 'fait'). Au lieu d'essayer de faire des rapprochements motivés par l'intuition (Blanche-Benveniste, 2008), les participants ont souvent choisi de traduire le mot opaque en s'appuyant sur les premières lettres de celui-ci.

Pour ce qui est de l'ancienne version de l'épreuve lexicale (v. 3.1.3), nous avons prévu d'inclure une cinquième catégorie – consacrée aux faux-amis – dont le but était de montrer combien le

contexte peut aider à réinterpréter le sens du mot. Cependant, nous nous sommes débarrassé de ce groupe, car il aurait augmenté considérablement la pesanteur totale de l'épreuve et il n'était pas comparable aux autres catégories que nous avons établies. Pour en donner un exemple, les décisions traductionnelles prises par les sujets d'expérience auraient été très difficiles à analyser : ne pas attribuer de points dans la première partie pour une mauvaise traduction serait injuste, à cause du manque du contexte. Sans contexte, l'on pourrait traduire '*nombre*' par « *nombre* », ce qui serait correct dans le cas du français ; néanmoins, ce que les participants ignoreraient, c'est que ce terme apparaît dans un contexte espagnol (*¡En el **nombre** del Criador, que no pase por al!*) où il signifie « *nom* ».

Il y existe aussi quelques aspects auxquels nous n'avons pas prêté beaucoup d'attention. D'abord, construire une entrevue adaptée à nos besoins n'était pas possible avant que le recueil des résultats n'eût été achevé. C'est pour cela que nous l'avons remplacée par demander la langue apparaissant dans le contexte. Pourtant, l'enregistrer ne nous semblait pas très pertinent devant nos questions de recherche. Nous relevons cependant que la ressemblance des deux stades du français (ancien vs moderne) était bien aperçue, mais de telle manière qu'il y avait beaucoup d'hésitation de deviner la langue correctement. Presqu'un quart des participants avaient proposé que ces deux phrases – « *De la citet l'une **meitet** est sue* » et « *Laisum le plait e si **preium** le rei que Guenelun cleimt quite ceste feiz* » – étaient rédigées en latin. Pour finir, la considération de l'anglais en tant qu'un élément délimitant notre choix de mots est restée trop superficielle, car du moins trois mots – *tuneire*, *preium*, *vizinhas* – étaient reconnaissables à travers la *lingua franca* du monde (*thunder*, *pray*, *vicinity*) ; ce dernier étant identifié par FR10 à l'aide de l'anglais et non pas sa langue maternelle, le français.

En guise de conclusion, notre recherche offre une base bien justifiée pour les moyens de scruter l'intercompréhension dans les langues apparentées historiques. Elle présente explicitement comment il faut procéder dans la création d'une épreuve adaptée à l'étude comparative des langues. Nous avons différencié les résultats selon les langues de provenance (AF vs les autres parlers romans médiévaux), mais ce n'est qu'en raison du choix de matériau que nous avons agi de telle manière. Des études approfondies dans le domaine de la psycholinguistique ou la phonétique pourraient répandre encore plus de lumière sur les détails de cette thématique multidimensionnelle. Tout compris, les résultats de notre étude insistent indubitablement sur l'importance de l'apprentissage général des langues étrangères, car même la connaissance d'une seule permet d'en comprendre plusieurs autres.

## Références






- Andrieux-Reix, N. (1993). *Ancien et moyen français. Exercices de phonétique*. Paris : Presses Universitaires de France.
- Beck, I. L., Perfetti, C. A. & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521. Disponible sur : <https://psycnet.apa.org/doi/10.1037/0022-0663.74.4.506>
- Bédier, J. (1922). *La Chanson de Roland* (traduction). Paris : L'Édition d'Art Henri Piazza.
- Bezooijen, R. van & Gooskens, C. (2007). Interlingual text comprehension: Linguistic and extralinguistic determinants. Dans J. D. ten Thije & L. Zeevaert (dir.), *Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts*, 6, 249-264. Amsterdam : Benjamins.
- Blanche-Benveniste, C. (2008). Comment retrouver l'expérience des anciens voyageurs en terres de langues romanes ? Dans V. Conti & F. Grin (dir.), *S'entendre entre langues voisines : vers l'intercompréhension* (p. 33-51). Chêne-Bourg : Éditions Georg.
- Blanche-Benveniste, C. & Valli, A. (1997). Une grammaire pour lire en quatre langues. Dans C. Blanche-Benveniste & A. Valli (dir.), *L'intercompréhension : le cas des langues romanes. Le français dans le monde. Recherches et applications*, 22, 33-37. Paris : Hachette.
- Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris : Éditions Hatier/Didier.
- Borin, L. (2013). Introduction: The why and how measuring linguistic differences. Dans L. Borin & A. Saxena (dir.), *Approaches to Measuring Linguistic Differences* (p. 1-28). Berlin : Walter de Gruyter.
- Coirier, P., Gaonac'h, D. & Passerault J-M. (1996). *Psycholinguistique textuelle. Approche cognitive de la compréhension et de la production des textes*. Paris : Armand Colin/Masson.
- Collins Dictionary. Disponible sur : [Collins Online Dictionary | Definitions, Thesaurus and Translations \(collinsdictionary.com\)](https://www.collinsdictionary.com)
- Deutsch, A. & Rayner, K. (1999). Initial Fixation Location Effects in Reading Hebrew Words, *Language and Cognitive Processes*, 14(4), 393-421. Disponible sur : <https://doi.org/10.1080/016909699386284>
- Doetjes, G. (2010). *Akkommodation und Sprachverstehen in der interskandinavischen Kommunikation*. Hambourg : Université de Hambourg.
- EUR-Lex. (2016). Règlement Général sur la Protection des Données. Disponible sur : [EUR-Lex - 32016R0679 - EN - EUR-Lex \(europa.eu\)](https://eur-lex.europa.eu/32016R0679)


- Frick, M., Grünthal, R. & Praakli, K. (2018). A linguistic encounter between neighbors and relatives: Reconnecting Estonian and Finnish in terms of contemporary multilingualism. *Multilingual Practices in Finno-Ugric Communities* (p. 163-196).
- Herman, J. (1967a). *Le latin vulgaire*. Paris : Presses Universitaires de France.
- Herman, J. (1967b). *Précis d'histoire de la langue française*. Budapest : Tankönyvkiadó.
- Infopédia. Disponible sur : [infopedia.pt](http://infopedia.pt) - Porto Editora
- Institut des langues de la Finlande. (2021). *Kielitoimiston sanakirja*. Dictionnaire rédigé par Kotimaisten kielten keskus. Disponible sur : [Kielitoimiston sanakirja](http://kielitoimiston.sanakirja.fi)
- Janson, T. (2012). *The History of Languages. An Introduction*. Oxford : Oxford University Press.
- Kameenui, E. J., Carnine, D. W. & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly*, 17, 367-388. Disponible sur : <https://doi.org/10.2307/747525>
- Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, 6 (*Vocabulary acquisition*), 10-20.
- Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading*, 15(2), 95-103. Disponible sur : <https://psycnet.apa.org/doi/10.1111/j.1467-9817.1992.tb00025.x>
- Laufer, B. (1996). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. Dans J. Coady & T. Huckin (dir.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy* (p. 20-34). Cambridge : Cambridge University Press.
- Lutjeharms, M. (1997). Reading skills for weak learners. Dans G. Egloff & A. Fitzpatrick (dir.), *Languages for Work and Life: The Council of Europe and vocational oriented language learning* (p. 164-177). Strasbourg : Publications du Conseil de l'Europe.
- Lutjeharms, M. (2007). Processing levels in foreign-language reading. Dans J. D. ten Thije & L. Zeevaert (dir.), *Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts*, 6, 265-284. Amsterdam : Benjamins.
- McCann, W. J., Klein, H. G. & Stegmann, T. D. (2003). *EuroComRom - The Seven Sieves: How to read all the Romance languages right away*. Aix-la-Chapelle : Shaker Verlag.
- Murillo Puyal, J., Piccaluga, M. & Harmegnies, B. (2008). Pour un français partenaire dans l'aire romane : aspects psycholinguistiques et didactiques de l'intercompréhension entre locuteurs de langues latines. Dans V. Conti & F. Grin (dir.), *S'entendre entre langues voisines : vers l'intercompréhension* (p. 341-370). Chêne-Bourg : Éditions Georg.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12-25.

- Oller, J. & Tullius, J. R. (1973). Reading skills of non-native speakers of English. *International Review of Applied Linguistics in Language Teaching*, 11, 69-80. Disponible sur : <https://doi.org/10.1515/iral.1973.11.1-4.69>
- Palokangas, A. (2018). *Viron yleisten sananmuotojen ymmärtäminen*. (Mémoire de master). Université de Jyväskylä.
- Piccaluga, M. & Harmegnies, B. (2008). Aux sources de l'intercompréhension. Facteurs psycholinguistiques de la qualité interprétative en langues proches. Dans V. Conti & F. Grin (dir.), *S'entendre entre langues voisines : vers l'intercompréhension* (p. 169-195). Chêne-Bourg : Éditions Georg.
- Posner, R. (1996). *The Romance languages*. Cambridge Language Surveys. Cambridge : Cambridge University Press.
- Pottier, B. (1997). La parenté des langues romanes. Dans C. Blanche-Benveniste & A. Valli (dir.), *L'intercompréhension : le cas des langues romanes. Le français dans le monde. Recherches et applications*, 22, 75-82. Paris : Hachette.
- Rastier, F. (1991). *Sémantique et recherches cognitives*. Paris : Presses Universitaires de France.
- Ringbom, H. (1987). *The Role of the First Language in Foreign Language Learning*. Clevedon : Multilingual Matters.
- Schmitt Jensen, J. (1997). L'expérience danoise et les langues romanes. Dans C. Blanche-Benveniste & A. Valli (dir.), *L'intercompréhension : le cas des langues romanes. Le français dans le monde. Recherches et applications*, 22, 95-108. Paris : Hachette.
- Simone, R. (1997). Langues romanes de toute l'Europe, unissez-vous ! Dans C. Blanche-Benveniste & A. Valli (dir.), *L'intercompréhension : le cas des langues romanes. Le français dans le monde. Recherches et applications*, 22, 25-32. Paris : Hachette.
- Stahl, S. (1983). Differential word knowledge and reading comprehension. *Journal of Reading Behaviour*, 15(4), 33-50. Disponible sur : <https://doi.org/10.1080/10862968309547495>
- Steffensen, M. S., & Joag-Dev, C. (1984). Cultural knowledge and reading. Dans J. C. Alderson & A. H. Urquhart (dir.), *Reading in a foreign language* (p. 48-61). Londres : Longman.
- Stegmann, T. D. (2005). EuroComRom – Les sept tamis pour le plurilinguisme réceptif. *Synergies Italie* (p. 35-42). Sylvains-Lès-Moulins : GERFLINT.
- Sundqvist, L. (1986). *Lexical inferencing among Swedish- and Finnish-speaking primary-school pupils*. (Mémoire de master inédit). Université d'Åbo Akademi.
- Zeevaert, L. & ten Thije, J. D. (2007). Introduction. Dans J. D. ten Thije & L. Zeevaert (dir.), *Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts*, 6, 1-22. Amsterdam : Benjamins.

# Annexes

## Annexe 1. Questionnaire préliminaire

Esikatsela & testaa    Sivu 1 / 2 Vie kysely:  

**UNIVERSITY OF TURKU**  
**Questionnaire**

Ceci est un questionnaire à compléter pour participer à un test, lié à mon mémoire de master qui examine l'intercompréhension des langues romanes du point de vue des locuteurs du français. Les réponses seront traitées de manière strictement anonyme et les informations personnelles seront définitivement supprimées à la fin de cette année (2022), selon RGPD.

**1. Règlement général sur la protection des données (RGPD) \***

J'accepte que l'on utilise mes réponses dans un mémoire de master sous forme anonyme

Je refuse que l'on utilise mes réponses dans un mémoire de master sous forme anonyme

**2. Informations personnelles**

Prénom \*

Nom de famille \*

Adresse électronique \*

Nationalité \*

Langue(s) maternelle(s) \*

**3. Année de naissance \***

Sélectionner

**4. Sexe \***

Sélectionner

**5. Éducation \***  
Mentionnez ici votre dernier diplôme. P. ex. lycée, université, IUT...


**6. Domaine d'études ou d'expertise \***  
P. ex. linguistique, pharmacie, informatique...

**7. Profession \***  
P. ex. étudiant, enseignant, technicien, ingénieur...

1 de 2

Esikatsele & testaa Työpöytä Tabletti Mobiililaitte

Sivu 2 / 2 Vie kysely: PDF Word


**UNIVERSITY OF TURKU**

### Questionnaire

**8. Quelles autres langues romanes connaissez-vous et à quel niveau ?  
Faites l'évaluation vous-même. (0 = aucune connaissance, 5 = correspond aux compétences en langue maternelle/niveau C2)**

	0	1	2	3	4	5
Espagnol *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Italien *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Portugais *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Catalan *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Roumain *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Autre, laquelle? <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**9. Avez-vous étudié le latin ? \***

Oui, pendant combien d'années ?

Non

**10. Avez-vous suivi un cours sur...? \***

L'ancien français

Le moyen français

La linguistique comparée

Une thématique similaire, laquelle ?

Non, je n'ai suivi aucun cours de ce genre

La littérature classique

La poésie

**11. Si vous le souhaitez, vous pouvez commenter librement comment vous apprenez les langues**

**12. Voulez-vous participer à une interview ? \***

Oui

Non

**13. Êtes-vous intéressé(e) à savoir les résultats de cette étude ? \***

Oui

Non

2 de 2



## Annexe 2. Épreuve lexicale (avec réponses)

À gauche sont situés les trente mots du matériau (dans la deuxième image ils sont entourés par du contexte, soulignés). L'on voit aussi les réponses de deux participants. À droite se trouve la colonne de notation, contenant les points attribués selon les traductions données.

	Vos traductions	64/150
meitet	<u>mettre</u>	
fiume	<u>fumée</u>	
pie dra	Pierre	5
sofferto	Souffrir	4
capra	<u>chèvre</u>	5
vuldreie	Volonté	3
preium	Premier	
tuneire	Tenir	
velz	Veau	
altre	<u>autre</u>	5
lontane	Lointaine	4
conhecem	Connaissance	3
ventos	Venteux	3
eissut	Eut	
meïsme	Les mêmes	4
vizinhas	Vision	
ascolte	Ecouter	4
parfunz	Parfum	
arvores	<u>arborées</u>	3
curune	Testicule	
habedes	Nous avions	4
esteiles	Atteindre	
mugier	Mugir	
culcez	Coucher	4
toluz	Tenue	
viene	Viens	4
podriemos	Nous pourrions	5
amsdous	???	
sustienget	<u>soutenu</u>	4
amarellós	Cerises	

	Vos traductions	C'est quelle langue ?	87/150
1. De la citet l'une <u>meitet</u> est sue	<u>rencontrer</u>	<u>portugais</u>	
2. Vidi genti a la riva d'un gran <u>fiume</u>	fumée	italien	
3. Del agua feziste vino y de la <u>pie dra</u> pan	pâte	espagnol	
4. E come noi lo mal ch'avem <u>sofferto</u> perdoniamo a ciascuno	souffre	portugais	4
5. Quando 'l corno de la <u>capra</u> del ciel col sol si tocca	bout	portugais	
6. Si grant doel ai que jo ne <u>vuldreie</u> estre!	<u>voudrais</u>	<u>roumain</u>	5
7. Laisum le plait e si <u>preium</u> le rei que Guenelun cleimt quite ceste feiz	<u>prévoit</u>	<u>roumain</u>	
8. Orez i ad de <u>tuneire</u> e de vent, pluies e gresilz desmesurement	Tonnerre	<u>portugais</u>	5
9. Carles li magnes <u>velz</u> est e redotez: recreanz ert de sa guerre mener	<u>droit</u>	<u>portugais</u>	
10. L'un gist sur l' <u>altre</u> e envers e adenz	entrée	catalan	
11. Quant' è che tu venisti a piè del monte per le <u>lontane</u> acque?	lointain	portugais	4
12. Mais bravos e mais feros se <u>conhecem</u> pela fama nas obras e nos feitos	connaissent	catalan	5
13. Lá nas covas marítimas entrarão fugindo á tempestade e <u>ventos</u> duros	vents	espagnol	5
14. Païen d'Arabe des nefes se sunt <u>eissut</u> , puis sunt muntez es chevaux e es muls	<u>descendus</u>	catalan	5
15. Mais lui <u>meïsme</u> ne volt mettre en ubli	<u>même</u>	<u>catalan</u>	5
16. Vio ter a muitas villas suas <u>vizinhas</u> , que por armas tu, Sancho, humildes tinhas	proches	catalan	5
17. O dolce padre mio, se tu m' <u>ascolte</u> , io ti dirò	consultes	italien	1
18. Passent cez puis e cez roches plus haltes, cez vals <u>parfunz</u> , cez destreiz anguisables	<u>parfums</u>	<u>catalan</u>	
19. Estava a terra em montes revestida, de verdes ervas e <u>arvores</u> floridas	arbres	espagnol	5
20. Puis si li tolent sun sceptre e sa <u>curune</u>	<u>couronne</u>	<u>italien</u>	5
21. D'esto que <u>habedes</u> fecho vos no perderedes nada	avais	espagnol	4
22. Clere est la lune e les <u>esteiles</u> flambient	<u>étoiles</u>	<u>portugais</u>	5
23. Por amor de mi <u>mugier</u> y de mis fijas amas	mère	espagnol	3
24. Passet li jurz, la nuit est aserie: <u>culcez</u> s'est li reis en sa cambre voltice	Celui qui	<u>roumain</u>	
25. Li reis Marsilie est de guere vencud: vos li avez tuz ses castels <u>toluz</u>	<u>détruit</u>	<u>catalan</u>	2
26. ¿No te <u>viene</u> en miente en València lo del leon?	viens	espagnol	4
27. Si no, en esta tierra angosta no <u>podriemos</u> vivir	pourrions	espagnol	5
28. Cuntre le ciel <u>amsdous</u> ses mains ad juinz, si priet Deu que pareis li duinst	baisse	<u>roumain</u>	
29. N'en avrai ja ki <u>sustienget</u> m'onur; suz ciel ne quid averi ami un sul	<u>soutient</u>	<u>roumain</u>	5
30. Cheios de terra e crespos os cabelos, a boca negra, os dentes <u>amarellós</u>	jaunes	catalan	5



### Annexe 3. Liste simplifiée des mots du matériau

<b>Mot du matériau</b>	<b>Traduction</b>	<b>Mot du matériau</b>	<b>Traduction</b>
<b>Altre</b>	Autre	<b>Meitet</b>	Moitié
<b>Amarellous</b>	Jaunes	<b>Mugier</b>	Femme
<b>Amsdous</b>	(Tous) les deux	<b>Parfunz</b>	Profonds
<b>Arvores</b>	Arbres	<b>Piedra</b>	Pierre
<b>Ascolte</b>	Écoutes	<b>Podriemos</b>	Pourrons, pourrions
<b>Capra</b>	Chèvre	<b>Preium</b>	Prions
<b>Conhecem</b>	Connaissent	<b>Sofferto</b>	Souffert
<b>Culcez</b>	Couché	<b>Sustienget</b>	Soutient, soutienne
<b>Curune</b>	Couronne	<b>Toluz</b>	Pris, enlevés, conquis
<b>Eissut</b>	Sortis, descendus	<b>Tuneire</b>	Tonnerre
<b>Esteiles</b>	Etoiles	<b>Velz</b>	Vieux
<b>Fiume</b>	Rivière, fleuve	<b>Ventos</b>	Vents
<b>Habedes</b>	Avez	<b>Viene</b>	Vient
<b>Lontane</b>	Lointaines, éloignées	<b>Vizinhas</b>	Voisines, voisins, proches
<b>Meïsme</b>	Même	<b>Vuldreie</b>	Voudrais, voudrai

## Annexe 4. Les regroupements selon les compétences linguistiques

La distribution des participants dans les quatre groupes, accompagnée par la profession de chacun. Les nuances de vert indiquent la connaissance (1-5) des langues romanes. La colonne « Autre(s) » comporte le catalan, le provençal et le roumain, tous mélangés. La colonne de la profession est surlignée si le participant a étudié l'AF, ce même vaut pour le latin dans la colonne à droite.

G1 : Faible connaissance (7)		Espagnol	Italien	Portugais	Autre(s)	Latin
FR8	Étudiant					
FR9	Sans emploi					
FR11	Infirmière					
FR14	Enseignante FLE					
FI7	Étudiante					
FI8	Enseignante					
FI10	Étudiante					
G2 : Une langue (9)		Espagnol	Italien	Portugais	Autre(s)	Latin
FR1	Ingénieur					
FR2	Étudiante alternante					
FR4	Étudiante					
FR10	Étudiante					
FR12	Étudiante					
FR13	Étudiante					
FI1	Étudiante					
FI4	Sans emploi					
FI5	Enseignante					
G3 : Deux langues (7)		Espagnol	Italien	Portugais	Autre(s)	Latin
FR5	Enseignant-chercheur					
FR6	Enseignante FLE					
FI2	Chercheuse					
FI3	Enseignante					
FI6	Enseignante					
FI11	Étudiante					
FI14	Étudiant diplômé					
G4 : Plusieurs langues (7)		Espagnol	Italien	Portugais	Autre(s)	Latin
FR3	Entrepreneur					
FR7	Enseignant					
FR15	Enseignant FLE					
FI9	Expert en législation					
FI12	Étudiante					
FI13	Étudiante					
FI15	Enseignant					

## Annexe 5. Résumé en finnois – Suomenkielinen tiivistelmä

### 1. Johdanto

Tässä pro gradu -tutkielmassa tarkastellaan keskiaikaisten romaanisten kielten – ranskan, espanjan, italian ja portugalilaisien – sanojen tunnistusta. Työn tavoitteena on selvittää, millainen vaikutus kontekstilla, äidinkielellä, romaanisella kielitaidolla sekä muokkausetaisyydellä on sanojen tunnistamisessa: edellä mainittuja käsitteitä lähestytään ensisijaisesti ranskankielisestä perspektiivistä ja muinaisranskan aineistolle annetaankin täten eniten painoarvoa. Tutkimuskysymykset ovat:

- 1) *Miten lauseyhteys auttaa eri äidinkieltä puhuvia tulkitsemaan sanan merkitystä?*
- 2) *Miten muinaisranskan sanat tunnistetaan suhteessa muiden keskiaikaisten kielten sanoihin, muokkausetaisyyden perusteella?*
- 3) *Millainen rooli romaanisten kielten osaamisella on sanantunnistuksessa?*

Näihin kysymyksiin pyritään löytämään vastauksia 30 sanaa käsittävän sanastokokeen avulla, johon osallistui sekä ranskaa että suomea äidinkielenään puhuvia henkilöitä. Tutkimus on siis kvalitatiivinen ja se sijoittuu pääasiassa leksikologian ja psykolingvistiikan tieteenaloihin (Murillo Puyal ym., 2008). Keskenään sukua olevien kielten vertailu hyödyttää uuden kielen oppimista sekä yhteisen kieli- ja kulttuuri-identiteetin vahvistumista (McCann ym., 2003; Zeevaert & ten Thije, 2007). Oletuksena on, että sanojen tunnistus on helpointa tilanteessa, jossa niitä ympäröi konteksti ja ne ovat muoto-opillisesti lähellä nykyranskan vastinetta. Myös romaaniseen kielikuntaan kuuluvan äidinkielen sekä kielikunnassa omaksutun laajemman kielitaidon otaksutaan parantavan sanojen tunnistamista. Teorioiden mukaan aiemmin hankittu tietämys kieliopillisista, kirjoitusasullisista sekä foneettisista muutoksista edesauttaa sanojen ja lauseyhteyksien hahmottamista (McCann ym., 2003; Murillo Puyal ym., 2008).

### 2. Teoriatausta

Euroopassa on aina esiintynyt monikielisyyttä: yksikielisyyden periaate syntyi vasta kansallisvaltioiden luonnin yhteydessä (Zeevaert & ten Thije, 2007). Vaikka monikielisyyden arvostus on jälleen kasvanut, riittää vieraan kielen ymmärtämiseksi myös osittainen kielitaito,

jota perinteinen monikielinen opetus ei aina huomioi (Blanche-Benveniste, 2008). Zeevaartin ja ten Thijen mukaan vastavuoroinen kommunikaatio onnistuu silloinkin, kun kaksi osapuolta keskustelevat keskenään omilla äidinkielillään: tätä ilmiötä kutsutaan *reseptiiviseksi monikielisyydeksi*. Tutkijat pitävät tärkeänä sen luokittelamista omaksi tieteenalaksi kansainvälistymisestä ja globalisaatiosta aiheutuneiden muutosten vuoksi.

Kohentuneen luetunymmärtämisen taidon on havaittu edesauttavan sanastontuntemuksen kehittymistä (Laufer, 1996). Samaa tarkoittavien sukulaissanojen runsaan läsnäolon on todettu auttavan tekstin tulkitsemisessa, mutta myös vähäisemmässä käytössä olevia synonyymejä voidaan hyödyntää (Bezooijen & Gooskens, 2007). Lukijan kielitaidon ollessa alhainen, vieraat sanat aiheuttavat lähinnä ylikuormittumista. Keskinertainen kielitaito mahdollistaa huomion keskittymisen kieliopillisiin rakenteisiin ja tunnuksiin, kun taas korkea kielitaito synnyttää mielenkiintoa kielten keskinäistä vertailua kohtaan. (Blanche-Benveniste & Valli, 1997.) Lukijan taitotasosta huolimatta on riskisanojen tunnistamiseen kiinnitettävä poikkeuksellista huomiota: väärin tulkittu sana voi johtaa kokonaisen lauseyhteyden vääristymiseen (Laufer, 1996; Lutjeharms, 2007).

Kun lukija katsoo mitä tahansa sanalista, hän ei välttämättä muista sanoja ulkoa, vaan pyrkii luokittelemaan niitä temaattisesti (Bogaards, 1994). Jokainen sana tallentuu ensin episodiseen muistiin ja vasta tietynlaisen prosessoinnin (kuten kääntämisen) jälkeen sana siirtyy säilömuistiin osaksi *mentaalileksikkoa* (Lutjeharms, 2007). Sanojen temaattisen luokittelun avulla voidaan analysoida mentaalileksikon rakennetta. Silmänliiketutkimuksissa on havaittu, että katse siirtyy samalla tavalla kahden toisilleen sukua olevien kielten tekstejä luettaessa (Lutjeharms, 2007) ja keskittyy yksittäisessä sanassa ensiksi sen alun ja puolenvälin välimaastoon (Deutsch & Rayner, 1999).

Laajempaan tekstinymmärrykseen liittyen asiayhteyden rooli on ratkaiseva, ja juuri se antaakin yksittäisille sanoille merkityksen (Rastier, 1991; Bogaards, 1994). Laufer (1996) mainitsee termin ”kynnyssanavarasto” (*threshold vocabulary*), jonka rajana pidetään 3000 sanaryhmää (yhteen ryhmään sisältyy mm. sanan eri taivutusmuodot ja johdannaiset; Nation, 1983). Kynnyksen alle jäävän lukijan on vaikea hahmottaa vieraskielisessä tekstissä tiedon hierarkiaa, ja heikoiten ymmärretyt osuudet saatetaan helposti tulkita turhaksi informaatioksi. Mitä enemmän tekstissä on lukijalle ennalta tuttuja sanoja, sitä todennäköisemmin ne tukevat toinen toisiaan (ja siten kontekstia kokonaisuudessaan): toisaalta, lukija saattaa myös tyytyä tulkitsemaan sanan omalla tavallaan, mikäli tulkinta istuu kontekstiin. (Laufer, 1996.)

Sundqvistin (1986) tutkimuksen perusteella myös kielisukulaisuus helpottaa sanojen ymmärtämistä, ja tämä onkin erityisen ilmeistä etenkin romaanisten kielten kohdalla, sillä monet nykykielten kansainvälisistä lainasanoista juontuvat latinasta sekä sen tytärikielistä (McCann ym., 2003). Latinan vaikutus Euroopassa säilyi pitkään, mutta se ei estänyt puhuttuja murteita kehittymästä itsenäisiksi kieliksi. Kirjallista ilmaisua ajatellen tapahtui 1000-luvun Ranskassa käänteentekevä muutos, kun uusi yhteiskuntaluokka (aatelisto) halusi uskonnollisten tekstien sijasta lukea kaunokirjallisuutta, omalla kielellään. Tämä *metalingvistinen* muutos levisi sittemmin muualle Eurooppaan. Kielten erilleen luokittelu perustuu rakenteellisten erojen lisäksi paitsi vallitsevaan ajattelutapaan (esimerkiksi kreikka ja kiina ovat vuosituhansienkin jälkeen yhä ”elossa”) myös poliittiskulttuuriseen tahtoon. (McCann ym., 2003; Janson, 2012.)

Tieteenalana sukulaiskiarten vertailu on jo vanha, mutta sitä hyödynnetään erittäin vähän nykykielten opetuksessa eikä vertailun periaatteita tai sisäistä hierarkiaa juuri käsitellä: kielten suurimmat erot löytyvät tyypillisesti ääntämyksestä, erot ovat vähäisempiä sanaston tasolla ja pienimpiä kielioppiin liittyvissä asioissa (Posner, 1996; Blanche-Benveniste, 2008). EuroCom-metodi (McCann ym., 2003) keskittyy nimenomaan keskinäisen vertailun yksityiskohtiin sekä osittaisen kielitaidon kehittämiseen (vrt. *reseptiivinen monikielisyys*). Tämän tutkielman kannalta metodin olennaisimmiksi piirteiksi luetaan yhteisen sanaston ja syntaksin lisäksi systemaattiset äännevastaavuudet, ja tietoisuus näistä vastaavuuksista helpottaa sanojen käännösprosessia (McCann ym., 2003). Vastaavuuksien on todettu myös auttavan uuden kielen oppimista oman äidinkielen pohjalta (Simone, 1997), mutta niiden hyödyntäminen on toisaalta jätetty tietoisesti huomioimattakin (Blanche-Benveniste, 2008).

Romaanisten kielten kirjoitusasu on ranskaa lukuun ottamatta verrattain säännönmukainen, mutta varsinkin palataalikonsonanttien [ɲ], [ʎ], [ʃ] ja [ʒ] merkitsemisessä on kielten välillä laajemminkin eroavaisuuksia, mikä täten vaikeuttaa sanojen tunnistamista (Pottier, 1997; McCann ym., 2003). Vaikka latinan lainasanojen osuus ranskan leksikosta on huomattava, ja kielen ”romaaninen olemus” siten poikkeuksellisen ilmeinen, on kielievoluutio ollut niin perinpohjaista, että lainattu sanasto on helppoa erottaa peritystä sanastosta (Posner, 1996). Valtaosa kielievoluutiosta on tapahtunut jo melko varhain, sillä Janson (2012) huomauttaa muinaisranskan jakavan monia samoja piirteitä nykyranskan kanssa. On syytä pitää mielessä, että ranskan ääntämys on ottanut vielä tuon ajankohdan jälkeenkin suuria harppauksia, ja että romaanisten kielten keskinäisessä (kirjallisessa) ymmärtämisessä tukeudutaan nimenomaan

ranskan kirjoitusasuun, joka ei ole juuri muuttunut 1200-luvun jälkeen (Herman, 1967b; McCann ym., 2003).

### 3. Aineistonkeruumenetelmä

Kansainvälisiä lainasanoja välttääkseni valitsin aineistoksi neljä muinaisromaanisilla kielillä kirjoitettua kansalliseeposta, lueteltu ilmestymisjärjestyksessä: *Rolandin laulu* (*Chanson de Roland*; Ranska), *Cidin laulu* (*Cantar de mio Cid*; Espanja), *Jumalainen näytelmä* (*Divina Commedia*; Italia) sekä *Os Lusíadas* (Portugali). Kirjoitusasua ajatellen *Rolandin laulu* on kauimpana nykykielestään ja *Os Lusíadas* lähimpänä nykykieltään. Näistä teoksista valikoitiin osin sattumanvaraisesti, osin järjestelmällisesti ensin 40 sanaa lauseyhteyksineen sanastokokeen ensimmäistä versiota varten, mutta viralliseen versioon sanoja kerättiin 30. Sanojen valintaa ohjasi niiden etymologia, sanaluokka sekä niitä ympäröivä konteksti.

Valitut termit aseteltiin neljään luokkaan niiden läpinäkyvyyden perusteella (läpinäkyvät, melko läpinäkyvät, melko läpinäkymättömät ja läpinäkymättömät; Bezooijen & Gooskens, 2007). Läpinäkyvyyden selvitykseen käytettiin Levenšteinin etäisyyttä eli muokkausetäisyyttä (Borin, 2013). Kun aineistosta poimittu termi halutaan muuntaa nykyranskan vastineeksi, täytyy kirjaimia joko poistaa, lisätä tai niiden paikkaa vaihtaa. Jokaista ortografista muutosta nimitetään operaatioksi (*cost*), jonka arvo on tyypillisesti 1. Sanan ensimmäisen kirjaimen otaksuttiin kuitenkin olevan merkitykseltään painokkaampi, joten tällöin annettiin arvoksi 2. Arvot laskettiin yhteen ja jaettiin kirjainten kokonaismäärällä (*alignment*). Lopputuloksena saatiin prosenttiarvo, joka kertoi sanan läpinäkyvyysasteen. Pääluokittelun yhteyteen lisättiin vielä nykyranskaan etymologista yhteyttä vailla olevien sanojen ryhmä, jolla testattiin aineiston sanojen ymmärtämistä muiden romaanisten kielten osaamisen pohjalta.

Aikaisemmassa versiossa sanojen määrät kunkin aineiston ja kunkin läpinäkyvyysluokan välillä olivat tasaiset, mutta koska muinaisranskan painoarvoa haluttiin korottaa, uuden version 30 sanasta puolet koostettiin pelkästään muinaisranskan aineistosta ja toisen puoliskon muodostivat kolme muuta aineistoa yhdessä. Pääasiallinen syy kokeen muuttamiselle oli sen helpottaminen koehenkilön näkökulmasta: kahden pilotoijan suorituksista voitiin päätellä kokeen olleen liian vaativa yleistä levitystä varten. Näin ollen uuteen versioon valittiin paitsi läpinäkyvämpiä, myös yleisemmässä käytössä olevia sanoja. Kontekstin merkitystä tunnistamisessa kyettiin puolestaan tutkimaan jakamalla koe kahteen

osaan: ensimmäisessä osassa sanat olivat yksin, toisessa osassa taasen lauseyhteyden ympäröiminä.

Koehenkilöiksi kutsuttiin ranskaa ja suomea äidinkielenään puhuvia henkilöitä. Kaikkien kokeeseen osallistuvien tuli ensiksi täyttää taustakysely, jossa kartoitettiin perustietojen lisäksi myös kielitaitoa romaanisissa kielissä. Vastaukset käsiteltiin anonyymisti ja niitä hyödynnettiin ryhmittelyperusteissa tuloksia analysoitaessa. Ranskalaisten osuus järjestettiin keväällä 2022, suomalaisten osuus puolestaan saman vuoden syksynä. Suomalaisista koehenkilöistä lähes kaikki olivat joko ranskan kielen opiskelijoita yliopistossa tai opettivat kieltä jossakin laitoksessa. Ranskalaisten profiilikuvaus oli opiskelualojen valossa paljon kirjavampi ja ammattikunnatkin ulottuivat selvästi kielitieteiden ulkopuolelle (koehenkilöiden joukossa oli esimerkiksi insinööri, yrittäjä sekä sairaanhoitaja). Muutama koehenkilö jouduttiin hylkäämään ennen analyysin aloittamista: ensisijainen syy tälle oli termien puutteellinen kääntäminen kokeessa – kokelaat jättivät jommassakummassa kokeen osiossa kääntämättä yli puolet pyydetyistä kohdista. Kokeeseen osallistui siis hyväksytysti 15 ranskankielistä sekä 15 suomenkielistä koehenkilöä, yhteensä osanottajia oli täten 30.

Koetulosten pisteyttämiseksi käytettiin Palokankaan (2018) luoman asteikon yksinkertaistettua versiota, joka on kuusiportainen. Kokelaan vastaus oli viiden pisteen arvoinen, mikäli se osui täysin oikeaan: myös täysin samaa tarkoittava synonyymi kelpuutettiin. Neljä pistettä sai vastauksesta, joka oli oikea, mutta väärin taivutettu tai merkitykseltään lievästi poikkeava. Kolmeen pisteeseen vaadittiin selkeää samuuden havaitsemista, vaikka vastaus itsessään olisikin ollut väärin: vastauksen tuli siis kuulua merkitysopillisesti samaan kategoriaan oikean käännösvastineensa kanssa. Kahteen pisteeseen riitti, kun sanalla oli selkeä etymologinen tai semanttinen viittaus käännösvastineeseen, edellyttäen sanaluokan olevan myös oikea. Yksi piste annettiin silloin, jos tämä viittaus oli heikko tai se oli väärässä sanaluokassa. Väärä vastaus oikein taivutettuna mahdollisti myös yhden pisteen saamisen: olihan lauseyhteydestä käynyt ilmi sanan syntaktinen funktio. Täysin virheellisestä vastauksesta sekä vastaamatta jättämisestä ei annettu lainkaan pisteitä. Koska kokeen kummassakin osiossa oli 30 sanaa, maksimipistemääräksi saatiin 150. Kun tämä kerrottiin koehenkilöiden määrällä (30), merkitsi se sitä, että pisteitä oli yhteensä jaossa 4500 per osio.

#### **4. Tulokset ja johtopäätökset**

Tutkimuksessa haettiin vastauksia kolmeen tutkimuskysymykseen sanastokokeen avulla. Ensimmäinen tutkimuskysymys pyrki kartoittamaan, miten konteksti ja kokelaan äidinkieli vaikuttivat kokeen tuloksiin. Sanastokoe koostui kahdesta osasta ja näiden toisiinsa vertaamisella saatiin selville asiayhteyden merkitys. Kontekstin läsnäolo paransi tuloksia lähes poikkeuksetta. Erotus oli suurin niiden sanojen kohdalla, joilla ei ollut etymologista yhteyttä nykyranskan sanoihin. Muilta osin ei ollut viitettä siihen, että sanojen läpinäkyvyys tai läpinäkymättömyys olisi vaikuttanut tunnistettavuuden kasvuun lauseyhteyden ollessa esillä. Tämän jälkeen kokelaat jaoteltiin äidinkielen (ranska ja suomi) perusteella kahteen ryhmään, ja näiden välillä toteutettiin statistinen vertailu läpinäkyvyysluokkien osalta. Yleisesti ottaen ranskalaiset pärjäsivät hieman suomalaisia paremmin vertailussa, mutta luokkien väliset erot olivat monesti erittäin minimaalisia ja molemmat ryhmät onnistuivat jossakin kategoriassa selkeästi toista paremmin. Mitään selvää kuviota ei ollut nähtävillä, mikä osoittaakin sen, että laajempi otanta olisi ollut tarpeen. Luovuin koehenkilöiden jaottelusta äidinkielen mukaan ja seuraavia kohtia tarkasteltiin koko ryhmän tasolla.

Toisella kysymyksellä selvitettiin korpustenvälisten erojen lisäksi muokkausetäisyyden vaikutusta sanojen tunnistamisessa. Muokkausetäisyyttä eli Levenšteinin etäisyyttä hyödynnettiin termien asettamisessa järjestykseen: täten myös läpinäkyvyysluokat voitiin havainnollistaa taulukoissa paremmin. Jo aineistonkeruuluvussa linjattiin, että muinaisranskan korpus erotetaan muiden keskiaikaisten kielten aineistoista, ja että sen painoarvoa työssä korotetaan. Näin ollen yhtä aineistoa pystyttiin vertaamaan kolmeen aineistoon. Selvisi, että koehenkilöt tunnistivat selvästi muinaisranskan aineistoa paremmin muiden romaanisten kielten läpinäkymättömät sekä etymologista yhteyttä vailla olevat sanat. Ero kaventui sanojen läpinäkyvyysasteen ollessa korkea. Tunnistaminen noudatti jotakuinkin läpinäkyvyysasteen vaihtelua, mutta erot muiden romaanisten kielten sanoissa poistuivat käytännössä kokonaan niiden asetuttua lauseyhteyteen. Muinaisranskan sanojen tunnistuksessa läpinäkymättömyys aiheutti jopa lauseyhteydessä mittavia haasteita.

Kolmas kysymys keskittyi tarkastelemaan tulosten eroavaisuuksia luokittelemalla kokelaat neljään ryhmään taustakyselyssä ilmoitetun nykyromaanisten kielten kielitaidon perusteella. Kielitaidoksi laskettiin luokittelussa osattujen kielten määrä; taitojen laajuudella ei ollut merkitystä muissa ryhmissä kuin G1 (heikko osaaminen), johon niputettiin sekä muita romaanisia kieliä kuin ranskaa taitamattomat että yhtä toista romaanista kieltä hyvin heikosti taitavat kokelaat. Yksittäisen ryhmän kokonaisvahvuus oli seitsemän tai yhdeksän henkeä.



Tuloksista kävi ilmi, että laajempi nykyromaanisten kielten taitaminen helpottaa sanojen tunnistamista, vaikka sana- ja ryhmäkohtaista vaihtelua esiintyi myös runsaasti. Tunnistuksen parantuminen kielitaidon perusteella oli lineaarisinta muiden romaanisten kielten kontekstittomassa aineistossa. Lopuksi analysoitiin sitä, vaikuttavatko aiemmat muinaisranskan opinnot muinaisranskan sanojen tunnistukseen. Muinaisranskaa opiskelleiden kokelaiden vähyyden huomioon ottaen vertailusta rajattiin pois ryhmien G1 ja G2 jäsenet. Laajimman kielitaidon omaavissa ryhmissä G3 ja G4 muinaisranskaa opiskelleita (7) oli yhtä paljon kuin sitä opiskelemattomia (7). Kun molempien osioiden kokonaistulokset laskettiin yhteen, ei ryhmien välillä ollut juurikaan eroja: yksittäisten sanojen kohdalla oli kuitenkin vaihtelua. Muinaisranskaa opiskelleet menestyivät opiskelemattomia paremmin ainoastaan sanojen ollessa lauseyhteydessään; jälkimmäiset puolestaan tunnistivat sanat keskimäärin paremmin ilman kontekstia.

Tulosten tarkempaa analysointia varten toteutettiin tammikuussa 2023 haastattelu kymmenellä sanastokokeeseen osallistuneella koehenkilöllä, joista seitsemän oli suomalaisia ja kolme ranskalaisia. Haastattelu rajattiin koskemaan seitsemää sanaa, joista viisi (*amsdous*, *meitet*, *pie**dra*, *velz*, *vizinhas*) oli tunnistettu selvästi heikommin kuin mitä muokkausetäisyys antoi olettaa; kaksi (*podriemos*, *vuldreie*) puolestaan oli tunnistettu odotettua paremmin. Kommentointi oli hyvin vapaamuotoista, mutta monia risteäviä huomioita nousi esiin. Tärkeimmät ymmärtämistä vaikeuttavat seikat koskivat niin sanan kirjoitusasua (ensimmäisen tavun korostunut merkitys, ajatuksen katkeaminen erikoisten kirjainten vuoksi), sanan assosioinnin vaikeutta (mielikuvan puute sanan mahdollisesta merkityksestä, huono istuvuus lauseyhteyteen, kontekstissa esiintyvien ”helppojen” sanojen aikaansaamat virheelliset tulkinnat) kuin myös omiin taitoihin liittyviä virhearvioita (liiallinen luotto omaan kielitaitoon sekä toisaalta myös epävarmuus omasta osaamisesta). Ymmärtämistä puolestaan edesauttoivat sanan assosiointi (suoraviivainen konteksti, verrattavuus jo johonkin aiemmin opittuun sanaan) sekä sanan alkukirjainten samankaltaisuus nykyranskan vastineen kanssa.

Voidaan siis päätellä, että sukulaiskieliä vertailevan sanastokokeen laatimiseen tarvitaan paitsi runsaasti aikaa, myös tarkkaan suunniteltu sekä hyvin jäsennelty lähestymistapa. Muiden muinaisromaanisten kielten kuin ranskan aineistoja koottaessa ei keskitytty riittävästi häivyttämään nykykielten osaamisen vaikutusta, eivätkä aineistot tämän takia olleet vertailukelpoisia. Vankempi tukeutuminen Bezooijenin ja Gooskensin luokitteluun olisi voinut selkiinnyttää muutoksetäisyyden vaikutusta huonommin testissä sanoja tunnistaneiden

kokelaiden osalta. Sanojen ensimmäisen tavun painoarvoa olisi pitänyt nostaa, eikä vain ensimmäisen kirjaimen. Teoriaviitekehityksessä historiallisten ja runomuotoisten kontekstien käsittely jää erittäin vähäiseksi, joten jatkotutkimuksia ajatellen niihin olisi syytä kiinnittää enemmän huomiota. Kokelaiden kommentteista käy ilmi nimenomaan se, että tiettyjen lauseyhteyksien rakenteellinen vaikeus hankaloitti huomattavasti yksittäisen sanan ymmärtämistä. Perusteellisempi psykolingvistinen analyysi tai foneettinen lähestymistapa tutkimusongelmaan voisivat tarjota runsaasti uusia näkökulmia aiheeseen myös opetusmielessä. Mielenkiintoista olisi myös tarkastella, parantuisivatko kokelaan tulokset ajankäytön lisääntymisen myötä. Tutkielma toi kaikesta huolimatta selkeästi esiin sen, että konteksti ja laaja-alainen kielitaito helpottavat vieraiden, toisilleen sukua olevien kielimuotojen sanojen tunnistamista, ja että ortografisella etäisyydellä on myös osuutensa hahmotusprosessissa.