



**HAL**  
open science

# L'approximation dans la retranscription des toponymes à travers les langues

Saandia Ali, Jonas Löffström, Margarita Muñoz, Betina Schnabel-Le Corre

► **To cite this version:**

Saandia Ali, Jonas Löffström, Margarita Muñoz, Betina Schnabel-Le Corre. L'approximation dans la retranscription des toponymes à travers les langues. Travaux linguistiques du CerLiCO, inPress. hal-01616801

**HAL Id: hal-01616801**

**<https://univ-tlse2.hal.science/hal-01616801>**

Submitted on 14 Oct 2017

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

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

# **L'approximation dans la retranscription des toponymes à travers les langues**

Saandia Ali\*

Jonas Löffström

Margarita Muñoz Garcia\*

Betina Schnabel-Le Corre\*

## **Résumé**

Les toponymes (endonymes et exonymes) ne correspondent pas toujours au système linguistique communément connu et assimilé par les locuteurs. Les orthographes déviantes et les variations segmentales et suprasegmentales régionales et dialectales engendrent des approximations dans la production (prononciation et représentation graphique d'un toponyme). En classe de langues étrangères, l'impact de ces approximations sur la perception auditive chez les apprenants devrait être pris en considération. Dans cet article nous présentons les résultats d'une expérience pilote de perception auditive réalisée auprès d'un groupe d'étudiants francophones, à partir d'un corpus en quatre langues (anglais, espagnol, allemand et suédois). Nous analysons les résultats de l'expérience en fonction de la langue étudiée. Nous examinons également les pourcentages de bonnes réponses des apprenants en fonction du degré d'approximation observé dans leurs productions.

## **Abstract**

Toponyms (endonyms and exonyms) do not always match the linguistic system, which is commonly known and assimilated by its native speakers. Orthographic deviations and segmental and supra-segmental variations due to regional accents and dialects often cause approximations in speech production (pronunciation and graphic representation of a toponym). In foreign language acquisition, the impact of these approximations on the perception of toponyms should be taken into account. In this paper, we will present the results of a pilot experiment of oral perception, which was carried out with a group of French speaking learners by the means of a corpus in four languages (English, German, Spanish, Swedish). Our study brings up the differences of the results according to the examined language. Furthermore we point out the percentage of correct answers proportionately to the degree of approximation produced by the learners.

## **1. Introduction**

Cet article présente les premiers résultats d'un travail de recherche entrepris dans le cadre d'un projet multilingue au sein de l'équipe LIDILE, portant sur l'anglais, l'allemand, l'espagnol et le suédois et rassemblant des chercheurs spécialistes de toponymie et des phonéticiens. Une partie de ce travail se centre autour des phénomènes d'approximation lors de la perception des toponymes en langue 2 (L2) et sur leur retranscription.

L'approximation est un phénomène bien connu en toponymie. Partout dans le monde, le contact entre les langues s'est établi par la conquête d'autres territoires ou la découverte du monde. Les grands conquérants ou voyageurs ont été confrontés à des peuples autochtones et ils ont essayé de noter approximativement dans leur langue maternelle les noms de lieux

qu'on leur indiquait dans une langue inconnue. C'est par exemple le cas en Bretagne où les Français ont transformé des noms bretons comme *Landreger* et *Montroulez* respectivement en *Tréguier* et *Morlaix*.

De plus, les toponymes nous intéressent pour leur spécificité dans le système linguistique d'une langue. Ceux-ci, en tant qu'une sous-catégorie des noms propres, ne correspondent pas toujours au système linguistique connu par les locuteurs et encore moins par les locuteurs non natifs (apprenants). Ils se distinguent par leur structure morphologique/morpho-syntaxique ou bien sur le plan phonétique/phonologique (Löfström&Schnabel, 2010 ; Nübling, 2012).

L'orthographe déviante et les variations segmentales et/ou suprasegmentales, qu'elles soient régionales ou dialectales, engendrent des approximations dans la production et dans la perception des toponymes.

Comment prononcer un nom de village comme « Niederkriegsheim » en français sans connaître les pratiques de prononciation locales? Comment orthographier un nom comme [e :bɛs] si on ne l'a jamais vu écrit « Eybens » ? Comment savoir prononcer correctement en allemand *Bochum* ou en anglais *Arundel*, lorsqu'il y a différentes réalisations dans une même langue ? La prononciation de *Bochum* est soumise à des variations segmentales sociolectales entre [ɔ] et [o:]. On trouve également des variations suprasegmentales dans l'accentuation d'*Arundel* : soit sur la première syllabe ['aerəndəl], tel qu'il est prononcé à Londres, soit sur la deuxième syllabe [ə'ɾʌndəl], à l'extérieur de Londres.

Ces phénomènes se retrouvent dans la production des endonymes<sup>1</sup> (Kadmon, 1995) d'origine étrangère, réalisés par les membres d'autres communautés et ce malgré les efforts de normalisation menés par les groupes d'experts auprès de l'ONU (UNGEGN<sup>2</sup>).

Notre intérêt pour ce sujet vient aussi d'une prise de conscience des difficultés causées par les toponymes à l'oral : ceux-ci peuvent entraver la compréhension et engendrer des approximations entre ce qui est entendu et la représentation que l'on s'en fait. L'objectif de ce travail est d'abord de pouvoir analyser les caractéristiques phonétiques spécifiques des toponymes et leur impact sur la compréhension de l'oral en vue de proposer de nouvelles stratégies didactiques pour l'enseignement.

Si la connaissance des règles de correspondance graphie-phonie est reconnue comme importante en langue étrangère, la spécificité des toponymes en la matière demande à être prise en considération. De plus, ces difficultés persistent même à un niveau avancé chez les apprenants. Il faut préciser que le terme « production » englobe ici à la fois la prononciation ou réalisation phonétique et la représentation graphique d'un toponyme. Nous nous concentrerons donc sur la réalisation phonétique de certains toponymes que nous allons étudier et décrire et sur leur perception auditive par des apprenants.

L'étude de l'approximation dans la représentation graphique d'un toponyme en langue étrangère est considérée comme une première étape permettant d'identifier les difficultés rencontrées par les apprenants. Il s'agira d'observer les capacités des apprenants à comprendre les toponymes auxquels ils seront exposés, donc à les identifier en tant que lieu. Ensuite il s'agit de voir si les apprenants ont capables d'en donner une représentation

---

<sup>1</sup> Appellation toponymique exprimée dans la forme orthographique et dans la langue locale de la région où l'entité dénommée est située, indépendamment du système d'écriture.

\* LIDILE EA 3874

<sup>2</sup> United Nations Group of Experts on Geographic Names

orthographique, témoin de leur connaissance des correspondances graphie-phonie dans une langue donnée.

Ce phénomène est étudié à partir de corpus constitués en anglais, en espagnol, en allemand, et en suédois. Dans cet article, nous décrivons les étapes expérimentales réalisées en commençant par la constitution d'un corpus de toponymes, puis la classification de ces toponymes sur le plan phonétique (classification nous permettant de formuler de premières hypothèses quant à l'effet sur la compréhension du toponyme), l'élaboration de tests de perception basés sur les toponymes retenus, l'enregistrement et la passation de tests auprès d'un public d'étudiants francophones.

## 2. Expérience

### 2.1.1 Le corpus

Pour déterminer notre corpus de base nous avons utilisé la liste des 193 pays officiellement reconnus d'après des listes de l'ONU<sup>3</sup>, et leurs capitales, tout en ajoutant quelques villes très connues, comme New York pour les États-Unis et Barcelone pour l'Espagne. Pour la plupart d'entre eux il s'agit d'exonymes<sup>4</sup> (Kadmon, 1997) dans une langue donnée, comme Barcelone en français (*Barcelona* étant l'endonyme espagnol) mais également d'endonymes comme *New York*.

Cependant, les règles phonétiques propres à chaque langue constituent le premier obstacle à une utilisation appropriée de l'endonyme. La production orale d'un endonyme étranger devient presque automatiquement un exonyme, parce que les locuteurs l'adaptent au système phonétique de leur langue maternelle, comme le montrent les exemples suivants<sup>5</sup>:

Endonyme *Paris* (fr) [pa'ʁi] -> exonymes *Paris* (all) [pa'ʁi :s] <sup>1</sup> *Paris* (angl) ['pærɪs]

Endonyme *Dublin* (angl) [dʌblɪn] -> exonyme *Dublin* (frç) [dyblɛ̃]

Comme l'a constaté Otto Back en 2002: « Etant donné que toutes les langues ont leurs propres stratégies pour adapter les noms d'origine étrangère le nombre d'exonymes correspondant à un seul endonyme est théoriquement illimité ».

### 2.1.2 Les critères de sélection d'oïkonymes

Nous avons établi un corpus de 189 oïkonymes (noms de pays et d'agglomérations) en français et leurs formes correspondantes dans les quatre langues. Les critères de sélection ont bien été définis d'après des considérations purement linguistiques comprenant en majorité le rapport entre les niveaux graphique et phonétique d'une langue.

#### □ Les endonymes

Nous avons gardé tous les toponymes qui correspondent en français et au niveau graphique, aux endonymes. Le transfert d'un endonyme dans une langue donnée vers les

<sup>3</sup> <http://www.un.org/en/members/>

<sup>4</sup> Nom géographique utilisé dans une langue pour désigner un lieu situé en dehors du territoire dont cette langue est la langue officielle.

<sup>5</sup> Dans cet article, les noms de langues seront abrégés allemand = all, anglais = angl, espagnol = esp, français = frç, suédois = sué, islandais = isl.

quatre langues étudiées engendre la création d'exophones bien différents, plus au moins éloignés de l'original dans la langue source.

ایران en persan => Iran [i'ra :n] (all), [i'ra:n] (angl), [iran] (esp) [irã] (frç)

ירושלם en hébreu => Jerusalem [je'ru:salem] (all), [dʒə'ru:sələm] (angl),

Jérusalem [ʒery:za'lem] (frç),

Le toponyme en écriture non-latine constitue une difficulté supplémentaire car leur translittération en écriture latine introduit déjà des interprétations phonétiques plus ou moins divergentes de la langue d'origine, comme le montre l'exemple du toponyme qui s'est établi pour la capitale de la Chine, *Beijing*. Ce toponyme, exigé de la part de la Chine comme endonyme tel qu'il est écrit en pinyin<sup>6</sup> n'est pas prononcé de la même manière en anglais, en allemand, en espagnol, en français, en suédois ou bien en d'autres langues. Nous nous sommes basés sur les formes officielles en écriture latine telles qu'elles apparaissent dans la liste de l'ONU.

□ Les traductions partielles ou complètes

Nous avons écarté tous les toponymes résultant d'une exonymisation par une traduction partielle ou complète (Schnabel & Löffström 2010) de l'endonyme.

a. Traduction partielle : Un toponyme comme *North America* est traduit partiellement, à savoir son constituant 'Nord' est traduit dans quasiment toutes les langues, pendant que le constituant propre 'Amérique' est seulement adapté aux règles inhérentes à chaque langue (variation ica / ika / ique) :

i. Endonyme : *North America* (angl) => exonymes : *Nord/amerika* (all, sué), *Norður-Ameríka* (isl), *Amérique du Nord* (frç), *América del Norte* (esp).

b. Traduction complète : Elle est seulement possible pour des toponymes complexes transparents ayant une structure où chaque constituant correspond à un générique.

i. *Nederland* (endonyme ndl) => *Pays bas* (exonyme frç), *Niederlande* (all), *The Netherlands* (angl), *Países Bajos* (esp), *Nederländerna* (sué).

ii. *United States of America* (endonyme angl) => *Vereinigte Staaten von Amerika* (all), *Estado Unidos de América* (esp), *Etats unis d'Amérique* (frç), *Amerikas förenta stater* (sué)

□ Formes fortement déviantes

Nous n'avons pas non plus retenu les exonymes dont la forme est trop éloignée de l'endonyme sans être pour autant une traduction, comme par exemple *Deutschland* (endonyme allemand) => *Germany* (angl), *Alemania* (esp), *Saksa* (fin), *Allemagne* (frç), *Niemcy* (pol), *Tyskland* (sué).

La traduction - comme moyen d'exonymisation - n'est pas toujours appliquée par toutes les langues. Il n'est pas rare que le même toponyme soit traduit dans une langue et gardé sous la

---

<sup>6</sup> Le pinyin est un système de romanisation du chinois mandarin, promu officiellement par la Chine

forme orthographique de l'endonyme dans une autre langue, comme le montre l'exemple *Nederland* (*supra*) ou encore les différentes formes du toponyme *Montagnes rocheuses*:

*Rocky Mountains* (endonymes anglais) => *Rocky Mountains* (all), *Montagnes rocheuses* (frç), *Klippiga bergen* (sué)

### 2.1.3 Classement en cinq catégories selon les degrés de divergence,

A partir de ces 189 toponymes en français et de leurs formes correspondantes dans les quatre langues, nous avons dégagé 5 catégories selon le degré de divergence, mis en lien avec les hypothèses émises :

- Catégorie 0 : Dans la première catégorie sont rassemblés les exonymes dont la forme orthographique est identique (congruence complète) ou quasi-identique avec le toponyme français. Les variations se trouvent essentiellement au niveau suprasegmental entre plusieurs langues, mais touchent également le niveau segmental.
  - Bali, Amsterdam, Oslo (les toponymes sont à 100% identiques dans les 5 langues)
  - Afghanistan, Canada, Jérusalem : l'exonyme peut varier dans les signes diacritiques (i.e. accent aigu en français dans Jérusalem), la valeur phonétique identique pour différentes lettres dans les différentes langues (i.e. le 'c' dans Canada a la même valeur phonétique que le 'k' en allemand et en suédois) ou encore quelques lettres non prononcées (i.e. Afghanistan s'écrit sans 'h' en espagnol).
- Catégorie 1 : Dans la deuxième catégorie se retrouvent les toponymes pour lesquels il y a congruence graphique entre plusieurs langues avec des variations phonétiques « segmentales » et/ou « suprasegmentales »
  - Angola, Berlin, Iran
- Catégorie 2 : Dans cette catégorie se trouvent des toponymes ayant des différences mineures en orthographe avec des variations phonétiques « segmentales » et/ou « suprasegmentales »
  - Malte (*Malta* en allemand, anglais, espagnole et suédois)
  - La Gambie (*Gambia* en allemand, anglais, espagnole et suédois)
  - Athènes (*Athen* en allemand [a'te :n] et suédois [a'te :n], *Athens* en anglais ['æθɪnz], *Atenas* en espagnol) [a'te nas]
- Catégorie 3 : la quatrième catégorie contient des toponymes montrant des variations systématiques, le plus souvent au niveau morphologique :
  - terminaison –ie en français, -ien en allemand et parfois en suédois , - ia ou -y en anglais et –ia en espagnol :  
*Italie* (frç), *Italien* (all), *Italy* (ang), *Italia* (esp), *Italien* (sué),  
*Bolivie* (frç), *Bolivie* (all), *Bolivia* (esp), *Bolivia* (sué),  
*Jordanie* (frç), *Jordanien* (all), *Jordan* (ang), *Jordania* (esp), *Jordanien* (sué),

- **Catégorie 4** : La dernière catégorie est consacrée aux toponymes dont l'orthographe et/ou la phonétique s'éloigne considérablement du toponyme français, mais qui ont les mêmes origines :
  - *Hongrie, Ungarn, Hungary, Hungaria, Hungría, Ungern*
  - *Chypre, Zypern, Cyprus, Chipre, Cypern*
  - *Moscou (Москва en russe et écriture cyrillique) => Moskau (all.), Moscow (ang) Moscú (esp), Moskva (suéd.)*

## 2.2 Test expérimental

Sous ce point, nous présenterons les résultats de l'expérience pilote réalisée auprès d'un groupe d'étudiants francophones.

### 2.2.1 Hypothèses

Nous émettons trois hypothèses de départ qui déterminent notre test expérimental. Elles sont établies par rapport aux difficultés que les sujets rencontrent pour identifier le toponyme en fonction de la déviation<sup>7</sup> avec leur langue maternelle (français) :

#### **Hypothèse 1 (approximation par rapport à la langue maternelle)**

Plus la déviation est importante, plus le sujet aura du mal à l'identifier, en d'autres termes, plus l'approximation du toponyme est proche de leur langue maternelle moins les sujets rencontreront des difficultés pour leur identification.

De la même manière, la variation systématique entre deux toponymes (par exemple *Italie* (toponyme français) > *Italien* (all, sué), *Italy* (angl), *Italia* (esp) ; *Croatie* (toponyme frç) > *Kroatien* (all, sué), *Croatia* (angl), *Croacia* (esp)) favorisera la compréhension et en même temps l'approximation au niveau orthographique.

#### **Hypothèse 2 (L2)**

Les sujets auront plus de difficultés à identifier les toponymes des L2 non romanes (anglais, allemand, suédois) que des L2 de la même famille, c'est-à-dire l'espagnol.

#### **Hypothèse 3**

L'approximation est plus prononcée lorsque le toponyme n'est pas connu par les sujets. Les connaissances générales (notamment du nom du lieu) et la proximité culturelle facilitent la compréhension du toponyme (Kromp, 2008). Plus un toponyme est connu (présent dans actualité ou grâce aux connaissances culturelles des apprenants), plus les apprenants seront en mesure de le connaître (Meyer, 2008) et de le reproduire malgré les écarts orthographiques constatés par rapport au (par exemple *Warschau* en allemand pour Varsovie en français).

### 2. Procédure expérimentale

Il s'agit d'une expérience de perception auditive, où les sujets écoutent des enregistrements effectués en laboratoire, ce qui nous permet de vérifier le degré de difficulté pour reconnaître un toponyme.

---

<sup>7</sup> En phonétique, la déviation est mesurée par rapport au nombre de phonèmes par unité lexicale qui sont différents entre deux ou plusieurs langues (Gooskens, 2007). L'exemple de l'exonyme français Moscou plus haut sur cette page montre plus de déviation phonétique avec les exonymes allemand et le suédois ([u/a<sup>o</sup>] frç/all ; [u/va] frç/sué) qu'avec l'espagnole.

### 2.2.3 Sujets

Les sujets sont des francophones, étudiants de l'Université Rennes 2 dans la filière L.E.A.<sup>8</sup>, troisième année du semestre 2, qui correspond à un niveau B2, selon le Cadre Européen Commun de Référence pour l'Apprentissage / Enseignement des Langues.

Notre échantillon de sujets pour la première expérience se compose de 17 sujets pour l'espagnol, 14 pour l'allemand et 12 pour l'anglais.

Quant à la deuxième partie de l'expérience, nous disposons de 10 sujets qui ont réalisé le test en espagnol, 15 en allemand et 11 en suédois.

### 2.2.4 Stimuli

Nous avons établi un corpus de stimuli en trois langues (allemand, anglais, et espagnol) à partir d'exemples issus de notre classification de toponymes en 5 catégories d'exonymisation pour la première partie de l'expérience et une quatrième langue a été ajoutée (le suédois) pour la deuxième partie de l'expérience.

Pour chaque type nous avons choisi 5 toponymes, 25 au total pour chacune des langues.

### 2.2.5 Procédure de passation des tests

Nous avons procédé à l'enregistrement des *stimuli* du corpus à l'aide du logiciel Audacity. Chaque *stimulus* est intégré dans une phrase authentique, extraite de la presse ou des articles d'Internet et, parfois simplifiée. Le toponyme se trouve toujours enchâssé dans la phrase, jamais en position initiale ou finale. La phrase est lue par une locutrice native avec une prononciation neutre dans les quatre langues (Bielefeld près d'Hanovre pour l'allemand, Chicago pour l'anglais, Salamanca pour l'espagnol et par un locuteur natif à Göteborg pour le suédois).

Nous avons soumis le test sous deux conditions différentes.

Pour la première partie de l'expérience, le test s'est déroulé dans un laboratoire de langues, SANAKO, de l'Université de Rennes 2. Chaque sujet écoute individuellement la liste de *stimuli* à l'aide d'un casque. Il écoute chaque phrase deux fois. Entre chaque phrase, il y a un intervalle de 12 secondes afin de répondre aux questions par écrit.

Les sujets répondent à deux questions par phrase, du type « 5 W » : *qui, où, comment, combien, quoi, etc.* Dans une des deux questions, la réponse attendue correspond au toponyme contenu dans la phrase. Nous avons décidé de poser les questions en français, leur langue maternelle, pour nous assurer une bonne compréhension de la question et donc être sûrs que la réponse visée était le toponyme (Athias, 1995). Les sujets doivent répondre aux questions dans la langue cible (allemand, anglais, espagnol). Tout cela est indiqué dans les consignes données avant la passation du test.

La deuxième partie de l'expérience a été réalisée dans une salle de cours avec notamment deux différences. Premièrement, les sujets écoutent les phrases dans une salle de cours et non dans un laboratoire de langues à travers des haut-parleurs comme dans les conditions d'un cours de compréhension auditive. Deuxièmement, chaque *stimulus* n'est présenté qu'une fois, c'est-à-dire qu'ils n'écoutent la phrase qu'une seule fois. Entre chaque *stimulus* il y a également 12 secondes d'intervalle.

Par exemple :

---

<sup>8</sup> Langues Étrangères Appliquées, niveau Licence (L3).



**a) Zypern (allemand) :**

- Sechs Flüchtlinge sind von der Mittelmeerinsel Zypern bis ins heilige Land geschwommen.
  - Combien de réfugiés ont regagné la Terre Sainte à la nage ?
  - D'où sont-ils partis pour réussir leur traversée de la Méditerranée ?

**b) Cyprus (anglais):**

- Forty years ago, on 20 July 1974, Turkish troops landed on Cyprus, in response to a Greek-led coup on the island.
  - Que s'est-il passé le 20 juillet 1974 ?
  - Où s'est déroulé le conflit ?

**c) Chipre (espagnol) :**

- Estalla una bomba casera junto a la vivienda de un árbitro en Chipre. Es el quinto ataque en lo que va de temporada.
  - Où a explosé la bombe ?
  - Combien d'attaques se sont déjà perpétrés ?

**d) Cypern (suédois) :**

- Tidigare utvanns stora mängder mineral på Cypern och koppar exporterades redan under bronsåldern men numera spelar gruvindustrin en underordnad roll.
  - Quelle activité économique est décrite dans le texte?
  - L'activité concerne quelle région?

### 3. Résultats

Nous prenons en considération le type de toponyme par rapport à notre classification et donc la déviation que celui-ci présente par rapport au français et nous observons les différences ou les similitudes des résultats selon la langue étudiée. Nous examinons également les pourcentages de bonnes réponses en fonction du degré d'approximation que présente le toponyme.

Pour le traitement de données, nous comptons un total de 2150 observations pour la première partie de l'expérience (43 sujets qui répondent à deux questions par phrase) et 1850 pour la deuxième partie de l'expérience (37 sujets qui répondent à deux questions par phrase).

Nous avons analysé les résultats en fonction des pourcentages de bonnes réponses ou d'erreurs pour réaliser les tests statistiques pertinents.

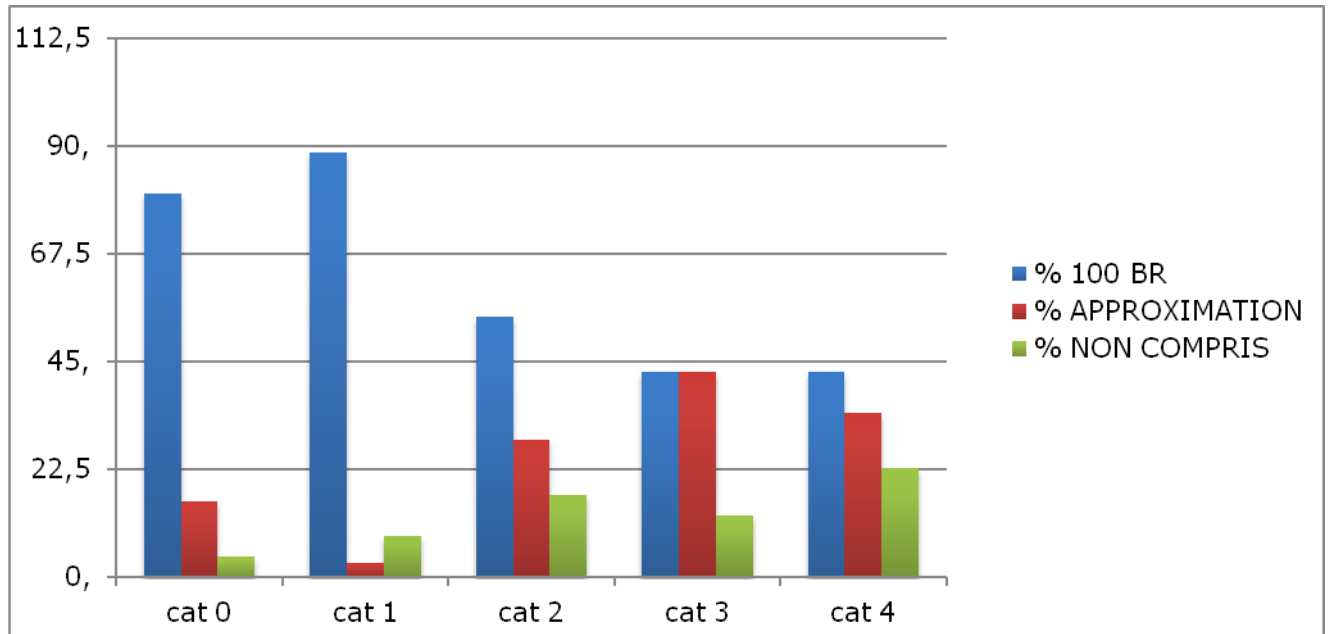
Afin d'analyser les résultats globaux pour chaque langue et par toponyme, nous étudions quatre facteurs :

1. Nombre de bonnes réponses
2. Nombre d'approximations
3. Nombre de 'non-compris'
4. Nombre de 'bien compris' (= BR + Approximation)

#### 3.1 Première série de tests

Pour pouvoir exploiter au mieux nos expériences, nous avons compté pour chaque catégorie le nombre de toponymes qui étaient bien compris et correctement orthographiés (100 BR en bleu) par les étudiants lors des deux séries de test, le nombre de toponymes qui étaient bien compris mais mal orthographiés (approximation en rouge) et le nombre de toponymes non compris

### 3.1.1 Les résultats en Allemand



Graphique 1 : Pourcentages de bonnes réponses, approximations et *stimuli* non compris par les apprenants de l'allemand dans chacune des 5 catégories pour le premier test.

Le graphique montre bien que les apprenants de l'allemand ont bien compris les stimuli des catégories 0 et 1 et que ces deux catégories engendrent peu d'approximations et de stimuli non compris. Il est néanmoins à noter que la compréhension des toponymes de la catégorie 0 est plus faible que pour la catégorie 1 et que le nombre d'approximations est plus élevé. Ce score est assez inattendu, mais s'explique par la mauvaise compréhension d'un seul *stimulus* (le toponyme Haiti) qui descend significativement la moyenne.

En règle générale et de manière tout à fait attendue, les approximations augmentent quand l'orthographe des toponymes dans les langues étrangères se distingue de l'orthographe française (28% en catégorie 2, 42% en catégorie 3, et 34% en catégorie 4). Cette observation confirme d'ailleurs notre première hypothèse, à savoir que les variations systématiques morpho-phonétiques favorisent la compréhension, mais également l'approximation en orthographe, comme le montre l'exemple des toponymes pour la Slovaquie : *\*Slovakei*, *\*Slovaquei*, *\*Slovakia*, *\*Slovakey*, l'orthographe correcte en allemand étant *Slowakei*. Le pourcentage d'approximation pour ce *stimulus* est de 92,5% contre 7,5 % de bonnes réponses.

Le pourcentage de 'non-compris' est plus faible en catégorie 3, catégorie qui contient avant tout des variations systématiques (comme le montrent les exemples ci-dessous). Ces variations systématiques facilitent la compréhension comparée à la catégorie 2, mais favorisent également les approximations :

**Français**

**Allemand**

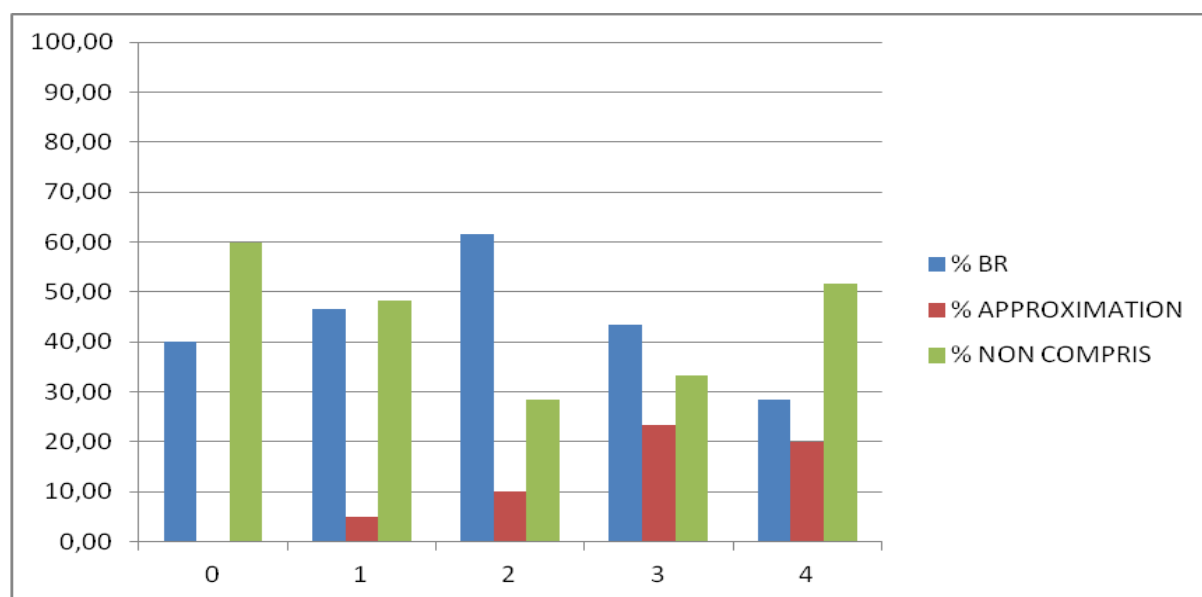
|           |                  |
|-----------|------------------|
| Slovaquie | <i>Slowakei</i>  |
| Géorgie   | <i>Georgien</i>  |
| Éthiopie  | <i>Äthiopien</i> |

Le fort pourcentage des ‘non-compris’ dans la catégorie 4 est dû à la grande distance phonétique entre le toponyme français et le toponyme allemand

Zypern (Chypre) = 64 % de non compris, 22% d’approximation lors du test 1

Warschau (Varsovie) = 73 % de non compris, 15% d’approximation lors du test 1.

### 3.1.2 Résultats globaux en anglais



Graphique 2 : Pourcentages de bonnes réponses, approximations et *stimuli* non compris par les apprenants de l’anglais dans chacune des 5 catégories pour le premier test.

Les résultats globaux montrent que le pourcentage de bonnes réponses le plus élevé est obtenu en catégorie 2 (61,67 %) et en catégorie 1 (46,67 %). L’hypothèse 1 n’est donc pas parfaitement confirmée. Le score élevé de la catégorie 2 s’explique probablement par la familiarité de certains toponymes de cette dernière catégorie (comme : China, Croatia, Egypt )

La catégorie 0 obtient un nombre de bonnes réponses plus faibles que ce qui était attendu (40%). Deux cas de figure se présentent : on observe soit une compréhension complète des toponymes (voir Canada et Oslo) soit une non-compréhension totale (pour Népal et Haiti).

Ainsi dans le cas de ces derniers, malgré une congruence graphique presque parfaite, les deux toponymes sont non compris. On suppose que cela est dû au fort effet de l’accentuation et de la prononciation de la voyelle dans ces toponymes (/nə'pɔ:l/ et /'heɪ.ti/).

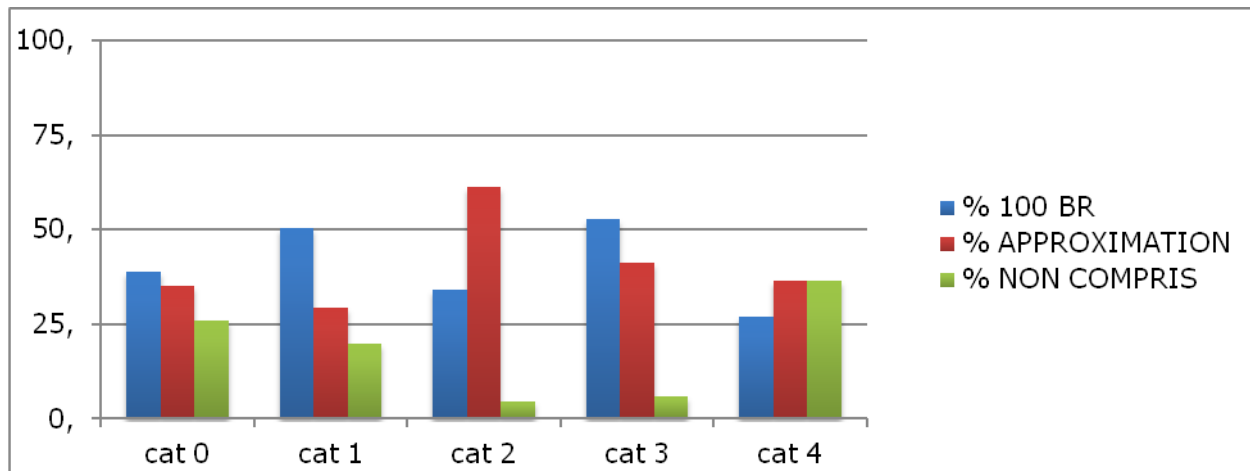
Enfin, on trouve moins de bonnes réponses dans la catégorie la plus complexe (28,33%).

L’hypothèse 3 se confirme pour la catégorie 4 : le phénomène de non compréhension est le plus important dû à l’effet des toponymes peu connus (par exemple Warsaw, Ghent).

En ce qui concerne l’approximation le pourcentage augmente avec la complexité des toponymes (de 0% à 23,37%) et est favorisé par la présence de variations systématiques : les toponymes de la catégorie 3 sont mieux compris, mais sont orthographiés avec beaucoup plus d’approximations (ex : Slovaqia, Slovaquia, Slovaquie, Sovaquia).

Les toponymes les plus problématiques sont Nepal, Israel, Cyprus, Georgia et Warsaw.

### 3.1.3 Résultats globaux en espagnol



Graphique 3: Pourcentages de bonnes réponses, approximations et *stimuli* non compris par les apprenants de l'espagnol dans chacune des 5 catégories pour le premier test.

Globalement on observe une bonne compréhension du toponyme dans les catégories 2 (95%) et 3 (94%). L'hypothèse 1 ne s'est pas confirmée, peut-être à cause des très faibles résultats obtenus pour un des toponymes « *Bali* », pour lequel nous avons observé 70% de réponses non comprises.

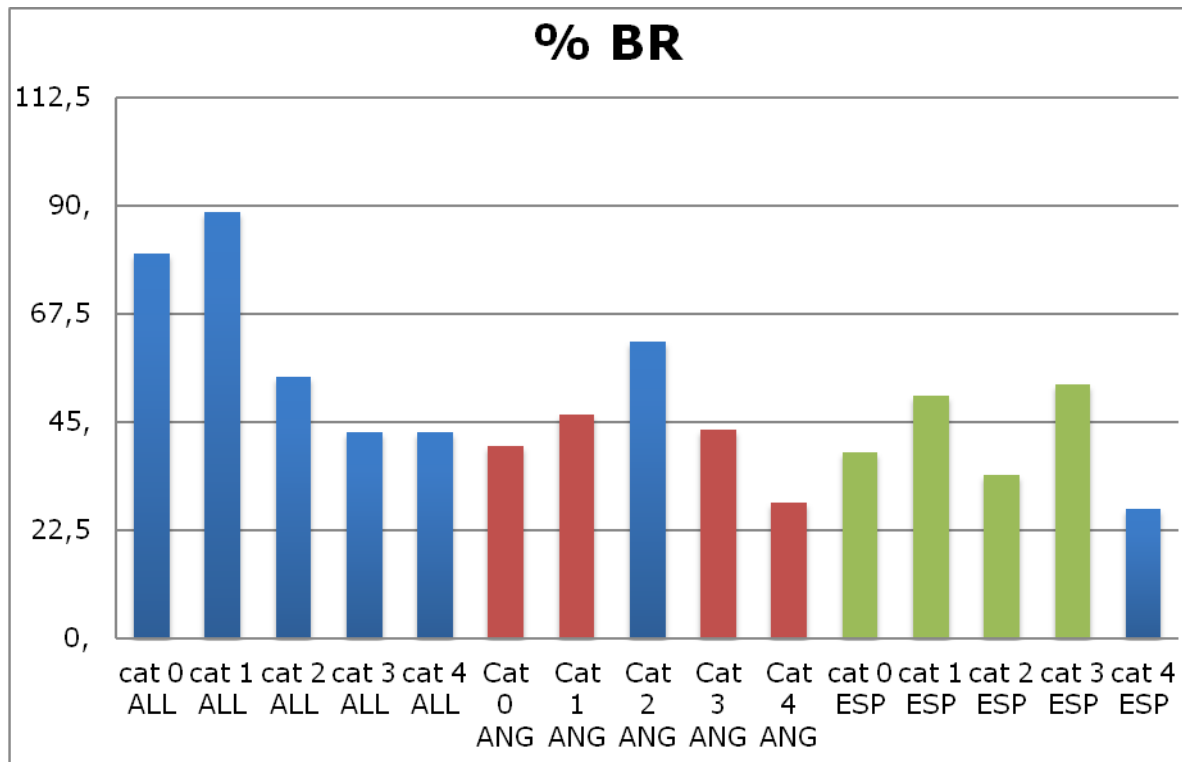
Nous constatons le pourcentage de toponymes non compris le plus élevé dans la catégorie 4, avec 36,47% de réponses. Dans ce cas, notre première hypothèse est confirmée: plus la déviation est importante, plus les sujets trouvent des difficultés à identifier le toponyme.

On observe un nombre d'approximations plus important dans la catégorie 2 (61%). En effet, les variations systématiques morpho-phonétiques favorisent la compréhension mais également l'approximation en orthographe: ex. : \**Etiopia* (8 occurrences), \**Etyopia* (2 occurrences), \**Ethiopia* (1 occurrence), \**Ethopia* (1 occurrence) \**Etiópa* (1 occurrence) \**Ethipía* (1 occurrence) pour le toponyme correctement orthographié en espagnol *Etiopía* ou \**Slovaquia* (6 occurrences) \**Slovaquiá* (2 occurrences) \**Eslovakia* (1 occurrence) pour *Eslovaquia*

## 3.2. Résultats pour les trois langues comparées

### 3.2.1. Comparaison de pourcentages de bonnes réponses pour les trois langues

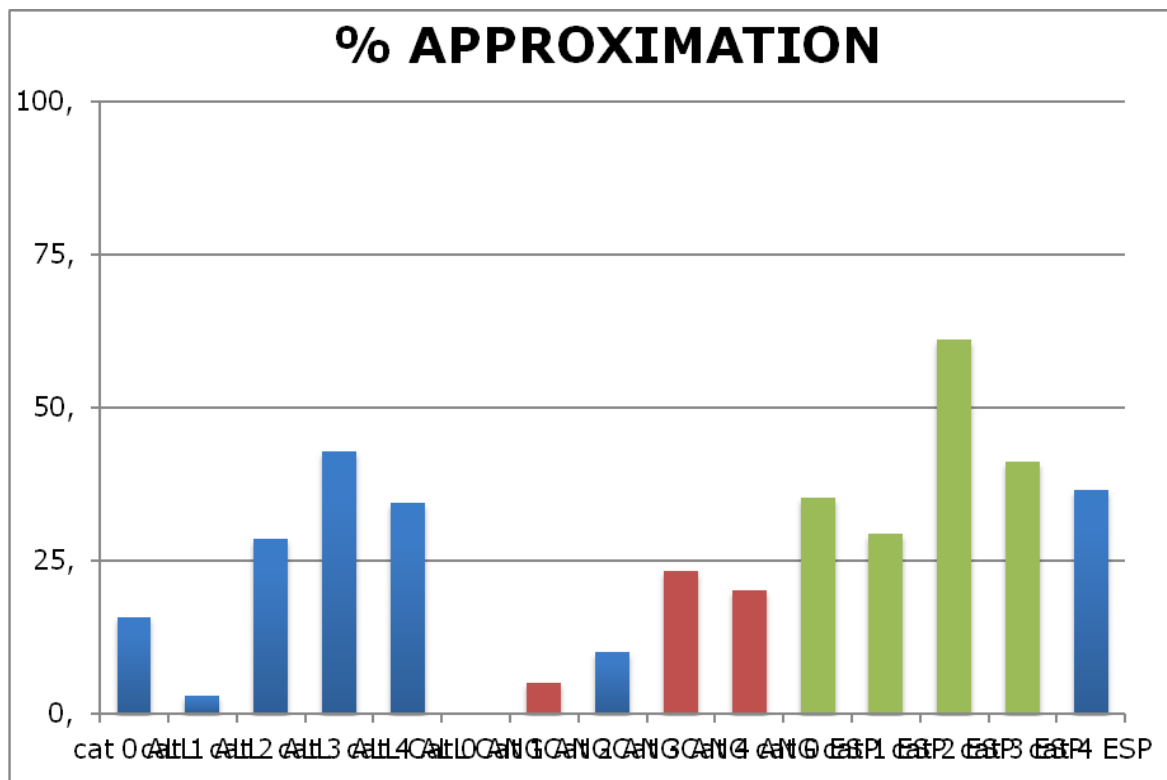
Nous observons de grandes différences entre les trois langues. Premièrement, en allemand les résultats sont plus en conformité avec les hypothèses, avec 61,7% de bonnes réponses. En anglais, le résultat de bonnes réponses descend à 44%. On trouve le meilleur score de bonnes réponses en catégorie 2, sans doute dû à la familiarité et à la connaissance des toponymes de la part des sujets (par exemple *China*, *Egypt*). Finalement, en espagnol, les résultats mettent en évidence le score le plus faible, 40,7%, laissant apparaître une grande irrégularité entre les catégories par rapport aux deux autres langues. Les résultats confirment notre première hypothèse, mettant en lumière les faibles scores de bonne compréhension pour les trois langues.



**Graphique 4 :** Pourcentages de bonnes réponses données par les apprenants d'allemand, d'anglais et d'espagnol dans chacune des 5 catégories pour le premier test.

### 3.2.2. Comparaison de pourcentages des approximations dans les toponymes par langue

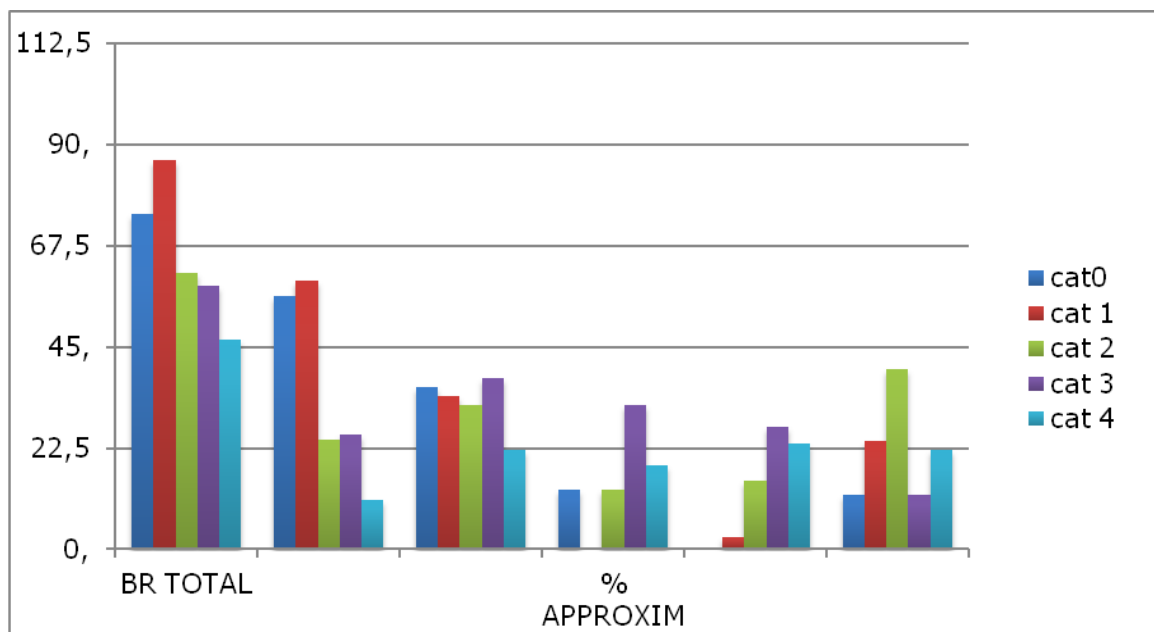
Nous avons pu observer 24,9 % d'approximations pour l'allemand, 11,7% d'approximations pour l'anglais et 40,7% d'approximations pour l'espagnol. Notre première hypothèse se voit confirmée car nous avons observé un nombre beaucoup plus important d'approximations orthographiques en espagnol, langue romane comme le français. La proximité avec le français aide à la compréhension, mais induit également plus d'approximations orthographiques. En revanche, en anglais et en allemand les écarts avec le français conduisent plutôt à une incompréhension nette (le toponyme n'est pas reconnu du tout) et réduit donc les approximations orthographiques. Nous avons pu constater des exemples en l'anglais, comme \*Valley (pour Bali) ou \*Serious (pour Syria) ou en allemand \*Vallmos-Aier (Buenos Aires).



Graphique 5 : Pourcentages d’approximations dans les toponymes donnés par les apprenants d’allemand, d’anglais et d’espagnol dans chacune des 5 catégories pour le premier test.

### 3.3 Deuxième série de tests

Pour cette deuxième expérience les mêmes *stimuli* en langues allemande, espagnole et suédoise ont été testés dans les conditions d’une salle de classe.



Graphique 6 : Pourcentages de bonnes réponses, approximations et *stimuli* non compris dans chacune des 5 catégories pour le deuxième test.

Les résultats montrent des tendances comparables entre les deux langues germaniques, l’allemand et le suédois. Le pourcentage de compréhension et de retranscription correcte

atteint 75% à 85% en allemand et 58% à 60% en suédois pour les catégories 0 et 1 et diminue significativement dans les trois autres catégories.

En espagnol par contre, les résultats se trouvent à un niveau assez bas et montrent peu de variation selon les catégories.

Les taux d'approximations augmentent pour les deux langues germaniques en catégories 3 et 4, avec cependant un pic en catégories 3.

Les approximations pour les *stimuli* espagnols sont élevées dans toutes les catégories.

Ces résultats confirment les observations faites à partir du premier test et notre hypothèse, qui dit que plus la proximité linguistique est grande plus les apprenants tendent à faire des approximations en L2.

#### **4. Conclusion**

Les expériences menées dans ce travail mettent en valeur des disparités entre les langues étudiées, et plus particulièrement entre les langues romanes et non romanes en matière d'approximation.

En espagnol, la similarité a entraîné plus de difficultés au niveau de la retranscription graphique des toponymes. L'approximation dans la retranscription est donc plus significative lorsqu'il y a proximité entre les langues, comme c'est le cas pour l'espagnol et le français.

L'approximation dans les langues non romanes, telles que l'anglais, l'allemand et le suédois, est plus importante pour les toponymes caractérisés par des variations graphiques systématiques.

Nos résultats portent à penser que les apprenants n'appliquent pas les règles orthographiques de la L2, pourtant supposées connues au niveau B2, pour retranscrire correctement les toponymes.

Par exemple le phonème /z/ est mal transcrit en allemand (dans Zyrien, \*Zürien), des signes diacritiques non existants sont ajoutés en L2 (« ë » en espagnol: « \*Israël » par « Israel ») ou des graphèmes tels que « Quia » sont remplacés par « kia » en anglais: « \*Slovaquia » par « \*Slovakia », « \*Slovaquei » en allemand par « Slowakei » et inversement en espagnol: « \*Eslovakia » pour « Eslovaquia ».

Nous envisageons d'abord élargir cette expérimentation à un plus grand nombre de sujets, tout en introduisant aussi un nombre plus important de *stimuli*. Dans cette expérimentation, nous inclurons des apprenants d'un niveau plus élevé (C1-C2), ainsi que d'autres langues, notamment une autre langue romane.

Enfin, une seconde étape consistera à ouvrir l'expérimentation à des tests de production orale, tels que des tâches de lecture, des tâches d'imitation ou des tâches de production orale spontanée. Nous aimerions étudier les capacités des sujets à prononcer des toponymes de façon appropriée tout en analysant le lien entre leurs compétences sur le plan de la production et sur celui de la perception.

## 5. Références

- Athias Jacques (1996). *Linguistique et didactique de l'allemand*. Masson, Paris 1995.
- Back, Otto (2002). *Übersetzbare Eigennamen. Eine synchronische Untersuchung von interlingualer Allonymie und Exonymie*. Vienne : Edition Praesens
- Charlotte Gooskens (2007). "The Contribution of Linguistic Factors to the Intelligibility of Closely Related Languages" in *Journal of Multilingual and Multicultural Development*, Vol. 28, No. 6, pp. 445-466.
- Kadmon, Naftali. 1997. *Glossaire de la terminologie toponymique*. (Version française de Glossary of Toponymic Terminology). <http://www.ngi.be/FR/glossaire/glossfr-inf.htm>
- Löfström, Jonas & Betina Schnabel-Le Corre (2010). Comment analyser et comparer les toponymes de différentes langues dans une perspective synchronique. *Nouvelle Revue Onomastique* 52, 2010.
- Schnabel-Le Corre B. & J. Löfstöm (2010). « Identification et accentuation des toponymes en contexte: l'exemple de l'allemand et du suédois » in Pepin N. & E. De Stefani, *Eigennamen in der gesprochenen Sprache*. Narr Verlag, Tübingen.
- Kromp Ilona (2008). *Eigennamen in der deutschen und polnischen Kinderliteratur unter textlinguistischem und translatorischen Aspekt*. Peter Lang, Frankfurt a.Main.
- Meyer, Bernd (2008). Interpretating Proper Names: Different Interventions in Simultaneous and Consecutive Interpreting? trans-kom, 2008 (=http://www.trans-kom.eu)
- Nübling, Damaris (2012). *Namen. Eine Einführung in die Onomastique*. Narr Studienbücher, Narr Francke Attempto, Tübingen.
- UNGEGEN, United Nations Group of Experts on Geographical Names. <http://unstats.un.org/unsd/geoinfo/UNGEGN/publications.html>