

Le Vocabulaire fondamental du plurilinguisme roman (VFPR), EuroComDidact ToGo et l'apprentissage

Franz-Joseph Meißner (Gießen)

Brève présentation du *Vocabulaire fondamental du Plurilinguisme roman*

Le VFPR (*Core Vocabulary of Romance Plurilingualism* [CVRP], *Kernwortschatz der romanischen Mehrsprachigkeit* [KRM]) est probablement la première banque de données plurilingue d'une orientation pédagogique basée sur la fréquence lexicale dont les enregistrements découlent des comptages de la lexicologie computationnelle. Ainsi, l'inventaire du VFPR dépend en grande partie des indications des *Frequency dictionaries* au sous-titre *Core vocabulary for learners* de la maison d'édition Routledge, disponibles pour l'espagnol, le français et le portugais et d'autres langues. Par ce biais, le VFPR est – indirectement – alimenté des vastes corpus comptant des millions et des millions de *tokens* et d'occurrences. Ces *core vocabularies*, confectionnés sous l'égide de Mark Davies (2006) et réunissant des lemmes d'un rang de fréquence <5000 (cf. Davies & Preto-Bay 2008 ; Lonsdale & Le Bras 2009), se sont montrés particulièrement appropriés¹ pour confectionner un inventaire fondamental plurilingue de mots du même rang de fréquence (dorénavant rf). Malheureusement, un tel corpus n'existe pas pour l'italien. Ainsi on a dû recourir au *Lessico di frequenza dell'italiano parlato* (LFIP) de De Mauro et son équipe (1993). Notons que le LFIP se sert d'une définition du *parlato* assez vaste. Par conséquent, on y trouve le segment lexical qui correspond à la définition du *vocabulaire central* (*Zentralwortschatz*) fournie par Zimmer (1990 : 75 et suiv.) d'un rf~40.000 (à peu près) à partir duquel une liste au rf <7000 a été confectionnée. Vu le discours des italianistes sur la question s'il y a un *italiano parlato unitario* et son impact national, on comprend pourquoi le LFIP a été façonné ainsi. L'élargissement de l'intervalle saisi nous a mené à introduire une quatrième catégorie réduite à l'italien. Les intervalles sont échelonnés de façon suivante :

- I) inventaire de base <2000
- II) inventaire fondamental <5000
- III) inventaire fondamental <7000 (exclusivement pour l'italien)
- IV) inventaire >5000 ou >7000

Malgré l'incontestable fiabilité des listes Routledge, le sélectionneur choisi rf<5000 excluait des vocables très présents dans classes de langue (par ex. *tableau noir*, *craie*, *retroprojecteur*...). Ceci explique la consultation de listes spécifiques qui reflètent l'expertise d'auteurs pédagogiques. Pour remédier à cet inconvénient, on a recouru aux vocabulaires de base (*Grundwortschätze*) de la maison Klett dont la composition suit les mêmes principes pour les différentes langues-cible. Ainsi, l'inventaire du VFPR et surtout le segment rf<2000 a été enrichi de lemmes issus des listes KLETT.

L'originalité de la banque de données du VFPR découle de sa composition pluri-langues (français, italien, portugais, espagnol auxquels s'ajoutent l'allemand et l'anglais comme langues désambiguïsantes). Une liste ÉTYMONS complète les rubriques LANGUES. Pour permettre un triage instantané, déclenché par les besoins linguistiques et pédagogiques de l'utilisateur, plus d'une cinquantaine de sélecteurs ont été développés et intégrés.

¹ Voir Davies pour la liste espagnole (2006), par exemple : "Based on a twenty million word corpus evenly divided between spoken, fiction and non-fiction texts from both Spain and Latin America, the Dictionary provides a detailed frequency-based list, as well as alphabetical and part-of-speech indexes to ensure maximum ease of access to the information and efficiency of use."

Didactiquement, la structure et l'inventaire du VFPR se situent dans le cadre de l'approche intercompréhensive (famille romane). Autrement dit, cette liste sert en premier lieu à la promotion d'une compétence de lecture plurilingue. Un vocabulaire de production aurait bien sûr exigé l'intégration d'autres informations (prononciation, définition, exemples cotextuels, indication d'usage, etc.). Sont visés comme destinataires les romanophones et non-romanophones maîtrisant au moins une langue romane et/ou l'anglais au niveau B1 du *Cadre européen commun pour les langues* (CERL) ou au-delà, désireux d'élargir leur compétence de lecture à une ou plusieurs langues romanes. L'orientation vers ces deux groupes signifiait que le CVPR devrait répondre à trois profils de compétences linguistiques :

- les locuteurs natifs d'une langue romane ayant, en plus, probablement une certaine connaissance de l'anglais et peut-être d'une deuxième langue romane ou/et le latin. Vu que les langues romanes sont massivement interconnectées aux niveaux du lexique et de la grammaire, ces *native speakers* en profitent pour effectuer (facilement) des transferts d'identification dans d'autres langues romanes.
- les locuteurs non-natifs d'une langue romane mais maîtrisant au moins une langue romane au niveau B1 du CERL. Pour ce groupe, leur langue-pont romane a le statut d'une *interlangue*² ou langue de l'apprenant (tout comme la langue-cible)
- les locuteurs sans connaissance d'une langue romane, mais qui ont une bonne maîtrise de l'anglais.³

Au total, la liste électronique quadrilingue compte à peu près 28.000 lemmes appartenant aux langues-cible réunis en 9.555 séries composées d'équivalents sémantiques auxquels s'ajoutent 15.000 lemmes pris des langues dites désambiguïsantes. La présentation en séries et le manque d'intersynonymie expliquent pourquoi les mêmes lemmes peuvent apparaître dans plusieurs séries. C'est pourquoi le VFPR connaît le marqueur « formes récurrentes intra-langues ». Le marquage permet l'identification et la computation des enregistrements récurrents ainsi que des formes uniques. Notons que les formes non-récurrentes des langues-cible totalisent par langue-cible en moyenne plus de 7.000 lemmes sur 9.500. Pour identifier le nombre de formes que l'élève désireux d'acquérir un niveau de lecture B2 doit apprendre, on ne se tourne pas aux mots mais aux interlexes (cf. ci-dessous).

La banque de données du VFPR est un instrument flexible permettant de dresser des listes différentes, façonnées aux besoins différents des destinataires. Du côté des enseignants, la composition de telles listes peut suivre les intérêts assez concrets des élèves qui demandent un input plus approprié. Par conséquent, on ajoute un quatrième groupe :

² Une *langue de l'apprenant* ou *interlangue* désigne un statut transitoire de compétence de langue, caractérisée par plusieurs critères : en comparaison avec une langue maternelle, l'interlangue est (1) transitoire et idiosyncratique, (2) restreinte et limitée, (3) systématique, (4) instable, (5) hautement évolutive et (6) portant des « traces » des langues impliquées (Meißner 2017).

³ Qu'on permette une parenthèse. Des notions comme *langues romanes*, *philologie romane* désignent une classification scientifique. Mais attention, penser (comme c'est souvent le cas) que c'est une langue romane seule qui déclenche des transferts d'identification dans une autre langue romane, mène à l'erreur. D'après les computations des lemmes réunis dans l'inventaire du VFPR, nettement plus de 50 pour cent des lemmes entièrement ou partiellement pan-romans (congruences de forme) se retrouvent en anglais. Seulement 12,5 pour cent est la part des séries exclusivement romanes. En réalité, pour effectuer un transfert d'identification dans une langue D, l'apprenant exploite des schémas cognitifs dans une langue A, B ou C, c'est-à-dire là où il croit trouver des ressemblances inter-langues, que ce soit dans une langue romane étrangère, sa langue maternelle ou l'anglais. De nombreuses recherches empiriques analysant les processus mentaux ayant trait aux transferts d'identification inter-langues le prouvent, et des études interlexicologiques sur les ressemblances de forme et de contenu en expliquent les raisons.

- les enseignants désireux de confectionner un input optimisé parce qu'adapté aux besoins de leurs étudiants.

La création du VFPR et du VBPR

Dans le cadre du projet européen MIRIADI⁴, l'ancienne équipe de l'université de Giessen s'était proposé de compiler un inventaire des *mots opaques* pour le vocabulaire roman d'un rf<5000. Car ce sont les mots non-transparents qui posent problème à l'intercompréhension.

Au préalable, ce projet demandait trois démarches :

- a) la confection d'une liste quadrilingue fiable, comprenant la totalité des vocabulaires fondamentaux (rf<5000) ou de base (rf<2000) des quatre langues-cible,
- b) la création de séries comprenant les équivalents de sens des quatre langues-cible, à savoir le français, l'espagnol, l'italien et le portugais,
- c) la mise sur pied d'une structure permettant la caractérisation des séries selon des critères lexicologiques et pédagogiques.
- d) Dès le départ, il était clair que les destinataires étaient aussi bien des locuteurs maternels de langues romanes que des non-romanophones ayant une solide maîtrise d'au moins une langue romane. Ceci explique le rajout de l'anglais et de l'allemand comme langues de désambiguïsation.

Comme le programme de MIRIADI était orienté vers la création de supports numériques dans le domaine de l'approche intercompréhensive, le partenaire de Giessen proposa

- e) la confection d'une appli d'apprentissage alimentée par une banque de données (qui alors était encore à compiler) réunissant les mots opaques les plus fréquents de quatre grandes langues romanes.

La structure de base de la banque de données et de l'appli était claire dès le début. Elle prévoyait l'arrangement parallèle des lemmes, ce qui provoque l'identification immédiate d'un lemme-cible à partir de la forme visuelle d'un lemme de départ (ou de plusieurs) (voir figure 1).

Se rendant compte assez tôt que le cadre temporel et financier (qui ne pouvait être augmenté) était insuffisant pour créer et une telle liste et un didacticiel adéquat, les moyens obtenus étaient dirigés exclusivement vers le développement d'une appli embrassant un « vocabulaire de base » (VBPR), c'est-à-dire pour un segment du rf<2000. À la différence de l'appli *Eurocomdidact ToGo* alimentée par l'inventaire entier du VFPR, cette appli est alimentée exclusivement par le VBPR. Elle est accessible dans une version bêta sous <https://miriadi.eurocomdidact.eu>.

La fabrication du VFPR demandait des ressources et des expertises considérables dans les domaines lexicographique aussi bien qu'informatique. Tout compté, les heures de travail investies se montent à nettement plus de 16.000. Afin de protéger la banque de données contre les actes de piratage, elle est localisée à un endroit protégé⁵.

⁴ MIRIADI est l'acronyme pour « Mutualisation et Innovation pour une Réseau de l'Intercompréhension à Distance » (<https://www.miriadi.net/>).

⁵ Pour répondre au maximum aux intérêts de la recherche et à la transparence, la fabrication du VFPR et de son inventaire ont été décrits dans deux publications de manière détaillée (Meissner 2015 ; 2018). Le lecteur intéressé trouvera des informations supplémentaires sur le site www.eurocomdidact.eu.

ABRI	RIPARO	abrigo	abrigo	shelter, cover, refuge
ACHEVER	portare a fine	ACABAR	ACABAR	achieve
AGRÉABLE	GRADEVOLE	AGRADÁVEL	AGRADABLE	agreeable
AIGUILLE	lancetta	AGULHA	AGUJA	needle
AINSI	COSÌ	ASSIM	ASÍ	thus, so
ALLEMAND	TEDESCO	ALEMÃO	ALEMÁN	German
ÂME	ANIMA	ALMA	ALMA	soul
APERCEVOIR	ACCORGERE	aperceber	PERCIBIR	perceive
avenue	VIALE	AVENIDA	AVENIDA	avenue
AVION	AEREO	AVIÃO	AVIÓN	airplane, aircraft
botte	stivale	BOTA	bota	boot
CADRE	CORNICE	QUADRO	CUADRO	frame
CHARGE	CARICA	CARGA	CARGO	cargo
CHARGE	CARICO	CARGA	carga	charge, load
CHAUFFEUR	conduttore	chauffer	chófer	driver, chauffeur
chèque	ASSEGNO	CHEQUE	CHEQUE	check
choir	CADERE	CAIR	CAER	fall
COMPARER	PARAGONARE	COMPARAR	COMPARAR	compare
GOUTTE	GOCCIA	GOTA	GOTA	drop
dédain	DISPREZZO	DESDÉM	DESDÉN	disdain

Figure 1 : Sélection des opacités italiennes du VFRP (extrait)

Les enregistrements en majuscules indiquent les lemmes issus des vocabulaires de base de la maison KLETT.

Inventer une structure pour un vocabulaire fondamental quadrilingue

La confection d'un tel inventaire pédagogique quadrilingue était une nouveauté à maints égards, et sa composition constituait un certain défi lexicographique. Car il ne s'agissait pas d'une pure et simple traduction d'un vocabulaire en trois, respectivement en cinq langues supplémentaires. Une telle démarche n'aurait pas respecté le rapport intime civilisation-langue, unique pour chacune d'entre elles. Par conséquent, il fallait compiler l'inventaire en ayant recours aux listes appropriées, confectionnées pour chaque langue-cible. Par conséquent, le répertoire du VFRP présente une sélection fidèle au statut sémantique et fréquentiel des lemmes dans chacune des langues-cible.

L'énumération réunit les démarches nécessaires au niveau de la planification et de la réalisation du VFRP :

- PRISE DES ENREGISTREMENTS
- saisissement des lemmes dans un vocabulaire de fréquence (<5000),
- ... du français
- ... de l'italien (<5000, respectivement <7000⁶)
- ... du portugais
- ... de l'espagnol
- COMPOSITION DE SÉRIES POUR CHAQUE LEMME TROUVÉ
- identification des équivalents sémantiques pour chaque lemme. Étant donné que le rang de fréquence des équivalents n'est pas forcément le même, il a souvent été nécessaire d'ajouter

⁶ La liste choisie dans de De Mauro incorpore des vocables ayant un rf<7000.

d'autres vocables qui ne suffisent pas au critère fréquentiel. Ceci explique pourquoi le nombre des séries dépasse celui des lemmes.

- SPÉCIFICATION ET INTÉGRATION DE CHAMPS DE SÉLECTION DANS LA BANQUE DE DONNÉES
- définir et établir une cinquantaine de sélecteurs (trouver des marqueurs permettant la restructuration immédiate des listes visualisées)
- DÉVELOPPEMENT D'UNE SYSTÉMATIQUÉ PERMETTANT LA CATÉGORISATION DES LEMMES SELON LEUR « DEGRÉS DE TRANSPARENCE INTERLEXICALE » (classe 1 à 4) et remplissage du champ « DEGRÉ- TRANSPARENCE ». Rappelons : Une série peut être « transparente au niveau des quatre langues-cible », par ex. *absence, assenza, ausência, ausencia, absence, Absenz* ; « de trois » ou de « deux langues-cible », par ex : *travail, lavoro, trabalho, faena* ou bien *mantien, sostentamento, sustento, mantenimiento* ou « entièrement opaque », par ex. *boucher, macellaio, açougueiro, carnicero, brique, laterizio, tijolo, ladrillo, en. brick*.
- CRÉATION ET REMPLISSAGE D'UN CHAMP DE SÉLECTION « ÉTYMON » (inachevé).
- CRÉATION ET REMPLISSAGE D'UN CHAMPS NOMBRE-RADICAUX. La classification donnait des indications diachroniques assez fiables pour catégoriser les classes de transparence.
- RAJOUT DES RUBRIQUES DE DÉSAMBIGUÏSATION POUR L'ALLEMAND ET L'ANGLAIS. Pour ces deux langues, les critères de la correspondance fréquentielle n'ont pas été pris en compte. En cas de congruence de forme, le mot était inséré dans la rubrique ANGLAIS ou ALLEMAND, par ex. fr. *chaîne*, it. *catena*, pt. *cadeia*, esp. *cadena*, angl. *chain, catena*.
- insertion des équivalents allemands et anglais dans les séries
- ENREGISTREMENT DE SYNONYMES ou parfois d'antonymes pour permettre le réseautage sémantique inter-sériel et inter-langues (inachevé). Dans la série *accorder, accordare, acordar, acordar*, l'utilisateur trouvera des renvois à l'esp. *concordar* et au pg. *otorgar*. Les séries correspondantes le renvoient en plus à d'autres séries délivrant d'autres lemmes.
- CRÉATION ET REMPLISSAGE D'UNE RUBRIQUE « INTERLIGALEX » (en voie d'être achevée)
- création et remplissage d'autres champs de sélection (ne seront plus explicités ici)

Quelques principes de la confection du VFPR explicités

Quant à la composition des séries d'équivalents sémantiques, on a retenu six critères suivants :

(1) « Rang de fréquence et de dispersion ». Cet indicateur couvre la fréquence absolue des *tokens* lemmatisés (par ex. *je vais, tu vas, iront > aller*). Il se peut bien sûr que plusieurs lemmes arrivent à un même rang. La dispersion traduit la présence d'un lemme dans différentes variétés diatopiques ou diastratiques. Cet indicateur est d'une grande importance au fur et à mesure qu'une langue est polycentrique. Ainsi l'espagnol « de ambos hemisferios » connaît des variétés cubaines, argentines, mexicaines, etc. ou de l'espagnol péninsulaire. Les mots ayant droit de cité dans le VFPR doivent, dans ce cas, faire partie de l'espagnol international (ou « atlantique »). Des régionalismes ne sont pas admis.

En effet, la congruence de forme est loin d'être sans équivoque. Où commence la ressemblance des signifiés – et où finit-elle (voir pour ce point : Müller-Lancé 2003 : 159 et suiv.) ?⁷

⁷ Qu'on permette une deuxième parenthèse. Depuis les débuts d'une didactique du plurilinguisme, on sait que la congruence de forme entre des mots d'une langue A et d'une langue B ne déclenche pas automatiquement un transfert d'identification inter-langues (Bieritz 1974). Car aucun pronostic exclusivement lexicologique ne peut prendre en compte des variables personnelles – motivation, résilience, connaissance de langues appartenant à la même famille que la langue-cible, expériences au niveau de l'acquisition de langues étrangères, etc.

(2) « Transparence/opacité inter-langues infra-sérielles ». Ce critère concernant ressemblances et analogies s’est avéré crucial pour déclencher un transfert d’identification inter-langues. Qu’on compare la série mélangée ou hybride :

MOUILLÉ	UMIDO	MOLHADO	MOJADO
---------	-------	---------	--------

Figure 2 : Lemme italien sortant de la congruence de forme sérielle

Évidemment, la forme italienne *umido* fait exception dans la série. Autrement dit, elle est sériellement opaque. Mais où en est-on quant à *mouillé* à la différence de *molhado*, *mojado* ? Il est probable que l’identification du mot français ne réussira pas facilement⁸ parce qu’il compte une syllabe de moins (cf. Jamet 2007 : 157⁹) et que le nexus *-[ill]/je/-* est assez loin de ses correspondants dans les deux langues, *-[lh]/ʒ/* et *[j]/χ/-*. Les données relevées chez des apprenants indiquent que la définition de la congruence de forme ne doit pas forcément se baser sur une étymologie (historiquement) correcte mais sur la capacité estimée de l’apprenant à effectuer un transfert d’identification inter-langues. Néanmoins l’étymologie fournit un critère important, car, en général, différents étymons signalent différentes congruences de forme (*abattere*, *abbattere*, *derrubar*, *derribar*¹⁰). L’exception concerne les « accidents phonétiques » comme dans es. *vino* (bebida)≠(forme du verbe *venir*) ; fr. *voler* < VOLARE (dans l’air)≠(débaucher), expliquent Dauzat et al. (1971 : 799), est un cas à part, étant donné que la dernière signification découle de la première. Il va sans dire que de tels discernements ne sont pas à la portée des apprenants.

Regardons deux séries partiellement opaques, partiellement transparentes :

BRUIT	RUMORE	RUÍDO	RUIDO
-------	--------	-------	-------

Figure 3 : Série mélangée (F≠I≠P=S)

poupée	pupa	BONECA	MUÑECA
--------	------	--------	--------

Figure 4 : Série mélangée (F=I≠P≠S)

La première série démontre apparemment deux convergences et deux divergences de forme dans une distribution variée :

Néanmoins, nombreuses sont les études empiriques qui ne laissent pas le moindre doute sur le fait que la ressemblance inter-langues des signifiants joue un rôle crucial dans l’apprentissage du vocabulaire d’une langue étrangère (De Florio-Hansen 1994). Les études portant sur le « démasquage » des mots « difficiles » dans des langues « inconnues » soulignent unanimement que la connaissance de quelques régularités inter-phonologiques améliorent le taux d’identification de manière significative⁷. Connaître de telles règles facilite l’appartenance de *mouillé* – *molhado* et *mojado* à la même famille. Quant à l’attribution d’une série à tel ou tel degré de transparence, la connaissance supposée de l’utilisateur peut déclencher un changement de classe.

⁸ La mise en série a pour conséquence que le type du transfert d’identification intra-sériel est décontextualisé et que la possibilité de « intelligent guessing » à partir des cotextes n’est pas donnée. Dans un premier temps, ce type de visualisation entraîne surtout ce que Klein & Stegmann appellent le « démasquage » des cognats de langue en langue à l’aide de la connaissance des régularités interphonologiques.

⁹ Parmi les congénères non-identifiés par les « oreilles italiennes » on trouve : *onze/undici*, *héros/eroe*, *empire/impero*, *roman/romanzo* et les terminaisons en nasale : *médecin/medico*, *développement/sviluppo*, *incidence/incidenza*, *entraînement/allenamento* et des compositions complexes : *n’importe quoi*, *d’ailleurs*, *en revanche*, *quel que soit...*

¹⁰ Voir Corominas 1980 : “DERRIBAR, 1202, probablemente derivado de *riba* con el sentido de ‘hacer caer de un ribazo’, de donde ‘echar al suelo a una persona o un animal’ y luego ‘demolir una construcción’.” – Dauzat et al. 1971: « ... du lat. pop. **abattère...* ».

	FR	IT	PT	ES
pour le FR	0	-	-	-
pour l'IT	-	0	-	-
pour le PT	-	-	+	+
pour l'ES	-	-	+	+

Tableau 1 : Distribution des transparences infra-sérielle

Comme celle de la série de *mouillé*, cette distribution, elle aussi, est douteuse étant donné qu'il a une certaine ressemblance entre fr. *bruit* et les congénères *ruído*, *ruido*. Dauzat (et al. 1971 : 114) expliquent la forme française à l'aide du latin populaire **brūgĕre*, croisement de RUGIRE (rugir) et *BRAGERE (*braire*).

Les études empiriques attestent que de nombreux apprenants ont du mal à réussir un transfert d'identification dans les cas où la graphie des couples comparés diffère en position initiale, même si les formes sont assez proches l'une de l'autre (*four*, *forno*, *horno* ; *fer*, *ferro*, *hierro* ; es./pt. *cabra*, it. *capra*, fr. *chèvre*). L'exemple démontre la difficulté qu'on rencontre au niveau de la catégorisation « transparent à première vue – peu transparent ». Un autre facteur concerne l'ordre et les contextes dans lesquels un apprenant rencontre un lemme à identifier ; un troisième la forme sous laquelle il l'aperçoit. Ainsi la racine *contin-* est la base pour la série graphique *continuer* – *continuar* – *continuar* – *continuar* – (*to*) *continuo* complètement congruente, laquelle se distingue cependant des formes orales : /fr. kɔ̃tinje, angl. kən'tinju(:), es. kontinuár, it. kontinu'are/. Il arrive qu'un mot parlé est plus facilement à identifier que sa forme écrite.

(3) Le cas de fr. *bruit* - *ruído*... concerne aussi « la gradation sérielle de la transparence inter-langues ». Comme on vient de le voir (figure 3) pour le couple espagnol/portugais *ruído/ruido*, il y a une congruence de forme de presque cent pour cent. Celle de fr. *poupée* et it. *pupa* cependant semble moins grande. Si, par contre, l'apprenant-utilisateur sait que l'orthographe française traduit le son /u/ par le digraphe [ou] (*boucher*, *trouver*, *ou*, *où*), les formes fr. *poup-* et it. *pup-* finissent par atteindre une convergence formelle quasi-totale. Au niveau des graphèmes, il établira un rapport similaire à l'égard de [-ill-, -gli-, -lh- et -j-] : *mouillé*, *moglie*, *mulher*, *mujer* ; *accueillir*, *accogliere*, *acolher*, *acojer*. Les exemples démontrent une fois de plus que la connaissance des régularités interphonologiques change la donne non seulement au niveau de l'apprenant, mais aussi pour la classification des degrés de transparence lemmatique. L'identification de la correspondance fr. [ou]/it. [u] dans *poupée-pupa* suscite une autre conclusion. Elle concerne des formes morphologiquement proches et assez fréquentes, comme celles de *trouver/trovare*, *prouver/probare*, *doux/dolce*. L'exemple montre que la connaissance d'une règle phonologique provoque la connaissance d'autres règles. – Les exemples discutés prouvent que hormis la ressemblance des lexèmes comparés, deux facteurs sont essentiels pour la réussite de ce transfert : du côté apprenant, son expérience en intercompréhension, sa motivation, sa résilience aux difficultés ayant trait à l'intercompréhension. Grosso modo, on dira (ce qui est assez trivial) : plus un individu connaît des schémas interphonologiques correspondants plus il est à même d'identifier de nouvelles bases de transfert.

Évidemment, comme les destinataires de l'appli *EuroComDidact ToGo* ne sont pas les linguistes-experts, mais des apprenants ou des enseignants cherchant des appuis pour leur cours, une connaissance en phonétique historique ou en phonologie comparative romane ne peut être présupposée (voir Klein & Reissner 2003). Comme on vient de le dire pour les italo-phones écoutant le

français, ce n'est pas toujours une filiation étymologique correcte qui décidera de l'attribution d'un signifiant à tel ou tel degré de transparence ou d'opacité, mais une compétence côté utilisateur présumée de « démasquer » le lemme en question.

(4) Problèmes sémantiques : « polysémie et homonymie ». Disons-le d'emblée : la synonymie au sens strict n'existe que dans les cas des monosémies, c'est-à-dire dans les langues de spécialité scientifiques. Ainsi l'explique Schaefer (1990, p. 71) :

Une équivalence de signification n'existe que dans les cas où un lexème A correspond à un lexème B au niveau paradigmatique, syntagmatique et diasystématique (diachronique, diatopique, diastratique, diaconnotatif, diatechnique, dianormatif et diafréquent) - au moins partiellement. (trad. Meissner)

Dire qu'il n'y a pas de synonymie au sens très strict signifie aussi que l'intersynonymie n'existe pas non plus. Certes, une telle conclusion est linguistiquement correcte, mais elle ne correspond pas aux expériences pratiques et aux données des recherches empiriques. Car, d'après l'expérience d'innombrables apprenants de langues étrangères et le rôle fondamental de l'intercompréhension qui a toujours accompagné le discours interculturel et polyglotte, les ressemblances inter-langues et les « intersynonymes approximatifs » sont un facteur efficace pour faciliter l'acquisition d'une langue.

Mais comment identifier les équivalences de signification inter-langues ? Nous explicitons cette étape de la construction de séries « inter-synonymiques » en partant de la série française *râler*, *rous-péter*, *gronder*, *gueuler*, *faire la gueule* comme exemple. Évidemment, la proximité sémantique de ces verbes s'explique par le partage d'un même « noyau sémantique commun » : 'exprimer son désaccord de manière plus ou moins ostentatoire'. Bien sûr, ce noyau sémantique commun admet des écarts de signification relevés dans des analyses sémiques. Ainsi à la différence de *râler*, le verbe *gueuler* ne peut jamais cacher la sémie 'bruit, bouche' qu'il partage avec *gueuleton*, *je te casse la gueule*, *le gueulard*, *ça gueule*, *gueulement*, *gueuletonner*, *gueule de bois*, *gueule de loup* dont les associations suscitées concernent les références 'visage', 'crier', 'bouche et manger', 'malaise à force d'avoir trop bu', 'fleur à la forme d'une bouche' (all. *Löwenmaul*). *Râler* par contre signale plutôt 'exprimer un désaccord', le canal utilisé étant secondaire. Malgré ces écarts de signification, l'identification du noyau sémantique permet la concrétisation du sens occurrence du mot dans son cotexte. Quand on regarde la série italienne *brontolare*, *belare*, *criticare*, *mormorare*, *bofonchiare*, *lamentare*, *borbottare*..., on finit par retrouver non seulement le même noyau sémantique mais en grande partie aussi les mêmes sémies. À comparer les « parasynonymes », on rencontre par exemple l'équivalent italien de fr. *gueuler* : *borbottare*. Tous deux s'associent à la sémie 'faire du bruit'. Mais regardons les verbes es. *refunfuñar*, it. *lamentarsi*, fr. *râler*, pt. *resmungar* (angl. *to grumble*, all. *motzen*) dans leurs cotextes (incorporés dans des concordanciers) :

*El charlatán me clavó su mirada y quiso refunfuñar, aunque sus palabras sonaron cordiales.
... seu filho começa a resmungar que é o único na escola (e no bairro) que...
Il se contenta de râler encore plus fort, la bouche grande ouverte.
... 4 in cui quest'ultimo sembra lamentarsi di tanti mediocri traduttori dei suoi romanzi...*

Quand on connaît un de ses verbes, on finit facilement par comprendre les autres à partir de leurs cotextes (à condition qu'on dispose d'une maîtrise B1 du CERL (ou plus) de la langue-cible .

Les exemples montrent que tous les verbes signifient 'se plaindre'. Au point de vue des connotations diastratique et dianormative, l'italien *lamentarsi* est le plus neutre, les textes français, espagnol et portugais indiquent que les verbes appartiennent plutôt aux registres oraux (non-conventionnels, mais pourtant de plus en plus présents dans la langue écrite). L'exemple élucide les limites d'une liste plurilingue de parasynonymes sérialisée qui ne peut résoudre le problème évoqué par Schaefer.

Le processus décrit est assez trivial étant donné que nous l'effectuons dans nos pratiques communicatives incessamment. On l'observe dans chaque cas où une signification généralement connue (lexicalisée) est transformée en signification occurrence. Comme on ne veut pas pratiquer une sorte de sismographie de signification, la fabrication de séries plurilingues demande l'identification d'un même noyau sémantique comprenant tous les lemmes d'une série donnée.

Retenons : Seulement si les lemmes ont un même noyau sémantique, ils entrent dans une même série. Pour la composition des séries, cela implique qu'on suit le principe « un noyau sémantique – une série, deux noyaux – deux séries, etc. ». Ceci explique pourquoi le nombre des séries dépasse celui des lemmes, mais aussi qu'une même forme (signifiant) peut apparaître dans deux ou trois séries.

(5) « La désambiguïsation de forme et de contenu par des synonymes non-romans ». Étant donné que ni la structure du VFPR ni de l'appli ne connaissent des définitions, et que les apprenants – hormis les locuteurs natifs romanophones – n'ont pas encore durablement intégré le répertoire lexical central d'une langue romane dans leur lexique mental plurilingue, les langues dites de désambiguïsation servent à clarifier le noyau sémantique de la série. Tandis que les blancs des langues-cible exigent l'insertion d'une forme particulière, les champs de désambiguïsation véhiculent souvent plusieurs mots (par ex. pour es. *derribar* : angl. *pull down, beat*, all. *niederschlagen, niederhauen, umreißen*). Les équivalents allemands ou anglais serviront donc non seulement à faciliter la mémorisation des lemmes en langues-cible, mais aussi à identifier un sens occurrence. Et n'oublions pas que la visualisation sérielle des lemmes romans soutient, elle aussi, la mémorisation dans les cas où l'utilisateur en connaît déjà un.

La psychologie de l'apprentissage prise en compte dans les applis du VFPR

Il a été mentionné que le VFPR et ses applis se situent dans la didactique de l'intercompréhension. Il s'agit d'une approche transversale qui ne vise pas à remplacer les didactiques traditionnelles (de l'anglais, du français, communicative, etc.) mais qui se propose de les enrichir d'aspects pluriels tels qu'ils se trouvent explicités dans le CARAP (Candelier et al. 2007 ; pour une version allégée : Siebeneicher-Brito & Meißner 2018). Parce que la méthode intercompréhensive se focalise très concrètement sur les structures linguistiques, leur analyse et l'apprentissage autonome, cette didactique se situe au cœur des différents secteurs d'une didactique intégrée ou du plurilinguisme. Toujours concernées sont l'activation des connaissances disponibles et pertinentes, la construction d'une grammaire d'hypothèses de la langue-cible et, en même temps, plurilingue à l'égard d'une systématité inter-langues. La rapidité stupéfiante avec laquelle la méthode mène à une compétence de lecture en une langue « inconnue mais virtuellement transparente » se trouve explicitée dans cette phrase du psychologue américain David Ausubel (1968, p. vi) :

If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor in influencing learning is what the learner already knows.

Le format de la sérialisation, intégré dans les applis, correspond tout à fait à l'importance des préconnaissances soulignée par le courant dominant de la psychologie de l'apprentissage. Ainsi l'arrangement de plusieurs lemmes équivalents, très souvent aussi congruents de forme côte à côte provoque l'établissement immédiat d'associations entre les mots désambiguïsants d'une série et leur correspondant en langue cible.

Les applis : « *Vocabulaire de base du VFPR* » et « *EuroComDidact ToGo* »

L'intro fonctionnel du *Vocabulaire de base* explique très bien comment fonctionne l'appli :



Figure 5 : Ouverture du Vocabulaire de Base (VBPR)

Français

L'appli promeut l'acquisition d'une compétence de lecture plurilingue romane.
Choisis, s'il te plaît, un des programmes suivants :

- 1.) Programme concernant une langue-cible particulière.
 - a. Clique sur le drapeau symbolisant la langue que tu veux apprendre.
 - b. Commence l'exercice. Remplis les cases laissées en blanc. Si ta proposition n'est pas correcte, le programme montrera la bonne solution.
 - c. Le programme retiendra les tâches que tu as déjà résolues. Dans une session suivante, tu recevras d'autres tâches.
- Ou :
- 2.) Programme concernant le plurilinguisme roman.
Ce programme fonctionne de la même manière. À la seule différence que tu n'as pas à définir une langue-cible particulière. Comme le programme veut t'aider à comprendre plus d'une seule langue romane, il génère des cases vides en différentes langues. Remplis les lacunes.

Choisis ta langue de destination et clique sur la flèche





   

Figure 6 : Intro du VBPR

Une fois qu'il a fixé la langue du guidage (*choose your language*), l'utilisateur suivra les points indiqués dans l'intro et remplira les lacunes.

deutsch	english	français	español	português	italiano
Einbildungskraft	imagination		imaginación	imaginação	immaginazione
Unruhe	trouble		inquietud	inquietação	inquietudine
Kleid, Kleidung	dress, robe		ropa	roupa	abito
Strom	current		corriente	corrente	corrente
(auf, um)hängen	hang		colgar	pendurar	appendere

Figure 7 : La langue-cible est le français

Dans les cas où il a choisi « plurilinguisme roman », l'appli présente les blancs de manière aléatoire. Le destinataire germanophone par exemple trouvera des bases de transfert dans l'adjectif *imaginär* et le substantif *Robe*, également dans *Garderobe*, emprunt du français, mais les congruences de forme anglaises (*to imagine, wardrobe*) ne lui sont très certainement pas inconnues.

Rappelons que la visualisation des équivalents facilitera la mémorisation ; surtout dans les cas où les signifiants correspondent aux régularités morphologiques pan-romanes.

deutsch	english	français	español	português	italiano	
um zu, sodass	for, to	afin que	(a ~ de)	(a ~ de)	fim	affinchè
bitter	bitter	amer	amargo	amargo	amaro	amaro
geboren	born		nato	nato	nato	nato
Erklärung	declaration	déclaration	declaración	de	declaração	dichiarazione
Pflege,	maintaining, holdig	maintien	mantenimiento	ma		

Figure 8 : Fiche pour faciliter l'acquisition d'une compétence de lecture pan-romane

À la différence des formats destinés à l'apprentissage de lecture d'une seule langue-cible, le choix de l'exercice « plurilingue » demande le remplissage de blancs alternant entre langues.

L'appli « mémorise » les activités de l'apprenant, ce qui assure la continuité du guidage personnalisé au-delà des séances individuelles. En même temps, les « erreurs » sont enregistrées et les « tâches non résolues » réapparaîtront au cours des séances qui suivent.

Grosso modo, *EuroComDidact ToGo* suit la même philosophie et présente la même structure que le *Vocabulaire de Base*.



Figure 9 : Intro d'EuroComDidact ToGo

Mais comme on vient de lire, son répertoire, fort de 9.555 séries, est nettement plus important. Le didacticiel permet une visualisation variable qui, selon les intérêts de l'apprenant, inclut ou exclut l'affichage des langues-cible (voir Figures 11 et 12).

Tandis que le VBPR ne fonctionne que sous Windows, EuroComDidact ToGo soutient Windows, Android et iOS.



Figure 10 : Choix de la langue du guidage



Figure 11 : Choix de la langue de départ et de la langue-cible



Figure 12 : exclusion des langues « sans intérêt »



Figure 11 : La langue-cible est le français. La bonne recherche du lemme-cible est facilitée par l'indication de la lettre initiale.

La Liste des Formes verbales opaques du plurilinguisme roman (FVOPR)

Que les formes lemmatiques occultent celles des *tokens* relevés fait obstacle à l'intercompréhension dans le domaine des formes verbales conjuguées (type : *aller – vais, va, vont, irons* es. *querer – quiso, quisiera*, etc.). Pour atténuer ce désavantage, une *liste des verbes opaques du plurilinguisme roman* (VoPR) a été établie et publiée sur le site <https://www.eurocomdidact.eu/de/verben>. Cette liste suit également le principe du filtrage des irrégularités morphologiques. Y sont incorporés : les radicaux et les écarts, à savoir les formes dites irrégulières (grisées). L'effet synergique du filtrage est considérable.

En mettant en relief les divergences de formes, la liste VFPR-VERBES donne aux apprenants les radicaux à partir desquels ils arrivent facilement à identifier n'importe quelle forme conjuguée.

Le potentiel pédagogique du VFPR et d'EuroComDidact

Pour mesurer le potentiel pédagogique des applis, il ne suffit pas de les regarder isolément, car leur usage s'intègre dans un arrangement d'apprentissage plus large. À force de travailler avec l'appli, c'est-à-dire en comparant les formes et leurs significations, les élèves développent des stratégies d'apprentissage de toute sorte. En ayant recours aux différents outils externes : grammaires de consultation, dictionnaires en papier ou électroniques, concordanciers, les possibilités de l'internet en général, etc., ils élaborent leur capacité à désambigüiser les structures floues de leur interlangues. C'est aussi ici que la didactique de l'intercompréhension « élargie » rencontre l'autonomisation de l'apprenant ! Somme toute, les données recueillies dans les cours appliquant la méthode intercompréhensive montrent presque sans exception que la voie intercompréhensive va de pair avec la croissance de la « compétence d'apprendre les langues » (*Sprachlernkompetenz*).

Mais pour décrire le potentiel pédagogique du VFPR et des applis, il ne suffit pas de se contenter d'une perspective de l'apprentissage. Le VFPR offre d'importants avantages à l'enseignement quant à l'input et sa présentation. Il est évident que cette optimisation se traduit en un meilleur guidage effectué par les tâches et les exercices. Voilà quelques avantages concernant un ciblage efficace¹¹ dû à l'exploitation des avantages de la banque de données et/ou de ses applis :

- la compilation de l'input lexical fondamental sur le lexique central des quatre plus grandes langues romanes « dont l'étudiant a vraiment besoin », basée sur la lexicographie computationnelle d'où la haute fiabilité du VFPR :

Why not simply rely on the vocabulary lists in a course textbook? The short answer is that also a typical textbook provides some thematically-related vocabulary in each chapter (food, illnesses, transportation, clothing, etc.), there is never any indication of which of these words the student is likely to encounter in actual conversation or texts. In fact, sometimes the words are so infrequent in actual texts that the student may never encounter them in the "real world". (Davies 2006: 1)

- des comparaisons lexicales bilingues sélectionnées ou sérielles selon les pré-savoirs supposés des apprenants et les langues-cible
- des contrastes bilingues au niveau d'une suite donnée de caractères choisis en fonction de l'interphonologie, par ex. pour les phonèmes $[-l\text{h}]/\text{z}/ \sim [j]/\text{x}/$ - dans la combinaison espagnol-portugaise ou français-espagnol-italienne par exemple
- optimisation de l'input par l'alimentation des exercices adapté aux besoins des apprenants

¹¹ Il est prévu d'exemplifier les nouveaux formats d'exercice dans une publication dédiée à la thématique.

- saisie d'un input captant les noyaux au niveau des congruences de forme et les équivalences de contenu qui sont, toutes deux, à l'origine du réseautage des langues européennes
- construction de nouveaux formats d'exercice en didactiques de l'intercompréhension
- construction de nouveaux formats d'exercice à partir d'une liste embrassant les interligalexes verbaux des langues romanes
- nouveaux formats d'exercice alimentés par les interligalexes
- nouveaux format d'exercice alimentés par des combinaison de lettres.

Sur le site www.eurocomdidact.eu, on trouvera des conseils pour exploiter au mieux le potentiel du VFPR et de ses applis.

Références

- Ausubel, D. (1968) : *Educational Psychology. A Cognitive View*. New York: Harcourt Brace & World.
- Battaglia, S. (dir.) (1961-1998) : *Grande dizionario della lingua italiana*, Firenze: Unione tipografico - Editrice Torinese.
- Bieritz, W.-D. (1974) : *Semantischer Transfer auf verwandte Fremdsprachen. Die Bedeutungerschließung der Inhaltswörter des spanischen Grundwortschatzes durch Schüler und Studenten mit lateinischen und französischen Vorkenntnissen*, Bochum: Seminar für Sprachlehrforschung.
- Candelier, M., Camilleri-Grima, A., Castellotti, V., de Pietro, J.-F., Lörincz, I., Meissner, F.-J., Schröder-Sura, A., Noguerol, A. & Molinié, M. (2007) : *À travers les Langues et les Cultures/Across Languages and Cultures. CARAP Cadre de Référence pour les Approches Plurielles des Langues et des Cultures*. Graz: CELV/Strasbourg: Conseil de l'Europe.
- Corominas, J. (1980) : *Breve diccionario etimológico de la lengua castellana*. Tercera edición muy revisada y mejorada. Madrid : Gredos.
- Dauzat, A., Dubois, J. & Mitterrand, H. (1971) : *Nouveau dictionnaire étymologique et historique*. Paris : Larousse.
- Davies, M. & Preto-Bay, A. M. R. (2007) : *A Frequency Dictionary of Portuguese*. London/New York: Routledge.
- Davies, M. (2006) : *A Frequency Dictionary of Spanish. Core vocabulary for learners*. London/New York: Routledge.
- De Florio-Hansen, Inez (1994) : *Vom Reden über Wörter. Vokabelerklärungen im Italienischunterricht mit Erwachsenen*. Tübingen: Narr.
- De Mauro, T., Mancini, F., Vedovelli, M. & Voghera M. (1993) : *Lessico di frequenza dell'italiano parlato. Prefazione di Pierluigi Ridolfi*. s.l.: ETASLIBRI.
- Free Dictionary Farlex. (<http://www.thefreedictionary.com>).
- Jamet, M.-C. (2007) : *À l'écoute du français. La compréhension orale dans le cadre de l'intercompréhension des langues romanes*. Tübingen: Narr.
- Klein, H. G. & Stegmann, T. D. (2000) : *EuroComRom. Die sieben Siebe. Romanische Sprachen sofort lesen können*. Aachen : Shaker.

- Klein, H. G. & Reissner, C. (2003): *EuroComRom. Die historischen Grundlagen der romanischen Interkomprehension*. Aachen: Shaker.
- KLETT = Nickolaus, G. (1967) : *Grund- und Aufbauwortschatz Französisch*. Stuttgart: Klett. (Des listes parallèles ont été publiées pour un certain nombre de langues dont l'espagnol [Heupel, 1966], l'italien [Giovannelli, 1971] et le portugais [Amorim-Braun, 1996]).
- Lonsdale, D. & Le Bras, Y. (2009) : *A Frequency Dictionary of French. Core Vocabulary for Learners*. New York/London : Routledge.
- Lutjeharms, M. (2010) : Der Leseprozess in Mutter- und Fremdsprache. In: Lutjeharms, M. & Schmidt, C. (Hrsg.) : *Lesekompetenz in Erst-, Zweit- und Fremdsprache*. Tübingen: Narr, 11-26.
- Meißner, F.-J. (2004) : Introduction à la didactique de l'eurocompréhension. In : F.-J. Meißner, C. Meissner, H. G. Klein. & T. D. Stegmann : *EuroComRom. Lire les langues romanes dès le départ*. Avec CD-ROM. Aachen : Shaker.
- Meißner, F.-J. (2016) : *Der Kernwortschatz der romanischen Mehrsprachigkeit. Didaktische, lexikologische, lexikographische Überlegungen zu Erstellung, Präsentation, Anwendungen einer elektronischen Mehrsprachenwortliste und von Lernapps zur romanischen Mehrsprachigkeit. GiF:on 7*
http://geb.uni-giessen.de/geb/volltexte/2016/11950/pdf/GiFon_7.pdf.
- Meißner, F.-J. (2017) : Interlanguage. In C. Surkamp (Hrsg.) : *Metzler Lexikon Fremdsprachendidaktik. Ansätze – Methoden – Grundbegriffe*. Stuttgart : Metzler (2. Aufl.)
- Meißner, F.-J. (2018) : *Die Vermessung des Kernwortschatzes der romanischen Mehrsprachigkeit. Eine didaktische Analyse zur interlingualen Transparenz- und Frequenzforschung*. GiF :on 11
<http://nbn-resolving.de/urn:nbn:de:hebis:26-opus-135898>.
- Müller-Lancé, J. (2003) : *Der Wortschatz romanischer Sprachen im Tertiärsprachenerwerb. Lernerstrategien am Beispiel des Spanischen, Italienischen und Katalanischen*. Tübingen : Stauffenberg.
- Schaeder, B. (1990) : Das Problem der Äquivalenz aus der Sicht der Internationalismenforschung. In : P. Braun, B. Schaeder & J. Volmert (Hrsg.) : *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie*, Tübingen : Niemeyer, 63-73.
- Siebeneicher-Brito, K. & Meißner, F.-J. (2019): *Os descritores dos 'soft skills' no Quadro de Referência para a Abordagem Plural de Línguas e Culturas (QuRAPLeC/CARAP). Um handout prático para o ensino baseado em competências focalizando a promoção da aprendizagem de línguas, intercultural e plurilíngue. GiF:on 12.*
<http://geb.uni-giessen.de/geb/volltexte/2018/13773/>
- Zimmer, R. (1990) : *Äquivalenzen zwischen Deutsch und Französisch. Theorie - Korpus - Indizes. Ein Kontextwörterbuch*. Tübingen : Niemeyer.