Short presentation of the
Core Vocabulary of Romance Plurilingualism and *EuroComDidact ToGo*

The following description is addressed to linguists, language teachers, students of Romance languages, English or other languages and advanced foreign language learners[1].

<div align="center">*</div>

> "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*."*
> (David Ausubel: *Educational Psychology. A Cognitive View*. New York, 1968: vi)

The *Core Vocabulary of Romance Plurilingualism* (CVRP) is a lexical database which feeds the *EuroComDidact ToGo* apps' inventories. Its 28,000 form divergent target lemmas correlate the most frequent words ('intersynonyms') of French, Italian, Portuguese and Spanish in 9,550 plurilingual series. For reasons of disambiguation, the CVRP gives, furthermore, the Romance lemmas' equivalents in English and German (and Latin or other ancient idioms wherefrom the CVRP lemmas derive [2]). The electronic architecture allows immediate regrouping and displaying of contents, selected and composed in accordance to didactically relevant criteria, such as the learner's mother tongue or a given bridging language combined with a special target language(s), the degree of interlingual form (signifier) transparency [3] and semantic adequacy, et cetera. Thanks to the variable content arrangements, *EuroComDidact ToGo* can respond to special didactic functions or learners' needs.

Selection of the vocabulary: The CVRP list has mainly been compiled from electronic frequency glossaries whose repertoires are composed from numerous text corpora of millions and millions of word-forms (tokens). These core vocabularies – i.e. the *Routledge Frequency Dictionaries* of French, Portuguese and Spanish (for ex. Davies 2006) – comprise lemmas of a frequency range below 5,000[4]; or 7,000 with reference to the *Lessico di frequenza dell'Italiano parlato* (De Mauro et al. 1993)[5]. The compilations ensure highest reliability. In a second step,

---

[1] A closer description concerning the selection is provided by Meißner (2016).

[2] The Latin list has not been completed yet.

[3] The database assigns its lemmas to 5 'interlingual transparency cohorts': T4 (*abattre-abbattere-abater-abatir*), T3 (*abri-riparo-abrigo-abrigo*), T2 (*affaires-affari-negócio-negocio*), T1 (*ajouter-aggiungere-acrescentar-añadir*), T0 (*balai-ramazza-vassoura-escoba*).

[4] Frequency marks must be identified for each target language. It occurs that the equivalent of a French word with a frequency range of <2,000 corresponds to a range of >2,000 or even >5,000 in language Italian or Portuguese.

[5] Of course, lexical counts show that different words can have same frequency ranks. Thus, in the *Français Élémentaire,* 22 lemmas share the frequency position of 30 (Gougenheim et al. 1956: 78). A third criterion is a word's spread in the dia-register of a language. Is a lemma part of the spoken or the written language? Is it part of both? What register is concerned: standard, colloquial, proximity or formal speech, etc.? The diatopic argument is particularly relevant to polycentric languages of global spread like English or Spanish. To get registered as a core vocabulary lemma, the word must be found in numerous texts of different sorts, topics, situations, and in various relevant diastratic and diatopic registers as well. A further criterion is semantic

the CVRP inventory has been completed by the traditional basic vocabulary lists (*Grundwortschätze*) of the Klett publishing company (for ex. Nickolaus 1967). These lists incorporate entries of pedagogic relevance (*classroom, blackboard, chalk,* etc.). Mostly, such words and collocations escape from computation below a frequency rank of 5,000 or 7,000. That is why both, the CVRP-list and the apps, link the validity of corpus linguistic based evidence about lexical frequency, frequency range as well as diatopic, diastratic and diaphasic spread with intercultural and educational communication relevance-criteria provided by pedagogic expertise. It goes without saying that the apps benefit from that.

Construction and extension of the series: More than bilingual lexicography, plurilingual dictionary making encounters thorny questions affecting synonymy, homonymy, polymorphy of correlating lemmas. As there is no unique, clear and always applicable rule for assigning corresponding word forms and precise significations between various languages, the CVRP tries to compensate such ambiguous 'intersynonymies' by presenting the entries concerned (duplicates) in various series and serial combinations. The examples about the semantic nucleuses of 'deep, below, under, beneath' and 'bird' (both sorted to French) provide an explanation:

ID	Fr	It	Pg	Sp	En
192	AU-DESSOUS	SOTTO	DEBAIXO	DEBAJO	below
260	BAS	BASSO	BAIXO	BAJO	deep, low
2036	SOUS	SOTTO	SOB	DEBAJO DE	under
3590	BAS (EN ~)	GIÙ	ABAIXO	ABAJO	beneath, under
7902	sous	SOTTO	EMBAIXO	DEBAJO	under, beneath

1: Serial variation concerning the semantic nucleus 'below, under, down, beneath'

2468	OISEAU	UCCELLINO	AVE	ave	bird	Vogel
1485	OISEAU	UCCELLO	PÁSSARO	PÁJARO	bird	Vogel

2: Serial variation concerning the equivalents of 'bird'

The duplicates explain why the number of interlingual series is quite higher than the number of (non repeated) lemmas or entries.

Lexical disambiguation in plurilingual word lists – a structure for fostering pluri-language awareness:[6]: For foreign language learning purposes, traditional frequency word lists are mostly bilingual and related to only one target language. The fundamental structures of their articles resemble mathematical equations (word of language A=word of language B). Entry by entry, signification by signification, the lists try to deliver best possible and precise semantic correspondences. Plurilingual vocabulary arrangement must proceed differently. Although intersynonymy does not exist in a strict sens, as Schaeder put it (1990), semantic adequacy between the target lemmas is delivered by a (common) semantic nucleus shared by all the members of a plurilingual series. To provide an example: Pt. *trabalho/labor*, it.

---

coverage (*dog* vs. *poodle*; *to speak* vs. *to gossip, to whisper*). Generally, the CVRP follows the criteria of the pedagogical frequency lexicography (Davies 2006: 6).

[6] To learn more about intercomprehensively based learning and teaching, see Meißner (2016b).

lavoro/travaglio, fr. *travail/labeur* have the same signification like en. *work*, ge. *Arbeit*, swed. *arbete...* Pt. *labor/trabalho,* en. *labour/am. labor* mean 'hard work', 'travail *pénible'*, *effort*, sp. *esfuerzo…* other parasynonyms are fr. *peine*-en. *paine*-sp. *pena…* Obviously, the common nucleus is situated in the relationship between en. '*work-effort-labor-paine* and their semantic neighbours'. Consequently, plurilingual lexicography must be content with providing users with a semantic nucleus. Accordingly, it is up to the users themselves to find (construct) the precise occurent meaning in a given target language text. We do not leave this aspect without highlighting that the described plurilingual lexicographic micro-structure invites users to initiate the same mental processing as that they use for unmasking unknown words in foreign language texts.

Interlingual transparency: Unmasking cognates in new target languages needs to be considered a fundamental key qualification in intercomprehension and intercomprehension related didactics. As the spontaneous identification of new vocables must be considered as a *conditio sine qua non* for the comprehension of a target language text, the CVRP database subdivides its entire inventory into five transparency cohorts encompassing series of full transparency (full serial form congruency) to those of full opacity. The principal distribution criteria, i.e. the interlingual transparency benchmarking, is given by the relative sequence of identical letters in regard to the length of interlingually correspondent words. Beside other selecting effects, the interlingual transparency index (ITP) is worth while to construct multiple formats of lexical learning materials.

The following list shows the transparency reach between lemmas of a same series. Identical letters symbolize transparency (form convergencies) between the words of different target languages; divergent letters indicate intransparency or opacity caused by a different concrete language, hyphens signalize target languages different from the serially regular language. Full serial transparency is delivered when the four lemmas concerned have a same common transfer base; full serial opacity is given when there is no transfer base. In accordance with these criteria, the distribution of transparency/opacity series allows to establish five cohorts going for T(ranparency)4 to T0.

As far as counting transparency or opacity categories is intended, serial counts and counts without dublicates should be distinguished. Furthermore, the indication "three transfer bases" must be differenciated into opacity and transparency counts. Thus, the opacity occurrency for French is at serial counts is 725, the transparency value is (1533-725=808; for more: Meißner 2021c))

FR	IT	PT	SP	interlingual transparency spread	number of transfer bases	opacity spread
a	a	a	a	4	1	0
-	a	a	a	3	2+x	1
a	-	a	a	3	2+x	1
a	a	-	a	3	2+x	1
a	a	a	-	3	2+x	1
a	a	b	b	2	2x2	0
a	b	a	b	2	2x2	0
a	b	b	a	2	2x2	0
b	a	a	b	2	2x2	0
b	b	a	a	2	2x2	0
a	a	b	c	1	3	3
a	b	c	a	1	3	3
a	c	b	a	1	3	3
a	b	b	c	1	3	3
a	c	c	b	1	3	3
a	b	a	c	1	3	3
a	c	a	b	1	3	3
b	b	a	c	1	3	3
b	a	c	b	1	3	3
b	b	c	a	1	3	3
a	c	b	b	1	3	3
c	c	a	b	1	3	3
c	a	b	c	1	3	3
a	b	c	c	1	3	3
c	a	a	b	1	3	3
c	b	b	a	1	3	3
a	b	c	d	0	4	4

*1: Sample of serial transparency/opacity degrees*

Interligalexes: Another innovation beside the ITP concerns the *interligalex* concept (Meißner 2019) and the practical realization in the form of (Romance) interligalex lists. Interligalexes are identical with what the intercomprehension approach usually calls the 'interlingual bases of transfer'. With regard to Rey-Deboves notion *ligalex* (2004)*,* we define interligalexes those lexical form elements that can be found within words of different languages. The CVRP-interligalex column lists the graphematical nuclei of the whole CVRP-inventory. In matters of learning efficiency, the intercomprehension approach tries to capitalize the learners' previous relevant knowledge in order to facilitate the acquisition of (further) languages. Whereas the vocabulary list obtains its efficiency effects by comparing entire word forms within series (Meißner 2021a) , the interligalex list boosts such effects by remarkably reducing the number

of transfer bases and simultaneously increasing the number of words identified. Counts within the CVRP's inventory show that about 8,000 interligalexes (in their expanded form) cover more than 30,000 lemmas. The interlingual identification rate of less expanded forms would be significantly higher.

The limits of plurilingual and serial cross-language presentation: Vocabulary presentation for productive purposes requires more and other explanations than for receptive ones. As the CVRP-list focuses exclusively on fostering plurilingual reading comprehension, it does not give enough information about how a lemma fits in with a co-text of a given language (valence, coverage, registers, collocations: *cheveux rares/schütteres Haar/thinning* or *thin hair*; *Wiedersehn macht Freude/ça s'appelle Reviens/I want to get it back*, interdictions, etc.). Instead of trying to create lexical associations within only one target language system, it intends to facilitate cross-language word recognition caused by phonetic, graphic or semantic resemblances.

Learning with the apps: It goes without saying that the efficiency of learning depends on the learners' own mental activities. Simply regarding the apps, this means that learners focus on comparing forms or signifiers, significations and functions of new linguistic schemes with those they already are familiar with. Of course, zapping or not leaving enough time to our brain for fixing in mind the properties of the equivalents severely hampers language acquisition. To avoid these risks, foreign language acquisition pedagogy recommends activities like note taking, comparing the words' definitions and uses in various dictionaries, and other awareness raising strategies (Meißner 2021b).

The structure of the exercises: As already said, the arrangement of six lemmas of six languages side by side invites learners to compare word forms and significations. Especially the existence of duplicates triggers cross-language association building not only within a series, but between different series as well. As comparing words and analyzing their function in different languages is a very powerful strategy of vocabulary memorizing, learners will pay attention not only to the target word but to its equivalents in other languages too. Needless to say that this attitude effectively fosters foreign language learning competence.

2: A serial arrangement of lexical equivalents

To support vocabulary association beyond intraserial comparisons, the CVRP architecture has additionally implemented a column 'notes'. The field will allow users to find information about

intersynonymies, interhomonymies, 'false friends', etc. The rubric has not yet been integrated into the apps' structure up to now.

An algorithm effecting the individualization of learning: During learning sessions, the exercises, supplied by the 9,550 CVRP series, are being continuously rearranged. Thus, learners will encounter the cloze exercises in randomized sequences. As the app automatically memorizes the user's device IP, consecutive sessions are possible which exclude tasks that already had been solved. At the same time, this individualization takes account of 'erroneous' insertions (not expected by the app's programming) or of missing answers. Users will meet these concrete tasks again later on.

Teaching with the CVRP-list and its apps: The inventory can be used to shape "lexical input relevance" for teaching (and learning) the four target languages or only one of them. In other words, it allows selecting a core lexical input in accordance to a syllabus or to a learner's individual needs (previous knowledge, learning objectives). Consequently, in addition to learners, the addressees of the CVRP are designers of all kinds of didactic materials, teachers and learners.

The following table shows how – by grouping the apps' input – EuroComDidact ToGo can respond to various educational questions responding to individual linguistic needs. To provide some examples, we give five requests: (1) Give me the non-transparent Italian words of of Italian (target language) which German (Turkish, etc.) students who have a B1[7] command of French and English in addition to their mother tongue cannot decode (*barzeletta*-fr. *blague,* sp. *chiste*-pt. *piada/brincadeira*-en. *joke*-ge*. Witz*). The response is delivered by App no. 3. (2) The app 'Visualizing Portuguese cognates in contrast with Spanish' (not listed in the table) provides 1933 form-congruent series; (3) the app 'form convergences for French and Italian but divergent from Portuguese and Spanish' (not listed in the table) spits out 295 series. (4) 'Show form divergences between Portuguese and Spanish gives 2892 series. (5) In total, 5642 of the 9553 core vocabulary series show an overall Romance form convergence. It is clear, that word lists like these allow selecting the lexical input in accordance to the learners' relevant previous knowledge in terms of forms and significations to their learning objectives. In a teaching perspective (or in that of self-guided learning), apps like these are tools for composing exercises or tests. In every case, they stimulate language learning awareness. (6) The *Seven Sieves* (McCann et al. 1999) give a complete description of Romance phonological and orthographic regularities. Looking for the string /bia/ in the Italian column yields the correspondences *bianco-blanc-branco, sabbia-sable, bibbia–bible…, biasimare-blâmer, nebbia*-sp. *niebla*; the Portuguese string /br/ in *branco* gives it. *bianco*-sp. *blanco, brando*-sp. *blando, nobre*-it. *nobile*, fr./sp. *noble, obrigar-obliger-obbligare-obligar*, et cetera*.* Such selectors identify the validity extent of interphonological rules. This kind of knowledge delivers benchmarks for constructing exercises, new types of exercises included, designed for training the identification competence of interlingual phonological correspondences between Romance languages.

---

[7] The performance level is related to the *Common European Framework of Languages* (Council of Europe, 2001).

The CVRP verb-lists: Reading a 'new' language, learners do not meet lemmas but tokens, that means all kinds of word forms. Especially in the field of verbs, students encounter important morphological irregularities which can severely hamper comprehension. The CVRP verb-lists highlight forms that deviate from the corresponding infinitives or from regular morphological (conjugated) patterns (https://eurocomdidact.eu). In short, there are two descriptive patterns: a more complete one showing all deviations of an infinitive concerned, and a shortened one that exclusively provides radical variation. Whereas in the list of 'verbal radicals only', the lemma *ÊTRE* mentions *suis, es, est, sommes, êtes, sont et étai-, éti-, fu-, fû-, été,* sp. *QUERER*: *quiero, quieres,* (*queremos*)*, quisé, quiso, quiera, quisiera, querido, queriendo, PLACER* gives *plazco,* (*places*)*, plazqua, placiendo, LEVANTAR* remains without any conjugated form because of its morphological regularity. A longer list (B) gives the forms of the infinitive, indicative present: 1^{rst}, 3rd and 4th person; imperfect; imperfect simple/passé simple; future; subjunctive present; s. imperfect; imperative past and present; participle or gerund. The fundamental criteria to select the descriptive categories follow pedagogical evidences rather than linguistic systematicity.

The common original inventory is based on what is given by the corresponding target language dictionaries and word lists.

The EuroComDidact ToGo-apps: The CVRP database allows to filter its contents in accordance to pedagogic goals or to the learning needs of special addressees. As aforementioned, the following list does not imply all possibilities of regrouping and contrasting the CVRP lexical material. Special requests can be addressed to team@eurocomdidact.eu.

CVRP Learning apps	*filter/addressees/pedagogic objectives*
1. Acquiring the entire **Romance Core Vocabulary** (freq. range <5,000) (RCV)	DISPLAY (FR, IT, PT, SP, EN, DE). No filter. Bringing plurilingual reading skills to more perfection (can be used in combination with the *Seven Sieves*) (McCann et al. 1999; Meißner et al. 2004).
2. … the entire **Romance Basic Vocabulary** (freq. range <2,000) (RBV)	DISPLAY (FR, IT, PT, SP, EN, DE). Filter "GW". Developing plurilingual reading skill in the four TL within the scope of the basic vocabulary.
3. … the total **Opaque Core-Vocabulary of Romance languages (ORCV)**	DISPLAY (FR, IT, PT, SP, EN, DE). Start filter (FR≠IT≠PT≠SP) plus (the opaque forms of the T3-cohorts [*-F+*-I+*-P+*-S]). For users interested in an overall Romance languages reading competence; especially for native speakers of a Romance language.
4. … the **Opaque Core-Vocabulary** of **Portuguese (OPCV) (or another target language)** regarding English and German and other Romance languages	DISPLAY (FR, IT, PT, SP, EN, DE). Start filter (PT≠FR≠IT≠SP) plus (opaque Portuguese lemmas of the T3-cohort), or set filter on a special contrast, (PT≠SP) for example. This item will only show all the Spanish words whose form diverge from their Portuguese equivalents, even If they are not part of the list of opaque forms (F≠I≠P≠S).

5.	... the **Italian Core Vocabulary** combined with French, Portuguese, Spanish, English and German **(ICV)**	DISPLAY (FR, IT, PT, SP, EN, DE). No filter. Sorting by Italian headwords. The interlingual contrast can also be specified with regard to specific TLs ("show me the Italian words which can be identified on the basis of their French… equivalents"). Learners primarily interested in developing reading competence of Italian.
6.	... the **French Core Vocabulary (FCV)** visualized with Italian, Portugues, Spanish, English and German	DISPLAY (FR, IT, PT, SP, EN, DE). Sorting by French headwords; no filter.Learners with operative knowledge in at least one Romance language and/or English wishing …to develop reading skill of French; … to develop interphonological rules with French and English.
7.	... the **Spanish Core Vocabulary (SCV)**…	same like 5, 6, 8
8.	... the **Core-Vocabulary of Portuguese (PCV)**…	Most learners of Portuguese master Spanish to a certain degree, this contrast is particularly promising.
9.	Visualizing **Romance cognates** in contrast with English and German	DISPLAY (PT, SP, EN, DE). Start filter (PT=FR=IT=SP) plus the transparent lemmas of the T3, T2 and T1 cohorts. The respective items foster general Romance interphonological competence with special regard to this or that targuet language.
10.	**Romance cognates with English (RCE)**	DISPLAY (PT, SP, EN, DE). Start filter (*E). Given that interlingual identification transfer creates bidirectional memorization effects between a bridging and the target languages, the RCE-app consolidates English competence as well.
11.	Using m**inimal lexical elements for constructing Romance thesauri**	Assignment of interligalexes. For obtaining highest lexical association scores
12.	**CVRP Etyma-List**	Provides the etyma (raw list in progress)
13.	**False friends lists** for special language contrasts	(projected)
16.	**Lists of verb variation**	Access over eurocomdidact.eu.

F.-J. Meißner, Giessen, 2021 05 04

References:

Council of Europe (ed.) (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Strasbourg: Council of Europe.

Davies, M. (2006). *Frequency Dictionary of Spanish.* New York/London: Routledge.

De Mauro, T./Mancini, F./Vedovelli, M./Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato.* Prefazione di Pierluigi Ridolfi. s.l.: ETASLIBRI.

Gougenheim, G./Michéa, R./Rivenc, P./Sauvageot, A. (1956). *L'élaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.

McCann, W./Klein, H.G./Stegmann, T. (1999). *EuroComRom - The Seven Sieves: How to read all the Romance languages right away.* Aachen: Shaker.

Meißner, F.-J. (2004). Introduction à la didactique de l'eurocompréhension. In: Meißner, F.-J./Meißner, C./Klein, H. G./Stegmann, T. D.: *EuroComRom - les sept tamis. Lire les langues romanes dès le départ.* Aix-la-Chapelle: Shaker, 7-140.

Meißner, F.-J. (2011). Teaching and learning intercomprehension: a way to learner autonomy. In: De Florio-Hansen, Ines (ed.), *Toward Multilingualism and the Inclusion of Cultural Diversity*. Kassel: Kassel University Press, 37-58.

Meißner, F.-J. (2016a). *Der Kernwortschatz der romanischen Mehrsprachigkeit (KRM). Didaktische, lexikologische, lexikographische Überlegungen zu Erstellung, Präsentation, Anwendungen einer elektronischen Wortliste und von Lernapps zur romanischen Mehrsprachigkeit.* Giessener Fremdsprachendidaktik: online 7. Available under: http://nbn-resolving.de/urn:nbn:de:hebis:26-opus-1119506.

Meißner, F.-J. (2016b): The Intercomprehension Approach explained in 32 points. Available under: http://eurocomdidact.eu.

Meißner, F.-J. (2019): The interligalexes of the Core Vocabulary of Romance Plurilingualism (CVRP) and their potential effects on plurilingual learning economy. In: *Rivista di Psicolinguistica Applicata/Journal of Applied Linguistics* 19(2), 31-46.

Meißner, F.-J. (2021a): *The Vocabulary of Romance Plurilingualism. Vol. 1: French and Italian word list. Vol. 2: Portuguese and Spanish word list*. (GiF:on 18.1 & 18.2): http://geb.uni-giessen.de/geb/volltexte/2021/16034/; http://geb.uni-giessen.de/geb/volltexte/2021/16035/.

Meißner, F.-J. (2021b): Das didaktische Potential mehrsprachig-wortserieller Einsatzübungen (EuroComDidact ToGo). In: Grünewald, A./Tassinari, G./Nock-Ziegler, S./Wieland, K. (Hrsg.): *Fremdsprachendidaktik als Wissenschaft und Ausbildungsdisziplin*. Frankfurt: Peter Lang (im Druck).

Meißner, F.-J. (2021c): The Interlingual Transfer Potential of the *Core Vocabulary of Romance Plurilingualism (CVRP)* – statistically. In: *Zeitschrift für interkulturellen Fremdsprachenunterricht.* (submitted).

Nickolaus, G. (1963). *Grund und Aufbauwortschatz Französisch*. Stuttgart: Klett (bis 1970, 14 editions).

Rey-Debove, J. (2004) : De l'étymologie à l'analyse méthodique des mots : vers une nouvelle pédagogie. In: *Le Robert Brio. Analyse des mots et régularités du lexique*. Paris: Le Robert, VII-XI.

Schaeder, B. (1990): Das Problem der Äquivalenz aus der Sicht der Internationalismenforschung. In: Braun, P./Schaeder, B./Volmert, J. (Hgg.) (1990): *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie.* Tübingen: Niemeyer, 63-73.

10/05/2021