5th International Conference on Corpus Linguistics (CILC2013)

# Not to Teach but Give Insights: Corpus-Based Approach in Portuguese-English and Portuguese-Russian Cross-linguistic Error Correction

Nataliya Godinho Soares Vieira*

*FCSH-NOVA University of Lisbon, Avenida de Berna, 26, Lisbon, 1069-061, Portugal*

## Abstract

Foreign language learners make mechanical errors caused by cross-linguistic influence. It comes from trying to extract information from a native language (L1) and apply it in a second or foreign language (L2/FL). There are not many didactic materials proposing a set of techniques that train the skills to keep L1 and L2/FL apart and help students to reduce cross-linguistic errors. In alternative to a traditional rule-based teaching, a corpus-based approach provides a framework for exploring online corpora to give insights into authentic language use. The paper describes the types of errors caused by cross-linguistic Portuguese-English and Portuguese-Russian influence, and illustrates how some monolingual and bilingual online corpora may be rich sources for raising learner's awareness to face such linguistic obstacles.

## 1. Introduction

Acquiring a foreign language L1speakers unconsciously or consciously try to extract information from their native language and apply it in L2/FL. It is a natural process. That is why the matters related to crosslinguistic influence (Sharwood Smith and Kellerman, 1986) or language transfer (Odlin, 1989) in a foreign language acquisition are always adequate for scientific investigations.

---

\* Corresponding author. Tel.: +351968373270.
 *E-mail address:* nataligsv@fcsh.unl.pt

Crosslinguistic influence "…resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin, 1989). Crosslinguistic influence takes constant place on the principal levels of the language system. They are lexis and grammar.

At the level of vocabulary language transfer depends on some familiar features between L1 and L2/FL. L1 speakers may mix "true friends" with "false friends" of L2/FL. For instance, in spite of the fact that English is the West Germanic language, and Portuguese represents the family of the Romance languages, these both languages are close because of Latin borrowings that may provoke language transfer.

On the level of grammar foreign language learners may apply their previous linguistic knowledge translating from L1 to L2/FL some structures literally (word for word). There is a dangerous of assuming this fact as a rule and making common crosslinguistic errors.

It is also possible to identify some peculiarities of crosslinguistic influence in L2/FL pronunciation. Foreign language learners naturally modify the production of some sounds of a studied language (L2/FL) according to L1 experience. It can be demonstrated in the conditions of acquiring a foreign language from not a similar language family as, for example, acquiring Russian (FL), a Slavic Language, by native speakers of Portuguese (L1) on the initial stages.

There are not many textbooks or other didactic materials aimed at solving the problems of crosslinguistic error correction at different levels of language proficiency. The availability of a significant number of electronic corpora may help to bridge the gap between the lack of resources and to avoid the problem of using out-of-date, artificial foreign language examples: "Corpus-informed teaching materials provide students with examples of real language use, helping learners to know how to use language that is appropriate in different contexts" (Reppen, 2010: 20); "Corpus-informed materials motivate because teachers and learners can be sure that the language they are practicing is modern, used in everyday situations, targeted to situations they are likely to find themselves in, and corresponds to what they will hear and see in real conversations, movies, radio and TV shows, newspapers, books, Internet texts, and magazines" (McCarthy, 2004: 17).

Production of corpus-based pedagogic materials is based on the creativity, because electronic corpora contain just a set of concordances and are not able to explain the rules or regularities of their usage in different contexts. However, applying corpora in pedagogical purposes, teachers may open not only new ways of crosslinguistic error correction, but have a good possibility to provide language teaching-learning with anytime available and quality resources – electronic corpora.

## 2. Methodological recommendations

A corpus-based approach in teaching foreign languages has its roots in data-driven learning or DDL (Johns, 1991; 2000). DDL is so called "discovery learning" where "the task of a learner is to "discover" the foreign language, and the task of the language teacher is to provide a context in which  the learner can develop strategies for discovery  – strategies through which he/she can "learn how to learn" (Johns, 1991: 1).

The basic problem is that not every teaching context allows the implementation of DDL, because DDL is more appropriate for advanced learners who are ready to be "detectives" discovering facts about the language. "Discovery Learning … is most suitable for very advanced learners who are filling in gaps in their knowledge rather than laying down the foundations" (Hunston, 2002). For this reason, in the context of a corpus-based approach it would be appropriate to add a concept of "discovery teaching" that provides observation of teaching contexts to find the ways for applying of discovery learning or to create some corpus-based didactic materials.  The principles of discovery teaching may look like an action research cycle (Wallace, 1997). The process of creation of corpus-based data-driven exercises is an example of discovery teaching. To prepare corpus-based  data-driven exercises, teachers have to  follow some steps: 1) to analyze a teaching context (including  a lesson content,  all  available resources; learners´ age, levels of language proficiency (CEFR,  2001), the gaps in language knowledge that are typical for every particular group of students,  and learner's styles, 2) to identify what to explore with corpora (for example, linguistic data or metadata), 3) to choose  some necessary corpora or a corpus (for this purpose, teachers  have to experiment with different  types of  corpora); 4) to  consider how to design corpus-based data-driven exercises; 5) to apply the prepared corpus-based data-driven exercises in practice; 6) to observe the results and make reflections;

7) to correct and improve the exercises (if necessary); 8) to apply the exercises once more; 9) to observe the results and make reflections.

At the elementary levels of language proficiency, it is not necessary to explain how to work with electronic corpora. Corpora are tools in teachers´ hands. Only teachers decide how to apply these tools and how to convert the information into usable didactic materials. For beginner language learners it is suitable to prepare training corpus-based data-driven exercises that consist of concordances that are deliberately selected by teachers for training of specific language skills. Training corpus-based data-driven exercises may look like filling the gaps exercises, matching or multi-choice exercises, etc. Producing training corpus-based data-driven exercises, teachers may use extra ICT-based resources (for example, Hot Potatoes, Teachers Pet, Power Point, etc.) or just print handouts. Training corpus-based exercises do not provide discovery learning directly through on line corpora. These types of exercises make language learners discover some patterns of authentic language use through already retrieved information from on line corpora.

At the intermediate and advanced levels of language proficiency teachers may engage learners into "discovering learning" applying discovering corpus-based data-driven exercises. Discovering corpus-based data-driven exercises provide observations of concordances directly from online corpora for discovering some facts of a foreign language (for instance, from linguistic data to metadata). Discovering corpus-based data-driven exercises also may be prepared by teachers for training of some specific language skills, but learners have to be able to retrieve and analyze the information directly from online corpora to come to the necessary conclusions.

## 3. Corpus-based approach in Portuguese-English and Portuguese-Russian crosslinguistic error correction

### 3.1. Grammar crosslinguistic errors (teaching English (L2/FL) to native speakers of Portuguese (L1)

The most frequent Portuguese-English crosslinguistic errors are related to the usage of English prepositions, because it is not correct to translate some prepositions from Portuguese into English literally. For instance, some native Portuguese speakers (L1) may say: "They are on the beach in the Algarve", because a Portuguese correct variant is "Eles estão na praia no Algarve".

The list of concordances of the British National Corpus demonstrates that there are three possible combinations "a preposition + a noun "beach" that English speakers (L1) use in an authentic discourse. They are "at the beach", "in the beach", and "on the beach". For producing of a training corpus-based data-driven exercise, teachers may prepare some handouts with gap filling exercises, based on the concordances retrieved from the British National Corpus (<http://www.natcorp.ox.ac.uk/>). For instance, there is an example of a gap filling training corpus-based data-driven exercise that contains some chosen concordances with different prepositions before a noun "beach" retrieved from the British National Corpus (Table 1).

Table 1. A gap filling training corpus-based data-driven exercise.

| | |
|---|---|
| 1. | Summer is summer … the beach. |
| 2. | His father wasn't down … the beach to greet him. |
| 3. | One day … the beach, they encountered a beautiful blonde starlet called Lila Leeds, something of a Lana Turner look alike. |
| 4. | That, gentle people, was hardly a day … the beach. |
| 5. | Twenty years ago the city's population never thought of spending a day … the beach. |
| 6. | In high-frequency waves the water motion appears to be nearly circular so that, when the waves break, the mass of water is directed downwards … the beach. |
| 7. | Food served … the Beach Terrace Bar and Restaurant is excellent and made all the more enjoyable by our visitors at breakfast. |

A discovering corpus-based data-driven exercise may be aimed at exploring the concordances with prepositions before a noun "beach" and explaining the differences of their usage in the contexts.

### 3.2. Lexical crosslinguistic errors (teaching English (L2/FL) to native speakers of Portuguese (L1)

Portuguese native speakers (L1) mix "true friends" (lexical units which mean almost precisely the same in the other language) with "false friends" (lexical units which suggest a wrong meaning in translation). For instance, a Portuguese speaker (L1) is making a crosslinguistic lexical error saying the following: "One conductor, making taxi service, accelerated up to 150 km/h to try to get in time". In English "conductor" does not mean "driver", but this lexical crosslinguistic error is an example of the use of a false lexical unit based on an incorrect translation from Portuguese (L1): "Um conductor, a fazer serviço de táxi, acelerou até aos 150 km/h para tentar chegar à hora certa". For preventing this Portuguese (L1)-English (L2/FL) crosslinguistic lexical error, teachers may produce a multi-choice training corpus-based data-driven exercise, retrieving concordances that contains the nouns "driver" and "conductor" from the British National Corpus (Table 2).

Table 2. A multi-choice training corpus-based data-driven exercise.

| | | |
|---|---|---|
| Andre Previn is now better known as an orchestral … | a) conductor | b) driver |
| The … William may well have had a few sleepless nights but must have been proud of his singers. | a) driver | b) conductor |
| A radio called Tony and an ambulance … who had just picked up a man with bandaged fingers who was suffering from exposure. | a) conductor | b) driver |
| It has been said that he understood the psychology of an orchestra better than almost any other ... | a) driver | b) conductor |

Discovering corpus-based data-driven exercises for lexical crosslinguistic error correction provide autonomous searching for correct translation equivalents, for example, in the COMPARA – Portuguese-English parallel corpus (<http://www.linguateca.pt/COMPARA/psimples.php>).

### 3.3. Phonological crosslinguistic errors (teaching Russian (L2/FL) to native speakers of Portuguese ( L1)

At the initial stage of learning Russian (L2/FL) adult native speakers of Portuguese (L1) modify the production of some consonant clusters and do not distinguish the degrees of vowel reduction of the studied language. It demonstrates the effect of phonological transfer that "involves using the phonological status of a segment or a contrast in the L1 during production or perception of a segment or contrast deemed equivalent in the L2, or transferring syllable structure or phonotactic constraints from the L1 for use during production or perception of the L2" (MacLeod, 2008: 8).

Russian initial, mideals and final clusters may consist of more than three consonants that are not typical for Portuguese. So learners' production, based on the L1 experience, break up the consonant clusters: ´здравствуйте´ [ˈdrɑsvujti], ´консульство´ [ˈkonsulsvɔ], etc.

Comparative with Portuguese, Russian has a greater degree of vowel reduction. Russian unstressed vowel sounds vary extensively their reduction depend on positions (Table 3). For instance, "the realization of /a, o/ in syllables in various positions relative to lexical stress…surface as [o, a] under stress, merge as [a] in the first pretonic syllable, and raise to schwa in the second pretonic syllable" (Barnes, 2007: 3).

Table 3. Some examples of vowel reduction in Russian.

| | Stressed | Degree 1 Reduction | Degree 2 Reduction |
|---|---|---|---|
| /o/ | дом (house) | домашний (domestic) | домовитый (housewifely) |

| | [ˈdom] | [daˈmaʃnij] | [dəmɐˈvitij] |
|---|---|---|---|
| /a/ | сад (garden, noun) | садовый (garden, adjective) | садовод (gardener) |
| | [ˈsat] | [sɐˈdovɨj] | [sədɐˈvod] |

Native speakers of Portuguese (L1) may pronounce Russian vowel sounds without altering the degrees of vowel reduction that make the language production not natural.

Producing training and discovering corpus-based data-driven exercises aimed at correcting the Portuguese-Russian phonological crosslinguistic errors teachers may apply the Russian National Corpus <http://www.ruscorpora.ru/index.html>. For example, there is a training corpus-based data-driven exercise based on the retrieved and selected concordances from the Russian National Corpus within a general search (Table 4).

Table 4. A training corpus-based data-driven exercise.

| Read aloud the following phrases: |
|---|
| Все трое вошли и поздоровались. - Здравствуйте! Очень рады вас видеть! |
| ― Здравствуй, Максим. ― Здравствуйте, маэстро. ― Твой импресарио прислал предложения, но у меня на фестивале нет таких денег. |
| Прорываемся к английскому и шведскому консульству, хотим принять участие в организации Дней культуры этих государств в Москве. |
| На следующий день, согласно прежде полученному и теперь подтверждённому приглашению, мы перебрались из отеля в турецкое консульство. |
| Вторая половина 1930-х годов, а особенно послевоенное время характеризуются стремлением к домашнему уюту, приватности, спокойствию, основанному на стабильности быта. |
| Чтобы избежать многих проблем, предусмотрительный садовод заранее произведёт замену грунта на более плодородный, добавит в посадочные ямы деревьев и кустарников комплекс удобрений. |

The multimedia subcorpus of the Russian National Corpus is a good database for discovering corpus-based data-driven exercises that may help to solve the problems of Portuguese-Russian phonological crosslinguistic error correction (Vieira, 2012b). This subcorpus provides numerous episodes from Russian documentary or feature films that are suitable for searching of difficult for pronunciation Russian sounds, and listening to them in the context of an authentic Russian discourse.

## 4. Conclusions

A corpus-based approach has many advantages in the practice of foreign language teaching. One of them is a possibility to produce corpus-based data-driven didactic materials for different pedagogical purposes and levels of language proficiency. The illustrated examples of training and discovering corpus-based data-driven exercises may be called "a scientific product" of modern foreign language pedagogy that provide a good didactic equipment for solving the problems of lack of specific didactic materials, for example, focused on crosslinguistic error correction.

A corpus-based approach involves teachers in discovery teaching through preparing corpus-based data-driven materials as well as learners – in discovery learning through exploring electronic corpora and doing linguistic research.

## References

Anderson, W., & Corbett, J. (2009). *Exploring English with Online Corpora*. Basingstoke: Palgrave MacMillan.
Barnes, J. (2007). *Phonetics and Phonology in Russian Unstressed Vowel Reduction: A Study in Hyperarticulation*. Boston University: Department of Modern Foreign Languages and Literatures. Available at

<http://www.bu.edu/linguistics/UG/barnes/Barnes%20VR%20final%20draft.pdf> [accessed March 2013].

*British National Corpus* (BNC). Available at <http://www.natcorp.ox.ac.uk/>.

*Common European Framework of Reference for Languages. Learning, Teaching, Assessment* (2001). Available at
    <http://www.coe.int/T/DG4/Portfolio/documents/Framework_EN.pdf> [accessed October 2012].

*COMPARA. Portuguese-English parallel corpus.* Available at <http://www.linguateca.pt/COMPARA/psimples.php>.

Huston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Johns, T. (1991). Should you be persuaded – two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 1–16). Birmingham: The University of Birmingham, Centre for English Language Studies.

Johns, T. (2000). Data-driven Learning: The Perpetual Challenge. In B. Ketteman, G. Marco (Eds.), *Teaching and Learning by Doing Corpus Analysis*, *Proceedings of the Fourth International Conference on Teaching and Language Corpora*, (pp. 107 –117). Graz 19 – 24 July.

McCarthy, M. (2004). Touchstone: from corpus to course book (Cambridge: Cambridge University Press). Available at
    <http://www.cambridge.org/elt/teacher-support/pdf/McCarthy-Touchstone-Corpus.pdf> [accessed 20 August 2012].

MacLeod, B. (2008). The Acquisition of Vocalic Sequences by English-Speaking L2 Learners of Spanish. *Toronto Working Papers in Linguistics*. Available at <http://twpl.library.utoronto.ca/index.php/twpl/article/view/6535/3528> [accessed January 2013].

Odlin, T. (1989). *Language Transfer: crosslinguistic influence in language learning*. Cambridge: Cambridge University Press.

Reppen, R. (2010). *Using Corpora in the Language Classroom*. New York: Cambridge University Press.

*Russian National Corpus*. Available at <http://www.ruscorpora.ru/index.html>.

Sharwood Smith, M., & Kellerman, E. (1986). Crosslinguistic influence in second language: an introduction. In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic Influence in Second Language Acquisition* (pp. 1–9.). Oxford: Pergamon.

The International Phonetic Alphabet (revised to 2005). *The International Phonetic Association (IPA)*. Available at
    <http://www.langsci.ucl.ac.uk/ipa/ipachart.html> [accessed January 2013].

Vieira, N. (2012a). A comprehensive database for successful communication. *IBER – International Business and Economics Review*, 3 (pp. 186 – 195). Lisboa: CIGEST.

Vieira, N. (2013). Corpus-based data-driven exercises for exploring grammar through authentic texts. *Conference Proceedings of III International Symposium on Lexical Typology* (pp. 101 – 105). Granada: University of Granada.

Wallace, M. J. (1997). *Action Research for Language Teachers*. Cambridge: Cambridge University Press.

Виейра, Н. (2012b). Мультимедийный подкорпус на уроках русского языка как иностранного. *Современные гуманитарные исследования*, 4 (с. 25 – 27). Москва: Спутник+ [Vieira, N. (2012b). The Multimedia Subcorpus on the lessons of Russian as a Foreign Language. *Modern Researches in Humanities* , 4, (pp. 25 – 27). Moscow: Sputnik +].