

Analysis of Compositions by B1 Level (Threshold) Francophone Learners of Catalan: Typology of Errors and Correspondences

Joan Casademont, Anna
Université TÉLUQ

1. Introduction

In recent years, interactions between individuals have multiplied, highlighting the important role communication plays with people who have different first languages (L1) and, therefore, the importance of being able to communicate in different languages. The range of situations requiring communication and the traditional needs of the population have diversified dramatically. Professionals working in the linguistic field are no longer the sole persons interested in language proficiency, so too are those who want to communicate at an academic, political, economic, cultural, social, recreational, or other levels (Tarone 2015, 448). Moreover, states and communities increasingly encourage the acquisition of languages other than the first language, as it promotes growth and development (Eurostat 2016).

Therefore, in order to help learners to master different languages more effectively, researchers and practitioners are continually seeking more effective teaching strategies. Linked to the idea of growth and development, our main motivation here is contributing to specifically support Catalan teachers' decisions in terms of pedagogy and feedback in writing, using linguistic analysis of students' errors to provide teachers with an overall portrait of their learners.

Using a corpus made up from compositions written by L1 French learners of L2/L3 Catalan, texts were part of the official Catalan language examination of the Ramon Llull Institute conducted in Montreal between 2009 and 2016. This study aims to analyze errors found in the writing of learners of Catalan in order to elaborate on pedagogical comments or observations and eventually find pedagogical approaches that will help learners avoid them. Our goal is to help develop a learner's transitional communicative competence into an even more grounded proficiency in the L2/L3 Catalan. Therefore, errors are seen as an opportunity to improve strategies learners could use when communicating in an L2 or additional language (Astolfi 1997, 2015).

Drawing on the CAF model [*complexity, accuracy, fluency*] (Skehan (1998), Ellis (2003, 2008), Ellis and Barkhuizen (2005), Housen et al. (2012), and others) and particularly on the *accuracy* component, we are interested in the lexical and grammatical level as well as the expressive and communicative level of learners' productions. Therefore, we adopt a holistic view of accuracy, including what Germain and Netten (2002) refer to as linguistic accuracy (grammatical and discursive) and pragmatic

accuracy (functional and sociocultural). Following Celce-Murcia's (2007) broad model on communicative competence, a dynamic vision between different types of competences: linguistic, formulaic, sociocultural, interactive, discursive and strategic, has been employed.

As presented in further detail below, for this study, we modeled a taxonomy that would allow us to examine errors from different perspectives: linguistic aspect, modification type, interlinguistic influences, intralinguistic causes and communicative consequences. Thus, we used both a descriptive and etiological approach.

An important note regarding the corpus of our project (see next section for more details) is that Catalan language learning at the university level is done through a communicative task-based approach (Ellis 2003, and others). In other words, contextual scenarios derived from different communicative situations are used to ensure effective communication in accordance with pre-established objectives. The topic of the compositions found in our corpus, a letter describing habits and environment to a friend while living abroad, reflects this approach.

It should also be noted that a few preliminary analytical studies have already been conducted on the corpus presented here. The first is a study that reports several quantitative and qualitative findings in terms of linguistic descriptions: linguistic aspect and modification type (AUTHOR, 2020a). The second study focuses on analyzing the errors detected from an etiological standpoint (interlinguistic influences and intralinguistic causes) (AUTHOR, 2020b). In both cases, the analyses allow us to present a few trends and generalizations and to propose potential pedagogical approaches to address the recurring errors made by students.

In this article, we will use a multiple correspondence analysis (MCA) at the error level to identify associations between the variables that characterize the detected errors.

Specifically regarding interlinguistic or crosslinguistic influences in additional language acquisition, Orcasitas-Vicandi (2019) considers that a "better awareness of the associations between the languages learners know could help them and their teachers to make more conscious, appropriate and effective connections among the languages in the classroom context". We strongly believe that this awareness and explicit knowledge is extremely useful, not only when considering interlinguistic influences, but also in other perspectives from which to consider errors.

This article will be structured as follows: first, we will present the corpus and the methodology used for the analyses; afterwards, we will explain the framework we work with and different analysis and results carried out; we will end by discussing our main findings, pointing the limits of this research and drawing some conclusions and pedagogical proposals.

2. Corpus and methodology

Before considering the framework of this research in detail, we will present here information on the corpus analyzed and the main instruments used for it.

For the analysis presented in this article, we used a corpus of compositions written by intermediate learners from the Ramon Llull Institute's official Catalan language examination of the Common European Framework of Reference for Languages (CEFR)¹. Specifically, these are the B1 examinations conducted in Montreal between 2009 and 2016.²

The compositions we analyzed are one of the Ramon Llull Institute's official exam questions, located in the written production section, and have been made anonymous. They are open-ended compositions with an underlying topic: students were asked to write an informal letter to a friend describing their experiences during an internship abroad (accommodation, classes or internships, unique local features, people they have met, etc.) and invite the recipient to visit them.

The corpus includes sociolinguistic data for each composition in order to limit the number of variables in the linguistic analyses of the texts. For the analyses presented here, we sorted through the 22 original compositions and chose only compositions written by learners whose first language is exclusively French. This gave us a final total of 17³ compositions for the corpus of this research.⁴ Afterwards, each

¹ The CEFR was created from a document published by the Council of Europe (2001). This document defines the levels of language fluency in terms of proficiency in different competences. These levels currently set the standard for language acquisition and teaching in several countries, regardless of the language studied or the organization responsible for assessing language proficiency. The B1 level (*Threshold*) marks the beginning of autonomy in the learned language. At this level, oral and written comprehension, oral and written production, and oral interaction are skills identified by the CEFR. A companion volume to the CEFR with new descriptors was published in 2018.

² This corpus is part of a larger collaborative one (BLABLA Corpus) created from different sources, different L1, L2 and Ln, as well as different levels and tasks. All documents of BLABLA Corpus are annotated using the same methodology, allowing the researchers participating in its construction to approach SLA from different perspectives and different combinations of variables.

³ Among the 17 participants from our corpus, 9 were men and 8 were women. With respect to age at the time of the examination, most participants were between the ages of 25 and 35.

⁴ Even though the corpus was previously anonymized by the Institut Ramon Llull, the institution had compiled different sociolinguistic data linked to the code of each composition by using the information given by participants in their pre-inscription form to the exam. By using this anonymized data, we were able to trace participants L1 and other learned languages, participants proficiency in Catalan, age and place where the exam was taken. Before doing so, we applied for an Ethical Conduct for Research Involving Humans certification, which was granted for the first time in 2019 by the ethical committee of Université TÉLUQ. Some of the annotation of the BLABLA Corpus has been financed by the research funding program FAR1 (Université TÉLUQ).

composition was manually marked by the researchers and 237 errors of different types were found.

The description of the errors was done using an ad hoc classification system based on the bibliography consulted on errors analysis and that we explain more in detail in section 3 of this article.

For this research, we carried out a manual annotation using the software Nvivo and, then, we considered different types of statistical analyses using SPSS software:⁵

1. First, for each error category, we summarized the number of errors made for each subtype using frequencies and percentages out of the total 237 errors found in the corpus.
2. Cross-frequency tables were then generated to study the relationship between two types of errors at the same time.
3. Spearman⁶ correlations were also generated to measure the association between the number of each type of error per student.
4. Finally, a multiple correspondence analysis (MCA)⁷ allowed us to identify links between categorical variables at the error level; the results of the latter analysis will be discussed here. In our study, there are four variables, and each represents one of the four principal types of information gathered for each error: error aspect, modification type, interlinguistic influence, and intralinguistic cause.

In section 4, we present a few quantitative results and several qualitative analyses based on the data collected using the aforementioned method. We will focus on the values associated with the linguistic level in which errors occur, the phenomenon linked to the error, and the possible origins for the error (interlinguistic influences and intralinguistic causes). In particular, we will examine results gathered through the observations of the correspondences of categorical variables at the error linguistic aspect.

⁵ It should be noted that none of the analyses carried out assume normality and we did not detect any outliers.

⁶ The Spearman correlation was used instead of the Pearson as the number of errors was a discrete variable and not a continuous variable.

⁷ Multiple correspondence analysis (MCA) is a data analysis technique for nominal categorical data, used to detect and represent underlying structures in a data set. It is a method permitting to study the association between more than two variables. It creates two dimensions representing the maximum of information of all the qualitative variables, so that those dimensions present data as points in a graphic using Cartesian coordinates.

3. Framework

In the field of second/third/additional language acquisition, error analysis (EA) has long been a field that encompasses a wide range of studies and research. According to Erdogan (2005, 269), there are three common characteristics that would provide a comprehensive description of EA-related research in second language acquisition:

- EA identifies strategies used by language learners.
- EA seeks to explain why learners make mistakes.
- EA identifies common challenges in language acquisition and assists in the development of resources and materials to address these issues.

In 1967, Stephen Pit Corder was already distinguishing between systematic errors or *errors* (deviations from the norm) and production errors or *mistakes* (occurring non-systematically, caused by fatigue, lack of concentration, etc.) (Fernández Jódar 2006, 13).

It is necessary to mention the term *interlanguage* (Selinker 1972), as it relates to errors. Interlanguage is the intermediate stage in the language learning process, which often bears traces of the L1. Fernández Jódar defines it as follows:

The IL [interlanguage], however, is a mental language system created by the learner to express in an L2 what he/she could express in his/her L1. To do this, he/she uses the L1 (transfers) or makes assumptions in the L2 that may be correct or incorrect. This mental language system is not considered erroneous but is an emerging language system between the L1 and the L2 with features of both and with its own unique features. It reflects the learner's transitional communicative competence. (Fernández Jódar 2006, 10; our translation)

A major focus in the field of EA is the establishment of error typologies or taxonomies, as it is an essential step in order to further process corpus data and obtain results. At this stage, depending on the research objectives, researchers usually look at errors in several linguistic levels simultaneously or they concentrate their efforts on specific aspects, such as morphology or grammatical number. Additionally, specialists choose different classification criteria, in accordance with the purpose of their research and propose or use taxonomies allowing them to study relevant elements. Moreover, some studies reflect, either generally or abstractly, on errors while others base their research on the analysis of one or more languages.

In our case, we propose an original annotation using a broad perspective including a description of the linguistic aspects and modification types as well as the annotation of the communicative consequences and error etiology (interlinguistic and intralinguistic), in order to approach errors from different perspectives. Table 1 presents the different categories that were considered in our framework, along with the nodes associated with them and some examples of the sources of inspiration when establishing them.

Table 1
Framework: Categories and Nodes

Categories	Nodes	Sources
Linguistic Aspects	Orthographic, Morphologic, Lexical-Semantic, Syntactic, Pragmatic, Cohesive and Coherence, Typographic	e.g. Corder, 1973; Ellis, 1997; Alexopoulou, 2006.
Modification Types	Omission, Overinclusion, Misselection, Misordering, Blends	e.g. James, 1998; Fernández Jódar, 2006; Alexopoulou, 2006.
Interlinguistic Influences	L1, L1 and Other L, Other L(s)	e.g. Dulay, Burt and Krashen, 1982; James, 1998; Fernández Jódar, 2006; Alexopoulou, 2006.
Intralinguistic Causes	False Analogy, Non-Application of the Exception of Rules, Uncomplete Application of Rules	e.g. Dulay, Burt and Krashen, 1982; James, 1998; Fernández Jódar, 2006; Alexopoulou, 2006.
Communicative Consequences	Causing, Not causing miscommunication	e.g. Burt, 1971; Dulay, Burt and Krashen, 1982; Gozali, 2018.

To summarize, we present here the most relevant fields of description for the quantitative and qualitative analyses that we discuss in this article, accompanied by a brief description:⁸

- LINGUISTIC ASPECT:
 - Orthographic error: error in how a word is written.
 - Morphological error: error in word formation, affixes and root word.
 - Syntactic error: error concerning the presence or absence of mandatory elements and their sequence in a sentence.
 - Lexical-semantic error: error in the use of a word to express a specific meaning; use of a foreign word.
 - Cohesive and coherence error: error in any unit above the sentence level, where it can affect the sequencing of sentences and the coherence and/or cohesion of the text.

⁸ For a more detailed and comprehensive overview of the values linked to the description of the linguistic level and the phenomenon, see Appendix 1.

- Pragmatic error: error in discourse conventions (conventions, implications, innuendos, etc.).
- Typographic error.
- MODIFICATION TYPE:
 - Omission: necessary element is missing.
 - Overinclusion: extra element is present.
 - Misselection: incorrect element is used.
 - Misordering: elements are correct but in the wrong sequence.
 - Blends: various modifications.
- INTERLINGUISTIC INFLUENCES:
 - L1.
 - Other L.
 - L1 and Other L.
 - Interlinguistic N.A.
- INTRALINGUISTIC CAUSES: interference with what has already been learned in the target language:
 - False analogy: applying a rule when it does not apply; using a word or expression that is incorrect in the given context.
 - Incomplete application of the rules.
 - Non application of the exceptions of rules.
 - Intralinguistic N.A.
- COMMUNICATIVE CONSEQUENCES:
 - Causing miscommunication.
 - Not causing miscommunication.

4. Analysis and results

In this section, we present the main findings as well as a qualitative analysis of them.

4.1. MCA I: Presentation of results

As mentioned above, by using a MCA, the recurrent associations between variables make it possible to create dimensions, which in turn can be used to summarize the information contained in the data. Depending on the percentage of variance, each dimension can help explain more or less all the data in the corpus.

In the MCA carried out with our corpus, 46% of the variance in the data is explained by the first dimension and 36% by the second (figures in bold below). These two dimensions make it possible to summarize 83% of

the information contained within the data. The associations between the dimensions are therefore very informative (Table 2):

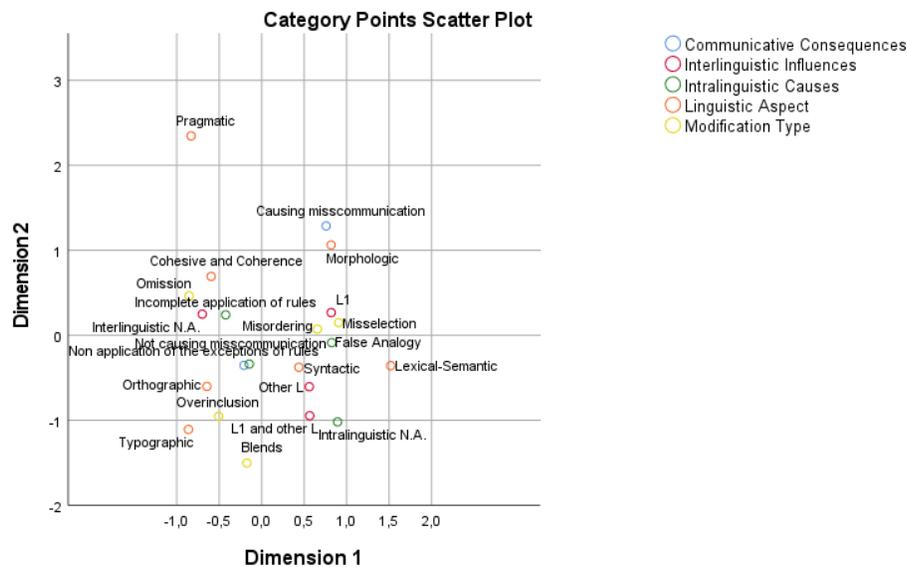
Table 2
Summary of Dimensions

Dimension	Cronbach's Alpha	Represented Variance		
		Total (Eigenvalue)	Inertia	% of variance
1	.711	2.318	.464	46.364
2	.566	1.828	.366	36.564
Total		4.146	.829	
Mean	.647 ^a	2.073	.415	41.464

Note. a. The Cronbach's alpha mean is based on the eigenvalue mean.

Using dimensions, we can represent each type and subtype of error in a scatter plot, in which the proximity of two points indicates that those types of errors (nodes of variables) often occur concurrently (Figure 1). Given the percentages of variance for each of the dimensions mentioned above, the distance between the points on the horizontal axis on the graphs is slightly more informative than the distance on the vertical axis:⁹

Figure 1
Overall Result of the MCA



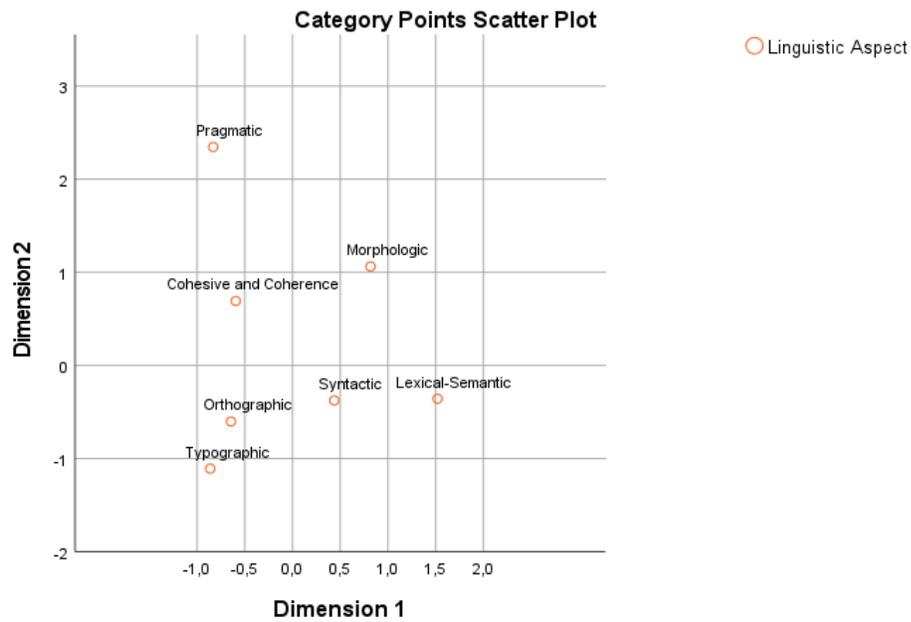
⁹ If we compare this graph to the cross-frequency tables, we would see that the points shown on the graph usually correspond to cells with high frequencies.

Before interpreting the results in the following section, we will first look at each variable independently.

Our first approach with the results of the MCA is to observe the positions of the different points of the same variable: the further the points are from one another, the more association information they provide.

The differences in the proximity of the points of a variable become visible when looking at the linguistic aspect variable in Figure 2:

Figure 2
Linguistic Aspect Variable Points

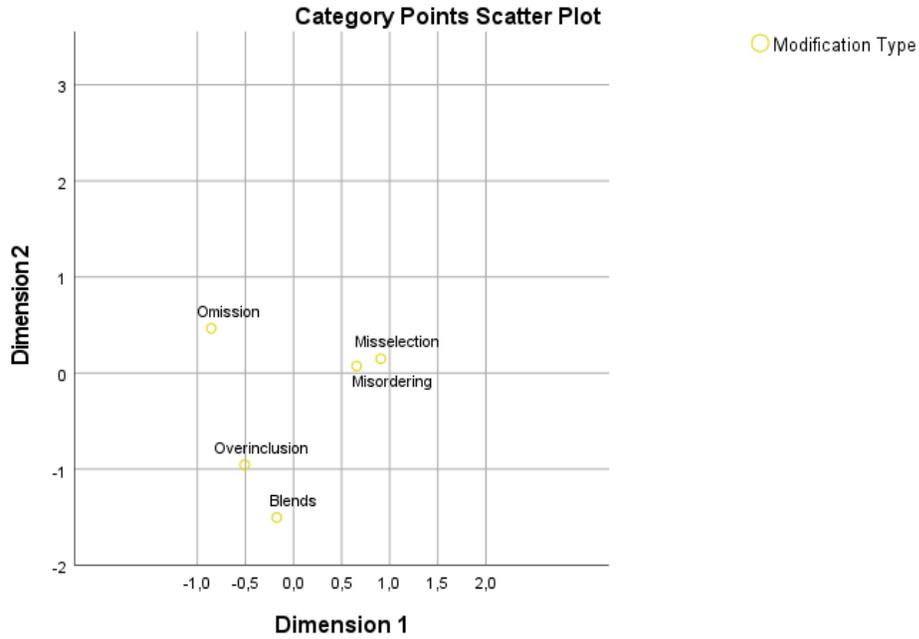


The points in Figure 2 are in more or less close proximity to one another. However, pragmatic errors appear to be farther from the other points. Therefore, these points should be easier to distinguish and characterize when compared with errors at other levels.

As of the variable of modification type (Figure 3), the points of this variable also appear to be more or less spread depending on the case:

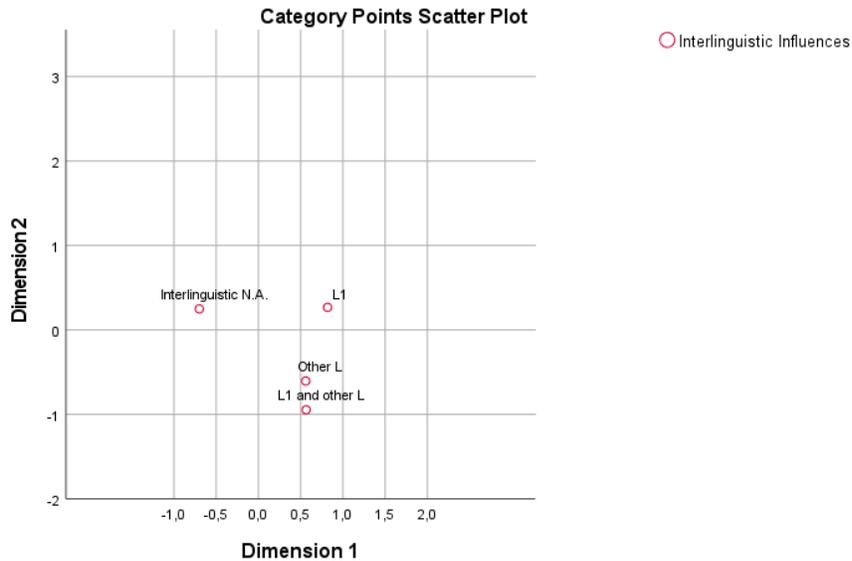
Figure 3
Modification Type Variable Points

In fact, the position of the omission value (Figure 3) appears to



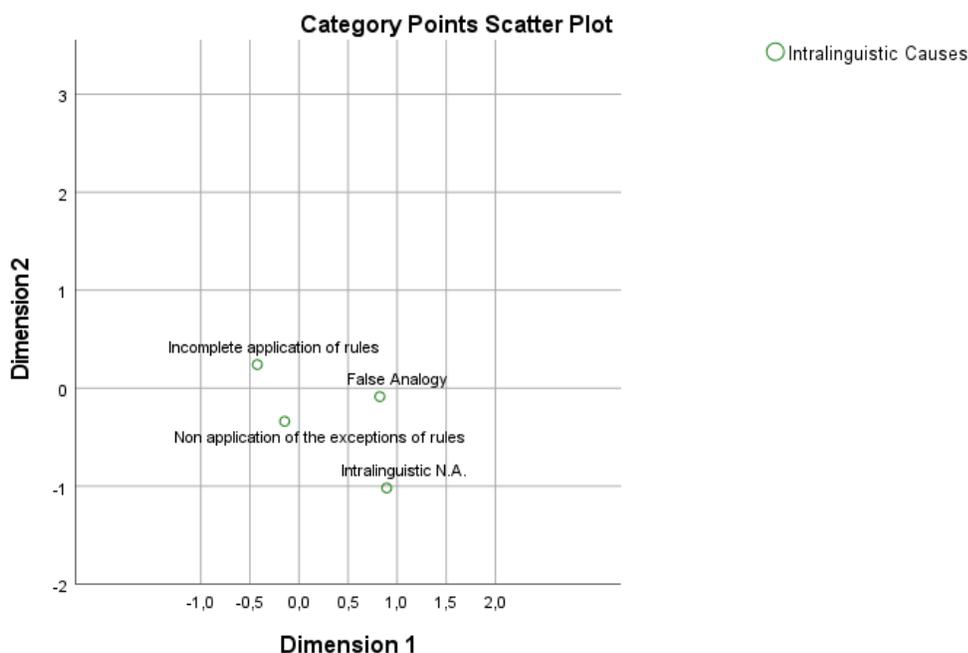
stand apart from the other values of the same variable. As for interlinguistic influences, in Figure 4 below, the points are closer to one another when compared with other variables seen above:

Figure 4
Interlinguistic Influences Variable Points



Therefore, if no other characteristics exist to characterize this type of error; they will not be useful to differentiate between these observations. In comparison, intralinguistic causes seem to be a bit more helpful when characterizing types of errors (Figure 5):

Figure 5



Intralinguistic Causes Variable Points

To summarize, we have looked at variables individually, and the results allowed us to hypothesize on their more or less helpfulness when characterizing errors found in our corpus. For example:

- As seen in Figure 1, interlinguistic influences points are in proximity with several other variable points and not just one or two. Therefore, the L1 and other Ls seem to play an important role in learners' errors: their influence appears to be quite global in that they don't seem to be more influential with specific types of errors or sub-types of errors.
- In Figure 2, we can observe that the pragmatic value seems to stand apart from the others. It would be interesting, for example, to study whether other concrete variables have a characteristic or value that could complete the characterization of these types of errors and clearly distinguish them from others. Something quite similar, in a less degree, could be said for the omission value in Figure 3.

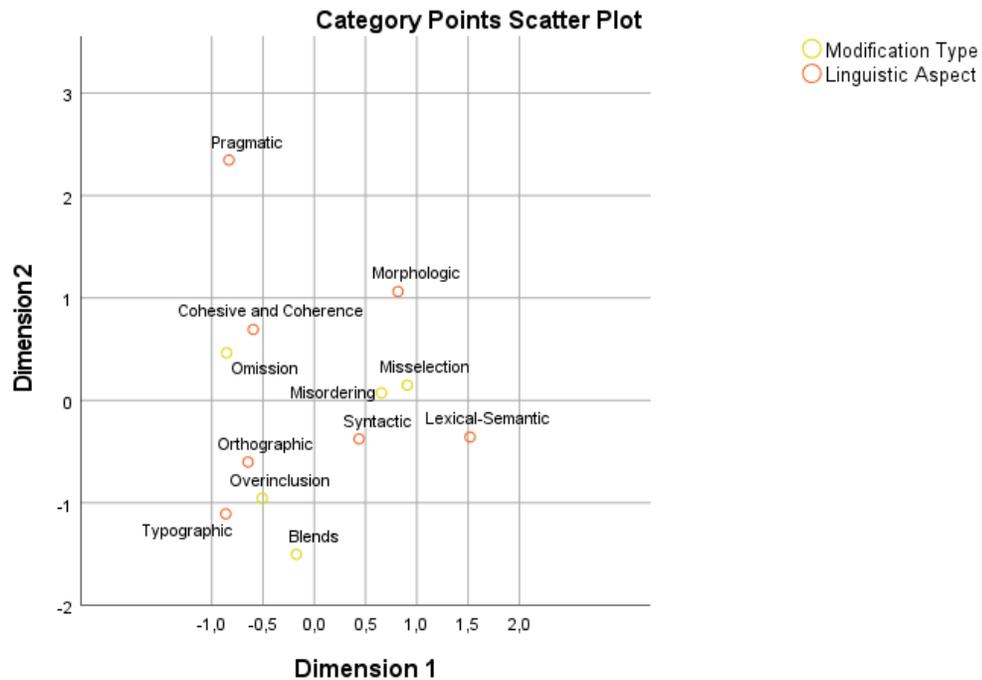
In the following section, we will attempt to interpret the gathered results, particularly with regard to possible links between the characteristics of the variables in an effort to suggest associations that could be informative and thus provide a clearer idea of the learners' performance. Examples from the corpus will also be used when discussing the different associations.

4.2. MCA II: Some associations

If we look at the possible associations between the characteristics of different variables, we should get an even more detailed picture of the learners' productions in our corpus.

Using the linguistic aspect variable as our basis, we will make comparisons with the modification type (Figure 6), the intralinguistic causes variable (Figure 7), and the interlinguistic influences variable (Figure 8): see graphs below.¹⁰

Figure 6
Combination of Linguistic Aspect and Modification Type Points



¹⁰ We will indicate in our description the approximate placement of the points in question on the axis of the graphs, as well as provide the graph number for reference purposes.

Figure 7
Combination of Linguistic Aspect and Intralinguistic Causes
 Points

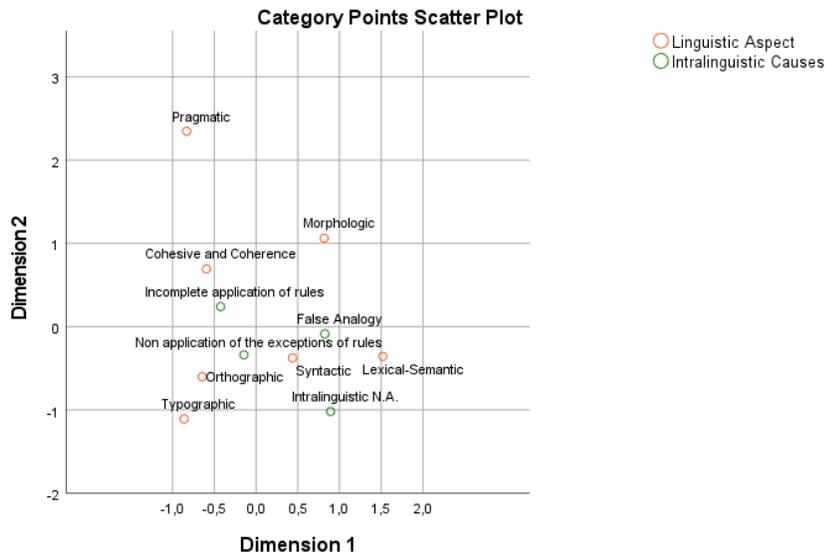
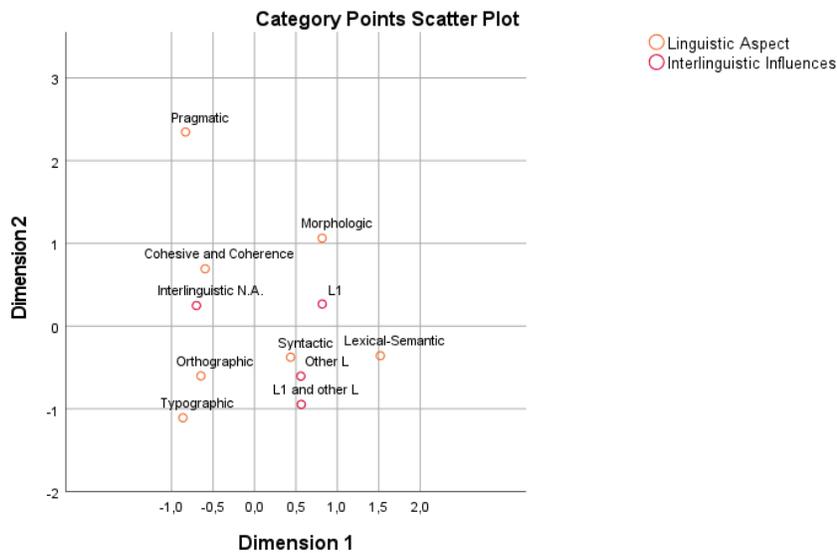


Figure 8
Combination of Linguistic Aspect and Interlinguistic Influences
 Points



Errors at the orthographic level

Errors at the orthographic level [near coordinates (-0.5,-0.5)] are found mainly in proximity to the phenomena of overinclusion and omission (Figure 6). Therefore, these two modification types would be common in

the corpus in the case of orthographic errors. Below is an example of overinclusion (*Oc. 1*) followed by an example of omission (*Oc. 2*):¹¹

Oc. 1: És força divertit, sempre diu acudits amb l'accent del nord del país, on vivia [vivia//lived] abans (EC: 114).

Oc. 2: Podriem anar junts al Parc de la Mauricie, un racó natural molt bonic amb gran varietat d'especies [espècies//species] d'ocells (EC: 118).

In the overinclusion example (*Oc. 1*) above, the use of an accent in **vivía* is incorrect. This error may be due to the influence of another learned language, in this case, Spanish [this interlingual variable point (Other L) is the one most in the proximity of the two anterior points on the horizontal axis: 0.5], even though interlinguistic influences points are closer to other linguistic aspects points as morphology, syntax and lexical-semantic errors (Figure 8).

In the omission example (*Oc. 2*), the accent is missing on the anterior penultimate syllable of the word **especies* (*espècies*); in this case, it appears as though the rules on accentuation have not been applied completely [this intralingual variable point (incomplete application of the rules) is also in proximity of the two anterior points on the horizontal axis: -0.5] (Figure 7).

Another trend that seems to cause omission errors at the orthographic level is the non-application of rule restrictions. The latter value is in proximity to some of the points discussed earlier (Figure 7). Take the following sentence (*Oc. 3*) for example:

Oc. 3: El més [mes//month] que ve faran el Faust de Gounod (EC: 203).

In the occurrence above, diacritical accent restrictions do not appear to have been applied. The accent in **més* is incorrect, giving it the meaning of "more" instead of "month", which in Catalan is written *mes* without a diacritical accent.

Errors at the morphologic level

If we look at the positions of other types of errors, it seems that at the morphologic level [near coordinates (0.7,1)], misselection errors are common (Figure 6) (*Oc. 4, Oc. 5*):

¹¹ The occurrences found in the corpus are numbered here according to their order of appearance. The number in parentheses at the end of each occurrence indicates the automated "Error Code" given when detecting and analyzing the errors in the corpus. The information in square brackets "[]" indicates the correct form of the detected error, followed by an English equivalent.

Oc. 4: És una ciutat del meu país que no coneixava [coneixia//knew] (EC: 210).

Oc. 5: El comparto [comparteixo//share] amb un estudiant mexicà molt trempat (EC: 70).

False analogy, also in proximity to the morphological point on the horizontal axis (Figure 7), could be a frequent cause for this type of error here. For example, in the occurrences above, there are conjugation errors with the words *conèixer* and *compartir*, where existing morphological conjugation forms are proposed though not correct in the case of those specific verbs. As for interlinguistic points, the influence of L1 seems to be especially important in this kind of errors (Figure 8), with the L1 point nearly at the same horizontal axis position as the morphologic point. The two latter examples could also be used to illustrate this tendency: there is no specifically irregular conjugation of the French verbs *connaitre* (*know*) and *partager* (*share*).

Errors at the syntactic level

If we examine the position of the types of errors at the syntactic level [near coordinates (0.5,-0.20)], misordering (*Oc. 6*) and misselection (*Oc. 7*) errors are expected to be frequent and seem to be also caused by the influence of other learned languages or other learned languages in combination with the L1 (Figure 8):

Oc. 6: Nicolas estudia història de Xina durant el XVIIIo segle [segle XVIII// XVIIIth century] (EC: 113).

Oc. 7: De fet, no et sembla una bona idea venir a Canadà en [el/al//the] desembre? (EC: 72).

In the first example above, the order "CENTURY NUMBER + *segle*" would be correct in French and English (L2 of the learner), but not in Catalan, where the pattern should be "*segle* + CENTURY NUMBER".

In the second example, the structure "*en* + MONTH" would be correct in French and Spanish (L2 of the learner), but the correct form in Catalan would be "*el/al* + MONTH".

Errors at the lexical-semantic level

Lexical-semantic errors [near coordinates (1.5,-0.25)] occur when the incorrect form is used (Figure 6). From an etiological standpoint, these errors do not seem to be caused by intralingual causes but by the influence of the L1 or another learned language, as seen in Figure 8 (*Oc. 8*):

Oc. 8: D'aquí a 30 minuts estaré [seré//will be] a l'escola (EC: 40).

As in the example above, most of these errors are caused by the incorrect use of the verbs "ser" and "estar" in Catalan, where the verb "estar" is used incorrectly at the expense of "ser". Taking into account that the learners in the corpus are francophones and that most of them have Spanish as a learned language, we believe this error can be explained by two complementing interlinguistic circumstances:

- The two verbs in Catalan have a single verb equivalent ("être") in French, the learners' first language.
- Learners who make these errors learned Spanish before, where the use of the verb "estar" is much more prominent than "ser".

Another example of a case at the lexical-semantic level with the influence of the L1 is seen in the following (*Oc. 9*), where the word **propera* ("proche" instead of "propre" in French) is used incorrectly:

Oc. 9: És una gran casa que té dues plantes, més aviat quadrada. Com que és espaiosa, hi tinc a meva propera [pròpia//own] habitació al segon pis (EC: 82).

Moreover, in this example, the use of **propera* could be also partially caused by a false analogy (the nearest intralinguistic value point to lexical-semantic point errors), contributing to the erroneous decision taken by the learner, since *propera* exists in Catalan though with a complete other meaning ("near").

A second recurring trend with lexical-semantic errors appears to be associated with misselection errors (Figure 6) in multi-lexical units, as is the case in the following occurrences (*Oc. 10*, *Oc. 11*), which also appear to be influenced by the L1 and/or another learned language (Figure 8):

Oc. 10: Et voldria convidar a Mont-real durant la fin de setmana [el cap de setmana//the weekend] que ve (EC: 188).

Oc. 11: Voldria que coneguis la meva família d'acoll [família d'acollida//host family] perquè és fantàstica! (EC: 87).

In the first example, **la fin de setmana* is not the correct form; the correct form is *el cap de setmana* (*fin de semaine* in French, *fin de semana* in Spanish). In the second example, the French term *famille d'accueil* seems to have influenced the learner's choice of words for **família d'acoll* instead of *família d'acollida*.

Errors at the cohesive and coherence level

Cohesive and coherence errors [coordinates (-0.5,0.7)] appear to be in proximity to omission (*Oc. 12*) and overinclusion (*Oc. 13*) errors mainly through the incomplete application of rules and without interlinguistic influences (Figure 6 and Figure 7):

Oc. 12: Es diu Carola [,,/] és Argentina i passem bons temps junts [...](EC: 311).

Oc. 13: Després, anirem al sud, a la ciutat de Kyeongju per visitar als temples budistes, i [∅ i//∅ and] tornarem a Seül per visitar la Capital (EC :206).

In the case of cohesive and coherence errors, most of them are related to punctuation problems, such an omission of a comma (*Oc. 12*) or the overinclusion of one (*Oc. 13*).

Errors at the pragmatic level

As for pragmatic errors, there are very few pragmatic errors [near coordinates (-0.7,2.5)] in the corpus; that said, it seems that they are mostly linked to errors on the genre format: omission of date, head and/or other compulsory parts of the letter genre (Figure 6). This type of errors is mainly considered to be free of interlinguistic influences and be caused by the incomplete application of rules.

Errors at the typographic level

Finally, there are also very few typographic errors [near coordinates (-1,-1)] in the corpus; that said, it seems that they tend to be linked to omission and, for the most part, to overinclusion errors (Figure 6), such as the use of a space before ":" as in French or the use of capital letters when it is not needed in days of the week or demonyms (*Oc. 14*):

Oc. 14: és una de les activitats d'oci mès populars entre els Coreans [coreans//Corean] (EC: 40).

In the example above, it seems that French and English (L2 of the learner) have influenced the wrong choice of using a capital letter.

5. Discussion and limits of this research

The multiple correspondence analysis (MCA) carried out allows us to establish general association patterns between the characteristics (nodes, points) of error variables made by intermediate learners of Catalan.

It is important to remember that, in this article, our focus was not on the specific descriptions of the different types of errors identified or the possible pedagogical approaches to counter them¹². Instead, we are interested in the trends detected by the MCA and presented herein, which allow us to make three general observations:

Observation 1: The interlinguistic etiological variable is present in learners' errors. However, the specific characteristics of this variable (namely the first language and learned language) are not distinct enough in our corpus to make associations that differentiate them. We observed that orthographic, morphologic, syntactic, and lexical-semantic errors seem to have an interlingual explanation for them. Nevertheless, it is unclear if the interlingual influence is from the L1 or another learned language (in this case, Spanish). Learners at the intermediate B1 level seem to draw on their knowledge of other languages, whether it is their L1 or another learned language. Is this type of interlinguistic influence more pronounced at the beginner level but gradually gives way to intralinguistic explanations in more advanced levels? This is what Alexopoulou (2005) suggests with regards to morphological errors and what our corpus appears to be more generally confirming. Does the lack of distinction between the influence of the L1 and other learned languages occur at all levels of learning or varies according to the learners' proficiency? We hypothesize that the influences overlap until a more advanced level of language proficiency is achieved. However, more research would be needed to confirm this hypothesis.

Observation 2: Overinclusion and omission are very prominent in cases of orthographic errors. Similarly, these errors are often associated with the incomplete application of the rules or the non-application of rule restrictions, as well as the influence of other learned languages, mainly Spanish, in this case. We also noticed that the majority of these errors were specific to rules on accentuation. Is this an issue independent of the learners' language and their learning level?

Moreover, the phonetic consequences of their errors do not appear to pose problems for learners at this stage. Does this change when a more advanced level is attained? We hypothesize that this is an issue independent of the learners' language and that the phonetic consequences become more pronounced as learners develop their level of language proficiency; therefore, errors of this type should decrease. As was noted above, more research would be required to confirm this hypothesis.

Observation 3: Misselection is prominent in cases of syntactic and lexical-semantic errors.

Similarly, these errors are often associated with interlinguistic influences, mainly other languages learned as well as L1 (see Observation 1 above). We also noticed that the majority of these errors were specific to the use of prepositions (syntactic level), and the erroneous use of verbs

¹² For information on these topics, it is possible to consult AUTHOR (2020a, 2020b).

"*ser*" and "*estar*" or false cognates (lexical-semantic level). In the specific case of misselection, we could ask ourselves if the type of other language learned could be considered as relevant in this case. This is what also suggests the findings in AUTHOR & Gagné (2020), where there seems to be important factors varying as a result of different learning paths of L1 French. This results would be consistent with the Rothman's Typological Primacy Model (TPM) since, at this intermediate level, interlinguistic or crosslinguistic influences "will come from the background language that the learner's internal mechanisms perceive" to be the most similar to the target language (Rothman, 2010, 2011, 2014, 2015).

The results of this analysis and the previous findings are not without limitation and allow us to raise new research questions. The association patterns identified to date are framed in the reduced context of the Quebec Francophonie, only among intermediate level (*Threshold*) students, and in a reduced sample of the corpus since it was the teaching context where the author usually practices that was chosen for these first experiments. As such, the next steps in our research are as follows:

- Increase the size and scope of the corpus I: same intermediate level, the addition of two sub-corpus with compositions from the same examinations written by French-speaking learners in Belgium and France. This step would allow us to confirm error trends and observe whether there are differences according to the learner's linguistic community.
- Increase the size and scope of the corpus II: within the same linguistic communities of the learners, addition of other sub-corpus with beginner and advanced level compositions. This step would allow us to observe possible evolutions in learners' errors.
- Increase the size and scope of the corpus III: within the same linguistic communities of learners, addition of other sub-corpus with other types of texts. This would allow us to analyze different constructs and see if errors are generalized in other contexts.
- Extend the corpus to other sociolinguistic combinations: in this case, mainly to learners of Catalan with different L1s and other languages learned.

6. Conclusions and some pedagogical proposals

Finally, it should be noted that Observation 1 presented above is different from Observations 2 and 3. The first observation focuses on one variable (interlinguistic influences) and its incidence on all types of errors, as the second and third observations focus on the strong links detected between

specific characteristics of different variables that are prominent in the corpus.

From a general point of view, our results confirm general trends in the field. We have indeed observed that error analysis in second language acquisition can reveal relevant, but varied, trends and generalizations that can improve our knowledge of (a) how learners progress in second-language acquisition, (b) strategies that learners seem to use in their productions, (c) characteristics of their errors that do not discriminate vis-à-vis other aspects, and (d) specific problems that seem to be the most recurrent. By using this type of data, we can then effectively design practical teaching strategies to help learners on their learning journey, in this case at the intermediate Catalan level.

In the analyses above, some recurrent more specific errors have been detected among tendencies:

- Orthographic level: errors on accentuation, linked to diphthongs or diacritic accents.
- Morphologic level: irregular common verbs which are usually regular in other known languages.
- Syntactic level: prepositions when compared to L1 and other Ls.
- Lexical-semantic level: the distinction between "*ser*" and "*estar*", as well as the use of cognates being false friends.
- Cohesive and coherence level: problems of punctuation, mainly commas, that sometimes change the sense of the whole paragraph.
- Pragmatic level: errors linked to the genre format of the written text.
- Typographic level: the space before ":" as well as the erroneous use of capital letters by interlinguistic influence.

These specific errors that were frequently found in our corpus could be useful to teachers when creating pedagogical activities since, though teachers tend to correct all errors at all levels, they are continually seeking more effective ways to support their learners in writing (Arntzen, Håkansson, Hjedle & Keßler, 2019).

Acknowledgements

I would like to sincerely thank Nancy Gagné (Université TÉLUQ) for her insightful comments, Anne-Sophie Julien (SCS–Université Laval) for her assistance with the statistical component and Emilie Gauthier for the English translation and her comments. Access to the corpus was made possible through the collaboration of the Ramon Llull Institute.

Works Cited

- Alexopoulou, A., 2005, "Aproximación al tratamiento del error en la clase de E/LE desde la perspectiva del análisis de errores", *Estudios de Lingüística Aplicada*, 23 (41), 101-125.
- Alexopoulou, A., 2006, "Los criterios descriptivo y etiológico en la clasificación de los errores del hablante no nativo: una nueva perspectiva", *Porta Linguarum*, 5, 17-35.
- Arntzen, R., Håkansson, G., Hjelde, A., & Keßler, J.-U., 2019, *Teachability and Learnability across Languages*, Amsterdam, Pays-Bas, John Benjamins.
- Astolfi, J.-P., 1997/2015, *L'erreur, un outil pour enseigner*, Issy-les-Moulineaux, France, ESF éditeur, 12^e édition
- Celce-Murcia, M., 2007, "Rethinking the Role of Communicative Competence in Language Teaching", Alcón Soler, E. et Safont Jordà, M. P., *Intercultural Language Use and Language Learning*, Dordrecht, Pays-Bas, Springer, 41-57.
- Conseil d'Europe, 2001, *Cadre européen commun de référence pour les langues*, Paris, France, Didier.
- Corder, S. P., 1967, "The Significance of Learners' Errors", *IRAL*, 5.4, 161-170.
- Dulay, H. Burt, M., Krashen, S., 1982, *Language Two*, Oxford, UK, Oxford University Press.
- Ellis, R., 1997, *Second Language Acquisition*, Oxford, UK, Oxford University Press.
- Ellis, R., 2003, *Task-based Language Learning and Teaching*, Oxford, UK, Oxford University Press.
- Ellis, R., 2008, *The Study of Second Language Acquisition*, Oxford, UK, Oxford University Press.
- Ellis, R. et Barkhuizen, G., 2005, *Analysing Learner Language*, Oxford, UK, Oxford University Press.
- Erdogan, V., 2005, "Contribution of Error Analysis to Foreign Language Teaching", *Mersin University Journal of the Faculty of Education*, 2, 261-270.
- Eurostat, 2016, "New 2016 foreign language learning statistics in the EU", <http://www.tellop.eu/new-2016-foreign-language-learning-statistics-in-the-eu> (checked 13-11-2019).
- Fernández Jódar, R., 2006, "Análisis de errores léxicos, morfosintácticos y gráficos en la lengua escrita de los aprendices polacos de español" (tesis doctoral), <http://www.educacionyfp.gob.es/dam/jcr:1709ed7b-e843-4f54-9413-237adb00b8f9/2007-bv-08-10raulfernandez-pdf.pdf> (checked 13-11-2019).
- Germain, C. ; Netten, J., 2002, "La précision et l'aisance en FLE/FL2 : définitions, types et implications pédagogiques", *La didactique des langues face aux cultures linguistiques et éducatives*, Paris, décembre 2002, http://www.mmecarr.ca/ICF/Paris_Precision-aisance.pdf (checked 13-11-2019).
- Gozali, I., 2018, "Local vs global errors: comprehensibility judgment on the speech of Indonesian students by native speakers", *Language in an Online and Offline*

- World (LOOW) 6: The Fortitude*, Universitas Kristen Petra Surabaya, May 2018.
- Housen, A.; Kuiken, F.; Vedder, I, 2012, *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*, Amsterdam, Pays-Bas, John Benjamins.
- Institut Ramon Llull, (Proves de català, corpus), www.llull.cat (last checked 12-07-2020).
- James, C., 1998, *Errors in Language Learning and Use*, London, UK, Longman.
- Joan Casademont, A., 2020a, "Types d'erreurs chez des francophones qui apprennent le catalan dans un examen de niveau B1 (*Seuil*)", *Íkala*, 25(2), 417-433.
- Joan Casademont, A., 2020b, "Estudio etiológico de los errores del alumnado francófono de catalán nivel B1", *Fuentes*, 22(1), 65-74.
- Joan Casademont, A., Gagné, N., 2020, "Crosslinguistic influences and L3/Ln teaching: A corpus-based analysis of L3/Ln Catalan intermediate learners' errors in writing", *TaLC2020: 14h Teaching and Language Corpora conference*, Universitat de Perpinyà, July 2020.
- Orcasitas-Vicandi, M., 2019, "Lexical crosslinguistic influence in Basque-Spanish bilinguals' English (L3) writing", *International Journal of Bilingual Education and Bilingualism*, 1-11. doi:10.1080/13670050.2019.1701979.
- Rothman, J., 2010, "On the typological economy of syntactic transfer: Word order and relative clause attachment preference in L3 Brazilian Portuguese", *International Review of Applied Linguistics (IRAL)*, 48 (1), 243–271.
- Rothman, J., 2011, "L3 syntactic transfer selectivity and typological determinacy: The typological primacy model", *Second Language Research*, 27(1), 107-127. Retrieved from www.jstor.org/stable/43103852
- Rothman, J., 2014, Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered" *Bilingualism: Language and Cognition*, 18. doi:10.1017/S136672891300059X
- Rothman, J., 2015, "Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered", *Bilingualism: Language and Cognition*, 18(2), 179-190. doi:10.1017/S136672891300059X.
- Selinker, L., 1972, "Interlanguage", *International Review of Applied Linguistics*, 10, 209-231.
- Skehan, P., 1998, *A Cognitive Approach to Language Learning*, Oxford, Oxford University Press.
- Tarone, E., 2015, "Second Language Acquisition in Applied Linguistics: 1925-2015 and beyond", *Applied Linguistics*, 36/4, 444-453.

LINGUISTIC ASPECT	MODIFICATION TYPE	SOME EXAMPLES
Orthographic Errors	Omission	<i>no tan com ella</i> [<i>no tant com ella</i>], <i>t'avindras</i> [<i>t'avindràs</i>].
	Overinclusion	<i>vé</i> [<i>ve</i>], <i>communicat</i> [<i>comunicat</i>], <i>hem sembla</i> [<i>em sembla</i>], <i>climat</i> [<i>clima</i>].
	Misselection	<i>nadar</i> [<i>nedar</i>], <i>la història</i> [<i>la història</i>], <i>avorits</i> [<i>avorrits</i>], <i>descant</i> [<i>descans</i>], <i>busquant</i> [<i>buscant</i>], <i>família d'acoll</i> [<i>família d'acollida</i>].
	Misordering	N.A.
Morphologic Errors	Omission	<i>assembla a un alemany</i> [<i>s'assembla a un alemany</i>].
	Overinclusion	<i>el pis es queda a la segona planta</i> [<i>el pis queda a la segona planta</i>].
	Misselection	<i>sortem</i> [<i>sortim</i>], <i>plujar</i> [<i>ploure</i>].
	Misordering	N.A.
Syntactic Errors	Omission	<i>venir visitar-me</i> [<i>venir a visitar-me</i>].
	Overinclusion	<i>et podria presentar al meu cap</i> [<i>et podria presentar el meu cap</i>].
	Misselection	<i>una per dormir i l'altre (habitació) per estudiar</i> [<i>una per dormir i l'altra per estudiar</i>], <i>la meva company de pis</i> [<i>la meva companya de pis</i>], <i>no he tingut que pagar</i> [<i>no he hagut de pagar</i>].
	Misordering	<i>m'agradaria et veure</i> [<i>m'agradaria veure't</i>].
Lexical-semantic Errors	Omission	N.A.
	Overinclusion	N.A.
	Misselection	<i>família de recepció</i> [<i>família d'acollida</i>], <i>diu acudits</i> [<i>explica acudits</i>], <i>fin de setmana</i> [<i>cap de setmana</i>], <i>el sol baixa</i> [<i>el sol es pon</i>], <i>estic a Londres</i> [<i>soc a Londres</i>], <i>els dies de la setmana</i> [<i>entre setmana</i>].
	Misordering	<i>la vella ciutat</i> [<i>la ciutat vella</i>].

Cohesive and Coherence Errors	Omission	<i>El menjar és molt dolent. La gent és molt acollidora. [El menjar és molt dolent, però la gent és molt acollidora.].</i>
	Overinclusion	<i>Finalment, has de venir a visitar-me. [Has de venir a visitar-me.].</i>
	Misselection	<i>He llogat un pis a l'Eixample. Aquest pis és molt maco i està ben comunicat. [He llogat un pis a l'Eixample. El pis és molt maco i està ben comunicat.], Acabo d'arribar a Rimouski. Vaig mudar-me aquí perquè vaig obtenir una beca a la universitat. [Acabo d'arribar a Rimouski. M'he mudat aquí perquè he obtingut una beca a la universitat.]. De primer, no m'agradava gens. [Primer / Al principi, no m'agradava gens.].</i>
	Misordering	<i>Hi ha pluja i núvols cada dia. La cuina anglesa és molt fastigosa alhora i la gent és una mica esquerpa. [Hi ha pluja i núvols cada dia. Alhora, la cuina anglesa és molt fastigosa i la gent és una mica esquerpa].</i>
Pragmatic Errors	Omission	No header in a letter.
	Overinclusion	Addition of a photograph to a CV (in Quebec).
	Misselection	Use of a formal or informal verbal form when not appropriate in the given communication context. <i>Ben atentament</i> (formal formulation in a letter or email to a friend).
	Misordering	Place and date at the end of a letter.
<u>Typographic Errors</u>	Omission	Digital corpus, space between paragraphs is missing.
	Overinclusion	Unnecessary capital letters.
	Misselection	Digital corpus, non-justified.
	Misordering	N.A.