

32nd International Conference of the Spanish Association of Applied Linguistics (AESLA):  
Language Industries and Social Change

## Mental processes in the oral production of non-native Spanish speakers: pauses and self-correction

Núria Enríquez<sup>a\*</sup>, Lourdes Díaz<sup>b</sup>, Mariona Taulé<sup>c</sup>

<sup>a</sup>*CLiC-University of Barcelona, Barcelona 08007, Spain*

<sup>b</sup>*CLiC- Pompeu Fabra university, Barcelona 08018, Spain*

---

### Abstract

In the field of teaching Spanish as a Foreign Language (SFL), textbooks and teaching materials often provide learners with language samples characterized by a lack of naturalness. We propose the use of a prototypical model of core competence, obtained from the analysis of communicative situations based on real corpora and the comparison of the same type of work with native and non-native speakers. The specific objective is the study of communication strategies related to pauses and self-correction in native and non-native speech, in order to analyse the repair strategies related to language processing

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Universidad Pablo de Olavide.

*Keywords:* Spanish L2 oral data; interactive tasks in L2; Pauses; self-correction, communicative strategies; language processing; L1/L2 corpora

---

### 1. Introduction

In this paper we present a pilot study on processing problem indicators in native and non-native Spanish oral production in a simulated transaction task. Pauses and self-correction are the processing indicators we have chosen for analysis. Normally, we do not pay attention to these indicators outside classroom discourse. However, they can be valuable indicators of processing problems or difficulties when performing a speaking task in the classroom. The specific objective of this work is the analysis of these indicators in oral production as a communication strategy and the differences between native and non-native production. Specifically, we analyse the context in which these pauses and self-corrections appear: before syntax problems (in interrogative sentence modality) and before lexical problems (in figures and numbers production).

\* Corresponding author. *E-mail address:* [nuriaenriquez@yahoo.es](mailto:nuriaenriquez@yahoo.es)

We compare the incidence and differences in pauses and self-correction between trained and untrained learners when carrying out the task, and the differences between these two groups of learners and native Spanish speakers.

Pauses and self-correction have been studied in longitudinal child acquisition studies in the classroom -in a similar way to García Mayo & Gavela, 2001; Perales & Cenoz, 1996- (Muñoz, 2003; Gost & Celaya, 2005), in the adolescent and adult English L2 classroom (García Mayo & Gavela, 2001), and as a translinguistic strategy used by multilingual speakers (Kellerman & Sharwood-Smith, 1986). Additionally, they have been analysed as mechanisms that signal self-editing of language production (repair and communication strategies) as in Piazza (1998). In general, in these works the focus is on collaboration between interlocutors to negotiate the meaning of utterances received or produced, or to perform a task together. In this study, however, we observe that in a task with two well defined and differentiated roles (vendor and customer), participants focus independently on their own production without interfering in the editing or (structural) repairing of the interlocutor's speech. Hence, meaning rather than form is prioritized in production work and interlocutors collaborate at a pragmatic and communicative level, but not at the formal level. In this context, pauses and self-correction are individual indicators of difficulties in language processing, task information management or linguistic content.

The hypotheses are: i) pauses and self-correction are associated with structural difficulties for learners, but not for native speakers, so contextual emergence patterns will be different; ii) familiarity with the situation (the prior preparation of the task) affects the patterns of occurrences of pauses and self-corrections. Therefore, learners with prior preparation should show less difficulty and a different distribution of indicators in their production.

To validate these hypotheses of behaviour in the given communicative situation (simulated commercial transaction), we used data extracted from the FerroviELE corpus (Caballero et al., 2012 and 2013). In section 1, the methodology used in the study is described, namely the corpus used and the categories established. The results and analysis of the data are presented in section 2. Finally, in section 3, conclusions are presented.

## 2. Methodology

This study is based on information extracted from the FerroviELE corpus containing dialogues by native and non-native Spanish speakers in simulated and non simulated train service transactions. The sample contains a total of 24 dialogues (6 native and 18 non-native) in which the interlocutors are adult individuals with similar ages, sexes and socio-cultural levels. All of them participated in a simulated interaction activity. Half of the participants had prepared previously the task (semi-controlled dialogue) while the other half had not.

Table 1: Characteristics of each group

Group	Dialogues	Interlocutors	Type of speech	interlocutor Features	Recordingdate
FerroviELE (No trained USA)	6	12	Semi-controlled dialogue (role-play)	Non-native, American, Young university, both sexes anglophones.	2012
FerroviELE (Trained USA)	6	12	Semi-controlled dialogue, prior preparation (role-play)	Non-native, American, Young university, both sexes anglophones.	2012
FerroviELE (Trained multilingual)	6	12	Semi-controlled dialogue, prior preparation (role-play)	Non-native, multilingual, Young university, both sexes anglophones	2012
FerroviELE (Native simulated)	6	12	Semi-controlled dialogue (role-play)	Natives, different ages, sociocultural levels, both sexes.	2012
TOTAL	24	48			

The learners in the first two groups of the table were English speaking US citizens at B1.2 - B2.1 level of E/LE according to the Common European Framework of Reference (2001). The multilingual group was made up of English, German and Asian language speaking Erasmus students. All were given a contact questionnaire to evaluate the degree of homogeneity in their contact with Spanish outside the classroom. Semi-controlled means that the participants in the interaction had been previously trained in the classroom through a practical simulation with supporting material on rail transport, prices, itineraries, services, payment methods, as well as in the usual manner of speech used in public services. Finally, the native control group was made up of native speakers with no connection with the world of teaching. The purpose of this last group was to be used as a control group to compare the differences and similarities between the interactions of native and non-native pairs, always in artificial simulation situations.

### 2.1. Transcription and encoding of the FerroviELEcorpus

FerroviELE consists of the orthographic transcription of audio files encoded with linguistic and extra-linguistic information in XML format. The transcription and encoding of the dialogues was performed using Transcriber (Barras et al., 2001). This tool allows the alignment between the audio tracks and orthographic transcription and for additional features including the encoding of participants information, communicative functions, and spoken phenomena (difficulty indicators).

Orthographic transcription was carried out following prescriptive standards. Onomatopoeia were also transcribed (um, em, etc.) as well as mispronounced and incorrect words (Eg.: *prefinta cama* instead of *prefiero cama*).

In the encoding process, the following information was tagged: i) the role of each speaker –the customer with the label <C> and the operator with the level <O>; ii) the gender <type: masculine/feminine/unknown>; iii) the language of the interlocutor <dialect: native/non-native> to identify whether it is his/her native or non-native language; and iv) pauses, self-corrections, overlaps and errors.

The Freelinganalyser was used (Padró & Stanilovsky, 2012) for the morphological analysis of the corpus, which was manually validated to ensure the reliability of the analysis. This process allowed the contextualization of pauses and self-corrections.

### 2.2. Pauses and self-corrections

By pause, we mean any silent interruption in spoken discourse, whether it is at the beginning, in middle or at the end of a turn-taking or intervention, regardless of its duration. In this pilot phase, the pauses were not measured. We distinguish between: silent pauses (transcribed with an ellipsis <...>) and filled pauses (transcribed with the corresponding onomatopoeia: <um, uh, uh, uh, uh, oh, uh, puah>). In this context it seemed relevant to start from the previous study of functions performed by Caballero et al. (2012, 2013).

By self-correction, we mean the corrections or rewordings made by learners in their (inter-language) production after making a mistake or what they perceive as a mistake. Like pauses, self-corrections can be located at any level: at the morphological, lexical, and syntactic, as well as in the management of the information or the task.

These mechanisms -pauses and self-corrections- are important because they allow us to appreciate that the student is aware of the grammatical options of the target language and she/he considers the options that exist in it. Among these strategies, self-corrections provide dual information: first, they indicate that learners are capable of language production in the L2 and, at the same time, students used correction methods as input and to reprocess their language production:

Self-corrections are indicative of the difficulties of processing, the stability or instability of the system at that given time and the consistency of the options shown in production (intra-subjective and inter-subjective variability in inter-languages). (Diaz, 2012).

To these difficulty indicators we have to add other hesitation phenomena such as rhythm disturbances and sentence length, which are not being studied at present.

From this point of view, all these phenomena can be considered to be measures of competence and indicators of processing overload or cognitive demand problems.

### 3. Data analysis

One important remark we should make is the difference in the type/token ratio in the productions of the different informant groups, which allows us to contextualize the weight of pauses and self-corrections in the production as a whole (see Table 2).

Table 2: Ratios type/token per group observed

Group	Types	Tokens
Not trained_USA	266	1117
Trained_USA	367	1324
Trained_multilingual	558	3035
Control natives_simulated	369	1221

This first quantitative analysis shows that there is a greater degree of similarity between the native group and the trained\_USA group than with the two other groups in terms of production. One possible explanation for this is that the trained\_multilingual group was part of an Erasmus program in a school of translation and has received intensive language training. All of the members of this group were multilingual speakers with an average of 3-4 languages learned in formal instruction. The way to tackle the production is different in the case of native monolingual speakers unaffiliated with the university.

The group of trained\_USA learners have a less multilingual profile (Spanish was, in most cases, their L2). Their training means that their profile is, in effect, closer to that of the native speaker group.

The language production of the untrained\_USA group is significantly lower because they do not benefit from the activation of knowledge and strategies.

These data are consistent with studies on foreign language learning by bilingual vs. monolingual subjects, according to which bilingual speakers learn faster and better than monolingual speakers (cf. Research in contexts of childhood and/or school bilingualism in Catalonia and the Basque Country, as in Muñoz (2000); Cenoz (2003)).

Regarding the indicators analysed, the following results in table 3 have been obtained:

Table 3: Total indicators

Group	Pauses	Self-corrections
Untrained_USA	99	5
Trained_USA	28	3
Trained_multilingual	138	19
Control native_simulated	12	1
Total	278	28

The figures refer globally to the following types of pauses and contexts:

- Pauses due to problems with the figures: all those that are related to prices or schedules (Eg.: *um ...cinc*).
- Pauses due to syntax problems: mainly errors of agreement and prepositions (Eg.: *um ... esta billete*).
- Pauses due to lexical difficulty: lack of vocabulary or usage errors (Eg.: *eh ... Alar*).
- Pauses before direct interrogative sentences<sup>2</sup> (Eg.: *um...qué precios*).
- Pauses before verbs: we have accounted for all pauses located before any verb form and tense (Eg.: *um ... quiero*).

- Pauses due to lack of information: all dialogues were simulated and none of the subjects actually worked in customer service, therefore all of them could have experienced problems in information management.

In the analysis of the native corpus, another variant has been introduced, which was not found in the learner corpus: the decision variable. Most pauses found in the native corpus appear to respond to a hesitation before making a decision, while in the non-native corpora it is due to the non-understanding of the concept, item or service offered.

In self-corrections, we used the same contextual variables used for pauses. The results obtained for each group is shown in figures 1 to 4.

The total number of pauses in the untrained\_USA group is 99 (Fig. 1). Pause appears more frequently before conjugated verb forms (28.28%). Secondly, 18.18 % of the pauses are caused by problems with syntax and information management.

In the trained\_USA group, the number of pauses is 29 (Fig. 2), well below that of the previous group. Again, we have higher incidence of pauses before verbs (34.48%), and pauses due to problems in information processing and syntax (20.69% each). In the trained\_multilingual group, the total number of pauses is 138 (Fig. 3) and the highest rate of pauses are related to information processing difficulties (31.88%), followed by syntax (26.09%) and conjugated verb (15.22%).

In the native group, there are 12 pauses (Fig. 4), which are due to problems with information management (41.67 %).

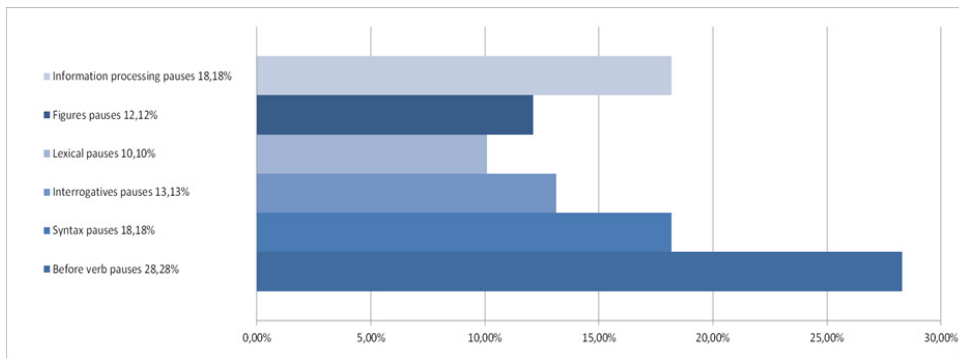


Fig. 1: Pauses in the group of Untrained\_USA learners

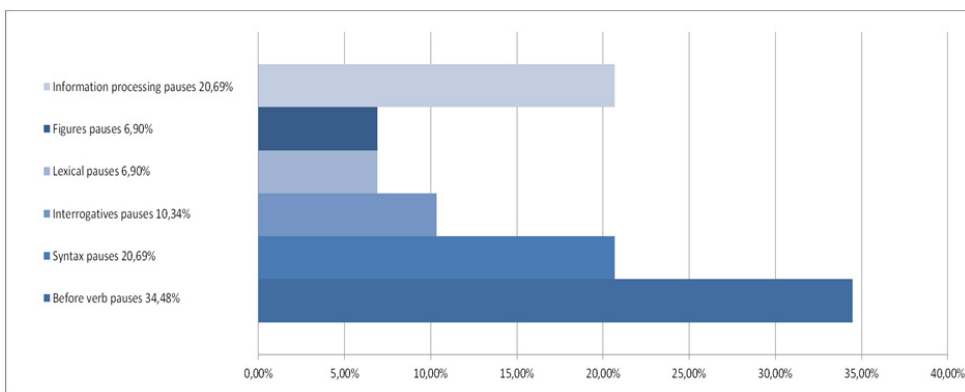


Fig. 2: Pauses in the group of Trained\_USA learners

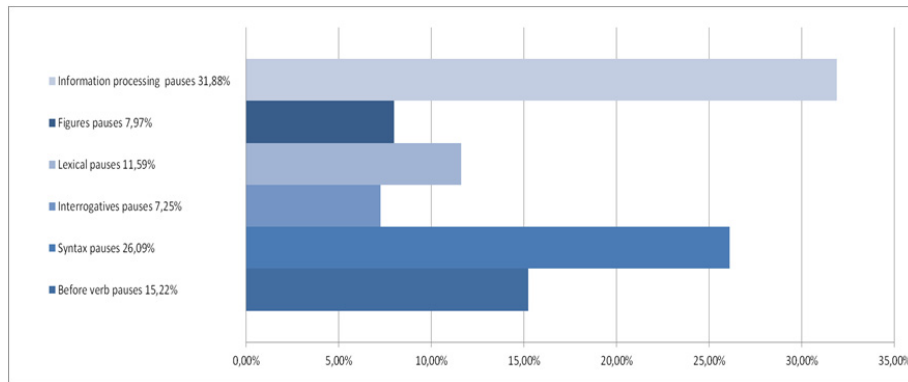


Fig. 3: Pauses in the group of Trained\_multilingual learners

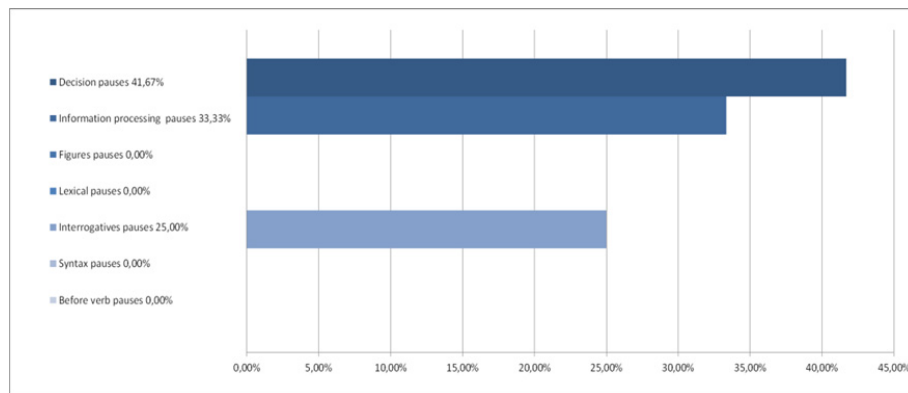


Fig. 4: Pauses in the Control\_native simulated group

Due to the small size of the sample and the low presence of occurrences of self-corrections, this category has been discarded from this pilot study. Table 3 shows the results.

#### 4. Conclusions

This study confirms the role of pauses as indicators of the difficulty of the task itself and the linguistic difficulty in carrying it out in L2, as stated in hypothesis (ii) above. The study also reveals significant differences between native and non-native speakers, as stated in hypothesis (i). Specifically, in native speakers, pauses are indicators of difficulty in management decisions, while in non-native speakers pauses indicate difficulties with syntax, verb inflection and information management. In addition, our study has also revealed that interrogative structures are a problem only in non-native production.

The frequency of self-corrections, in turn, is very low in all the groups studied due to the sample size. Therefore, data collection should be expanded in the future. The new data should allow for exploring separately the interventions of O (operator) and C(customer) along the transactions, in order to achieve a clear-cut distinction between difficulties due to unfamiliarity with the adopted role (seller) and strictly linguistic difficulties.

#### References

Barras, C. et al. (2001). Transcriber: Development and use of a tool for assisting speech corpora Production. *Speech Communication*, 33.

- Caballero, M., Díaz, L. & Taulé, M. (2012). Cómo prever y organizar los contenidos lingüísticos y sociopragmáticos de un escenario comunicativo en clase de E/LE a partir del uso de corpus, In Díaz, L. (Ed.) *Estudios sobre aprendizaje y adquisición de lenguas y sus contextos*, *Monográfico MarcoELE*, 12, 6-26.
- Caballero, M., L. Díaz, M. Taulé (2013). Using Corpora for Modeling Communicative Scenarios: Train Transport Domain in B1-B2 Level Non-Native Spanish Language Learners. *Revista Española de Lingüística Aplicada*, Vol. 26.
- Cenoz, J., C. Hoffmann (eds.) (2003). The Effect of Bilingualism on Third Language Acquisition. *The International Journal of Bilingualism* 7/1.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Díaz, L. & Aymerich, L. (1990). Repeticiones y reformulaciones en el discurso del profesor de ELE. *Cable*, 3: 39-44.
- Díaz, L. (2012). Las puertas al campo de la variabilidad no nativa: vacilaciones y repeticiones de los aprendientes de español L2 ¿indicadores de dificultad en el procesamiento de los pasados aspectuales del español?. In Díaz, L. (Ed.) *Estudios sobre aprendizaje y adquisición de lenguas y sus contextos*, *Monográfico MarcoELE*, Vol. 15, 38 -67.
- Faerch, K. & G. Kasper (1983/1989). *Strategies in Interlanguage Communication*. Oxford: Oxford University Press.
- García Mayo, M.P.& Gavela, M.B. (2001). Repair strategies among advanced EFL. In Muñoz, C. et al. (Eds). *Actas de AESLA 2000*, PPU: Barcelona.
- Johnstone, R. (ed.) (1994). Repetition in discourse. *Interdisciplinary Perspectives*. Vol.1. New Jersey: AblexPub.Co.
- Muñoz, C. (2000) Bilingualism and trilingualism in school students in Catalonia. English in Europe: the Acquisition of a Third Language, In Cenoz, J. & Jessner, U. (eds.). Clevedon: Multilingual Matters, 157–178.
- Padró, L. & Stanilovsky, E. (2012) FreeLing 3.0: Towards Wider Multilinguality, Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA. Istanbul, Turkey.
- Perales, J.& Cenoz, J. (1996). El significado de las pausas en la producción oral en segunda lengua. In Pérez, M. (ed.) *Estudios sobre la adquisición del castellano, catalán, euskera y gallego*. Santiago de Compostela: Servicio de Publicaciones, 689-694.