7th International Conference on Corpus Linguistics: Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond (CILC 2015)

The analysis of subjects in the oral and written production of L2 English learners: transfer and language typology

Sonja Mujcinovic*[1]

University of Valladolid Language Acquisition Lab (UVALAL), Facultad de Filosofía y Letras, 47009, Spain

Abstract

This study considers the oral and written production of L1 Danish L2 English learners and focuses on the analysis of sentential subjects and the issue of transfer. To deal with (the lack of) transfer in typologically close languages, it analyzes the English subjects produced by 20 primary school students divided in two groups depending on the time of exposure. The results show that the subjects produced are both grammatically correct and pragmatically adequate. Therefore, the analysis of transfer should look into language typology as a primary source for transfer, rather than into amount of exposure when typologically close languages are considered.

Keywords: language typology; second language acquisition; transfer; exposure; experimental tests

1. Introduction

The present study focuses on the analysis of sentential subjects and on the issue of transfer within the oral and written production of speakers with Danish as a first language (L1) who are learning English as their second language (L2). Its aim is to account for the importance of language typology, time of exposure and the possibility of transfer when two typologically similar languages are in contact. Furthermore, and given that both written and oral data are elicited (through an oral semi-guided interview and a written picture description task), it seeks to explore how different experimental tasks might influence the L2 English output in a specific grammatical area, i.e. sentential subjects.

2. Background

Previous studies on L2 acquisition have argued that typologically different languages often influence each other resulting in negative transfer (e.g. Odlin 1989, Meisel 2001, Pladevall Ballester 2012). In this respect, transfer is said (1) to have a specific directionality, since structures are often transferred from the L1 into the L2 and (2) to be linked to the amount of exposure so that the more exposure learners have the more native-like production they would have. However, one can hypothesize that, if the L1

Corresponding author: Sonja Mujcinovic. Tel.: 0034 983 423 000;
E-mail address: sonja@fing.uva.es

and the L2 exhibit similar grammatical properties, no transfer is expected in this particular area of grammar and amount of exposure should, therefore, play no role.

Sentential subjects have been widely studied both under linguistic theory (i.e. as in the null subject parameter) and in the case of language acquisition studies, both monolingual and bilingual, and in bilingual acquisition studies, both simultaneous bilingual (L1+L1) and sequential bilingual (L1+L2). Under the null subject parameter (Perlmutter, 1971; Jaeggli, 1981; Chomsky, 1981; and Jaeggli and Safir, 1989, among others), languages are typologically divided into two groups: [+ null subject] languages (i.e. Spanish and Bosnian) and [- null subject] languages (i.e. Danish and English). The examples bellow show how sentential subjects are presented in the different languages.

1. Tom and Mary are friends. **They** always play with a ball. (English)
2. Tom og Mary er venner. **De** leger altid med en bold. (Danish)
3. Tom y Mary son amigos. Siempre **Ø** juegan con un balón. (Spanish)
4. Tom i Mary su prijatelji. Uvjek **Ø** se igraju sa _ loptom. (Bosnian)

In the case of sentential subjects, most works on L2 acquisition deal with transfer between two typologically different languages (e.g. Brice and Rivero, 1996; Gottardo, Siegel, and Wade-Woolley, 2001; Gebauer, Zaunbauer, and Möller, 2013). In the case of L1 Spanish – L2 English, an overproduction of (illicit) null subjects in the L2 has been shown to occur and it has been attributed to the influence of the L1 where null subjects are a "legal" option (e.g. Montrul and Rodríguez Louro, 2006; Montrul, 2010). That is, given that null subjects are a grammatical option in Spanish (example 3), these speakers were transferring this grammatical property into English, a [-null subject] language (as in example 1 above). As a result, the English of these speakers contains a high amount of ungrammatical null subjects.

Although much work has been done on typologically different languages, studies discussing typologically similar languages are scarce (e.g. de Prada, 2009; Filiaci, 2010). Sauter (2002) reviews previous research on the null subject parameter where she argues that most research has actually been conducted on typologically different languages, where a [+ null subject] language is compared to a [- null subject] language and vice versa. Her review points to issues such as the following: (1) L2 Spanish learners start using null subjects before they have mastered the agreement morphology of the verb, (2) L2 Spanish learners from both [+ null subject] L1s and [- null subject] L1s omit both referential and non-referential subjects from the early stages of L2 acquisition and (3) L1 speakers of [+ null subject] languages who are learning a [- null subject] language omit both referential and non-referential subjects, which can be interpreted as transfer of the L1 possibility of null subjects. After looking at the research available, Sauter (2002) seeks to comprehend whether parameter resetting is possible in L2 acquisition and, if that is so, whether parameters actually transfer. The conclusion that she reaches is that the linguistic formulation of the null subject parameter is essential and any proposal related to parameter settings must involve a cluster of properties. In this respect, the few studies discussing typologically similar languages (de Prada, 2009; Filiaci, 2010) combine the basic property of the null subject parameter (i.e. the possibility of null subjects in finite sentences) and other related issues such as the lexico-semantic interface (de Prada, 2009) and pronominal antecedent retrieval (Filiaci, 2010).

De Prada (2009) deals with Catalan and Spanish and she argues that the discourse-pragmatic interface is more vulnerable to interlinguistic influence than the lexico-semantic interface. The main focus lies on the subject position in unergative and unaccusative constructions in Spanish, because in this context the lexico-semantic interface and the discourse-pragmatic interface can be studied. Both Spanish and Catalan are [+null subject] languages and in both the subject position is a syntactic reflex of predicate type and focus. The difference lays in the broad focus of unergative predicates; in Spanish preverbal subjects of unergative verbs are far less preferred than in Catalan. Everything considered, Spanish and Catalan have similar word order preferences seen from the discourse-pragmatic interface but differ in the word order preferences at the lexico-semantic interface. So, de Prada concludes that the results obtained are related to language pair similarity rather than to the vulnerability of interfaces.

Filiaci (2010) deals with the analysis of subjects by considering two closely related languages (Italian and Spanish); that is, languages that have a similar distribution in terms of the null subject parameter, as they are both [+null subject] languages. However, in spite of this similarity, Italian and Spanish exhibit a slight difference in retrieving the antecedent, especially when pragmatic aspects play a role in the interpretation. Filiaci argues that this difference is problematic as it is located at the interface between syntax and pragmatics, an interface that has been shown to be vulnerable when languages are in contact (Hulk and Müller, 2000; Müller and Hulk, 2001; Paradis and Navarro, 2003; Serratrice, Sorace and Paoli, 2004; Serratrice, Sorace, Filiaci, and Baldo, 2009, among others). In her study, she states that the difference between Italian and Spanish lies in the distribution between strong and weak pronouns and, in particular, between the types of antecedents these pronouns can have in the two languages. Both the Italian and the Spanish pronominal systems possess two series of overt pronouns: a strong one (*lui* and *lei* as the Italian 3rd person singular ones; and *el/ella* as the Spanish 3rd person singular ones) and a weak one (*egli/ella* in Italian; and *él/ella* in Spanish). However, while in Italian the weak pronoun can co-refer with discourse antecedents, the strong pronoun cannot. In Spanish the anaphoric behavior of the pronouns *él/ella* could then be understood as an indication of structural deficiency. The author argues for two possible approaches for this cross linguistic difference: (1) cross linguistic differences in the position of the antecedents and (2) cross linguistic differences in the nature of the personal pronouns. The data available show

that these cross linguistic differences in Italian and Spanish cannot be attributed to different syntactic positions occupied by the antecedents so she concludes that further investigation should look into both the Spanish and Italian the pronominal systems to find more conclusive results with respect to the nature of these pronouns.

These few studies about typologically similar languages (e.g. de Prada, 2009; Filiaci, 2010) have been concerned with [+null subject] languages such as Catalan, Spanish and Italian. Given the scarcity of works on typologically similar languages and, in particular, on studies that consider [-null subject] languages, the present study seeks to contribute to fill out this gap by addressing the similarities that English and Danish have regarding sentential subjects and how these similarities can be an important factor when dealing with transfer in oral and written tasks in the L2 English production of L1 Danish children.

3. Participants

The participants that took part in this study were 20 primary school students from Pedersborg Skole in Sorø, Denmark. They were selected on the basis of the time of exposure to English (either 2 or 4 years) and the following requirements: (1) both parents had to be L1 Danish speakers; (2) they should only have received instruction in English in educational settings and (3) they should not have been on a longer stays in an English speaking country. These 20 participants were divided in two groups according to the time of exposure and age (see table 1). All of the participants were asked to complete a background questionnaire which accounts for the above mentioned requirements.

Table 1: Participants

GROUP	PARTICIPANTS	AGE	L1	L2	EXPOSURE TO L2
# 1	10	11	Danish	English	2 years
# 2	10	13	Danish	English	4 years

All the participants from each group have been in the same class and received the same amount of input in English since they started primary school. Taking into account the above mentioned issues, no external factors were expected to influence the English production of these participants as there was a high homogeneity within each group.

4. Experiment

In order to obtain data the participants were asked to complete two types of experimental tests, an oral one and a written one.

The oral test was a semi-guided personal interview conducted in English, where the participants were asked general questions about their interests and free time activities. They were stimulated to talk and if they wanted to speak about other topics than the ones asked by the investigator, they were encouraged to do so. They were also able to ask for vocabulary and were given the required word in the uninflected form. The average time of each interview was between 10-15 minutes. The examples in 5 to 7 are shown in the exact form as the participant uttered them:

5a. **Mary giraffe** fall in love Tom elephant.
5b. **She** was ready to dry him.
5c. She smile and **Ø** say "Thank" to Tom Elephant.

6a. One time **I** go to swimming and one time **I** go to robot club.
6b. **Ø** Love to play together.

7a.   *INV: And you like that book?
        *PAR: I read a book to that's **Septimus Heap**.
        *INV: Septimus Heap?
        *PAR: yeah.
        *INV: what's that book about?
        *PAR: **it's** a fantasy book.

7b. **Mary giraffe and Tom elephant** play made on ball but the ball fall in the water and **Tom elephant** jump in the water and get ball.

7c. Mary giraffe and Tom elephant went to **Tom's house** to play with one of **Tom's new balls HE** got from a birthday present.

The written task was a picture sequence description that was based on the story *A1 Ball* available at The Edmonton Narrative Norms Instrument (ENNI) (Schneider, Dupé and Hayward, 2005). It depicts a story about two characters: Tom Giraffe and Mary Elephant. They story was adapted for this experiment so that the gender differences in Spanish did not interfere with the referential nature of pronouns: in Spanish giraffe is feminine and elephant is masculine. Thus the male Giraffe became Mary Giraffe and the female Elephant became Tom Elephant. The original and the adapted characters for the story appear in figure 1 and figure 2 respectively.
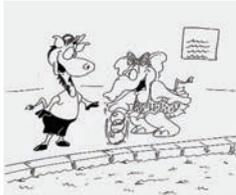

Figure 1: Original ENNI   *Figure 2: Adapted ENNI*

The participants were asked to write the story in English based on the pictures (a total of 5) and, as in the oral task, they were able to ask for vocabulary which was provided to them in an uninflected form. The time available for this task was 1 hour (including the task description and the introduction).

The data from both tasks were later transcribed and codified according to three criteria: (1) form (full DPs, as in 5a, pronouns, as in 5b, or null subjects, as in 5c); (2) grammaticality (correct, as in 6a, or incorrect, as in 6b) and (3) appropriateness in terms of referentiality (DPs for referent introduction, as in 7a, disambiguation, as in 7b, or emphasis and pronouns for referent maintenance, as in 7c).

5. Hypotheses

Given that both languages require their subjects to be overt, that is, they are both [-null subject] languages, null subjects in English are not expected to occur except for the grammatically adequate ones (for example when dealing with coordination as in "Tom jumped into the pool and Ø swam around"). Furthermore, this lack of non-native-like subjects is not supposed to be correlated with exposure so that, regardless of the degree of proficiency in the L2, no such subjects should occur.

6. Results and discussion

The results of this study are illustrated in the following tables (table 2, 3 and 4). The distribution of the overall production of the overt and null subjects in the written and the oral tasks, as indicated in table 2 below, shows a clear preference for the overt subjects (93.4% of all the subjects produced), as expected in a [- null subject] language.

Table 2: Subject production per task

	OVERT			**NULL**	
	Pronouns	DPs	Non- appropriate	Native	Non- native
Written task	38.2%	53%	26.5%	8%	0.8%
Oral task	87.4%	8.2%	0.8%	3.1%	1.3%
TOTAL	93.4%			6.6%	

Although overt subjects are favored in both tasks, a difference appears across tasks in the overt subject type that is favored. There seems to be a preference for the use of full DPs in the written task (53%) whereas there is an even stronger preference for the use of pronouns in the oral task (87.4 %). A low amount of null subjects (both native-like and non-native-like) is produced of which only 2.1% are non-native like structures. These results indicate that the participants under analysis use the different types of subjects correctly in the vast majority of the cases.

Tables 3 and table 4 report the mean number of different subjects and their distribution between the two tasks, written and oral respectively.

In the written task (table 3) a parallel distribution of the different types of pronouns is evident across the two groups. Although, group # 1 produces a lower amount of singular pronouns (22.3%) compared to group # 2 (29.2%), the differences are not significant. These results show a clear preference for overt pronouns as a very low number of null pronouns is produced by group # 1 (5.4 %) and by group # 2 (10.2 %), and out of these less than 1% are non-native like in both groups. A similar parallelism across the groups is also evident in the production of plural pronouns and all the different types of full DPs. Both groups produce a high number of only noun (N) structures (47.3 % by group # 1 and 43.1% by group # 2) which indicates a clear preference for proper names. Those structures where a pronoun should have been used instead of a noun are considered as non-appropriate (as in example 7b above). Also, in this case, there is a similar distribution across both groups. When analyzing the production of the null subjects, group # 1 produces 5.4% native like subjects of which only 0.9% are non-native like. Group #2 produces a little higher amount of native like null subjects (10.2 %) of which only 0.7% are non-native like.

Table 3: Subject production in the written task

	PRONOUNS			FULL DPS				NULL	
	Sing.	*Pl.*	*Non-appropriate*	*N*	*D + N*	*D + A + N*	*Non-appropriate*	*Native*	*Non-native*
Group # 1	22.3%	15.2%	4.5%	47.3%	8%	0.9%	24.1%	5.4%	0.9%
Group # 2	29.2%	10.2%	2.4%	43.1%	7.3%	0	22.6%	10.2%	0.7%
Total	26.1%	12.4%	3.2%	45%	7.6%	0.4%	23.3%	8.0%	0.8%

Table 4 shows the results for the oral task. As in the written task, there is also a clear preference for overt subjects, but in this case the preferences lies in the pronouns and not in the DPs. Group #1 produces 77.6 % singular pronouns and 10.2% plural pronouns of which only 1% are non-appropriate. Group # 2 shows a similar distribution: 74.8% are singular pronouns and 12.2% are plural pronouns of which only 0.4% are non-appropriate. Non-appropriate pronouns, in this case, are considered the ones with no clear referent (e.g. the referent of the pronoun cannot be recovered through context) or confusion in person and number (e.g. the use of a masculine pronoun when referring to feminine one). The production of DPs is very scarce, which was expected due to the type of task conducted. Nevertheless, their distribution is very similar across both groups. The production of null subjects is, as previously mentioned, infrequent and only very few of these null subjects (0.6% for group # 1 and 1.9% for group # 2) are considered non-native like structures.

Table 4: Subject production in the oral task

	PRONOUNS			FULL DPS				NULL	
	Sing.	*Pl.*	*Non-appropriate*	*N*	*D + N*	*D + A + N*	*Non- appropriate*	*Native*	*Non- native*
Group # 1	77.6%	10.2%	1.0%	1.5%	5.3%	2.2%	0.1%	2.4%	0.6%
Group # 2	74.8%	12.2%	0.4%	1.1%	2.3%	3.9%	0.1%	3.7%	1.9%
Total	76.2%	11.2%	0.7%	1.3%	3.8%	3.1%	0.1%	3.1%	1.3%

The results of the oral and written tasks show an overall native-like performance for both groups in the production of overt sentential subjects. However, their distribution is interesting to analyze, because full DPs were mainly used for both referent introduction and reference maintenance only in the written task, whereas pronouns were mainly used for referent maintenance both in the written and in the oral task. There was an important amount of inadequate full DPs (23.3%) in the written task which indicates that this overproduction is probably related to the fact that the participants wanted to avoid ambiguity and thus felt more comfortable using full DPs (53.0%) instead of pronouns (38.5%). This is so, even though the task was adapted in such a way that, if a pronoun was used, there would be no reference problems, as there was a clear distinction between the two characters in the story (Mary Giraffe (she) and Tom Elephant (he)). Thus, this overproduction of full DPs seems not to be related to referent clarification. It could rather be explained in terms of pure redundancy or a task effect that has been recurrent in written tasks where the participants were asked to narrate a story. This effect was not seen in the oral task where the high number of pronouns (87.4%) was expected, because the participants were describing different situations where the pronouns were used for referent maintenance constantly.

Another point that should be considered is the fact that the difference in both the overt and null subject production between the two groups and in both tasks was marginal. Therefore, there was no significant difference between the two groups. The similar production was evident not only in the specific tasks, but also across the different subject types.

Previous studies on L2 acquisition discussing null subject languages have argued for the importance of input and proficiency when dealing with transfer. As an example, Garcia Alcaraz and Bel (2010), who analyze L1 Arabic L2 Spanish speakers (both being [+null subject] languages), argue that the quality and quantity of input is fundamental in language learning, because it endows better competence and performance. However, the present study shows that this might not be the case when

typologically similar languages that are [-null] subject are in contact. This indicates that the time of exposure, and thus the proficiency in the L2, might not be the only factor playing a role in the acquisition of subjects in typologically similar languages.

The low amount of null subjects (8.8% in the written task and 4.4% in the oral task) is clearly related to language typology and typological closeness as well, since both English and Danish are [-null subject] languages with similar overt subject distribution. The high amount of overt subjects (full DPs and pronouns) can and should also be associated to the same fact, i.e. language typology.

Comparing the low amount of ungrammatical structures (0.8%) with the inadequate ones (26.5%), there seems to be a significant difference in the written task. The inadequate structures are typically the repetition of full DPs where a pronoun would have been a more suitable option. The overproduction of full DPs, as mentioned earlier, can be due to the task effect and, in particular, to using the characters' proper names.

The hypotheses that this study analyzed is, therefore, confirmed showing that a reduced number of ungrammatical subjects were produced and that overt subjects are clearly favored. Furthermore, it confirms that transfer is also related to language typology, since there is a homogenous distribution between the two groups, indicating that the time of exposure does not affect the production in this particular area of grammar.

7. Conclusion

What has been argued, when dealing with L2 acquisition, is that the degree of transfer from the L1 into the L2 depends on many factors such as morphological richness and complexity of a language, the sharing or not of the grammatical properties in question, time of exposure and language proficiency. The results of this study show that, when dealing with typologically similar languages, and, in particular, those that are morphologically poor (i.e. [-null subject]), the degree of transfer does not only depend on the factors previously mentioned, but also on language typology as a potential source of positive transfer. In order to support this theory further research must be conducted including similar language pairs and different experimental tasks.

Acknowledgements

References

Brice, A., and Rivero, Y. (1996). Language Transfer: First (Ll) and Second (L2) Proficiency of Bilingual Adolescent Students. *Per Linguam: A Journal of Language Learning, 12 (2)*, 1 - 16.

de Prada Pérez, A. (2009). *Subject Expression in Minorcan Spanish: Consequences of Contact with Catalan*. PhD dissertation. The Pennsylvania State University.

Filiaci, F. (2010). Null and Overt Subject Biases in Spanish and Italian: A Cross-Linguistic Comparison. In C. Borgonovo, M. Espanol-Echevarria, and P. Prevost (Eds.), *Selected Proceedings of the 12th Hispanic Linguistics* Symposium (pp. 171-182). Somerville, MA: *Cascadilla Press*.

Gebauer, S.K, Zaunbauer, A.C.M., and Möller, J. (2013). Cross-language Transfer in English Immersion Programs in Germany: Reading Comprehension and Reading Fluency. *Contemporary Educational Psychology, 38*, 64 – 74.

Gottardo, A. Y., B. Siegel, L. S., and Wade-Woolley, L. (2001). Factors Related to English Reading Performance in Children with Chinese as a First Language: More Evidence of Cross-Language Transfer of Phonological Processing. *Journal of Educational Psychology, 93 (3)*, 530 - 542.

Hulk, A, and Müller, N. (2000). Bilingual First Language Acquisition at the Interface between Syntax and Pragmatics. *Bilingualism: Language and Cognition 3 (3)*, 227 – 244.

Meisel, J.M. (2001). The Simultaneous Acquisition of Two First Languages: Early Differentiation and Subsequent Development of Grammars. In J. Cenoz, and F. Genesee (Eds), *Trends in Bilingual Acquisition* (pp. 11-41). Amsterdam: John Benjamins.

Montrul, S., and Rodríguez Louro, C. (2006). Beyond the Syntax of Null Subject Parameter: A Look at the Discourse-Pragmatic Distribution of Null and Overt Subjects by L2 Learners of Spanish." In L. Escobar, and V. Torrens (Eds.), *The Acquisition of Syntax in Romance Languages* (pp. 401–418). Amsterdam: John Benjamins.

Montrul, S. (2010). Dominant Language Transfer in Spanish L2 Learners and Heritage Speakers." Special Issue of *Second Language Research 26 (3)*, 293 - 327.

Müller, N., and Hulk, A. (2001). Crosslinguistic Influence in Bilingual Language Acquisition: Italian and French as Recipient Languages. *Bilingualism: Language and Cognition 4 (1)*, 1 – 21.

Odlin, T. (1989). *Language Transfer*. Cambridge: Cambridge University Press.

Paradis, J., and Navarro, S. (2003). Subject Realization and Crosslinguistic Interference in the Bilingual Acquisition of Spanish and English: What Is the Role of the Input? *Journal of Child Language 30*, 371 – 393.

Pladevall Ballester, E. (2012). Child L2 English Acquisition of Subject Properties in an Immersion Bilingual Context." *Second Language Research 28 (2)*, 217 - 241.

Sauter, K. (2002). Transfer and Access to Universal Grammar in Adult Second Language Acquisition. *Groningen Dissertations in Linguistics 41*. University of Groningen.

Schneider, P., Dubé, R. V., and Hayward, D. (2005). *The Edmonton Narrative Norms Instrument*. Retrieved from University of Alberta Faculty of Rehabilitation Medicine website: http://www.rehabresearch.ualberta.ca/enni.

Serratrice, L., Sorace, A., and Paoli, S. (2004). Cross Linguistic Influence at the Syntax-Pragmatics Interface: Subjects and Objects in Italian-English Bilingual and Monolingual Acquisition. *Bilingualism: Language and Cognition 7*, 183 – 205.

Serratrice, L., Sorace, A., Filiaci, L., and Baldo, M. (2009). Bilingual Children's Sensitivity to Specificity and Genericity: Evidence from Metalinguistic Awareness. *Bilingualism: Language and Cognition 12 (2)*, 239-257.