Linguistic Development in L2 Spanish: Creation and analysis of a learner corpus

Summary Report

Rosamond Mitchell, University of Southampton
Laura Dominguez, University of Southampton
Maria Arche, University of Southampton
Florence Myles, University of Newcastle
Emma Marsden, University of York

Introduction

This research project on the learning of Spanish as a second/ foreign language ran from April 2006 to March 2008, with funding from the Economic and Social Research Council (Award no. RES-000-23-1609). The project had two main aims: to create a small scale, high quality database of spoken learner Spanish, as a new resource for the study of second language acquisition, and to undertake a short programme of research, using the new database.

Procedures

Classroom learners in schools and universities in England, were recorded speaking Spanish using a series of specially designed elicitation tasks, including storytelling, picture description, discussion and individual interview. There were 20 learners at each of 3 levels: near-beginners (Year 9 students aged 13-14), intermediate students (A2 students aged 17-18), and fourth year university students. All of them were native English speakers. Depending on their level, each learner was audiorecorded undertaking between 3 and 5 oral tasks. They also completed computer based and paper based tasks which provided complementary data on aspects of their Spanish knowledge in particular areas where Spanish works differently from English (knowledge of the pronoun system, knowledge of Spanish word order). For comparison purposes, small numbers of native speakers were also recorded undertaking the same tasks.

The database

The resulting database contains 290 digital soundfiles (240 learner recordings, 50 native speaker recordings). These are accompanied by transcripts in CHILDES (Child Language Data Exchange System) format, which can be analysed with linguistic analysis software (CLAN). Some files have an extra layer of tagging which identifies parts of speech. A project website has been created and the material has been made freely

available in anonymised form through the website, for use by other second language acquisition researchers. The website can be viewed at www.splloc.soton.ac.uk.

The research programme

So far, the research team has investigated the acquisition of two central features of Spanish grammar which differ from English, i.e. word order in sentences (more fixed in English, more variable in Spanish), and the pronoun system (Spanish object pronouns are marked for gender, and generally precede the verb). The third substantive issue investigated to date is the development of Spanish vocabulary.

The results 1: learning pronouns

The investigation of learners' developing control of Spanish pronouns has been based on two tasks (a picture based speaking task, and a computer based comprehension task). The performance of learners at all three levels has been compared on this issue. The results show that beginner and intermediate learners tend to avoid the use of pronouns, preferring to use full noun phrases in their place. However, once they start to use pronouns, learners' usage tends to be accurate, and the advanced learners achieve high levels of accuracy (75 per cent in the production task, 90 per cent in the interpretation task). These findings are important for second language acquisition theory, because they are relevant to a central debate regarding the source(s) of learner errors, and the extent to which learners can ever develop a native-speaker-like grammar system. The findings support the view that learners can and do acquire a correct underlying representation of Spanish grammar, and early errors in their performance are most likely due to other problems such as processing limitations or communicative pressures.

The results 2: Spanish word order

The investigation of learners' developing control of Spanish word order is relevant to another central discussion in language learning theory. Spanish word order is variable, with subjects both preceding and following verbs, in varying circumstances. The different possibilities are controlled by two interacting factors: a) the grammar associated with particular types of verb, and b) the kind of information being conveyed (broad or narrow focus). Different explanations have been advanced previously for the difficulties learners encounter with this system. Does the problem lie in the underlying grammar system, or do the learners have difficulty in distinguishing between different kinds of information to be conveyed, and how these should influence their decisions about word order? The findings from our study suggest that learners go through a stage of overgeneralising just one of the grammatical options available in Spanish, before eventually using the full range of word order possibilities in a native-like way. The learners did not have 'information type' problems, as has sometimes been suggested in the literature.

The results 3: comparing vocabulary development in French and Spanish

Learners' Spanish vocabulary has been investigated through analysis of performance on an oral interview task, which combined picture description and personal conversation. A similar task was used with learners of French recorded in a sister project (details at www.flloc.soton.ac.uk), and research is being undertaken in parallel on both datasets. So far, results have been published regarding vocabulary development from beginner to intermediate level (Year 9 and Year 13 learners). The results provide information of a type not previously available about language learners in the UK educational system. Learner development in the two languages is strikingly similar, with similar numbers of

base words known at the two levels (though Spanish word knowledge was somewhat more diverse). In both languages, substantial gains were made from Year 9 to Year 13, in terms of the numbers of base words known, and also the range of inflections used. This part of the data was also tagged for parts of speech, which allowed further analysis of the word classes used at different levels. The Year 9 speakers' productions in both languages are noticeably noun-heavy, with verbs used proportionally much more by Year 13 speakers. Indicators of more complex language use, such as interrogative and relative pronouns and adverbs, were also more frequent in the Year 13 data. These findings have clear theoretical value (for our understanding of the relationship between vocabulary and grammar development, as well as of vocabulary development itself). They also have clear implications for curriculum design and for effective classroom pedagogy (e.g. the need to focus on development of verb knowledge).

Project dissemination

The corpus is being promoted and findings are being disseminated through a series of conference presentations in the UK, Europe and North America. A successful end of project seminar was also organized in Southampton in January 2008, and attracted researchers from the UK, Spain, Portugal and the USA with whom ongoing networking is planned. A successor project (SPLLOC 2) began in August 2008 and will allow further promotion of the corpus as well as continuation and extension of the substantive research programme.

Acknowledgements

The research team are very grateful to all the Year 9, year 13 and undergraduate learners who contributed speech data to the project. We are also very grateful to the institutions, teachers and other staff who supported the research and allowed us to work with their students.

Contact

For further details of this research including a list of publications to date see
www.splloc.soton.ac.uk .

To contact the research team, write to:

> Wendy Truscott, SPLLOC Project Administrator
>
> School of Humanities/ Modern Languages
>
> University of Southampton
>
> Southampton SO17 1BJ
>
> Email: w.truscott@soton.ac.uk
>
> Phone : 023 80599407