UNIVERSITAT DE BARCELONA

Factors affecting second language communication strategies use and development

Lidia Montero Micharet

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

FACTORS AFFECTING SECOND LANGUAGE COMMUNICATION STRATEGIES USE AND DEVELOPMENT

Tesis doctoral presentada por

Lidia Montero Micharet

como requisito para la obtención del título de

Doctora en Lingüística Aplicada

Programa de Doctorado: Estudios Lingüísticos, Literarios y Culturales

Departamento de Filología Inglesa y Alemana

Universitat de Barcelona

Septiembre de 2015

Directora: **Dra. Raquel Serrano Serrano**

Tutora: **Dra. Raquel Serrano Serrano**

ACKNOWLEDGEMENTS

Many people have contributed throughout these four years in one way or another to the study presented in the next few pages. This dissertation would not have been possible without all of their very generous help and continuous encouragement.

First and foremost, I would like to thank Dr. Raquel Serrano for all these years of my MA and doctoral studies. I cannot miss this chance to express my honest gratitude for her relentless support and impeccable dedication. It has been my pleasure to go on this journey knowing I could count on her attention whenever needed.

I am also truly grateful to Dr. Rosamond Mitchell for granting me the opportunity to learn from her insight and experience, and to renew my motivation among the PhD student community at the University of Southampton. I am likewise indebted to other researchers who shared their deep understanding of SLA and statistics and therefore also made a substantial qualitative difference in this dissertation: Dr. Joan Carles Mora, Antoni Ruiz, Dr. Jaume Llopis, Dr. Rob Schoonen and Dr. Luke Plonsky.

A very heartfelt thanks goes to all of my practically volunteer participants, to the friends and colleagues who accepted coffee and chocolate as payment for their time and effort (Aisha Al-Hafidh, Will Bromberg, Carla Coll-Vinent, Mireia Pagerols, Louise Palmour, Peter Reeves, Elena Safronova and Xinrui Yuan), and to those who bore the brunt of the collateral damage of this study by sharing their home with a PhD candidate (Adnan, Ana, Fer and Raúl).

Y por último y muy especialmente, quiero dar las gracias a mi familia por su apoyo incondicional en todo lo que me propongo.

ABSTRACT

The purpose of this dissertation is to contribute to the field of second language (L2) communication strategies (CS) and, particularly, to examine which factors or combinations of factors are involved in the use and development of effective CS in L2 speakers. According to the review of the relevant literature, there is a series of potentially affecting factors that may influence or be connected to CS use, though their role remains controversial. The research questions in the present study are thus formulated with reference to the following factors: learning context (study abroad or SA vs. at-home courses or AH), attention control, analytic ability, language learning strategies (LLS) and proficiency level. Particular attention is paid to learning context as an affecting factor via an exploration of which of the other factors predict CS effectiveness development in each of the contexts under study. The relevant constructs are measured for 65 adult Spanish intermediate to upper-intermediate learners of English as an L2. Within that sample, 33 learners participated in a SA programme in an English-speaking country for four months, while the remaining 32 took around 15 hours of university courses taught in English in Spanish universities (AH group) over the same time period. The use of CS in English L2 communication was assessed both at the beginning (T1) and at the end (T2) of the testing period in order to account for the effect of learning context on the development of effective CS. Effectiveness in CS implementation was measured by means of a double analysis approach, consisting of an in-situ assessment performed by raters, and the application of a systematic CS effectiveness measurement instrument, the mini-Delphi scale (Montero, Serrano & Llanes, 2013). The results of the present study suggest, first, that there is no significant effect of either SA or AH L2 learning context or of proficiency level on the development of effective CS, at least over a period of four months. Second, attention control, analytic ability and LLS do not seem to be connected to the use of effective CS in L2 communication. The correlation between proficiency level and effective CS use seems

to align with previous studies in indicating that more proficient L2 learners are more effective CS users. Finally, the results indicate that there are certain interactions between learning context and some of the other factors under analysis. More specifically, the results reveal that participants in the AH context with higher analytic ability develop their CS effectiveness to a greater extent than those with a lower analytic ability over a period of four months. Findings in the present study also suggest that SA students with higher attention control tend to develop more effective CS than those with lower attention control, although this connection is not as clear and therefore needs further research. This dissertation makes an original contribution to the CS research area by studying connections that have presumably not been investigated before, such as the possible influence of attention control and LLS on CS use, and also the possible interactions between L2 learning context (both SA and AH) and other factors with an effect on the development of effective CS. Additionally, this dissertation provides further empirical evidence to other controversial or understudied areas, such as the effect of learning context on CS effectiveness development and the connection between analytic ability and effective CS use.

RESUMEN

El propósito de esta tesis doctoral es investigar posibles conexiones entre una serie de factores y combinaciones de estos factores, y el uso y desarrollo de estrategias de comunicación (EC) eficaces para la comunicación en segundas lenguas (L2). Las preguntas de investigación de este estudio se basan en estudios previos que han sugerido posibles conexiones entre el uso de EC y los siguientes factores: contexto de aprendizaje de la L2 (estancia en el extranjero vs. cursos en el país de origen), control de la atención, habilidad analítica, estrategias de aprendizaje de lenguas (EAL) y nivel de competencia en la L2. Se presta especial atención al contexto de aprendizaje al

analizar cuáles de los otros factores predicen el desarrollo de EC eficaces en cada contexto. La muestra seleccionada consta de 65 universitarios hispanohablantes con un nivel de intermedio a intermedio-alto de inglés como L2. Entre ellos, 33 han participado en un programa de intercambio con universidades de países angloparlantes durante cuatro meses, mientras que los 32 restantes han tomado unas 15 horas semanales de cursos universitarios impartidos en lengua inglesa en el país de origen durante un periodo similar. Los resultados de este estudio indican que, aparentemente, no hay un efecto significativo de ninguno de los contextos de aprendizaje observados, ni del nivel de competencia inicial, ni del progreso realizado en el nivel de competencia en el desarrollo de EC eficaces en un periodo de cuatro meses. El control de la atención, la habilidad analítica y las EAL tampoco parecen estar conectadas con el uso eficaz de EC. La correlación entre nivel de competencia en L2 y eficacia en el uso de EC parece señalar, como han hecho estudios anteriores, que los hablantes de L2 con un nivel de competencia más alto utilizan EC con más eficacia. Finalmente, se observa que, en el contexto de cursos en el país de origen, aquellos estudiantes con más habilidad analítica son los que desarrollan EC más eficaces. También se revelan indicios, aunque menos claros, de que los estudiantes en el extranjero con más control de la atención podrían ser los que desarrollan EC más eficaces. Esta tesis doctoral contribuye al área de investigación de EC presentando conexiones presumiblemente no estudiadas antes (efecto del control de la atención y las EAL e interacciones entre factores) y aportando más pruebas empíricas a otras discusiones aún abiertas (efecto del contexto de aprendizaje y de la habilidad analítica).

Table of contents

List of tables

List of figures

1. Introduction

Communication strategies (CS) in second language (L2) communication, particularly oral communication, are an essential element of L2 communicative competence. The term CS is used to refer to the verbal or non-verbal tools L2 speakers turn to in order to compensate for their lexical shortcomings, i.e. to get their message across despite lacking the appropriate vocabulary they would need in a given communicative situation. Some examples of these tools are defining the intended concept, creating a new word based on the L2 speaker's first language (L1), using a lexical label that represents a semantically similar concept, using the corresponding L1 word or abandoning the message. So, if the intended concept were an elephant's trunk, the L2 speaker could attempt to convey meaning by saying "the thing elephants have to throw water", "a trump" (adjusting the pronunciation from the Spanish L1 word "*trompa*"), "a trumpet" (another label that refers to a sort of pipe through which air runs producing a certain sound), "a *trompa*", or "the elephant has a… I don't know. And then…" (thus abandoning the message). All these CS and many others are typically used by L2 speakers in order to attempt to communicate their intended message.

The research area regarding CS use is founded on the notion that these tools are inherent to L2 communication. They represent a key element of communicative competence in that they may enable L2 speakers to enhance their participation in L2 conversation beyond their L2 knowledge by providing them with alternative means to speak their mind. These tools are not exclusive to L2 speakers, though: native speakers use them too. CS are resorted to in situations in which the appropriate lexical label to the intended concept is unknown, a common scenario among L2 speakers, but also when the target word, even though it is stored in the speaker's mental lexicon, cannot be retrieved in a given communicative situation for different reasons, a scenario which applies to both native and L2 speakers. CS in L1 communication are then

typically used to overcome problems of lexical access, not lexical shortage. Therefore, native speakers use CS to a lesser extent than L2 speakers, since they have a more complete vocabulary range and therefore do not encounter as many communication break-downs as L2 speakers do. As Bialystok (1990: 116) stated, “communication strategies are an undeniable event of language use, their existence is a reliably documented aspect of communication, and their role in second language communication seems particularly salient”.

Although CS have already been thoroughly defined and classified (find extensive literature reviews in Dörnyei & Scott, 1997, and Oweis, 2013), the issue of why a particular L2 speaker chooses a specific CS type to bridge a given communicative gap remains unsettled. Corder (1983) claimed that when communicative means and ends do not correspond, L2 learners are left with two options: either adjust their message to the resources available, i.e. adjust their ends to their means (risk avoidance strategies), or alternatively attempt to expand their resources to realize their communicative intentions (resource expansion or risk-running strategies). There is a large repertoire of CS theoretically at the L2 learners’ disposal. Every time they encounter a break-down in communication, they make a choice within that repertoire and select the most appropriate and/or manageable solution to their communicative problem. This dissertation adds to the discussion of what makes learners select a specific CS type to attempt to overcome a communication gap and, particularly, how successful they are in doing so or, in other words, whether they manage to overcome it or not.

In order to shed some light upon these questions, the present dissertation has set as a main goal to identify possible factors or combinations of factors that may have an influence on how effectively L2 learners solve their communicative problems. More specifically, a series of potentially affecting factors have been identified in the literature, namely, language learning context (study abroad, SA, or at-home courses, AH), certain cognitive factors that constitute elements of language aptitude (attention control and

analytic ability), language learning strategies (LLS) and L2 proficiency level. These factors have been said to be somehow connected to the use and/or development of CS in L2 learners. Special attention is devoted to the effect of learning context on the development of effective CS use, hence the analysis of possible interactions between each of the learning contexts studied and all other factors observed. The present study has taken as its purpose to contribute to the disentangling of these connections, which in some cases are only theoretical but in others are the subject of controversy. This work therefore aims to explore how CS types are selected and how their use develops over a period of time and, ultimately, to contribute to the understanding of the linguistic phenomenon of CS in L2 communication.

This dissertation is structured in seven chapters. After this introductory chapter, chapter 2 will deal with the most relevant literature on CS and the potentially affecting factors under study. The chapter is divided into two sections: the first (section 2.1) focuses on the main construct of the present study, L2 CS, including different conceptualisations of CS, their relevance in speech production and underlying cognitive processes, taxonomies proposed to classify CS, the discussion on teachability of CS and some methodological challenges in reference to analysing CS use and CS effectiveness in L2 communication. The second section (2.2) covers the factors that may affect the use and development of effective CS, their definition, and how they have been previously measured and connected to CS in the literature, in addition to other factors that have been claimed to have a connection as well, although they are not examined in the present study.

Chapter 3 will formulate the research questions (RQ) based on the literature reviewed in chapter 2. Five different sections will respectively explain each of the five questions this dissertation attempts to contribute to answering. Each section will also include possible hypotheses and expectations based on previous related studies in the cases in which the relevant connection has a precedent in the literature. Questions 1-4 refer to the respective individual connections between each of the factors selected and

the use or development of effective CS, which means that there is a question on the effect of learning context on the development of effective CS (RQ 1), another question on cognitive factors (attention control and analytic ability, RQ 2), another one on the connection between LLS and effective CS use (RQ 3) and yet another on the effect of proficiency level and progress on effective CS use and CS effectiveness progress (RQ 4). The final RQ, number 5, is aimed at observing any possible interactions between learning context and all the other factors that affect CS effectiveness development.

Chapter 4 will describe the design of the methodology employed in the present study. The first section of this chapter (4.1) will focus on the participants in the study: the different recruitment processes used for the SA group and AH groups, the participants' linguistic profiles according to the information collected by means of initial questionnaires and the data collected on their linguistic experiences during the period spent in each learning context. The second section (4.2) will cover all the measurement instruments implemented, detailing their selection, administration, the type of results they are meant to provide and the data collection procedure followed. Finally, the third section (4.3) will explain the double methodological approach designed to analyse CS effectiveness, including the application of the mini-Delphi scale (Montero, Serrano & Llanes, 2013) and raters' assessment in context.

Chapter 5 will introduce the results of all the statistical analyses performed on the data collected. This chapter will also be divided into three different sections. The first section (5.1) will present the descriptive statistics and normality tests of all the variables involved in the study. The second section (5.2) will include the preliminary analyses performed prior to attempting to answer the RQs, i.e. the triangulation between both CS effectiveness analysis methods and an analysis of the comparability between learning contexts groups, since random assignment of participants to the different groups was not possible. To finish with, section 5.3 will report on the results of the inferential statistics intended to contribute to answering the relevant RQs.

Chapter 6 will offer the discussion on the results obtained and suggest possible interpretations of these results with the goal of formulating more accurate answers to the RQs, based on the particularities of the present study and in contrast with any previous research on each corresponding RQ. The chapter will include a section for each RQ, along with an initial section that will discuss certain validity and reliability issues that should be considered before embarking on the interpretation of the results.

The concluding chapter, number 7, will offer a final reflection on the study as a whole, clarify certain limitations of the methodology and suggest future research to continue the development of this area, as well as suggest some potential implications of the findings for L2 teaching and learning. After these seven chapters, the list of complete bibliographic references and the appendices referred to throughout the dissertation will be presented.

2. Literature review

This chapter covers the most relevant findings of prior research regarding the factors involved in the present study, i.e. communication strategies, language learning context, proficiency, cognitive factors (more specifically, attention control and analytic ability) and language learning strategies. Particular effort will be devoted to the definition of each of these constructs and to possible internal classifications, measurement instruments implemented to operationalise them and any possible connections, both among the factors themselves and between them and other aspects of language and second language learning. The chapter is organised in two major sections. The first focuses on the central object of study, L2 communication strategies. The second presents a series of factors potentially involved in the use and development of effective communication strategies.

2.1. Communication strategies

2.1.1. Definition

Several variants of the term *communication strategies* have been suggested in the literature, including *coping strategies* (Savignon, 1972), *communicational strategies* (Varadi, 1983), *communicative strategies* (Corder, 1983), *compensation strategies* (Harding, 1983) or *compensatory strategies* (Poulisse, 1990). Also, different definitions of CS have been proposed since the time in the early 1970s when Selinker (1972) first used the term and research interest in this aspect of SLA arose. The following table (adopted from Rababah's review article, 2002) and paragraphs present some of the

definitions of CS, formulated from a linguistic perspective, and common aspects found among such definitions.

Table 2.1: CS definitions (Rababah, 2002: 7-8)

Source	**Definition**
Tarone, 1977: 195	"Conscious communication strategies are used by an individual to overcome the crisis which occurs when language structures are inadequate to convey the individual's thought."
Tarone, 1980: 420	"A mutual attempt of two interlocutors to agree on a meaning in situations where requisite meaning structures do not seem to be shared."
Corder, 1981: 103; Corder, 1983: 16	"They are a systematic technique employed by a speaker to express his meaning when faced with some difficulty."
Tarone, 1981: 288	"Learners' attempt to bridge the gap between their linguistic competence in the target language and that of the target language interlocutors."
Faerch and Kasper, 1983: 36	"CSs are potentially conscious plans for solving what to an individual presents itself as a problem in reaching a particular communicative goal."
Wagner, 1983: 167	"Communication strategies predetermine the verbal planning, they serve the function of adjusting the plan to the situation, i.e. each individual utterance is to be seen as strategic. What is specific for IL users is that plans of action cannot be directly converted into verbal plans, because of gaps in the speaker's (and hearer's) linguistic repertoire. The primary function of communication strategies in the speech of IL users is to compensate for this deficit."
Stern, 1983: 411	"Communication strategies, i.e., techniques of coping with difficulties in communicating in an imperfectly known second language."

Harding, 1983: 1	"The domain of compensation strategies must be precisely defined. It is the domain of attempts made by non-native speakers of a language to remedy the disparity that exists between their communicative needs and the linguistic tools at their disposal."
Bialystok, 1983: 102-103	"…all attempts to manipulate a limited system in order to promote communication. Should learning result from the exercise, the strategy has also functioned as a learning strategy, but there is no inherent feature of the strategy itself which can determine which of these roles it will serve."
Poulisse, Bongaerts and Kellerman, 1984: 72; Poulisse, 1990: 88	"Compensatory strategies are strategies which a language user employs in order to achieve his intended meaning on becoming aware of problems arising during the planning phase of an utterance due to his own linguistic shortcomings."
Paribakht, 1985: 132	"Communication strategies (CS) have generally been defined as means that speakers use to solve their communicative problems."
Towell, 1987: 97	"The means used by a speaker to overcome a difficulty encountered whilst attempting to communicate in the foreign language."
Brown, 1987: 180	"The conscious employment by verbal or non-verbal mechanisms for communicating an idea when precise linguistic forms are for some reasons not available to the learner at that point in communication."

As Dörnyei and Scott (1997) pointed out in their review article, since the publication of these initial definitions of CS, researchers have proposed a series of extended views of the same concept: Tarone added an interactional perspective on CS with the idea that meaning-negotiation mechanisms, "intended to clarify intended meaning rather than simply correct linguistic form" (1980: 420), should be included as CS. Canale (1983) suggested that CS may include any attempt to achieve a communication goal, even mechanisms such as slowing speech for rhetorical effect. Dörnyei (1995) included stalling strategies (pause-fillers and hesitation gambits) since

they keep the communication channel open. And finally, Dörnyei and Scott (1995) "conceived CS to be the key units in a general description of problem-management in L2 communication" (Dörnyei and Scott, 1997: 179).

The most obvious aspect all these definitions of second language CS share, with the possible exception of Canale's (1983) generous extension of the concept, is that of problematicity. They all refer to a problem or difficulty in conveying meaning, in transmitting a message, and they assume that since they are considering CS in the context of second language communication the problem must be due to a lack of sufficient or adequate language proficiency. Other recurring factors are consciousness and intentionality (Bialystok, 1990): they imply that the speaker is aware to some extent of the fact that they are choosing an alternative way to express their message, i.e. conscious speech monitoring, and that they select a specific strategy from their available repertoire because they consider it will have the desired effect in solving the perceived communication gap.

Some researchers (Dörnyei, 1995; Dörnyei & Scott, 1995; Haastrup & Phillipson, 1983; Willems, 1987) have deemed it necessary to specify the kinds of problems CS are meant to deal with, and this has taken shape as a classification of problems into three groups, observed in Dörnyei and Scott's review (1997): own-performance problems (e.g. self-repair, self-editing), other-performance problems (meaning-negotiation mechanisms) and processing time pressure (fillers, hesitation devices). Not all CS researchers, however, consider all three groups as meeting their definitions of CS. For example, Tarone (1980) considered certain "processing time pressure" strategies as *production* strategies rather than *communication* strategies, since they are not involved in meaning negotiation. This is the case with the use of fillers and retrieval ("in an attempt to retrieve a lexical item, saying a series of incomplete or wrong forms or structures before reaching it", as defined by Dörnyei and Scott, 1997).

In reference to all three aspects common to the CS definitions, Bialystok (1990) made some observations as well in an attempt to come to a more nuanced definition. Her first comment was that CS can occur in the absence of problematicity and in a non-strategic way in ordinary communication. She tried to illustrate her point with the example of an explanation to replace 'roundabout': 'You take this street to the place where there is a round park in the centre and many roads come together'. From Bialystok's point of view, this utterance would be considered a CS if provided by a non-native speaker, but it should not be since a British native speaker could produce it too when talking to a North American visitor. The example offered, however, can be interpreted as a situation in which the speaker does not actually encounter a communication gap but they *perceive* a *potential* problem, even if such problem may not have occurred had the strategy not been implemented: the British speaker may use that explanation because, based on their knowledge of differences across English geographic variants, they foresee the possibility that the American will not understand 'roundabout', so this would be a strategy to prevent a potential communication gap. Actually, native speakers use CS too, even if they do so at a lower rate as compared to second language speakers, since communication break-downs are not as common when speakers communicate in their first language (Ellis, 1997; Kellerman, 1991; Yule & Tarone, 1990). Therefore, Bialystok's argument on this aspect does not completely rule out problematicity as a CS defining factor, but rather it seems to point to a nuanced observation that the communicative problem does not have to be real, but can merely be perceived as such, and it incidentally corroborates the fact that native speakers use CS to solve potential communicative problems as well.

Regarding the suggested defining factors of consciousness and intentionality, Bialystok suggests that if consciousness were criterial to the definition of CS, then children would be excluded as potential CS users since they are said not to consciously monitor their cognitive processing, although they do seem to resort to the same CS as adults (Bialystok, 1990; Montero et al., 2013). Besides, intentionality

presupposes consciousness, so children would be again excluded from CS use. Intentionality is also questioned because the research exploring links between specific CS and specific communicative situations has shown little proof of systematicity in the selection of CS. On the contrary, Dörnyei and Scott (1995) argue that, in some cases, it is in fact the very presence of consciousness that distinguishes a CS instance from a language production mistake.

Considering the lack of a final consensus on a clear unanimous definition for second language CS, the present study will base its reasoning, design, analysis and interpretation of results on the definition previously proposed in Montero et al. (2013: 2) after reviewing the definitions and criticism explained in this section: “tools implemented intentionally or unintentionally to solve successfully or unsuccessfully [see construct of effectiveness below] a real or a potential communication break-down due to a more or less consciously perceived lack of linguistic resources”. The concept of CS established in Montero et al. and the present piece of research disregards Canale’s (1983) “communication-enhancing strategies” and the strategies in the “processing time pressure” group mentioned above, in accordance with Tarone’s (1980) observation: these CS do not constitute attempts to convey meaning and get the message across to the interlocutor, but rather to keep the communication channel open.

2.1.2. CS as part of speech processing

The definitions considered in subsection 2.1.1 above make use of a linguistic approach to define CS. This subsection looks at how certain researchers (Bialystok, 1990; Poulisse, 1993, Littlemore, 2001) have attempted to explain CS use from a psychological perspective, examining the underlying cognitive processes involved in the use of CS that occur beneath the linguistic surface. Bialystok acknowledged the existence of these two possible approaches to the explanation of CS, one within the theory of communication and based on discourse analysis of conversations, and

another that views “CS as governed by the cognitive mechanisms of language processing” (1990: 116). This subsection presents some examples of the analysis of the language processing behind the use of CS in L2 speech production.

Bialystok connected her previously mentioned reluctance to fully accept traditional conceptualisations of CS with the psychological approach she took to the explanation of CS use, even if the latter approach does not necessarily contradict the former but rather elaborates on it. She divided language processing into two components, namely analysis of linguistic knowledge and control of linguistic processing, and she broke down the strategic use of language into two categories: analysis-based and control-based uses. In her own words, “The analysis-based strategy is an attempt to convey the structure of the intended concept by making explicit the relational defining features. Speakers examine (not necessarily consciously) their symbolic representations of conceptual and linguistic structures in order to select features that will most accurately define the intended meaning” (1990: 133). CS like circumlocution (“exemplifying, illustrating or describing the properties of the target object or action”, Dörnyei and Scott, 1997[1]) or word coinage (“creating a non-existing L2 word by applying a supposed L2 rule to an existing L2 word”) would be included in the category of analysis-based strategies since they entail expressions that include distinctive features of the intended meaning. As for control-based strategies, these are instances in which the intended message remains the same but the means of reference are altered, meaning that the speaker switches his attention from the linguistic system in use and resorts to a different symbolic reference system that may serve the same function. This would be the case of code-switching (“including L1/L3 words with L1/L3 pronunciation in L2 speech”) or mime (“describing whole concepts nonverbally, or accompanying a verbal strategy with a visual illustration”).

[1] All definitions of CS types in this dissertation appear as defined by Dörnyei and Scott (1997). See appendix A.

Poulisse's (1993) proposal, based in turn on Levelt's (1989) model of speech production, understood the cognitive processes underlying CS as three possible mental operations or strategy families: substitution, substitution plus and reconceptualization. "Substitution" entails replacing the intended target language item with another one, either from the L1 or the L2 (code-switching or approximation, "using a single alternative lexical item, such as a superordinate or a related term, which shares semantic features with the target word or structure"). "Substitution-plus" implies the use of an alternative lexical item but with some morphological or phonological alteration (word coinage or foreignising, "using a L1/L3 word by adjusting it to L2 phonology and/or morphology"). And "reconceptualization" means providing information such as the components, use, abilities or location of the intended concept (circumlocution).

Poulisse developed this classification as a reaction to the Nijmegen project's taxonomy (an extensive study on CS conducted by Bongaerts, Kellerman and Poulisse herself; Poulisse & Schils, 1989; Poulisse, 1990). This previous taxonomy also considered the mental processes involved in the use of CS and reduced them to two possible archistrategies: "conceptual" strategies, in which the speaker exploits conceptual knowledge to compensate for the missing lexical item, and "code" strategies, in which the speaker makes use of linguistic knowledge. Conceptual strategies are further classified as "holistic" (comparison-based, like approximation) or "analytic" (description-based, like circumlocution) strategies, while code strategies are divided into "morphological creativity" (word coinage) or "transfer" (foreignising, code-switching) strategies. Poulisse drew on research on the bilingual lexicon (Poulisse, 1993; Poulisse & Bongaerts, 1994) to argue that the Niijmegen taxonomy artificially separated unitary operations, maintaining that in fact choosing the equivalent L1 label for the intended concept (code-switching) or opting for a hypernym in the L2 (approximation) present no substantive processing differences. In Poulisse's words (1997: 63), "analytic strategies have become 'reconceptualization strategies', and both holistic and transfer strategies are now called 'substitution strategies'", although

foreignising, a transfer strategy, would fall under the 'substitution plus' category since it involves alteration of the L1/L3 word.

Littlemore (2001) connected the Nijmegen dichotomy between holistic and analytic strategies, within the conceptual archistrategy, to second language learners' cognitive styles. 'Cognitive style' was understood as "a pervasive attribute, shaping, at a number of levels, the way in which an individual performs any task" and "the way in which they process and retrieve information" (p. 244). Littlemore labelled each participant as displaying either a holistic, an analytic or a neutral cognitive style, determined by their result on Riding's (1991) *Cognitive Styles Analysis* test. Holistic individuals can perceive situations in their overall context, but they find it difficult to identify different components within that situation. Analytics can break information into parts and pay attention to detail but can get an unbalanced view of the situation as a whole because they may focus in on one particular component, disregarding the rest. The hypothesis under study therefore posited that holistic learners would tend to use comparison-based strategies, in which the intended concept is compared to other concepts that share certain features, whereas analytic learners would tend to describe the intended concept itself.

The results of Littlemore's study did in fact contribute to proving the existence of a relationship between the participants' cognitive style and their choices in CS. Learners showed a tendency towards the strategies matching their cognitive style, i.e. holistic learners tended to use more holistic strategies (comparison-based, e.g. approximation), and analytic learners used more analytic strategies (description-based, e.g. circumlocution), even though they all resorted to both kinds of strategies. This kind of findings, the connection between CS types and specific mental processes or even cognitive profiles, raises the issue of whether it is recommendable or even possible at all to teach CS in the L2 classroom. If the types of CS adult learners select depend to some extent on their cognitive styles, assuming their cognitive abilities are already developed and stable, can we teach them how to implement CS to compensate for

their L2 shortcomings? The unresolved debate is a longstanding one. In fact, the issue of the teachability of CS has been addressed by a variety of authors, Bialystok (1990), Kellerman (1991) and Dörnyei (1995) among them. This aspect of CS will be dealt with in subsection 2.1.4 below, after a review of the CS taxonomies proposed in the literature.

2.1.3. CS taxonomies

From the moment when CS started to capture the attention of the research community, not only were different definitions of the concept and analyses of the underlying mental processes published, but scholars also produced several CS taxonomies. Researchers involved in disentangling this part of L2 speech production developed a variety of classifications of CS types following different criteria, according to their linguistic or psychological perspectives, presented in subsections 2.1.1 and 2.1.2 respectively. In other words, each researcher's taxonomy represents their own understanding of what constitutes a CS. The following paragraphs will compare some of the most relevant taxonomies in the literature and then describe the one selected as the basis for the present study.

If we take a close look at some of the published taxonomies and pay attention to the individual CS types considered in them, disregarding the categories used, it is immediately obvious that, even if the wording of the definitions of CS may vary from one taxonomy to another, they clearly focus on the very same L2 phenomenon. The taxonomies reviewed share most of the CS types, which is to say that they identify the same kind of utterances in L2 speech production as CS instances, and they mostly label these utterances with the same CS type, even if slight differences in terminology may be observed across taxonomies. As Bialystok (1990: 61) puts it, "the variety of taxonomies proposed in the literature differ primarily in terminology and overall categorizing principle rather than in the substance of the specific strategies. If we ignore, then, differences in the structure of the taxonomies by abolishing the various

overall categories, then a core group of specific strategies that appear consistently across the taxonomies clearly emerges." The more substantive differences among taxonomies can therefore be found in the categories into which common CS types are classified by different researchers (see tables 2.2 and 2.3 for a summary of the most relevant taxonomies adapted from Dörnyei & Scott, 1997).

Table 2.2: CS taxonomies, adapted from Dörnyei and Scott (1997: 196), part 1

Reduction / achievement			**Source language**
Tarone, 1977	Faerch & Kasper, 1983	Willems, 1987	Bialystok, 1983
AVOIDANCE Topic avoidance Message abandonment PARAPHRASE Approximation Word coinage Circumlocution CONSCIOUS TRANSFER Lit. translation Lang. switch APPEAL FOR ASSISTANCE MIME	FORMAL REDUCTION Phonological Morphological Syntactic Lexical FUNCTIONAL REDUC. Actional reduc. Modal reduc. Reduc. propositional content - topic avoidance - mess. abandonment ACHIEVEMENT CS Compensatory CS - interlingual transfer - IL based CS * paraphrase * word coinage - cooperative CS - non-linguistic CS Retrieval CS	REDUCTION CS Formal reduc. phonological morphological syntactic lexical Functional reduc. Mess. abandonment Topic avoidance ACHIEVEMENT CS Paralinguistic CS Interlingual CS code-switching foreignising Intralingual CS approximation word coinage paraphrase self-repair appeals for assistance	L1-BASED CS Language switch Foreignising Transliteration L2-BASED CS Semantic contiguity Description Word coinage NON-LINGUISTIC CS

As presented in the first three columns of table 2.2, several taxonomies (Faerch & Kasper, 1983; Tarone, 1977; and Willems, 1987) identify a difference between on the one hand strategies that involve adapting one's message to one's resources by changing or reducing the message, and on the other strategies that attempt to transmit the intended message by manipulating the available resources. As a consequence, they split CS types into "reduction" and "avoidance" (or "risk-avoidance") strategies (terminology varies from one taxonomy to another) on the one side, and "achievement" or "risk-taking" strategies on the other side. The former category would include CS such as topic avoidance ("reducing the message by avoiding certain language structures or topics considered problematic languagewise or by leaving out some intended elements for a lack of linguistic resources") or message abandonment ("leaving a message unfinished because of some language difficulty"). The achievement strategies category would then apply to approximation, word coinage, circumlocution and code-switching among others.

Another approach taken to the categorization of CS types refers to the source language upon which strategies are based. Bialystok's 1983 taxonomy (before she developed her cognitive processing model and, therefore, a subsequent taxonomy; see below) is an example of application of this criterion, which has been followed by other subsequent researchers to classify CS instances identified in their data (Ghout-Khenoune, 2012; Liskin-Gasparro, 1996; Rababah, 2002; Rubio, 2007). This taxonomy considered three categories based on this principle: "L1-based strategies", like code-switching or foreignising; "L2-based strategies" (also known as interlanguage-based strategies), such as circumlocution or word coinage; and "non-linguistic strategies".

Table 2.3: CS taxonomies, adapted from Dörnyei and Scott (1997: 197), part 2

Mental processes			**Problem/ solution**	
Bialystok, 1990	Nijmegen	Poulisse, 1993	Dörnyei & Scott, 1995	
ANALYSIS-BASED CS Circumlocution Word coinage CONTROL-BASED CS Code-switching Mime	CONCEPTUAL CS Analytic circumlocution Holistic approximation LINGUISTIC /CODE CS Morphological creativity word coinage Transfer foreignising code-switching	SUBSTITUTION CS Code-switching Approximation SUBSTITUTION PLUS CS Word coinage Foreignising RECONCEP. CS Circumlocution	DIRECT CS Resource deficit CS Mess. abandonment circumlocution approximation code switching Own-performance CS self-rephrasing self-repair Other-perform.CS other-repair INTERACTIONAL CS Resource deficit CS appeals for help Own-performance CS comprehension check own-accuracy check	Other-perform. CS asking for repetition asking for clarification INDIRECT CS Processing time pressure use of fillers repetitions Own-performance CS strategy markers Other-perform. CS feigning understanding

Other researchers have created taxonomies based on language processing theories. They sought to create taxonomies that would meet three requirements: parsimony (minimum number of categories), generalizability (independence of particular languages, tasks and proficiency levels) and psychological plausibility (based on language processing theory). As already commented on in the previous subsection and presented in the first three columns in table 2.3, some of these attempts were made by Bialystok (1990), the Nijmegen Group (Kellerman, Ammerlaan, Bongaerts & Poulisse, 1990) and Poulisse (1993). Their taxonomies reflect their understanding of the mental processes involved in the use of CS. Therefore, as remarked above, Bialystok divided CS types into analysis-based strategies and control-based strategies,

following her model of cognitive language processing. Analysis-based strategies include defining features of the intended concept as a result of the speakers' analysis of its conceptual and linguistic structure, like circumlocution or word coinage. Control-based strategies imply the speaker resorting to a different symbolic system to convey meaning, like code-switching or mime. The Nijmegen Group represented their perception of CS processing by dividing strategies into two archistrategies: "conceptual" strategies, in which the speaker exploits conceptual knowledge to compensate for the missing lexical item, and "code" strategies, in which the speaker resorts to linguistic knowledge. Conceptual strategies are further classified as either "holistic" (comparison-based, like approximation) or "analytic" (description-based, like circumlocution) strategies, while code strategies are divided into "morphological creativity" (word coinage) and "transfer" (foreignising, code-switching) strategies. Later, Poulisse (1993), as mentioned above, put forth a three-category taxonomy that distinguished between "substitution strategies", "substitution-plus strategies" and "reconceptualization", based on Levelt's (1989) model of speech production. "Substitution" strategies, e.g. code-switching or approximation, replace the intended language item with another one, either from the L1 or the L2. "Substitution-plus" strategies, such as word coinage or foreignising, also replace the intended lexical label, but the alternative item involves some morphological or phonological alteration. Finally, "reconceptualization" entails referring to components, use, abilities or location of the intended concept (circumlocution).

Yet another approach to the classification of CS types is the one adopted by Dörnyei & Scott (1995). They grouped CS according to how they contribute to bridging communication gaps in L2 communication, which resulted in three possible categories. The first one was "direct strategies" (alternative means to get meaning across), which comprised most of the CS mentioned so far, like circumlocution, message abandonment, foreignising or mime. The second category was "indirect strategies" (tools to keep the communication channel open), which included CS such as the use of

fillers or feigning understanding. Finally the third category in this taxonomy was "interactional strategies" (cooperative problem-solving), e.g. appeals for help, comprehension checks, guessing. Moreover, within these three macrocategories, CS types were further classified according to the type of problem they were intended to solve: "resource deficit" (language production problem), "own-performance problem" (language production repair or check) or "other-performance problem" (comprehension problem). Interactional strategies also included the subcategory "processing time pressure" (use of fillers, repetitions) instead of the "resource deficit" subcategory. For instance, approximation would be a direct resource deficit-related CS, while self-repair would be a direct own-performance problem-related CS; in the same way, an appeal for help would be an interactional resource deficit CS, whereas an interpretive summary would be an interactional other-performance CS. A similar approach to CS identification and classification according to the type of communication problem being addressed was adopted in Dörnyei and Kormos (1998).

Identification of CS instances in the present study is based on Dörnyei and Scott's (1997) inventory of CS (see appendix A) due to the fact that it presents the most comprehensive selection to be found in the literature. This publication was a review article, and as part of the compilation of all relevant previously published information on CS, the authors developed an inventory that comprised CS types observed in nine different taxonomies without any categorisation (Bialystok, 1983, 1990; Dörnyei & Scott, 1995; Faerch & Kasper, 1983; the Nijmegen Group based on Kellerman, 1991, and Poulisse, 1987; Paribakht, 1985; Poulisse, 1993; Tarone, 1977; and Willems, 1987). This inventory included 33 CS types, plus variants of some of them, each of them illustrated with definitions and examples, along with an indication of which of the seven previous taxonomies considered them to be CS. The fact that it gathers together the taxonomies the most important CS researchers had produced up to that moment has granted it a privileged position among the taxonomies for CS identification that serve as points of reference for other studies in this field (Khan, 2010;

Lafford, 2004; Montero et al., 2013). By choosing this inventory as a reference, possible categorisations of CS types observed in the other taxonomies described above are implicitly disregarded in this piece of research.

2.1.4. Teachability of CS and communicative competence

As mentioned above at the end of subsection 2.1.2, there has been an ongoing debate in the CS research community about whether or not CS may be teachable. Kellerman (1991) distinguished between two different types of CS instruction, namely the *strong* and the *moderate* views of instruction. On the one hand, "the strong view of instruction", which is based on the taxonomic classifications of CS presented in the previous subsection, involves teaching L2 learners how to solve communication problems using CS, i.e. how to paraphrase, coin a new word, transfer, etc. Berry-Bravo's (1993) teaching proposal for circumlocution would fall under this line of action as it is intended to foster the use of CS in L2 communication from the initial level of proficiency. She suggested various practical activities to be carried out in the language classroom which may help students get used to defining the concept they intend to express by identifying its components or its connection to other objects or situations. With regard to this view of instruction, Bialystok (1990) again criticised the taxonomies used in such a methodology as based on language use instead of mental processes and observed that the success of this type of instruction had not been empirically proven. She was sceptical of the potential of teaching CS if it consisted only of presenting lists of possible tools to solve communication problems, without reference to the underlying connection between the problem and the strategy.

On the other hand, Kellerman's "moderate view of instruction" draws on mental operations rather than strategy description and presents strategies generally as possible solutions, so that the learners become aware of the options in their repertoire and the fact that communication break-downs can be solved. After revising both views suggested by Kellerman and contrasting them with her own language processing

model, Bialystok concluded that the way to improve the effective use of CS is to enhance the processing skills that drive them: analysis, through structural information about the language, and control, through practice of speaking, listening, reading and writing. In the end, Kellerman reached a similar conclusion after distinguishing between the two types of instruction.

Not long after this, Dörnyei (1995) presented his article “On the Teachability of Communication Strategies”. First, he commented on the debate as to whether or not CS should be taught: supporting the moderate view of instruction, as mentioned above, Kellerman (1991), Bialystok (1990) and Canale and Swain (1980); and supporting the strong view of instruction, Faerch and Kasper (1983, 1986), Haastrup and Phillipson (1983), Paribakht (1986), Savignon (1972, 1983, 1990) and Willems (1987). In Dörnyei’s opinion, much of the unresolved nature of the discussion was caused by three different facts: arguments on both sides were based on indirect evidence, variation within various CS in terms of their teachability and the notion of teaching itself, which is subject to interpretation (presenting new information vs. making learners practice using CS to solve communication break-downs). Then, in view of this lack of consensus, Dörnyei presented his own study on the matter. He tested the effects of training in the use of topic avoidance and replacement, circumlocution, and using fillers and hesitation devices in a study of 109 teenage English L2 learners’ quantity and quality of CS use, their speech rate and their attitudes towards the training itself and the usefulness of CS. The results suggested that “improvement in the quality of circumlocutions and in the quantity of fillers could indeed be attributed to the treatment; however, the same thing could not be confirmed about the quantity of circumlocutions” (p. 79). The author himself acknowledged that his results were not conclusive and further research was needed, but he still seemed to believe that explicit CS teaching was to some extent possible.

More recent contributions to this aspect of CS are studies like those by Maleki (2007) and Kongsom (2009). Maleki reported on a study in which a CS training

programme was implemented on a group of 30 EFL Iranian students as part of a language course, while a similar sample control group followed a regular language course without any specific focus on CS teaching. The CS taught in this study were approximation, circumlocution/paraphrase, word coinage, appeal for assistance, foreignizing and time-stalling devices. By the end of the course, the treatment group scored better on a Cambridge ESOL speaking test in all test parameters: vocabulary, discourse management, pronunciation and interactive communication. The author concluded from these results that CS training contributes to more effective use of CS and general language learning, since CS allowed learners to keep up a conversation for longer and thus benefit from additional modified input. Similarly, Kongsom's study offered further empirical evidence in favour of CS teachability. In this case it was a 12-week programme involving 12 students. The sample was smaller than Maleki's, but the research instruments implemented were more specifically focused on CS: a self-report strategy questionnaire, an attitudinal questionnaire (learners' attitude towards CS teaching), four different speaking tasks and retrospective protocols. According to the results obtained, participants' use and awareness of CS was increased as an outcome of the training period, they found CS useful and also showed positive feelings towards the instruction programme itself.

To sum up, the literature on a more theoretical level seems to support arguments against the specific instruction of CS, while empirical evidence suggests positive results of such training. Nonetheless, Oxford (2001: 170), although in reference to language learning strategies and not communication strategies, did make a fair point by claiming that "positive results about strategy instruction are pleasing; however, we might not have the complete picture because educational studies reporting ineffective treatments are rarely published." Therefore, it seems safer to accept that the matter remains unresolved.

After reviewing the most relevant literature published about CS, definitions, taxonomies, mental processes involved and teachability, one of the few aspects of this

object of study that seems to be an area of scholarly consensus is these strategies' role in L2 learners' communicative competence. Researchers may not have reached an agreement on exactly which utterances constitute a CS or whether they are teachable or learners should develop them spontaneously, but they do all agree on their existence and salient role in communication and, particularly, in L2 communication (Bialystok, 1990: 116). In fact, communicative competence models such as Canale and Swain's (1980) include strategic competence (ability to transmit information including the skilful use of CS to compensate for communication break-downs) as one of its three components, together with, and at the same level as, grammatical competence and sociolinguistic competence. Similar claims can be found in many other publications (Celce-Murcia, Dörnyei & Thurrell, 1995; Cohen, 1998): CS are essential for L2 learners to prove communicatively competent, even if, based on the findings gathered in this subsection, it is not clear from the literature how they are to be developed by L2 learners.

2.1.5. Methodological aspects of CS analysis

Research in the field of L2 CS has been developing over the last four decades and, apart from the problems with conceptualisation of the construct and possible classifications already covered in previous subsections, certain methodological issues regarding CS analysis must be taken into account before embarking upon the design of studies in this area. In this subsection two specific factors of CS research will be addressed, namely CS elicitation tasks and other instruments for data collection, and measurement of CS effectiveness. In each case, different approaches adopted in previous studies will be presented and compared.

2.1.5.1. CS data collection instruments

As Rababah's review (2002) observed, a variety of CS elicitation tasks have been implemented in previous CS research in order to collect relevant data for analysis, including picture description (Bialystok & Frohlich, 1980), picture reconstruction (Bialystok, 1983), video-taped conversation (Ghout-Khenoune, 2012; Haastrup & Phillipson, 1983), picture-based narration (Dechert, 1983; Montero et al., 2013; Raupach, 1983), instruction (Wagner, 1983) and interview (Lafford, 2004; Raupach, 1983; Rubio, 2007). These are all tasks designed to create a situation in which an L2 speaker may have to resort to CS to solve or avoid communication break-downs, i.e. they are intended to present the speakers with language obstacles in order to observe their reactions and the solutions they come up with to complete the task despite the difficulty encountered. An example of this would be providing the speaker with a comic strip which included difficult elements that are key to the story being represented and asking the speaker to narrate that story.

One possible criticism levelled against some of the CS elicitation tasks used in previous research is, as Rababah (2002) pointed out, that even though all of them successfully elicited strategic behaviour, some did not seem to represent real-life communication. Actually, any situation in which speakers know they are being tested, i.e. any experimental or quasi-experimental study, entails this factor of artificiality. Another point that might constitute a methodological dilemma when it comes to these tasks is that each of them presents different conditions for the speaker to display their repertoire of CS (task complexity and goals, interlocutors, time constraints), which might determine the conclusions to be drawn from the analysis of such speech productions.

Some studies have attempted to investigate how different tasks may elicit different strategic behaviours. Poulisse and Schils (1989) sampled three groups of 15 Dutch learners of English at different proficiency levels and tested them on different

tasks: a picture description task (including a concrete and an abstract picture), a story re-telling task and an oral interview with an L2 native speaker. The study showed that factors such as whether the task elicits a monologue or an interaction, the presence of absence of an interlocutor and whether there is a time restriction to fulfil the task all influenced CS selection. Similarly, DeKeyser (1991) identified different patterns in CS use depending on the task type. Actually, he compared the use of CS between picture descriptions and interviews on a sample of 12 American learners of Spanish (seven of them in a SA programme, five in AH courses) and observed, for example, more prominent use of circumlocution in picture description than was apparent in the interview task. Ghout-Khenoune (2012), in turn, compared the effects of picture description in pairs and free discussion in small groups on the use of CS in a sample of 16 Algerian students of English. This study nuanced that CS elicitation tasks had an effect on the quantity of CS but not on their quality: participants resorted more frequently to CS when performing the discussion as compared to the picture description task, but they used a similar proportion of CS types in the two tasks.

Bialystok (1990) tried to contribute to this discussion by arguing that, in line with her explanation of language processing and cognitive operations behind the selection of CS in L2 speakers, more demanding tasks call for more control-based strategies (code strategies, like code-switching or mime) rather than analysis-based strategies (e.g. circumlocution). This argument was based on Poulisse and Schils's (1989) findings regarding the different effects of a picture description task, a story re-telling task and an oral interview with an L2 native speaker on CS selection. Bialystok claimed that the tasks compared by Poulisse and Schils presented different levels of processing difficulty and that therein may lie the explanation for differences in CS selection patterns. An object description should place a lesser processing burden on the speaker since it entails "one single object, no discourse, no time constraints, no feedback, and no comprehension checks" (Bialystok, 1990: 134). It follows that fewer control strategies would be used to fulfil that task. In contrast, the other tasks should present a

higher processing demand: story re-telling requires memory, translation (participants had heard a story in Dutch and were asked to re-tell it in English) and coherent discourse structure; and interview adds to the burden of story-telling the difficulty of on-line processing and monitoring. This progressive increase in processing demand would imply a progressive increase in the need for control-based CS. Therefore, Bialystok suggested there is a qualitative difference in the data collected through different CS elicitation tasks.

Poulisse (1997), on her part, based her work on the same studies by the Nijmegen Group that Bialystok referred to (Poulisse & Schils, 1989) when she explained variability in CS use among different elicitation tasks from a pragmatic point of view. She drew on two general principles of referential communication: the Principles of Clarity and Economy. According to these principles, when speakers try to refer to a person, object or idea to communicate their messages, they do so by establishing a balance between being as informative as possible (clarity) while reducing the effort to be invested by themselves and their interlocutors to a minimum (economy). From Poulisse's point of view, these principles also apply to L2 communication and in the selection of CS to solve communication gaps: the clearest and most economical way to refer to a table when speaking English would be to say 'table' but, in the absence of the ideal lexical item, CS will be implemented following the same principles of communication.

Based on this assumption and the data collected by the Nijmegen Group, Poulisse stated that the conditions given in different types of CS elicitation tasks determine how informative speakers perceive they need to be and, therefore, how much effort is necessary to communicate their message. For example, in situations like the story re-telling or the interview tasks in Poulisse and Schils (1989), where an interlocutor is present and the speaker gets the chance to negotiate meaning, the interlocutor will provide the speaker with feedback. As a result, when the speaker encounters a communication break-down and they try a CS to overcome it, it will first

be an economical strategy like code-switching or approximation. Then, if the interlocutor shows they understand the intended message, it will mean that the speaker does not need to try to be more informative. Therefore, they will not resort to other more effort-consuming strategies like circumlocution to convey meaning. This kind of less economical CS will only be implemented if the interlocutor shows that they are not receiving the intended message correctly, which notifies the speaker they need to be more informative and invest more effort in order to communicate. However, in situations in which feedback from the listener is limited or non-existent, like the picture description with no interlocutor present tested in Poulisse and Schils (1989), speakers will tend to spare less effort, be less economical, and use as many resources as they have available to ensure the message gets through. As a consequence, CS tasks which involve monologue production or in which the researcher can control how much feedback participants receive should be the ones to elicit the most effort to communicate.

Possibly in view of the absence of a clear final consensus on this aspect too, some researchers have chosen to combine several tasks as a means of data triangulation to observe strategic behaviour from different perspectives and draw more informed conclusions, acknowledging that no instrument is free of disadvantages. This is the case of Khan (2010), who based her analysis of CS use on a combination of three different elicitation tasks: a picture story, an art description and an information-gap task (describing pictures of concrete objects or animals for the interlocutor to decide if they have the same or different pictures to the ones the speaker describes). Results in this study revealed that the information-gap task was the one that elicited the most CS instances in both high and low proficiency groups, since the goal of the task prevented them from avoiding difficult elements and thus forced them to find ways to refer to all the elements in the pictures. This task was adapted and implemented in the present study, but this was done while controlling for the feedback the speaker received, in accordance with Poulisse's (1997) observations presented above

regarding the effect of the interlocutor's feedback on the effort invested by the speaker to communicate their intended message (see 4.2.1).

Khan also contrasted the data from CS elicitation tasks with information gathered by means of questionnaires and stimulated recall. These two instruments constitute a complement to CS task elicitation (Kongsom, 2009; Nakatani, 2010) since they allow researchers to access the intentions underlying speakers' strategic behaviour. Therefore, they address the matter of consciousness, which is one of the problematic aspects of CS conceptualisation, as discussed in previous subsections. Such instruments may also be useful to distinguish between CS types in some cases in which several types may overlap in form even though the process behind them is different. For example, "message replacement", "re-structuring" and "self-repair" may adopt the same form in speech, i.e. a false start of a sentence, while they each constitute different CS types.

This subsection has covered the debate about possible differences among CS elicitation tasks used in this field of research. Most publications seem to agree on the existence of qualitative differences among the types of CS elicited by different tasks. However, other studies have found differences only in the quantity of CS used when comparing a variety of elicitation tasks. Based on Khan's findings and Poulisse's explanation of the Principles of Clarity and Economy and their task-dependent effect on CS use, the present study opted to use Khan's adapted information gap task, which involved picture description. This decision should imply that a higher number of CS instances (though fewer control-strategies, according to Bialystok) would be elicited and that participants would be placed in a situation with only limited feedback from the researcher, so they would feel the need to be as informative as possible.

2.1.5.2. CS effectiveness measurement

In previous sections CS conceptualisation, categorisation, teachability and elicitation for research purposes have been covered. There is extensive published literature on those aspects. Studies looking into CS effectiveness, however, seem to be more limited (Montero et al, 2013). Effectiveness is defined here as the degree of success achieved by L2 speakers in getting their intended messages through to their interlocutors when resorting to CS to solve or prevent a communication break-down. This conceptualisation relies on the assumption that, although the L2 speaker draws on CS when they perceive a certain lack of the necessary linguistic resources to convey their intended message, these might not always be enough to solve the communication gap. A specific CS type might solve the problem completely in some cases (the interlocutor is fully informed of the intended meaning), partially in others (they approach the idea but may not convey the exact meaning) or not contribute to support communication at all (the interlocutor remains uninformed).

The construct of CS effectiveness thus represents an inherently subjective and presumably context-dependent matter, which makes it rather challenging to measure. A few researchers have made attempts at operationalising this aspect of CS use through different approaches. These can be classified into two main groups, depending on the assessment methodology adopted: those regarding the context-dependent nature of CS and those attempting a systematic assessment of CS effectiveness, regardless of context. Haastrup and Phillipson (1983), Paribakht (1984) and Littlemore (2003) would fall under the former category while Bialystok (1983) and Montero et al. (2013) would belong to the latter.

Haastrup and Phillipson (1983) looked into CS use (specifically, achievement strategies) in interaction between eight Danish learners of English and English native speakers. The authors themselves carried out a qualitative analysis on each communication break-down within their collected data to decide whether the

interlocutors reached mutual comprehension or not in each case, which resulted in a dichotomous value, i.e. effective or non-effective. A parallel analysis was performed to classify the achievement strategies observed into L1-based strategies (e.g. foreignising), interlanguage-based strategies (circumlocution, word coinage, all-purpose words), cooperative strategies (help requests), non-verbal strategies (mime) and strategies aimed at solving retrieval problems (strategy markers in Dörnyei and Scott, 1997). Their results seemed to support the hypothesis of a continuum, parallel to language proficiency progress, ranging from mostly L1-based and less effective strategies on one extreme to interlanguage-based and most effective CS on the other. It should be noted here that only *achievement* strategies and not *risk-avoidance* strategies (e.g. message abandonment, omission) were considered in the analysis. In line with Haastrup and Phillipson's findings, other studies have used the categorisation between L1-based and L2 or interlanguage-based CS as a means to assess CS (Rubio, 2007).

Paribakht (1984) for her part based her analysis of CS effectiveness on concept-identification tasks and also focused on achievement strategies only. The sample consisted of 40 Persian students living in Canada at two different proficiency levels of English, plus a group of 20 native speakers of English. The researcher operationalised the construct of effective CS use as "the speed with which subjects could communicate their intended meanings (i.e., the average number of CS used per item)" (p.29). This had as a result a numeric value that was assigned to each communication break-down, implying that the lower the number of CS used, the more effective the speaker would be considered. Therefore, statistical analyses could be carried out. This methodology also implied that individual CS types were not correlated with effectiveness levels, only the amount of CS used to convey meaning of each single concept, so their observations cannot be implemented in further research. The quantitative analysis chapter in this study also included a construct of "success" (parallel to "effective CS use") of the participants in communicating their meanings.

However, the operationalization of this construct is not explained in the publication and the procedure section indicates that interaction between subjects and interlocutors continued in all cases until meaning was conveyed, so the “success” construct does not seem to refer to whether or not participants managed to communicate the intended meaning (identify the relevant concept).

Littlemore (2003) also assessed CS effectiveness in 82 French learners of English. Participants were asked to record themselves naming (or referring to in any other way) a set of 20 concrete items (out of which 15 were actual test items, two were practice items and three were filler items). A picture identification task with no interlocutor present was implemented to elicit CS use. CS instances were identified using an extended version of Poulisse’s (1993) taxonomy. The original taxonomy had encompassed only achievement strategies, classified into three categories: substitution, substitution plus and reconceptualization strategies. Littlemore’s study added to these “functional reduction strategies” (message abandonment and avoidance). CS effectiveness was conceptualised as a three-component construct, including “ease of comprehension”, “stylishness of expression” and “perceived proficiency”. Two raters provided a threefold assessment of each participant’s whole speech production under study (all 15 test-item descriptions together) according to this definition of effectiveness. Individual correlations were carried out between each of these components and each CS type.

Out of the three components of CS effectiveness conceptualised by Littlemore, only “ease of comprehension” seems to align with the concept of CS effectiveness observed in the present study. Ease of comprehension presented a significant positive correlation with substitution and reconceptualization strategies, but not with substitution plus strategies. More specifically, different forms of circumlocution (conventional analogical/metaphoric and literal comparison and description of components, function, activity or place) resulted in the strongest positive correlations, so these would be the strategies that contributed the most to conveying meaning. Functional reduction

strategies, on the contrary, showed a significant negative correlation with ease of comprehension, which would entail that they were obstacles to the raters' understanding the intended meaning. Although CS effectiveness in Littlemore (2003) was measured in context without attempting to create an instrument for systematic assessment, the table with the correlation values between each CS type and ease of comprehension provided in this publication could be implemented in assessing CS effectiveness in future studies.

Bialystok (1983) did aim at a systematic measure of CS effectiveness, even though the author acknowledged that no assessment criteria would apply to all communicative situations. L2 CS use was elicited by means of a picture reconstruction task. A group of 17 target-language native speakers was requested to individually rank order from "most effective" to "least effective" a selection of CS types represented on eight sets of cards grouped by target item. Each card contained a CS type illustrated by an example extracted from the picture reconstruction task. After checking inter-rater reliability, a ranked list of CS effectiveness was developed and, in accordance with Haastrup and Phillipson (1983), it placed most L1-based CS at the "least effective" end of the list, while different forms of circumlocution were deemed the most effective, in line with Littlemore's findings (2003).

Similarly, Montero et al. (2013) developed a systematic instrument for CS effectiveness measurement, based on a comic-strip narration task. In this case, the methodology followed was an adapted mini-Delphi method, i.e. a panel of three L2 native experts was presented with a selection of CS types (based on Dörnyei and Scott, 1997) with definitions and examples and asked to reach a consensus on whether each type would be effective, partially effective or non-effective in referential communication. These three possible labels were translated into numeric values (2, 1 and 0 points respectively), so that the resulting scale (see table 4.8 in subsection 4.3.1) could be implemented to assess CS effectiveness in other similar referential speech productions and result in a numeric value to upon which statistical analysis could be

carried out. In accordance with previous effectiveness assessments, the scale deems circumlocution to be effective and avoidance and L1-based strategies to be non-effective (e.g. message abandonment, code-switching) or only partially effective (e.g. message reduction, foreignising).

After reviewing all three CS effectiveness reference scales in this subsection, Bialystok's (1983), Littlemore's (2003) and Montero et al.'s (2013), some common points can be observed across them. The terminology implemented in each study is slightly different, and Bialystok's and Montero et al.'s definition of CS effectiveness does not include stylishness of expression or perceived proficiency: only ease of comprehension is considered to contribute to *effectiveness* in communication, i.e. to convey intended meaning, by these authors. Therefore, only this component of Littlemore's construct is taken into account in the following observations. All three effectiveness scales place circumlocution or paraphrase among the most effective CS, and they consider code-switching or foreignising ineffective or only partially effective. Bialystok's scale does not include avoidance strategies, e.g. message abandonment, while Littlemore and Montero et al. do and they both deem such CS non-effective. In conclusion, despite the diverse approaches to CS effectiveness assessment (Littlemore judged them in context while Bialystok and Montero et al. were based only on definitions and/or examples), there seems to be a common core of consensus as to which CS types contribute to solving communication gaps and which fail to get the message across. The present study tries to shed some light upon this hypothetical common basis for CS effectiveness in referential communication by implementing a double analysis approach: in-context assessment on the one side and application of the mini-Delphi scale in Montero et al. on the other, as will be further explained in the Methodology chapter.

2.2. Possible factors involved in the use and development of CS

The main purpose of the present study is to look into factors that might be involved in the selection and development on CS in L2 speakers, more specifically, in Spanish learners of English as an L2. The factors under study are learning context (study abroad, SA, vs. at home courses, AH), attention control, analytic ability, language learning strategies (LLS) and proficiency level. Of these potentially affecting factors, learning context and L2 proficiency level had been identified as such in previous publications (Guo, 2011). In fact, Guo concluded that "There is no single factor that has the explanatory adequacy. Therefore, researchers should take as many factors as possible into account when examining the effectiveness of CSs" (p. 99). The remaining factors investigated in this dissertation were selected based on theoretical rationale found in the literature and some connected empirical evidence. This is the case of cognitive factors (attention control and analytic ability) and LLS. Also, some other factors that have been said to explain CS selection in previous research will be reviewed in order to take these findings into account in the interpretation of the results in the present study. This section is divided into subsections to separate the information about each of the factors involved. Each subsection reviews previous literature on the conceptualization of the relevant construct (definition and relevance in L2 acquisition), operationalization (measurement instruments) and potential connection with CS use.

2.2.1. Learning context

The currently growing relevance of SA programmes as part of the economic and cultural globalisation process has given rise to a field of research on the effects of such programmes on SLA. Findings in this area seem to show a positive effect of SA experiences on L2 oral fluency (Freed, Segalowitz & Dewey, 2004; Llanes, 2010; Segalowitz & Freed, 2004) and vocabulary range (Llanes & Muñoz, 2009; Llanes,

2010) but fail to prove a positive effect in other areas of SLA (Llanes, 2011; Sanz, 2014); the existing literature is limited and findings often contradictory.

According to DeKeyser's (2014), Llanes's (2011) and Sanz's (2014) reviews of SA research, in order to draw general conclusions in this field of study, several methodological aspects need to be addressed, including the relevance of the length of stay: most studies focus on semester-long programmes (DeKeyser, 1991; Segalowitz et al., 2004), while in practice most SA programmes are shorter, up to eight weeks in duration (Cubillos, Chieffo & Fan, 2008; Llanes & Muñoz, 2009). According to Llanes, another criticism regarding SA research is the lack of a control group in some studies (Juan-Garau & Pérez-Vidal, 2007; Lennon, 1990) to ascertain whether the progress made is due to the experience abroad in comparison with other learning contexts. On the other hand, both DeKeyser (2014) and Sanz (2014) observe that comparison between SA and AH students is not appropriate because there is no random assignment to the groups and the type of students that choose one context or the other might be different. Other methodological issues in the field are the number of participants in some cases (DeKeyser, 1991: n=12; Lennon, 1990: n=4; Sasaki, 2004: n=11) and their age: most studies sample undergraduate students, while only a few look into teenagers' or children's gains in SA contexts, among them Llanes (2010, 2012), Llanes and Muñoz (2013), Montero et al. (2013) or Serrano, Tragant and Llanes (2014). Yet another gap in SA research refers to the participants' combination of languages, together with their institutional and social context: most studies sample American learners of Spanish or other languages (Freed et al., 2004; O'Brien, Segalowitz, Freed & Collentine, 2007), while again in practice English is the most broadly studied L2 in the world.

One of the aforementioned underdeveloped and contradictory areas within the study of the effects of SA experiences is the possible connection between learning context and the development of L2 CS (Montero et al., 2013). In fact, Canale and Swain (1980) believed that CS were most likely to develop in real-life communication

and not acquired in a classroom environment. The SA context provides learners the opportunity to use language in real communication: they need to manage to communicate their needs in order to fulfil them, and they also need to develop their communicative competence to establish social relationships and to participate actively in their academic or professional environment. Meanwhile, an AH course favours mainly simulated communication, often with other speakers who share the learner's L1, so the situations in which there is a real need to solve communication break-downs are limited. Some of the earliest evidence to be found in the literature of such correlation between the SA context and CS development is Raupach's (1983) study on a group of German learners of French who had spent a term in France: "Whereas the interviews following the stay abroad showed no appreciable progress in the learners' command of grammatical structures, there generally was a considerable change in the use of communication strategies" (p. 207).

Along the same lines, more recent studies such as Lafford (2004) and Rubio (2007) have observed differences in CS use between SA students and AH students. Both studies made use of oral proficiency interviews in order to elicit and analyse CS instances. Lafford (2004) sampled 26 SA and 20 AH American students of Spanish and compared their evolution in CS use over an academic semester. Results indicated a significant reduction of L1-based strategies among SA participants in comparison to AH students. Furthermore, both groups reduced the amount of CS used in the post-test in comparison to the pre-test, which could mean that both contexts contributed a reduction in the amount of communication break-downs, but this decrease was significantly greater in the SA group. Rubio (2007) for his part sampled six SA learners, six AH learners and four heritage speakers of Spanish as an L2 (L1 or dominant language being English in all cases). Again, SA participants seemed to be the group that resorted to L1-based strategies with the lowest frequency, followed by the group of AH students.

In contrast, DeKeyser (1991) maintained that learning context does not affect CS use. This author's publication reported on a previous study (1986) on the differences in CS use observed in a sample of seven SA and five AH American students of Spanish. Participants were asked to perform two different tasks, namely a picture description (monologue) and an interview (interaction). Opposite patterns were identified depending on the task, but the author claimed this was probably due to limitations in task design, i.e. different types of interlocutor and feedback. After reporting on a case-study which included two of the SA participants that showed different communicative behaviour, DeKeyser concluded that individual differences such as personality traits and not learning context were the cause of the patterns observed in CS use. Findings in Montero et al. (2013), based on a total sample of 95 Spanish-Catalan learners of English including both children (SA n=26, AH n=23) and undergraduate students (SA n=22, AH n=24), shed some additional light on this discussion. The study was based on a comic-strip narration task (monologue) performed by all participants at the beginning and at the end of the three-month testing period. In this case, CS use was analysed in terms of communicative effectiveness implementing the mini-Delphi scale developed by the same authors and described in 2.1.5.2 and 4.3.1. The results implied that learning context had an influence on children's development of CS effectiveness but not on adults', at least over a three-month period.

Another point made by DeKeyser (2014), Llanes (2011) and Sanz (2014) is the need for further research to look into a possible interaction between learning context and other factors in SLA, e.g. cognitive abilities. Some work has been done in this line, though with some contradictory results. For example, Segalowitz and Freed (2004) explored the effects of a semester abroad on oral proficiency and oral fluency in particular and the relation between these oral gains and specific cognitive abilities (speed of lexical access, efficiency of lexical access and attention control). The sample consisted of 18 AH and 22 SA English-speaking learners of Spanish. Their findings

showed a certain degree of interaction among oral, cognitive and contextual factors. Sunderman and Kroll (2009), who sampled 48 English-speaking learners of Spanish, also investigated the role of working memory in lexical comprehension and production in connection with learning context (AH plus SA experience, n=14, or only AH experience, n=34). The study concluded that participants under a certain threshold of working memory were unable to benefit from SA in terms of accurate production in the L2. Along the same lines, Tokowicz, Michael and Kroll (2004) looked into the connections among SA, working memory and translation errors. The study sampled 37 participants who were relatively proficient in English and Spanish (15 native Spanish speakers and 22 native English speakers) with varied SA experience. It was concluded that SA experience interacted with working memory in predicting the types of errors made during translation.

In contrast, O'Brien et al. (2007) sampled 18 AH and 25 SA American learners of Spanish and looked into a possible interaction between phonological memory and learning context (AH and SA) in oral fluency development over 13 weeks (a semester). Their results indicated that phonological memory predicted gains in five oral fluency measures irrespective of learning context. Similarly, findings in a recent study by Grey, Cox, Serafini and Sanz (2015) suggested that a five-week intensive SA experience fostered L2 lexical and grammatical development in 26 L1 English advanced L2 Spanish participants and that these gains were independent of variation in cognitive capacity. To sum up, even regarding the interaction of SA and cognitive abilities in the development of oral and lexical gains, the literature presents contradictory findings, which entails that further research is needed, as suggested by DeKeyser (2014), Llanes (2011) and Sanz (2014). In reference to the development of CS effectiveness, a contribution can be made in this respect by investigating a possible interaction among learning context, cognitive abilities and CS development, which might account for the different results reported in the literature regarding the effect of learning context on CS use.

In conclusion, the connection between learning context and CS use remains unresolved due to scarce and contradictory findings in this research area. The explanation behind this contradiction may be found in the interaction between learning context and other learners' individual differences. Therefore, the present study will focus on two aspects. First, the connections between each of the factors under study (learning context, proficiency, attention control, analytic ability and language learning strategies) and the use or development (depending on the case) of effective CS will be analysed. Second, the effect of all the other possible factors, explained in subsequent sections, on the development of CS effectiveness in each of the language learning contexts considered in this piece of research (AH and SA) will be explored. Furthermore, in response to some of the methodological gaps in the literature reviewed, this study will include an AH control group to compare with the SA participants and will sample a total of 65 participants, all of them Spanish learners of English. As DeKeyser and Sanz suggested, there may be differences between those who choose to be AH or SA students. Therefore, the groups will be compared in terms of the other factors measured in the present study (attention control, analytic ability, initial use of LLS, initial proficiency level and initial CS effectiveness; see section 5.2.2) in order to ensure that, at least in those aspects (though others like personality or motivation could play a role), they are not statistically different at the beginning of either context experience.

2.2.2. Cognitive factors (language aptitude)

Several researchers have looked into the cognitive processes behind the implementation of CS to solve break-downs in L2 communication (Bialystok, 1990; Poulisse, 1993), as mentioned in section 2.1.2. Some have even shown that the speaker's cognitive profile or learning style significantly influences the selection of certain types of strategies in a given communicative situation (Littlemore, 2001). The influence of cognitive characteristics on the development of a second language

constitutes the concept of language aptitude, a term used in SLA to refer to the learner's intellectual predisposition to acquire an additional language (Dörnyei, 2005). Since CS are a necessary element in the process of acquiring a language and play a central role in the development of communicative competence, it may be hypothesised that some factors of this language aptitude might be connected to the development and implementation of CS as well as to other aspects of SLA.

Language aptitude has been characterized as composed of various factors. For example, Carroll (1981) structured the construct in four constituent abilities: phonetic coding ability, grammatical sensitivity (which implies analytic ability), rote learning ability and inductive language learning ability. The latter is defined as "the ability to identify patterns of correspondences and relationships involving either meaning or grammatical form" (Carroll, 1973: 8), i.e. it involves, again, the analytic ability to infer grammar rules and semantic relationships, necessary to implement some CS types. More recent literature has focused its attention on working memory as a central and complex component of language aptitude. Working memory is defined as "the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities" (Baddeley, 2003: 189). As Dörnyei (2005: 56-57) explains, it is divided into four subsystems: the phonological loop, the visuospatial sketchpad, the episodic buffer and the central executive. First, the phonological loop is in charge of storing verbal and acoustic information temporarily and translating visual information into phonological form. Then, the visuospatial sketchpad integrates spatial, visual, and kinaesthetic information into a unified representation and translates verbal information into an image-based code. The episodic buffer in turn stores different modalities of information as a single, multi-faceted code or 'episode' and is connected to conscious awareness. Finally, the central executive is responsible for attention control and coordinating and integrating the information from the visuospatial sketchpad and the phonological loop together with long-term memory to perform complex cognitive tasks. The central executive is

involved in planning, decision making and problem solving, a fact that also links directly to the use of CS as problem-solving mechanisms (Dörnyei & Kormos, 1998).

The present study includes two of the language aptitude factors above, namely attention control and analytic ability, and it explores their possible connection to effective CS use and CS effectiveness development in the observed language learning contexts. This choice is based on the observations made in previous research literature on the possible cognitive processes involved in the selection of CS, a reflection of the works reviewed in section 2.1.2 (especially Bialystok's language processing model, 1990) and this extract from Kasper and Kellerman (1997: 6): "An analytic strategy by definition requires that the learner has an explicit understanding of the conceptual features of the intended referent, whereas the decision on the type of strategy to opt for and how to apply CS sequentially and to do all this in a timely and effective manner, are issues of processing control. Learners who use CS efficiently, then, display a high degree of processing control". The following subsections explain attention control and analytic ability and their possible connection to CS in more detail.

2.2.2.1. Attention control

As seen above, attention control (also called attentional control) is a function of the central executive, which in turn is one of the subsystems of working memory, and it has been characterized as playing an important role in SLA due to the fact that this cognitive ability is in charge of monitoring speech (Segalowitz & Freinkiel-Fishman, 2005). Attention control is actually a complex multidimensional construct which encompasses a series of cognitive functions (Baddeley, 1996). Stuss, Shallice, Alexander and Picton (1995, op.cit. Segalowitz & Freinkiel-Fishman, 2005) identified five components of this construct: monitoring, energizing, inhibiting, contention scheduling adjustment and if-then logic control. These five components combine in different ways to perform in "attention-demanding situations, including sustaining

attention in slow-changing situations in which vigilance is required, concentrating attention during fast paced, highly demanding activities, sharing attention when different cognitive activities must be executed at the same time, suppressing attention when inappropriate action schemata are automatically activated, and shifting attention focus when a complex activity frequently presents changing demands" (Segalowitz & Freinkiel-Fishman, 2005: 645).

Other research in this field (Baddeley, 1996; Miyake, Friedman, Emerson, Witzki & Howerter, 2000; St. Clair-Thompson & Gathercole, 2006) has put an emphasis on the relevance of three key executive functions of attention control, namely "updating", "inhibition" and "shifting". The report by St. Clair-Thompson and Gathercole (2006: 746) explains these three functions: shifting would involve "moving backwards and forwards between multiple tasks, operations or mental sets"; updating means monitoring and coding incoming information and revising information stored in working memory to replace no-longer-relevant information with relevant information; and inhibition implies deliberately inhibiting dominant automatic responses.

The instruments used to measure attention control are as varied and complex as the construct itself. In fact, different attention control tasks can tap into one or several functions within the attention control ability, although the most commonly explored, and therefore measured, in SLA are shifting and inhibition. It is also worth mentioning that attention control tests can be either language-based or non-language-based, and both types have been utilized in SLA studies. Some examples of attention control measurement instruments implemented in SLA research are the Trail Making Test (Bialystok, 2010; Isaacs & Trofimovich, 2011), which is non-language based and is designed to assess attention shifting, planning and inhibition; the Dimensional Change Card Sort Task (Bialystok & Martin, 2004), which tests inhibition and is also non-language-based; the metalinguistic categorisation task implemented by Segalowitz and Freed (2004), which is, as its name implies, language-based and measures shifting ability; the Stroop test (Zied et al., 2004), which is language-based and tests

inhibition; and the Flankers task (De Leeuw & Bogulski, 2015) and the Simon task (Cox, 2013), which measure inhibition and are non-language-based.

As a speech monitoring device and part of language aptitude, attention control has attracted particular interest from the SLA research community over the last few decades. Segalowitz and Freinkiel-Fishman (2005) claimed that all the attention-demanding situations mentioned above (Stuss et al., 1995) are part of the skilled use of language. Evidence of possible relationships between attention control and different aspects of SLA can be found in the literature. For instance, Segalowitz and Freinkiel-Fishman (2005) investigated the connection between attention control and general proficiency, operationalised in the study as efficiency of lexical access. They sampled 16 undergraduate L1 English L2 French speakers. The analyses performed indicated that attention control accounted for 59% of variance in proficiency and L2 attention control alone predicted 32%. Another example was provided by Goo (2012), who connected attention control with the effect of corrective feedback. This study compared the effect of recasts and metalinguistic feedback in the acquisition of the *that*-trace filter. The sample consisted of 83 Korean learners of English divided into two experimental groups (one per type of feedback) and one control group. The results indicated that attention control predicted significantly beneficial effects of recasts (implicit feedback) but not of metalinguistic feedback (explicit). Also, as mentioned in section 2.2.1 above, the interaction between attention control and L2 learning context in the development of oral fluency has been looked into by studies like the one reported in Segalowitz and Freed (2004).

The selection and implementation of CS in L2 communication might be another aspect of SLA with a certain relationship with attention control. This can be inferred from theoretical explanations of the language processing mechanisms behind CS use included in Bialystok (1990) and in Dörnyei and Kormos (1998). In fact, when Bialystok (1990) explained her language processing model, constituted by analysis of knowledge and control of processing, and her classification of CS into analysis-based and control-

based, she specified that the key component of processing control and control-based strategy use is selective attention, since the speaker moves their attention focus from the linguistic system in use to a different symbolic reference system that could serve the intended function. This statement can be interpreted as an indicator that not just attention control in general, but specifically attention shifting, might be one of the factors affecting CS selection.

It also bears mentioning that problem solving is among the functions of attention control (Dörnyei, 2005) and, based on the fact that CS are, by definition, a repertoire of tools to solve communicative problems, there seems to be a direct logical connection between attention control capacity and the selection of the appropriate communicative tools. Indeed, Shatz (1978) described referential communication as a complex cognitive problem and using strategies as a problem-solving activity. The author also mentioned, in reference to the cognitive aspect of using CS, that success in doing so can be predicted according to the speakers' "information processing capacity". Dörnyei and Kormos (1998) looked into the problem-solving mechanisms involved in L2 communication, which may be lexical, grammatical or phonological and work towards overcoming communication gaps due to lack of linguistic resources, which are characteristic of L2 speakers' speech production. The same publication raised the question of what could make the speaker choose one mechanism or another. Although later studies have looked into the connection between other cognitive factors (specifically, holistic and analytic cognitive styles) and CS selection (Littlemore, 2001), to the researcher's knowledge, no previous study has empirically connected attention control and CS use or CS effectiveness.

In conclusion, according to the theory that has been presented in this section, it can be speculated that attention control and CS selection could be somehow connected. The present study devotes part of its attention to the presumably unprecedented empirical correlation between attention control (specifically, attention shifting, as suggested by Bialystok, 1990) and CS effectiveness, in an attempt to

contribute to answering Dörnyei and Kormos's question. Additionally, its possible interaction with learning context (SA and AH) in L2 learners' development of such effectiveness will be observed.

2.2.2.2. Analytic ability

The term "analytic ability" refers to a cognitive factor that has been proven to be involved in language aptitude and that, as in the case of attention control, also contributes to problem solving (Grañena, 2013). The SLA research community has paid attention to this cognitive ability because it is thought to be responsible for L2 learners' grammatical sensitivity and inductive learning ability (Skehan, 1998) and related to verbal aptitude in general and even intelligence (DeKeyser & Koeth, 2011). More specifically, analytic ability seems to be involved in tasks like inferring grammatical rules and semantic relationships (Carroll, 1973).

Measurement of analytic ability in SLA studies has been attempted by means of several different instruments, often included as a subtest within a battery of tests intended to measure overall language aptitude. Examples of analytic ability measurement instruments include the Words in Sentences MLAT subtest, grammaticality judgment tests (GJT) and composites of tests that represent analytic ability. The Words in Sentences MLAT subtest (DeKeyser, 2000; Johnson & Newport, 1989) is a language-based test (it can be administered either in the test-takers' L1 or their L2) that measures grammatical sensitivity as a predictor of grammar learning. Each test item requires the test-taker to identify a word in a given sentence that bears the same syntactic function as a specific word in a previous sentence. DeKeyser (2000) also implemented a GJT, adapted from Johnson and Newport (1989). The new version consisted of 200 items and included four practice items at the beginning. Each item presented the test-taker with a sentence to judge as grammatically correct or incorrect.

Finally, an example of a composite of cognitive abilities that represent analytic ability can be found in Grañena (2013). This study looked specifically into the LLAMA aptitude test battery developed by Meara (2005). The LLAMA set is a language-based test battery but, unlike the Words in sentences or GJT presented above, it is L1-independent, i.e. items are based on British Columbian and Central-American indigenous languages, unlikely to be known to test-takers. It consists of four tests aimed at measuring four different components of language aptitude: LLAMA B is a vocabulary learning task, LLAMA D is a test of phonetic memory, LLAMA E is a test of sound-symbol correspondence and LLAMA F tests grammatical inferencing. In Grañena's report (2013), the results showed that three of the subtests of the battery measured the same underlying aptitude, which was interpreted by Grañena as analytic ability: the vocabulary learning test (LLAMA B), the sound-symbol correspondence test (LLAMA E) and the grammatical inferencing test (LLAMA F). However, Grañena (p. 199) added to this results report that LLAMA F was "the strongest loading on the component interpreted as aptitude for explicit learning".

Studies in the field of SLA research have provided results that indicate a connection between learners' analytic ability and a series of aspects of language learning. These aspects mostly revolve around the area of grammar learning and error correction, but findings about the correlation between analytic ability and general L2 proficiency have also been published. DeKeyser (2000) replicated the findings in Johnson and Newport (1989). The study, based on a sample of 57 Hungarian learners of English, connected analytic ability to adult learners' possibility to achieve near-native L2 competence, while analytic ability seems not to be such a predictor of children's ultimate attainment. Li (2013), whose study sampled 78 American learners of Chinese, analysed the interactions between implicit and explicit feedback and analytic ability (measured with the Words in Sentences test) and working memory. The results of this study showed that analytic ability is predictive of the effects of implicit feedback, while working memory seemed to predict the effects of explicit feedback. Yilmaz (2013)

carried out a study similar to Li's, this time on the role of analytic ability and working memory on explicit and implicit feedback, with a sample of 48 English-speaking beginning learners of Turkish. Participants were divided into two experimental groups (explicit feedback and recasts) and a control group (no feedback). In this case, the results seemed to indicate that analytic ability as measured by the LLAMA F test, as well as working memory, predicted the effect of explicit feedback but not the effect of recasts. Finally, Roehr (2008) sampled 60 English-speaking undergraduate learners of German to test their metalinguistic knowledge expressed as error correction and to explore the connection with analytic ability. A principal components analysis indicated that analytic ability and learners' ability to correct, describe and explain highlighted L2 errors may constitute components of the same construct.

The possible need for analytic ability in the selection and implementation of CS in L2 communication has also been suggested in previous publications. As mentioned previously, Bialystok (1990) regarded CS as either analysis-based or control-based in accordance with her theoretical language processing model. Analysis-based strategies entail that whenever the most direct label for a concept is unavailable (i.e. when facing a communication gap), the speaker provides distinctive information about the intended concept in order to attempt to get the message through to the interlocutor despite the lexical shortcomings. Bialystok argued here that the success of such a strategy "depends on the extent to which the concept is represented as analysed knowledge" (p. 132). According to her explanation, speakers analyse (although not necessarily consciously, as remarked in section 2.1.1) their previous knowledge of conceptual and linguistic structures and select features that most accurately represent the intended concept. Similarly, Kasper and Kellerman (1997) claimed that the implementation of certain CS (analytic strategies) required the learner to analyse and explicitly understand the conceptual features of the intended referent. As examples to illustrate how this ability connects to specific CS types, implementing "word-coinage" implies explicit knowledge and manipulation of L2 morphological rules, while "approximation" is

not feasible without analysing the semantic relationships or common features between the intended concept and other concepts to find a different yet related lexical item in the target language. Therefore, analytic ability is at work when such CS are employed in L2 communication, but probably not when other CS like message abandonment or code-switching are selected to solve the communication break-down.

Beyond the theoretical framework supporting a connection between analytic ability and CS use, Littlemore (2001, 2003) provided empirical evidence on this hypothesis. The study published in 2001 drew on the Nijmegen group taxonomy of strategies. As reviewed in section 2.1.2, this taxonomy classified CS into two possible archistrategies, based on the mental processes involved in their use: "conceptual" strategies, in which conceptual knowledge is exploited to compensate for lexical shortcomings, and "code" strategies, in which linguistic knowledge is applied. Conceptual strategies were further divided into "holistic" (comparison-based) or "analytic" (description-based) CS, while code strategies were classified as "morphological creativity" or "transfer" CS. In Littlemore's study, only conceptual strategies were considered, and it was hypothesised that, if analytic and holistic conceptual strategies reflected different cognitive processes, then they would be associated with different cognitive styles. Cognitive style was understood in these studies as the way in which a person processes information and completes cognitive tasks, taken as a reflection of their cognitive abilities. Littlemore measured participants' cognitive tendencies by means of Riding's (1991) *Cognitive Styles Analysis*, a computer-based test, and divided them into three groups according to their results: analytic (n=20), holistic (n=28) and neutral (n=34). Patterns in L2 CS use were therefore observed in participants with either an analytic or a holistic cognitive style, and the individuals did in fact show a tendency towards higher frequency of use of the strategies matching their cognitive profile. Littlemore (2003) in turn examined CS effectiveness and compared cognitive styles regarding effective CS use. Results indicated that participants with an analytic cognitive style, which in this publication was

labelled *ectenic* cognitive style, proved to be more effective in their use of CS than those with a holistic (re-named as *synoptic*) cognitive style. It should then follow that L2 speakers with greater analytic ability tend to choose more effective CS.

To sum up, both theoretical and empirical information in the literature seem to indicate a connection between analytic ability and certain patterns in CS use, more specifically the selection of CS that imply grammatical sensitivity or association between semantic features. Littlemore (2003) even suggests that this connection may have an influence on CS effectiveness. Therefore, the present study will correlate CS effectiveness with a measurement of analytic ability. Additionally, the effect of interaction between analytic ability and learning context will be tested.

2.2.3. Language learning strategies

The research area of language learning strategies (LLS) seems to share several features with that of CS. In fact, the two concepts have even been found to overlap to some extent (O'Malley & Chamot, 1990; Oxford, 1990). Dörnyei (2005) observed the lack of a unanimous definition of LLS proposed in the literature, which constitutes the first feature LLS and CS have in common, but Oxford's definition (1999: 518) may be taken here as a starting point: "specific actions, behaviors, steps, or techniques that students use to improve their own progress in developing skills in a second or foreign language. These strategies can facilitate the internalization, storage, retrieval, or use of the new language". In other words, LLS refer to the learners' attitude towards learning the language, i.e. their thoughts and actions with regards to improving their L2 proficiency. Cohen (1998) specified that learners *consciously* select these learning processes.

LLS have been also categorised in taxonomies in a similar way to CS. A well-known example of these taxonomies was published by Oxford (1990) and consisted of two macrocategories, direct and indirect strategies, each of them including three different strategy groups. An overview of this taxonomy is offered in table 2.4. O'Malley

and Chamot (1990) proposed a very similar taxonomy that only differed in that it included memory strategies under the same category as cognitive strategies (the former being a subclass of the latter) and grouped compensation, affective and social strategies under a broader category named "social/affective strategies". Dörnyei (2005) maintained that, although CS contribute to keeping conversation open and therefore grant L2 learners more opportunities for language acquisition, they are related to language *use* and not language *learning* and that these concepts should thus be kept separate. In addition, Hsiao and Oxford's (2002) factor analysis indicated that O'Malley and Chamot's taxonomy gained in explanatory power if social and affective strategies were split into separate categories.

Table 2.4: Oxford's (1990: 18-21) LLS taxonomy

Direct strategies	**Indirect strategies**
Memory strategies: storing and retrieving new information (e.g. creating mental linkages, applying images and sounds). Cognitive strategies: improving understanding and production of the L2 (analysing and reasoning, creating structure for input/output). Compensation strategies: overcoming deficiencies in the L2 (guessing meaning from context, CS).	Metacognitive strategies: controlling one's own cognition (planning and evaluating learning). Affective strategies: regulating emotions, motivations and attitudes (lowering anxiety, self-encouragement). Social strategies: interaction with others in L2 (asking questions, cooperating with others, cultural awareness).

Based on these publications and comments, Dörnyei (2005: 169) summarised a final taxonomy consisting of four categories: cognitive strategies, which involve "manipulation or transformation of the learning materials/input (e.g., repetition, summarizing, using images)"; metacognitive strategies, or "higher-order strategies

aimed at analyzing, monitoring, evaluating, planning and organizing one's own learning process"; social strategies, or "interpersonal behaviors aimed at increasing the amount of L2 communication and practice the learner undertakes (e.g., initiating interaction with native speakers, cooperating with peers); and affective strategies, or "taking control of the emotional (affective) conditions and experiences that shape one's subjective involvement in learning".

The use of LLS has been most commonly measured by means of Likert-scale questionnaires on which learners self-report on the frequency with which they implement each of the strategy items on a given inventory list. Two of the most recurrent measurement instruments in the LLS literature are the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia & McKeachie, 1991) and the Strategy Inventory for Language Learning (SILL; Oxford, 1990). The first encompasses both LLS and motivation represented in 81 items, 50 of which refer to LLS, and measured on a seven-point scale, whereas the SILL has two different versions, one including 80 LLS items for English speakers learning other target languages and another one with 50 items for learners of English as an L2, both of them based on a five-point scale. Other examples of LLS questionnaires can be found in Cohen and Chi (2002, the Language Strategy Use Inventory and Index or LSUII), Tseng, Dörnyei and Schmitt (2006, the Self-Regulating Capacity in Vocabulary Learning Scale or SRCvoc) and Tragant and Victori (2012).

The questionnaire developed by Tragant and Victori, which consisted of 55 LLS items, was later reduced by implementing an exploratory factor analysis and item analysis and the resulting 17-item scale was validated (Tragant, Thompson & Victori, 2013), so it was a much shorter test but remained equally appropriate to measure LLS based on learners' self-reporting. Questionnaire items are grouped according to their goal: improve general level of English, vocabulary learning, studying grammar, reading and writing. The results of the analysis suggested a two-factor structure which reflected "skills-based deep processing strategies" and "language study strategies". Examples of

skills-based deep processing strategies would be ‘When I see short texts in English, I try to figure out what they mean’ or ‘I review what I have written carefully’ (i.e. reading, writing and some of the improving general level of English LLS), while LLS like ‘I review what we have done in class or I test myself on my own’ or ‘I write summaries or outlines of the structures that we are learning in class’ (i.e. vocabulary learning, studying grammar and other improving general level of English LLS) were included in the language study cluster.

As one would imagine, the use of LLS has been studied in previous SLA research in connection to L2 proficiency (Ying-Chun, 2009; Wong & Nunan, 2011) as a way to investigate which LLS are the most effective ones, i.e. presumably the LLS used by successful learners should be the most useful in order to enhance L2 learning. Wong and Nunan (2011) sampled 110 undergraduate students in Hong Kong and connected their LLS use to their scores on a standard exam administered at the end of secondary school. The most popular LLS among successful learners (with a higher score on the test) were ‘I like to learn by watching/listening to native speakers’, ‘I like to learn English words by seeing them’ (this one was also popular among non-successful learners though), ‘At home, I like to learn by watching TV in English’, ‘In class, I like to learn by conversation’ and ‘I like to learn many new words’. In turn, Ying-Chun (2009) observed a sample of 418 EFL learners in Taiwan that took the EFL version of the SILL and concluded that the effective use of learning strategies may be influenced by cognitive abilities, the ones included in the present study in particular serving as possible factors promoting effective CS use. In fact, LLS that required attention control or analytic ability, i.e. cognitive and metacognitive strategies such as arranging and planning one’s learning and using analysing and reasoning skills, were employed by the most proficient learners, along with practicing pronunciation and speaking. Conversely, the least proficient seemed more often to make use of social and memory strategies.

Regarding the possible overlap between LLS and CS, apart from the LLS taxonomies that include CS (O'Malley & Chamot, 1990; Oxford, 1990), more recent studies have also suggested that in some cases the two concepts may occupy some common territory. LePichon, de Swart, Vorstman and van den Bergh (2010) investigated the effect of language learning context (formal instruction of additional language vs. acquiring two languages simultaneously before age 4 in a non-formal context) on strategic use of language in general, including both communication and learning strategies together, in a sample of 101 children. The results indicated that the children that had learned an additional language in a formal context used more frequent and diverse strategies. As seen in section 2.1.4, Maleki (2007) also found this overlap of the two concepts. This study, which tested the effect of a CS training programme on a group of 30 EFL Iranian students, concluded that CS training contributed to general language learning. CS helped learners to keep up a conversation for longer and thus benefit from additional input, meaning that using CS would qualify as an LLS. Similarly, Khan (2010) considered CS and LLS as separate concepts in general, but also made the subtler point that CS overlapped with LLS oriented toward improving speaking skills.

To conclude, previous researchers have considered LLS and CS as a single object of study (LePichon et al., 2010; O'Malley & Chamot, 1990; Oxford, 1990) and both of them are expressions of strategic behaviour in L2 learners, which implies that the cognitive processes involved in CS use may be somehow related to the processes behind LLS selection, as Ying-Chun suggested. Based on these findings and theoretical framework, a possible connection between LLS and CS will be considered in the present study. Additionally, an analysis of LLS, or in other words of learners' general attitude towards the learning process and the means at their disposal to expand their L2 knowledge may provide added information about extra-curricular contact with the language to be considered when looking into the learning context experience.

2.2.4. Proficiency level

Language proficiency can be defined as a speakers' "capacity to speak or perform a linguistic task (reading, writing or speaking) in an acquired language" (Anderson, 2012: 30). The construct of language proficiency has been measured with a wide variety of instruments in the field of SLA. Some pieces of research employ instruments that include all four language skills (speaking, listening, reading and writing) to determine the language level attained by the relevant participants, e.g. the International English Language Testing System or IELTS (Bahrani & Tam, 2012; Storch & Tapper, 2009) and the Test of English as a Foreign Language or TOEFL (Gobel & Kano, 2014; Tercanlioglu, 2004); while others implement a partial measure with the intention of representing general proficiency. These can include measures of grammar knowledge (Wayne, 2006) or vocabulary size (Nation & Beglar, 2007).

Publications like Milton's (2013) focus their attention on proving that one of these partial measures actually correlates with more complete measures, thus providing a validated solution for the measurement of language proficiency in a more cost-effective manner in further research. Specifically, Milton reviewed three different studies that show that vocabulary size correlates with other forms of language performance: First, Staehr (2008) tested the connection between vocabulary size and the skills of reading, writing and listening in Danish secondary-school learners of English as a foreign language. Vocabulary size was measured by means of the Vocabulary Levels Test (VLT), which assesses receptive knowledge of word meaning at the 2,000, 3,000, 5,000 and 10,000 levels of word frequency. Reading and listening were measured with multiple-choice tests and writing scores were based on an essay task. Binary logistic regressions showed that VLT scores explained 72% of reading variance, 52% of writing and 38% of listening.

Second, Milton, Wade and Hopkins (2010) correlated vocabulary size as measured by the X_lex (Meara, 2005; further explained in section 4.2.3) and A_lex

software (Milton & Hopkins, 2006) and all the IELTS sub-scores in 30 students at the intermediate and advanced level from different nationalities attending a pre-sessional course in the UK. Both the X_lex and the A_lex measure receptive vocabulary knowledge of up to 5,000 vocabulary items, though in the former the words are written while in the latter they are heard by test-takers. Each vocabulary size score was correlated with the relevant written or oral measures. The results showed statistically significant correlations between the A_lex and speaking scores (*r*=.71), and between the X_lex and reading and writing scores (*r*=.70 and .76).

Finally, Schoonen (2010) tested the influence of a combination of vocabulary size, depth and fluency measures to explain variation in performance in the language skills. Vocabulary size was measured using the VLT, speed of word recognition, and word retrieval was measured with computer-delivered tests, while questionnaires, grammar tests and reading and writing measures from previous studies were used for the other factors involved. The results of the vocabulary size and speed measures indicated significant positive correlations with reading and writing scores. Vocabulary size became a stronger predictor when combined with speed or fluency. As a conclusion from these three example studies, when investigating possible connections between proficiency and other aspects of SLA, measuring vocabulary size can be a less fine-tuned but still a rather practical way to collect data representative of overall language proficiency.

Among other aspects of SLA with connections to language proficiency, previous research has listed the use of CS. Examples supporting this connection can be found in Bialystok (1983), Haastrup and Phillipson (1983), Liskin-Gasparro (1996) and Paribakht (1985), all of which report on a parallel development of these two aspects of second language use, qualitative CS use and overall proficiency. Bialystok (1983) studied CS used by 16 high-school and 14 adult English-speaking learners of French in a picture description task. From the teenage group, six learners were in a higher level class than the rest, and all the adults were considered to have a more advanced level

of proficiency than the teenagers, so three groups of different proficiency levels were sampled. Apart from the systematic measure of CS effectiveness created in this study (as described in section 2.1.5.2), CS use was quantified and CS instances were classified into L1-based and L2-based strategies. No quantitative differences were observed in overall CS use across proficiency groups, but more advanced learners did make use of L2-based strategies more frequently and were more effective in their CS selection than lower level learners. As mentioned above in the discussion of CS effectiveness measures (section 2.1.5.2), Haastrup and Phillipson investigated the use of achievement CS in interaction between Danish learners of English and native speakers. Their results seemed to illustrate a continuum, parallel to increasing mastery of the target language, ranging from mostly L1-based and less effective strategies to interlanguage-based and more effective CS.

For her part, Liskin-Gasparro analysed CS use in oral proficiency interviews administered to two groups of American learners of Spanish in a summer immersion programme: the intermediate-high group with 17 participants and the advanced group with 13. CS instances, including both achievement and avoidance CS, were classified into L1-based and L2-based strategies. The results in this study seemed to indicate that proficiency level and frequency of CS use (particularly L1-based CS) are inversely proportional. The advanced participants resorted to L2-based CS significantly more than intermediate students when faced with communication gaps. Similar results were obtained in Paribakht's study (1985), which compared intermediate and advanced Persian students of English as an L2. Both groups used both L1-based and L2-based strategies, but higher level students tended more toward the use of L2-based CS while lower level learners resorted to L1-based CS more frequently.

Only minor contradictions or small adjustments have been proposed to this hypothesis. Liskin-Gasparro (1996) herself clarified that the inverse correlation found in her study only exists in intermediate to advanced learners, i.e. speakers that are not at either extreme of the proficiency continuum, since beginners lack the necessary

resources to implement CS and near-native speakers do not face as many communication break-downs, so they should not need them much more than a native speaker would. Also, Poulisse and Schils (1989) not only compared the effect of different CS elicitation tasks on CS use, as seen in 2.1.5.1, but also the effect of proficiency. The study sampled three groups of 15 Dutch learners of English at different proficiency levels and tested them on different tasks: a picture description task (including a concrete and an abstract picture), a story re-telling task and an oral interview with an L2 native speaker. Apart from the effect of differing task conditions, the results revealed that most proficient students used fewer CS than the least proficient, while the type of CS did not seem to be related to proficiency level. In other words, the differences found by Poulisse and Schils were only quantitative, not qualitative as in Bialystok (1983), Haastrup and Phillipson (1983), Liskin-Gasparro (1996) and Paribakht (1985).

Finally, Khan (2010) analysed CS use across three different elicitation tasks: a picture story, an art description and an information-gap task involving picture description, as mentioned in section 2.1.5.1, but she also investigated differences across proficiency levels. CS were identified and classified into compensation, interactional or metacognitive strategies. This CS identification was triangulated with the participants' self-reports on a questionnaire and stimulated recall sessions. The results indicated that task conditions exert more of an influence on CS than proficiency does. In general, few significant differences were found between high and low proficiency groups. The differences found were mostly in the use of compensation strategies. Regarding this type of CS, observations on patterns in CS use coincide with the generally accepted hypothesis in the case of picture story and art description tasks: lower level participants encountered more lexical problems and therefore used CS more frequently. However, the results indicated that differences across proficiency level groups in the information gap (picture description) task were not as remarkable as in the other publications here reviewed. According to Khan, this was probably due to the

fact that the information gap task was specifically designed to elicit compensation strategies, so the low level group maintained as much of a need for compensation strategies as they had had in the other tasks, while the high group increased their use of such CS. Regarding qualitative differences in CS use, the low group resorted to code-switching and clarification by code-switching (both L1-based CS) more than the high group, whereas the high group used restructuring (L2-based CS) and message abandonment more often. Only the message abandonment finding seems to contradict results of previous studies.

To sum up, the connection between overall language proficiency and the use of CS seems to be mostly agreed upon in the research community. In the present study, a proficiency measure was included in the original design only as way to help ensure comparability among participants. However, the opportunity will also be taken to look into this connection between proficiency and effective CS use in the hopes of contributing to the discussion. It must be noted here that not all the studies mentioned above observed CS effectiveness. In fact, only Bialystok and Haastrup and Phillipson did, whereas Liskin-Gasparro and Paribakht observed only source language of the CS implemented (L1/L2-based) and Khan and Poulisse and Schils compared CS types according to other categorisations. The present study will attempt to provide further empirical evidence on the CS effectiveness and proficiency level parallel continuum. This hypothesis might suggest that CS effectiveness develops spontaneously as proficiency increases, or perhaps it just implies that implementation of the most effective CS types, like circumlocution, entails higher language mastery than less effective CS. In addition to that, the effect of interaction between initial proficiency level and learning context on the development of CS effectiveness will also be examined.

2.2.5. Other factors

The research literature with regards to CS use has highlighted other factors that might affect CS selection on the part of L2 speakers. Guo (2011) reviewed a series of

potentially influential factors on CS use, which included learning context and L2 proficiency level, but also the speaker's gender, personality, L1 and age. The following paragraphs will cover some publications that have been found to shed light upon such possible connections with CS use. Even though the factors included in this subsection will not be tested in the present study, findings from other research may supplement and help interpret the results obtained here.

Regarding differences in CS use depending on gender, Ting and Kho (2009) revealed, based on data collected from 20 Malaysian undergraduates with English as an L2, that male and female speakers tended to resort to different CS types (e.g. females used more restructuring while males preferred approximation). Findings also indicated that speakers accommodated their use of CS to their interlocutor's gender, resulting in both male and female speakers decreasing the use of approximation and male learners increasing the use of restructuring, which the authors describe as a more careful formulation of the message. Similarly, Zeynep (1997) paired English native speakers with Turkish ESL learners and observed that all Turkish participants used more CS with female native interlocutors than with male interlocutors because the female native speakers were more cooperative and encouraging than the males. In addition, Zeynep observed the influence of learners' personality on their use of CS: not surprisingly, extroverted learners proved more successful in conversation; they were more willing to communicate and used more interactive strategies.

Other studies have looked into the effect of speakers' L1 on their use of L2 CS. An example of this connection can be found in Rababah and Bulut's (2007) study, which sampled 24 male Arabic learners with eight different native languages and concluded that differences in CS use could be attributed to L1 interference and also to educational and cultural background. And finally, the effect of age on the selection of CS has also been investigated in a few previous studies. Grañena (2006) studied the effect of age on the use of help requests in a narration task, and results indicated that older children resorted more frequently to explicit appeals for assistance while younger

children and adults preferred implicit signals such as hesitation and pauses, so the effect of age remained unclear. Oliver (1998) found that children aged 8-13 used the same interactional CS types as adults, but with different frequencies (adults preferred comprehension checks and children used more other-repetition). And findings in Montero et al. (2013) suggested that adults used more CS than children, did so more effectively and used fewer L1-based CS.

2.3. Summary

This chapter has reviewed the most relevant research findings related to the factors involved in the present study, i.e. communication strategies, language learning context, proficiency, cognitive factors (attention control and analytic ability) and language learning strategies. The first section of this chapter has presented the main findings and theoretical frameworks in CS research. CS, the tools implemented by L2 learners to overcome communication gaps caused by their linguistic shortcomings, have been broadly defined, identified and classified. However, the explanation for the selection made by the L2 speaker when encountering a break-down in communication to solve the problem remains unresolved, as does the issue of how or to what extent CS should be taught within the context of formal L2 instruction: explicitly, implicitly or not at all. In order to reach a conclusion on the latter question, information about which factors or combination of factors affect CS use and development seems essential. If the abilities, behaviours and conditions that foster the use and development of effective CS are disentangled, CS teaching (if possible and necessary) can be designed and integrated into language courses.

The second section of this chapter therefore focused on a series of factors that could potentially affect CS effectiveness: learning context (SA and AH), cognitive factors (attention control and analytic ability), LLS and proficiency level. Findings

regarding the effects of learning context (SA as compared to AH courses) on the development of effective CS are scarce and contradictory. With regard to the effects of attention control and analytic ability, the mostly theoretical information available seems to indicate that such cognitive factors may be connected to CS selection. LLS and CS have been found to overlap or even constitute a single object of study, and some findings have suggested that the two strategic behaviours may share some of the same underlying mental processes and cognitive abilities. Finally, L2 proficiency level has been mostly agreed to be linked to learners' quantity and quality of CS use, but this study will take the chance to contribute to the literature on this matter as well. The present study will therefore analyse the connections between each of these factors and effective CS use and development. Moreover, interactions among these factors will also be tested by observing whether attention control, analytic ability, LLS and proficiency level enhance CS development in each of the learning contexts under study (SA and AH). The following chapter will formulate the research questions of this study based on the literature reviewed.

3. Research questions and hypotheses

The purpose of this dissertation is to contribute to the understanding of how effective use of CS develops in L2 speakers by investigating some of the factors that may have an influence on the use of these strategies. CS have been broadly identified and classified in the field of SLA, both from a linguistic and a psychological perspective, and their relevance in L2 communication is undeniable (Dörnyei & Scott, 1997; Oweis, 2013). However, the question of how they develop so as to contribute effectively to communication, and thus whether or not they are teachable, remains unanswered. Contradictory findings have been reviewed as to the effect of learning context (specifically SA as compared to AH) on CS (e.g. Lafford, 2004, vs. Montero et al., 2013), little empirical evidence has been obtained about the relevance of the speakers' cognitive abilities (Littlemore, 2001, 2003) and no previous studies have been found on the correlation between LLS and CS, though they have in some cases been studied as a single phenomenon (LePichon et al., 2010). Furthermore, only Littlemore (2003) and Montero et al. (2013) have previously connected such factors specifically to CS effectiveness rather than to the quantity of CS used in general, to individual CS types or to L1/L2-based CS. Proficiency is generally agreed to influence CS use, but since a proficiency measure had been included in the study's design to help ensure comparability among participants, the opportunity was taken to replicate the observance of such a connection in the data collected for the present study.

All in all, these research gaps regarding the factors or combinations of factors that may contribute to the use and development of effective CS have yet to be filled. In order to contribute to this area, the abovementioned series of possible determinants of the CS use have been identified in the literature in previous theoretical and empirical pieces of research, and they include language learning context, cognitive factors (attention control and analytic ability), learning strategies and L2 proficiency. The

present study will look into their individual influence on CS and the effect of the interaction between learning context and all other factors on the development of effective CS. This chapter is devoted to a detailed explanation of the research questions and hypotheses formulated for the design of the study.

3.1. Is learning context a determining factor in the development of effective CS?

As reviewed in subsection 2.2.1, the effect of L2 learning context on changes in the use of CS, especially the effect of SA experiences, has been studied by a few authors, such as DeKeyser (1991), Lafford (2004), Rubio (2007) and Montero et al. (2013). However, the findings reported in those publications do not allow for a clear conclusion on the issue. Both Lafford and Rubio observed a lower rate of use of L1-based CS among SA students in comparison with AH students, while DeKeyser concluded, after observing patterns in CS use among SA and AH participants, that the differences were not due to learning context but to variations in the learners' personalities and tasks and in the interlocutors' characteristics, and results in Montero et al. indicated that learning context had an influence on children's development of CS effectiveness and reduction of L1-based CS but not on adults'.

With this research question, the present study will specifically look into the effect of four months of SA experience as opposed to AH courses in undergraduate students. This possible effect will be measured in terms of CS effectiveness, i.e. whether or to what extent the CS used by the participants prove successful in solving their communication problems, in getting the message through to the interlocutor. The hypothesis stated here is that the results obtained in this piece of research will replicate those reported in Montero et al. based on the fact that CS *effectiveness* will be analysed, without regard to whether the CS used are based on the L1 or the L2, as in

Lafford (2004) and Rubio (2007). Additionally, a double analysis approach will be implemented to measure CS effectiveness: the one included in Montero et al. (2013) and another complementary method, in order to obtain a broader perspective on the matter.

3.2. Are attention control and analytic ability involved in the selection of effective CS?

Based on the literature reviewed in subsection 2.2.2, the possibility of a connection between cognitive factors such as attention control and analytic ability and certain patterns in the use of CS is founded mainly on a theoretical basis. This basis relies on publications such as Bialystok (1990), Dörnyei and Kormos (1998), Kasper and Kellerman (1997) and Shatz (1978), which provide explanations of the mental processes involved in the selection of CS to solve break-downs. To the researcher's knowledge, no empirical evidence has been gathered on the influence of attention control in the selection of CS, but there seems to be a theoretical framework that points towards such a connection. And regarding analytic ability, the findings published by Littlemore have been the only ones to prove with empirical data that L2 speakers with an analytic (or ectenic) cognitive profile showed a certain pattern in the type of CS they resorted to (2001) and that they were more effective in the implementation of CS than learners with a holistic or synoptic cognitive style (2003).

This piece of research will attempt to shed some more light on this question by correlating measures of the two cognitive factors, attention control and analytic ability, to the use of effective CS in L2 communication. Despite the apparent lack of sufficient (or any, in the case of attention control) previous empirical evidence, it could be hypothesized that these two cognitive factors affect the use of CS on the grounds that all the theoretical and empirical information available seems to point towards this

connection. In other words, to the researcher's knowledge, no existing evidence, either empirical or otherwise, seems to refute this hypothesis.

3.3. Is there a connection between language learning strategies and the use of effective CS?

According to the bibliography reviewed in subsection 2.2.3, the relationship between the use of LLS and that of CS has not yet been explored. There is some evidence that similar cognitive processes may underlie the use of both CS and LLS, since Ying-Chun (2009) observed that cognitive abilities, particularly processes that require attention control and analytic ability, were connected to the use of LLS, and these cognitive factors have also been linked to the use of CS (see sections 2.2.2 and 3.2). In fact, previous authors have included CS as a part of LLS (O'Malley & Chamot, 1990; Oxford, 1990; see section 2.2.3), and others have suggested that the two concepts somehow overlap (Khan, 2010; LePichon et al., 2010; Maleki, 2007), e.g. CS may allow the L2 speaker to hold longer conversations and therefore be exposed to more L2 input, which would be a way to foster their language learning, according to Maleki (2007). However, Dörnyei (2005) maintained that even though CS contribute to keeping conversation open and thus give L2 learners more chances to acquire the target language, they are related to language *use* and not language *learning* and that these concepts should therefore be best studied separately.

Grounded in these ideas, this study considers the possibility that LLS and CS, though they are conceptually separate, may share some kind of connection. Evidently, they both constitute strategic behaviour with regard to additional languages, either learning or use. In addition to that, they might share some of the cognitive abilities (attention control and analytic ability) involved in the selection the speaker makes, since they are both intended to solve problems that result from lacking the required

linguistic resources. These findings lead the researcher to expect some type of connection between CS effectiveness and either the frequency of use of LLS overall or the use of specific LLS types, as measured by learners' self-report on a Likert-scale questionnaire.

3.4. Is proficiency level a determining factor in the choice of effective CS?

Level of L2 proficiency has been argued and empirically shown to correlate with the development of effective use of CS (Bialystok, 1983; Haastrup & Phillipson, 1983) and with a decrease in L1-based CS (Liskin-Gasparro, 1996; Paribakht, 1985), with minor exceptions. Haastrup and Phillipson observed that their findings illustrated a continuum, parallel to progress in L2 proficiency, ranging from mostly L1-based and less effective CS to interlanguage-based and more effective CS. And along same lines, both Liskin-Gasparro and Paribakht found the level of L2 proficiency and frequency of use of L1-based CS to be inversely proportional. On the other hand, regarding the minor exceptions mentioned, Liskin-Gasparro nuanced that beginners and near-native speakers should be excluded from such a continuum, since the former cannot implement CS with such limited linguistic resources and near-natives should not need CS any more than a native speaker would. Also, Poulisse and Schils (1989) found quantitative differences in CS use across proficiency levels, but not qualitative ones. Finally, one of the three CS elicitation tasks administered in Khan (2010), namely the information gap picture description task, did not prove this negative correlation between proficiency level and the use of compensation strategies, although the other two tasks did.

As mentioned in section 2.2.4, in the present study, a proficiency measure was contemplated in the original design only as way to contribute to ensuring comparability among participants. Nevertheless, the decision was made to take this chance to also

contribute to the discussion about this affecting factor in CS use, since the data would be available. Khan's information gap task will be administered as a CS elicitation task, and it happens to be the source of one of the few pieces of evidence to contradict the correlation between proficiency level and CS use. However, the CS measures implemented in the present study will be different to those used in Khan's study: the focus of the study is on CS effectiveness, not frequency of use. Therefore, the hypothesis is that the results obtained will align with the connection between proficiency and CS agreed upon by most of the previous research.

3.5. Can we define an ideal student profile for each learning context regarding development of CS effectiveness in terms of attention control, analytic ability, LLS and proficiency level?

One of the goals of this piece of research is to shed some light on whether learning context somehow interacts with the other factors listed in previous sections in this chapter (initial L2 proficiency, LLS, attention control and analytic ability) to influence the development of effective CS use, i.e. *progress* made by learners regarding CS effectiveness during the learning context experience (SA or AH). To the researcher's knowledge, there is no previous empirical evidence of such interaction or in fact of its absence. There are indeed contradictory findings in the literature with regard to the effect of learning context on the development of effective CS (see Lafford, 2004, and Rubio, 2007, vs. DeKeyser, 1991, and Montero et al., 2013), but these studies did not include any other factors apart from learning context, with the exception of Montero et al., which included age and showed that learning context had an effect on the development of CS effectiveness in child learners but not in adult learners. The contradictions found among different publications may stem from the possibility that learners' individual differences determine to what extent they benefit from each of the

learning contexts regarding CS use development. Actually, interactions between learning context and cognitive abilities have been found to have an effect on other aspects of SLA (Segalowitz & Freed, 2004; Sunderman & Kroll, 2009; Tokowicz et al., 2004).

This research question is thus intended to offer results on whether L2 learners' success in developing effective CS in a specific learning context (SA or AH) can be predicted by any of the other characteristics measured or a combination of some or all of them: attention control, analytic ability, LLS and initial proficiency level. If one or several of the factors considered were found to predict development of effective CS in one of the learning contexts included in the study but not in the other one, this result would help depict an ideal student profile for the relevant learning context. The fact that there is no precedent for this possible effect of interaction between learning context and the other factors on the development of effective CS use makes venturing hypotheses with regard to this question quite challenging, but some of the previously mentioned research findings may help shape certain expectations.

First, based on the findings provided by Segalowitz and Freed (2004), Sunderman and Kroll (2009) and Tokowicz et al. (2004), it seems that working memory and even its individual components (attention control) do interact with SA learning context to contribute to the development of different areas of SLA (oral skills in general, oral fluency, lexical production, and reducing translation errors). It follows that attention control, which is a component of working memory, may play a role in developing CS effectiveness as well in the SA context. Second, analytic ability is believed to predict aptitude for *explicit* learning (Grañena, 2013). AH students will receive L2 input mainly in a classroom environment, which might lead one to foresee that those with greater analytic ability could benefit more from this context also with regard to CS development. However, they will not receive explicit CS training and they will also attend other courses taught in English which do not entail explicit language instruction, so the researcher feels more reticent to expect any specific outcome in this respect.

Finally, initial proficiency level has been found to predict accuracy gains during a stay of six weeks in a SA context: DeKeyser (2010) measured both initial proficiency level and language aptitude on a sample of 16 American learners of Spanish and concluded that the best predictor of accuracy gains was initial proficiency and not aptitude. Therefore, initial proficiency level could be expected also to predict CS effectiveness gains in the SA context. No hypotheses are formulated on the possible interaction of learning context and LLS given that no other previous findings seem to help predict results with regard to these specific factors.

4. Methodology

This chapter presents an exhaustive description of every step taken and every decision made before and during the recruitment of participants, data collection, qualitative analysis of CS and operationalisation methods of CS effectiveness. This information is structured in three main sections: section 4.1. will describe the participants and learning contexts; in section 4.2., a description of the instruments implemented and the procedure followed will be provided; and the approaches taken for CS effectiveness analyses will be included in section 4.3. The details provided should allow for replication of the study and potential improvements on the methodology in future research.

4.1. Participants and learning contexts

This section provides information about the participants in the present study and the type and amount of L2 input they received in each of the learning contexts under analysis (study abroad or at home). First of all, the recruitment prerequisites and process will be explained in subsection 4.1.1.; then the participants' linguistic background and habits will be presented in subsection 4.1.2.; and finally, subsection 4.1.3 will cover the linguistic experience in each learning context as reported by the participants themselves.

4.1.1. Recruitment of participants

The participants in this study were volunteer Spanish undergraduate university students, either studying in an English-speaking country for four months as part of a university exchange programme (Study Abroad or SA group) or studying at their home

university and taking 10 to 15 hours a week of courses taught in English over those four months (At Home or AH group). In both groups, the prerequisites to participate were to be native speakers of Spanish (Catalan and Galician students, with two L1s, were also accepted), to speak English as an L2 and not to have lived in an English-speaking country for longer than a month, i.e. not to have experienced a similarly relevant L2 immersion stay before the testing period. Participants were to be tested at the beginning (T1) and at the end (T2) of the four-month period. The length of the testing period corresponds to the autumn university term, which includes three months of classes (September-December), the Christmas break and the exam season in January.

Recruitment for the SA group was carried out by asking the International Relations Offices of all state-funded universities in Spain (except for distance teaching universities) to share the call for participants with their students who planned to study abroad in the coming semester. Based on the information about the participants' home university, at least 11 out of 46 universities collaborated in the recruitment process. Those who decided to participate in the study wrote directly to the researcher. Volunteers were briefed from the beginning on the requirements for their collaboration in the study and as a compensation for their time if they completed all the tests they were promised information on their scores and a modest thank-you gift. Considering the nature of the research, it was assumed that many students interested in participating would be especially motivated in their language learning. In fact, 55% of the participants in the SA group were doing language-related university degrees: Philology, Translation or Second Language Teaching. The remaining 45% were students from very varied disciplines, such as Economics, History, Physics and Sport Sciences. Most participants spent their time abroad in universities in the United Kingdom, but there were also two participants who went to Ireland and one to the United States.

Participants in the AH group were all English Studies students at the University of Barcelona who chose to participate as an alternative learning activity within an Applied Linguistics course. In this case, participants were compensated for their time with information on their scores and course credit. A prerequisite for the participation in the present study was enrolment in several university courses (10 to 15 hours per week) taught in English during the testing period. That way, the amount of hours of L2 input in the academic environment would be similar in both the SA and the AH group. Attempts were made to recreate the multidisciplinary sample of SA participants in the AH group to guard against potential disparities in proficiency level or language aptitude between the two groups that would only appear at the moment of statistical data analysis after all data collection and processing were done. However, it is not common in the Spanish university system to take that many hours of courses taught in English at the undergraduate level. The vast majority of students attend classes in the official L1s (Spanish, Catalan or others, depending on the area) and then sometimes take extensive (3-4 hours per week) language courses. Rather exceptionally, the Autonomous University of Barcelona does offer an undergraduate degree in Business Administration taught entirely in English, and the faculty responsible for this degree was contacted and asked to share the call for participants with their 2nd and 3rd year students. Some lecturers did respond positively, but no students showed an interest in participating in the study. As a result, no at-home non-language oriented students could be recruited[2].

Two different rounds of recruitment and data collection took place in order to reach a higher number of participants: September 2012 to January 2013, and September 2013 to January 2014. The table below represents the number of valid

[2] Despite the differences in academic background, as will be shown in the results chapter, section 5.2.2, the participants in the two groups were found to be comparable (statistically not different) in terms of attention control, initial proficiency level, initial use of LLS and initial CS effectiveness.

participants who *completed* the study per learning context group (SA and AH) and academic year.

Table 4.1: Number of participants per data collection

Data collection	**AH**	**SA**
2012-2013	18	16
2013-2014	14	17
Total	32	33

Mortality, i.e. participants who were recruited and participated in the T1 data collection but failed to participate again in T2, in the first round was 33% (originally there were 55 participants), while in the second round it was only 8% (there were originally 36 participants). In addition to those who failed to complete the study, five participants were discarded because, despite the prerequisites stated in the recruitment call, they were later found to have spent periods longer than a month in English speaking countries. The final number of participants in this study was 65 (SA n=33; AH n=32).

4.1.2. Linguistic profile of participants

At the beginning of the testing period, participants in both groups were asked via a questionnaire (see appendices B and C) about their linguistic background and habits. Specifically, information was gathered about their L1s and L3s, whether or not they had made any short stays in foreign countries where they had communicated in English, if they had taken any extra English classes while they were at school, how often they usually read or watched television in English and if they had had any other contact with the language (partners, friends, family or professional relationships). The participants' answers to the initial questionnaire are presented in the following table.

For most items, the number of participants (n=) who reported each possible answer is indicated, while in others, like age range and mean, the numeric values refer to the answers given. This information is further explained in subsequent paragraphs, and the answers regarding previous short stays abroad are presented further below in a separate table.

Table 4.2: Initial questionnaire answers

Questionnaire items		**AH (n=32)**	**SA (n=33)**	**Total (n=65)**
Age	Mean	20	21	21
	Range	18-23	19-25	18-25
Gender	Male (n=)	2	16	18
	Female (n=)	30	17	47
L1	Spanish (n=)	1	13	14
	Spanish & Catalan (n=)	30	17	47
	Spanish & Galician (n=)	0	3	3
L3	No L3 (n=)	12	7	19
	Elementary (n=)	11	19	30
	Lower-intermediate (n=)	6	10	16
	Intermediate (n=)	6	10	16
	Upper-intermediate (n=)	1	1	2
	Advanced (n=)	1	0	1
Private classes	No (n=)	1	0	1
	Number of years (M)	5.7	4.6	5.1
	Range of years	1-15	1-15	1-15
Reading and	Less than once a week or never (n=)	3	11	14

watching TV products	Once/twice a week, 'often' (n=)	20	15	35
	Daily (n=)	8	7	15
Personal relationships or work	No (n=)	10	18	28
	Family (n=)	1	2	3
	Partners (n=)	4	0	4
	Friends (n=)	18	13	31
	Co-workers/clients (n=)	2	1	3

The questionnaire required participants to state their L1 or L1s and then, in a different section, to indicate if they spoke other languages besides their L1s and English, and to self-assess their proficiency level in those additional languages. Participants were considered bilingual when they either stated two different L1s in the relevant section of the questionnaire or if they stated only one and then added the other official language of their geographical area in the section on additional languages and stated there they spoke it at the upper-intermediate or advanced level. Considering participants in the AH group were all studying in Barcelona, it was only natural that most of them were Spanish-Catalan bilinguals. Actually, only one of them (3%) reported speaking Catalan below the upper-intermediate level. The SA group presents more variety in this sense but still, half the participants were also from Catalan universities and some from Galician universities, so there were also a high percentage of bilinguals in this group (61%).

As mentioned above, a section of the questionnaire invited the students to list any other languages they may speak apart from their L1/L1s and English (L3 or additional language) and to self-assess their general level of proficiency in each of those languages on a scale from 1 to 5 (1=elementary, 2=lower-intermediate, 3=intermediate, 4=upper intermediate, 5=advanced). In order to simplify this information, this summary will group elementary and lower-intermediate L3 level as

‘low-level additional language’ and intermediate to advanced L3 level as ‘high-level additional language’. In the AH group, 22% of participants reported having one or two high-level additional languages and 47% had one or two low-level additional languages. In the SA group, 30% of students seemed to have one or two high-level additional languages and 67% said they had one, two or even three low-level additional languages. The most multidisciplinary group, the SA learners, presented a higher interest in languages in general than the AH group, which consisted only of language students. In both groups, the most popular L3 was, by far, French, followed by German, Japanese, Italian and Arabic.

Answers regarding private English classes in academies or at home indicated a higher rate in the AH group: the average number of years of extra classes was 5.7 in the AH and 4.6 in the SA group. With regard to leisure activities, most participants in both AH and SA groups reported reading and/or watching series or films in English as a habit (weekly or even daily): 90% in the AH group and 67% in the SA group. Meanwhile, the answers on personal or professional relationships in which communication was held in English reflected that more AH participants had had such contact with the L2 (66%) than SA students (45%) before the testing period.

The questionnaire also included a section where participants were asked if they had been abroad and communicated in English before for a period of two or more weeks, though it should be noted that those having lived in an English-speaking country for longer than a month were barred from participating in the present study. Participants were also invited to indicate where each stay abroad had taken place and how long it had been. The following table includes information reported by the participants about the quantity of stays in either English-speaking countries or other countries in which they communicated in English, and the duration of these experiences. Even though this section of the questionnaire requested information about two-week or longer stays, some participants included information about one-week stays, and it has also been presented in the table.

Table 4.3: Previous short stays abroad

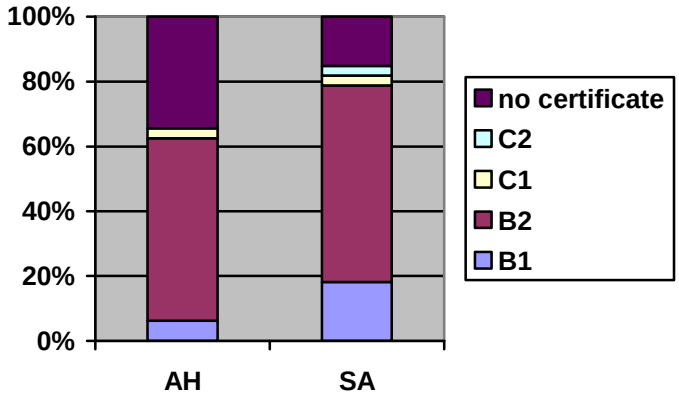
		AH (n=32)		SA (n=33)		Total (n=65)	
		English-speaking countries	**Other countries**	**English-speaking countries**	**Other countries**	**English-speaking countries**	**Other countries**
Quantity of stays	0 (n=)	14		7		21	
	1 (n=)	8	5	12	3	20	8
	2 (n=)	2	1	5	2	7	3
	3 (n=)	3	0	4	0	7	0
	4 (n=)	1	0	1	0	2	0
	>4 (n=)	1	0	2	0	3	0
Length of stay (weeks)	1 (n=)	3	4	4	1	7	5
	2 (n=)	11	1	9	4	20	5
	3 (n=)	6	1	17	0	23	1
	4 (n=)	10	0	19	2	29	2
	>4 (n=)		1		0		1

When asked about previous stays abroad during which they had used English to communicate, 17 (53%) of the participants in the AH group reported having had this experience, while the rest did not. Among those in the AH group with experience abroad, 47% described only one experience in English-speaking countries, with the majority of stays in such countries lasting either two (37%) or four weeks (33%). As for experiences in other countries where participants communicated in English, 29% of abroad-experienced AH students reported one such trip, which in 57% of the cases lasted for only a week. In comparison, the percentage of participants with experience abroad in the SA group was considerably higher (79%, i.e. 26 participants). Similarly to

the AH group, 46% of them reported one stay in English-speaking countries, but in this case, the experiences were longer in general than in the AH group: 35% of the stays reported lasted for three weeks, while 39% of them lasted for 4 weeks. With regard to stays in other countries, 11.5% of SA students with experience abroad reported one such stay, and 7.7% two stays. These stays were a little longer among SA students than among AH students: 57% of them were two weeks long.

To complement this linguistic profile, the participants were asked whether they had any standardised certificate that assessed their level. It has to be taken into account though, that they may have been awarded the certificate years before participating in the study, so their actual level could have been higher or lower at the moment of recruitment for this study. The following graph (figure 4.1) represents these certificates on the scale of the Common European Framework of Reference (CEFR; Council of Europe, 2001).

Figure 4.1: Certificates of CEFR levels



As the graph shows, most participants (62.5% in the AH group and 78.8% in the SA group) had been awarded either intermediate (B1) or upper-intermediate (B2) English proficiency certificates at some point in their lives before the testing period. Only one participant from each group had been awarded a C1 level certificate, and

another participant in the SA group had reached the C2 level[3] according to their reports. The remaining participants (31% in the AH group and 12% in the SA group) did not have any certificate to assess the level of proficiency they had reached.

As mentioned in section 2.2.5, certain individual differences such as speakers' gender, personality, L1 and age may have an influence on the use of CS in L2 learners. These factors will not be investigated in the present study. However, findings in this regard must be considered when describing the sample recruited for the present study and interpreting the results. Differences in CS use depending on speakers' L1 (Rababah & Bulut, 2007) and age (Grañena, 2006; Montero et al., 2013; Oliver, 1998) should have no implications for the present study, since participants were all undergraduate students (aged 18-25) and native speakers of Spanish and, in some cases, Catalan or Galician. For practical reasons, personality will not be accounted for. The sample includes both male (27%) and female speakers (73%). Even though the proportion of female speakers is noticeably higher, the interlocutor was a female researcher in all cases, which should control for the potential effect of gender by inspiring cooperation and encouragement to communicate in all participants, according to Ting and Kho (2009) and Zeynep (1997).

4.1.3. Linguistic experience in each learning context

At the end of the testing period, the participants took another questionnaire (see appendices D and E) about their contact with English (both groups) and with their L1 (SA group) during that time. Again, they were asked whether they had taken English classes outside the university (AH group), about their leisure activities, either organised (theatre, sports, etc.) or otherwise (reading and watching TV), about the amount of time they had spent outside their learning contexts, and about personal and professional

[3] As will be explained in section 4.2, participants took a proficiency test that measures vocabulary size, the XYlex test. The participant who reported holding a C2 level certificate scored 5,950 out of 10,000 possible points (estimated vocabulary size), just above the mean of the whole sample (M=5,814). Considering other participants got a much higher score (the highest result was 8,150), this student was not considered an outlier within the sample.

relationships in which they communicated in English. Table 4.4 includes their responses, which are further commented on below. Following this, answers regarding relationships are presented in a separate table.

Table 4.4: Final questionnaire answers

Questionnaire items		**AH (n=32)**	**SA (n=33)**
L1 contact	Daily (n=)		30
	< daily (n=)		2
Time spent outside L2 country (weeks)	0 (n=)		2
	1 (n=)		2
	2 (n=)		6
	3 (n=)		12
	4 (n=)		10
Housing	Hall of residence (n=)		16
	Shared flat (n=)		15
	Individual flat (n=)		1
Exposure to languages (reception)	English (M of %)		61
	L1(s) (M of %)		34
	Others (M of %)		5
Active use of languages (production)	English (M of %)		54
	L1(s) (M of %)		42
	Others (M of %)		4
Classes/week in L2 at the university	Number of hours (M)	13.5	10
L2 extra classes or	No	22	

language exchange	Yes, extensive (n=)	9	
Trips abroad using L2	No	27	
	Yes, 1 weekend	2	
	Yes, 4 weeks[4]	1	
Reading and watching TV products	Less than once a week or never (n=)	1	0
	Once/twice a week, 'often' (n=)	19	17
	Daily (n=)	12	15
Organised leisure or other activities in L2	No (n=)	29	7
	Sports (n=)	0	19
	Dancing lessons (n=)	0	5
	Theatre (n=)	2	0
	Cultural societies (n=)	0	6
	Research	0	1
	Work	1	3

The vast majority of participants (91%) in the SA group reported having had daily contact with their L1 during the testing period, either with their families or with other Spanish-speaking students. In addition, most of them (67%) had spent three or four weeks within that period outside the L2 country. When asked about their housing situation in the L2 country, 91% of SA participants stated they lived in either halls of residence or shared flats with other students. In addition to that, SA students were asked to estimate the percentage of time they had been exposed (reception) or actively using (production) the L2, their L1(s) and other possible languages, i.e. they were asked to split the 100% of their language exposure time and, separately, the 100% of their active language production time among English, their L1 and others. The table above shows the means of the answers provided by the participants. In general, they

[4] The student stated she had spent a month in Sweden for Christmas and that she had spoken 'mostly Swedish but also a little English'.

estimated they spent more than half of their time either receiving or producing English but, also a considerable amount of time using their L1. Not surprisingly in a linguistic immersion context, the estimation for exposure to English (61%) was slightly higher than the one for production (54%).

With the same purpose of gathering information about exposure to the L2 during those four months, the AH group was queried about possible extracurricular L2 classes or language-related activities (e.g. language exchanges) and about whether they had travelled abroad using the L2 during the testing period. With regard to the first of these questions, 28% of the AH students had English extra classes or language exchange activities during the testing period, which in all cases were extensive courses (between one and four hours per week). A much slimmer minority of AH participants (9%) left their AH context during the testing period. Three participants travelled abroad and used English to communicate: two of them only for a weekend and the other one for a longer period, although English was reportedly not the main language for communication during that stay.

Other questions were addressed to both the SA and AH groups in order to compare the amount and nature of L2 input they had received during the testing period. First of all, participants were requested to state the amount of hours per week of university lessons in English they had attended. The table above shows the mean of the responses in each group, which was a little higher in the AH group. The responses in the SA group presented higher variability since the students were doing different academic programmes at different universities. Then, participants in both context groups provided information about their leisure habits or work in English. On the one side, they were asked a series of questions regarding their consumption of cultural products (books, newspapers, films and TV series), and their answers were processed and summarised to allow for easier comparison of the frequency of exposure to such products in the two groups. As shown in the table above, the responses in the two groups were fairly similar: 97% of participants from each context group reported

reading or watching TV products either weekly or daily. On the other side, when asked about organised activities in English, the difference between the contexts was made obvious: only 9% of AH students participated in such activities, while 79% of SA students took part in the activities available in their living context, most of them (73%) in sport activities.

As mentioned before, participants in both groups were asked about their personal, academic (outside class hours) and professional relationships in English during the testing period. In fact, they were asked to name up to four people in each environment (personal and academic/professional) with whom they had communicated in English and, about each of these people, the participants had to state how often they saw them and what their nationality was. The table below presents the responses obtained from this section of the questionnaire. The amount of people they named for each context and how frequently they saw them, i.e. how frequently they communicated in English with them, are connected to those people's native language (assumed by their nationalities): 'L2' indicates English native speakers, 'L3' is for native speakers of other languages and 'L1' for Spaniards.

Table 4.5: Relationships in English during testing period

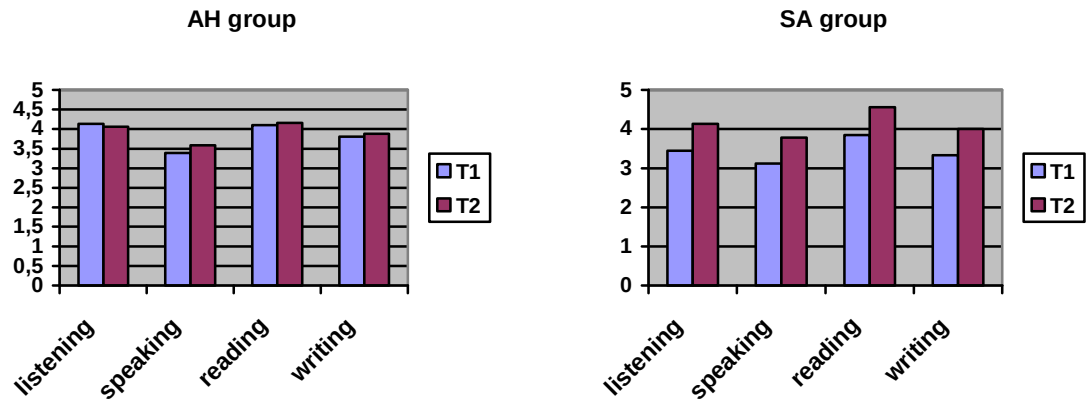
		AH (n=32)			**SA (n=33)**		
		L2	**L3**	**L1**	**L2**	**L3**	**L1**
Personal environment (amount of acquaintances)	0 (n=)	15			4		
	1 (n=)	9	5	3	10	3	4
	2 (n=)	2	0	2	5	10	1
	3 (n=)	0	1	0	2	4	0
	4 (n=)	0	0	0	2	6	0
University / work (amount of	0 (n=)	22			14		
	1 (n=)	3	2	4	5	7	1

		AH (n=32)			SA (n=33)		
		L2	**L3**	**L1**	**L2**	**L3**	**L1**
acquaintances)	2 (n=)	0	0	2	5	5	0
	3 (n=)	0	0	0	3	2	0
	4 (n=)	1	0	0	0	3	0
Frequency	Daily (n=)	0	2	4	24	49	2
	Weekly (n=)	8	7	6	27	37	3
	< Weekly (n=)	6	0	4	2	0	0
	Unspecified (n=)	6	1	1	3	6	1

As the table shows, 53% of AH students and 88% of SA students reported on personal relationships in which communication was held in English. Regarding relationships in the academic and professional environments, 31% of the AH group and 58% of the SA group presented information about such relationships in the L2. Not only did more SA participants report on their relationships, but they also listed more acquaintances and their encounters with them were more frequent. The AH group reported a total of 45 relationships, out of which 13% were daily contacts and 47% were weekly. Meanwhile, the SA group listed a total of 159 relationships, 47% of which were daily encounters and 42%, weekly. Predictably, the SA group had easier access to L2 native speakers, and they also listed numerous relationships with speakers of other languages.

Finally, all participants were also asked to self-assess their L2 listening, speaking, reading and writing level both before and after the testing period. Each participant rated every skill on a scale from 1 to 5 (1 = elementary, 2 = lower-intermediate, 3 = intermediate, 4 = upper intermediate, 5 = advanced). The following graphs (figure 4.2) show the means of the answers in each group per skill and per testing time (T1 and T2).

Figure 4.2: Graphs on self-reported proficiency level



An observable difference is presented on these graphs. In the AH group the difference between their T1 and T2 self-assessed reading and writing skills is rather modest (they rated themselves 0.06 and 0.07 points higher in the respective skills), while they perceived a bigger improvement in the speaking skills (0.2) and a small step back in their listening development (-0.07). On the contrary, the SA group observed improvement in all their skills between testing times, and evidently their perceived progress was more noticeable than that of their AH counterparts (each skill was rated in T2 between 0.66 and 0.71 points higher than in T1). It is also worth pointing out that the initial level was lower in the SA than in the AH group, according to this section of the questionnaire.

This subsection has provided all the information available about the participants in the present study, through the explanation of the recruitment process and participation requirements and through the linguistic profile the students presented on their initial questionnaires, and about their experiences in either the SA or the AH learning context. This information should help the reader contextualise and interpret the analysis and results that will be presented in subsequent sections and chapters.

4.2. Instruments and procedure

This section is devoted to the presentation of all the instruments and materials used for data collection in the present study and the description of the data collection procedure followed. The next subsections (4.2.1. to 4.2.5.) cover the questionnaires or tests used to measure each construct involved in the study and the final subsection 4.2.6 deals with the procedure implemented to collect the necessary data.

4.2.1. Communication strategies elicitation task

In order to analyse the use and development of L2 CS, the main construct in the present study, an information gap task was implemented. The data collected were audio recorded and later transcribed, both at the beginning and at the end of the testing period. The task used in T1 (appendix F) was adopted from Khan (2010) and served as an inspiration to create a new comparable set for T2 (appendix G). The task was designed to elicit compensation strategies through a referential communicative situation, where participants had to describe pictures to their interlocutor. Khan applied this task to pairs of participants who had to interact and negotiate meaning between them. However, in the case of this study, that implementation procedure was modified: the task would be carried out by each participant and the researcher, so the support the participant received in conveying meaning during the task would be controlled for, and the effectiveness in CS use of every individual participant would be more accurately assessed.

Each set consisted of 11 simple pictures of objects or animals copied in two versions (one for the participant and one for the researcher in this case). Each item could be either identical in both versions or present small differences. The participant was asked to describe them one by one in full detail so that the researcher could decide whether they both had exactly the same picture or if they were different in each

set. The researcher interacted with the participant by asking questions to try to force the participant to refer to every element in the picture. There was always at least one problematic element in each item, a word that was supposed to be unknown or difficult to retrieve for an intermediate or upper-intermediate L2 student, e.g. the *trunk* of the elephant, the *straw* in the glass, the different parts of a teapot (*spout*, *handle* and *lid*) or the *fence* of the house. This way, communication gaps were provoked and participants had to find a way to overcome them (CS).

The set of pictures for T2 was created with the goal of reproducing the characteristics of Khan's set, and both sets were piloted to check whether they caused similarly frequent communication gaps in upper-intermediate (B2) students. The pilot study involved nine university students taking part in B2-level English courses. They were requested to perform both T1 and T2 CS elicitation tasks to ensure that each individual item contained at least one problematic element at this level of proficiency and therefore provoked communication gaps and the use of CS. It took the participants between 16 and 25 minutes to complete both tasks. Half the participants described the T1 set of pictures first and then the T2 set, while the other half started with the T2 set and continued with T1. This procedure was followed in case describing 22 pictures could result in a repetitive task for the participants and they might perform differently at the beginning and towards the end of the process. By alternating the order of the tasks, the possible loss of concentration should not affect the comparison between the two sets of pictures in terms of the challenge they present to a B2 English learner.

The piloting of each individual item was to prove successful if at least seven out of the nine participants (78%) found a problematic element during the description. Indeed, 21 items passed this test: 16 items presented elements which proved problematic to all pilot study participants (100%), four items proved successful on 89% of the participants (8/9) and one item on 78% of the participants (7/9). The only exception was item number 3 in the T1 set, which could be objectively described as three triangles, two of which were partially covering the third one, or interpreted as a

small group of mountains, one of which seemed to be further away between the other two, which were at the front. Out of the nine participants, only six (67%) had to resort to CS to describe this item. However, foreseeing the possibility that some items would be discarded after data collection if they proved to be less productive on the actual participants of the study in order to focus the analysis on the most successful items, Khan's original set (T1) was presented to the participants since keeping item number 3 in the task would not compromise the results and more than half the pilot study participants had encountered communication gaps when describing it.

During the implementation of the CS elicitation task, the role of the interlocutor was, as mentioned above, to encourage the participants to be more informative and to ask questions about the details they would leave out in order to make sure they were merely overlooking them rather than avoiding them (e.g. 'what else can you see in the glass?' 'Nothing' 'No? Is it a cold drink or a hot drink?' 'Oh, there is ice'; or 'what else can you see around the top of the glass?' 'Yes, I know what you mean but I don't know the name'). If Khan's procedure of testing two participants at a time had been kept in this study, the questions asked might have been different and they might have included the 'difficult vocabulary' that the task needs to succeed as a CS elicitation task if one of the participants knew a difficult word. For example, a participant could have asked their partner 'is there a straw in your picture of the glass?' or 'is there any ice?' and, therefore, their partner would have had fewer opportunities for CS use since the necessary vocabulary had already been given to them. Here in contrast the researcher first let the participant describe the item completely according to their own criteria and then, if key elements were missing from their description, the researcher asked specific questions for every item to make sure that as many elements in the picture as possible were addressed in the descriptions and that all participants received similar prompts. The table below includes some examples of this protocol.

Table 4.6: Examples of item description

Item described	Transcript of the description
	***2219:** is a house is like a village house hmm is a...is a very is a...a family's house because is bigger but not too bigger to be a mansion or something like that it has two floors because i...i can see two levels of windows and...is a...a classic architecture and it has something like a garden...at his left at his right and...the door the door looks open ***INV:** what can you see around the garden? ***2219:** only the...only the woods which delimitate the garden ***INV:** ok can you describe the top of the house? ***2219:** yeah ehh...hmm there is... like...hmm it has a classic roof and it has a...a...the smoke...i can't remember how to say the...this kind of thing ehh the in the house they can they can do fire to...because they have something for the smoke i can't remember the word ***INV:** ok anything else? ***2219:** ehh the roof is like a triangle [trI] and...there are you can saw the door three windows and maybe an...a four window that is circle
	***2116:** the next i think that's a glass of something like a refreshment it's... it's not empty it's near well more or less near...the top the refreshment and there's something inside but i cannot hmm watch what it is i cannot see what it is and it has something to... you for drinking like...i don't know the name... i don't know ***INV:** what can it be that thing in the liquid? is it a cold drink or a hot drink? ***2116:** i think it's a cold drink that maybe it's ice what is it inside ***INV:** probably yeah and on the top of the glass? anything around there? ***2116:** yeah but i don't know what it is it's...i don't know what...how to describe it because i don't know it's like hmm i don't know like a leaf or something like when you put...ah no! maybe the lemon that you put in the refreshments in summer ***INV:** ok and then this thing you don't know the name of ***2116:** yes when you don't drink it from the glass there's this little thing that it's made of plastic and you drink from it
	***2108:** it is a a box a gift it is a gift as it has special a wrap paper and...and... a lace i think it is called and...and...it is black and white ***INV:** which part is black which part is white? ***2108:** ah yeah ehh the black part ehh it is the...the draw on the paper that are stars black stars...and...i don't know hmm there are one two

	three four five six seven eight nine ten eleven i can see twelve ***INV:** ok the rest of it is white? ***2108:** yes the...yes the...the lace is white ***INV:** what shape is the box? ***2108:** it is a square box

These three examples show cases in which the participant left some elements in the picture out of their initial description. There were other cases in which they directly addressed all elements and thus no questions from the researcher were needed. After each description was finished, the researcher told the participant whether they had identical or different pictures in accordance with the task instructions and asked the student to move on to the next item. Depending on the participant, this task took between 10 and 15 minutes.

4.2.2. Language use and learning strategies questionnaires

As described in subsection 4.1.2, participants answered a questionnaire at the end of each data collection time: in T1 the questions tried to depict the participant's linguistic profile and experience (L1s, number and proficiency level of L3s, previous experience and habits related to English learning and use), while in T2 the questions referred to the nature (activities and relationships) and frequency of contact with the English language during the testing period. Both questionnaires were adapted from models previously used by the GRAL group (Llanes, Tragant, Serrano, 2011).

In both data collection times, participants completed a questionnaire about their language learning strategies as well. This questionnaire was meant to provide information on participants' learning habits and to complement the cognitive profile drawn from the attention control and analytic ability tests with a measurement of their strategic competence as L2 learners. It was administered twice, in both T1 and T2, in order to observe whether or not these strategies would change during the testing period.

The LLS questionnaire administered was adopted from Tragant et al. (2013) and translated into Spanish (see appendix H). It consisted of 17 statements related to intentional language learning, including specific strategies to improve general level of English (e.g. 'When I hear someone speak in English, I make an effort to see what I can understand'), vocabulary learning (e.g. 'I write down the word together with an example sentence'), studying grammar (e.g. 'I review the structures mentally or out loud'), reading (e.g. 'I pay attention to the words that appear next to what I do not understand in a text') and writing (e.g. 'Before I start writing, I first think about what I want to say').

Tragant et al. (2013) reduced a 55-item LLS questionnaire to the 17-item version used in the present study by means of an exploratory factor analysis and item analysis, and then they validated the new scale on a sample of 1,425 students. The questionnaire does not include LLS oriented towards oral skills, speaking and listening, yet it should be representative of strategic behaviour and attitude towards language learning. The results suggest a two-factor structure within the 17 items reduced version. This structure reflects 'skills-based deep processing strategies' and 'language study strategies', which allow a distinction to be drawn between deep and surface clusters of strategies. Items such as 'When I see short texts in English, I try to figure out what they mean' or 'I carefully review what I have written' (i.e. reading, writing and some of the improving general level of English LLS) belong to the skills-based deep processing cluster, while items such as 'I review what we have done in class or I test myself on my own' or 'I write summaries or outlines of the structures that we are learning in class' (i.e. vocabulary learning, studying grammar and other improving general level of English LLS) fall under the language study cluster.

Participants had to assess the frequency with which they employed each of the 17 strategies by means of a six-point Likert scale (1 = 'never or hardly ever', 2 = 'very rarely', 3 = 'rarely', 4 = 'sometimes', 5 = 'often', 6 = 'always or almost always'). As will be explained in chapter 5, for the purpose of statistical analysis, strategies were

considered with a threefold approach: individually, i.e. observing the connection between each individual LLS and CS effectiveness; separating deep and surface LLS to test whether there is a connection between one of these types of LLS and CS effectiveness; and as a whole, i.e. adding the frequency of use of all the strategies to obtain a single numeric value per participant.

4.2.3. Proficiency test: X/Y_lex

The X_lex and Y_lex software (Meara, 2005; Meara & Miralpeix, 2006) measures vocabulary size in L2 learners. It is free software available to test English, French, Portuguese, Spanish and Swedish vocabulary size. According to the software manual (Meara, 2006), however, the development team particularly trusts the English version as a measurement for 'overall' proficiency:

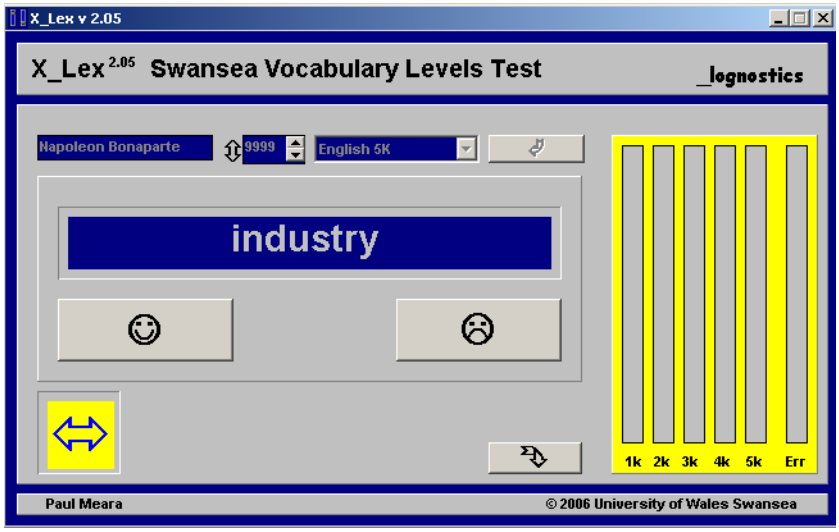
> […] for English there appears to be a reasonably good relationship between vocabulary size and other aspects of language competence, and as a result of using the tests with speakers from many different L1 backgrounds, and at many different levels of proficiency, we are able to use the X_Lex scores as a shorthand for overall proficiency levels. (Meara, 2006)

Despite vocabulary size being a partial measure of proficiency, the correlation Meara mentions between English X/Y_lex scores with other proficiency measures has been shown in other research (Milton et al., 2010) and vocabulary size has been resorted to before as a proficiency measure (Schoonen, 2010; Staehr, 2008). Based on this information, the X/Y_lex test was administered in the present study in order to compare the SA and AH groups and ensure they were not different at the beginning of the testing period in this respect. Additionally, it was later decided to use the participants' scores on this test to investigate connections between proficiency level as measured by vocabulary size and CS use and development. The scores could also

allow for classification of participants into proficiency groups to implement more varied statistical tests. Moreover, participants also took the X/Y_Lex in T2 so that it could be observed whether progress in vocabulary size as a measurement of progress in proficiency was connected to progress in CS effectiveness.

The X/Y_Lex is a Yes/No vocabulary test which presents participants with different words (50 in X_lex and 50 in Y_lex, which represent a vocabulary size of 5,000 words on each part of the test) and asks them if they know each word or not. Participants have to click on a button with a smiley face if they know the word or on another button with a sad face if they do not (see screenshot below in figure 4.3). Both parts of the test include 20 non-words, so that if test-takers say they know a word that does not exist, the result corrects itself by taking points off. The percentage of points taken from the raw score corresponds to the percentage of non-words the participant claims to know out of 20 non-words. For example, if a participant claims to know all the real words on one part of the test (raw score 5,000) but also indicates they know four non-words (4/20 = 20%), their corrected score will be 4,000 (5,000 – 20%). Both the raw score and the corrected score are provided at the end. Participants in the present study were told about this in advance, together with the instructions to take the test.

Figure 4.3: Screenshot of X_lex test item



The X_lex part of the test is aimed at intermediate L2 learners because it tests vocabulary with a higher frequency of use which they may have learned from the beginning to intermediate stages of the L2 learning process. Meanwhile, the Y_lex is addressed at advanced L2 learners since it presents vocabulary with lower frequency of use, words which are usually learnt at later stages in the L2 learning process than the vocabulary included in X_lex. Considering participants in the present study were university students, they were expected to be around the upper-intermediate level of English, so they took both parts of the test and the two scores were added together in order to obtain one single numeric value per participant regarding their estimated vocabulary size out of 10,000 words.

4.2.4. Attention control test: the Trail-Making Test

As explained in chapter 2, cognitive factors are being looked into in the present study as possible determinants of CS use and development. More specifically, this dissertation examines attention control and analytic ability, due to their possible connection to problem-solving competence. Attention control, which is a component of linguistic aptitude involved in monitoring speech, is 'mainly operationalized as the ability to rapidly shift attention to different levels of linguistic information' (García-Amaya & Darcy, 2013). It has been measured before through both language and non-language related tests: Dimensional Change Card Sort Task (Bialystok & Martin, 2004), the Stroop test (Zied et al., 2004), the Simon task (Cox, 2013) and the metalinguistic categorization task in Segalowitz and Freed (2004). The test chosen for the present study, though non-language related, has been implemented before in language acquisition studies (Bialystok, 2010; Isaacs & Trofimovich, 2011). The Trail-Making Test (TMT) actually involves visual perceptual ability and motor speed in both parts A and B, while part B also requires task shifting, planning and inhibition (Bialystok, 2010).

The TMT consists of two simple tasks (see appendix I): in the first one, participants are presented with a page full of numbered dots and they have to connect all the dots in ascending order drawing a single line from number 1 to number 25 while being timed. In the second task, participants are presented with a similar page where half the dots are numbered (1 to 13) while the remaining 12 dots are marked with letters (A to L), and they have to draw a similar line while being timed, but this time they have to alternate numbers and letters in alphabetical order (so it should go 1-A-2-B-3-C-4-D…). The result of the test is a numeric value obtained by subtracting the time needed for the first task from the time needed for the second one. That way, the extra time needed for attention shifting between two ordering criteria (numbers and the alphabet) reflects the attention control ability, where a lower value would mean more attention control because the introduction of a second criterion and the alternation between both is less disturbing than for those with lower attention control, who need more extra time for the second task. For example, if participant A took 33.47 seconds to complete the first task and 49.93 seconds to complete the second one, their score would be 16.46 (49.93 – 33.47). If participant B took 33.51 seconds in the first task and 39.93 in the second one, their score would be 6.42. This would mean that participant B showed better attention control than participant A.

The TMT was deemed the most appropriate instrument for the purposes and conditions of the present study. Other tests, such as the metalinguistic categorization task in Segalowitz and Freed (2004), rely on the implementation of specific software (e.g. Praat, E-Prime) that presents the items on screen and measures the speed and accuracy of the participants' responses. Considering the fact that data were to be collected via Skype (see subsection 4.2.6), if any special software was used, participants would have to set it up themselves on their own computers, which would entail longer data collection sessions and decrease the researcher's control over task administration. The TMT had been used before to measure attention control and connect it to language ability, and its simple administration procedure did fit the

videoconference conditions, although it had to be slightly adapted to computer logistics, i.e. the trail would be made drawing with a mouse on the screen instead of pen and paper, by means of Paint software (available in all Windows PCs). Participants would inform the researcher of the moment when they would start and finish drawing the lines in the right order so the tasks could be timed and they would save their drawings and send them to the researcher later, (together with the results of the other tests and the filled out questionnaires) to ensure that they had done the test correctly.

4.2.5. Analytic ability test: LLAMA F

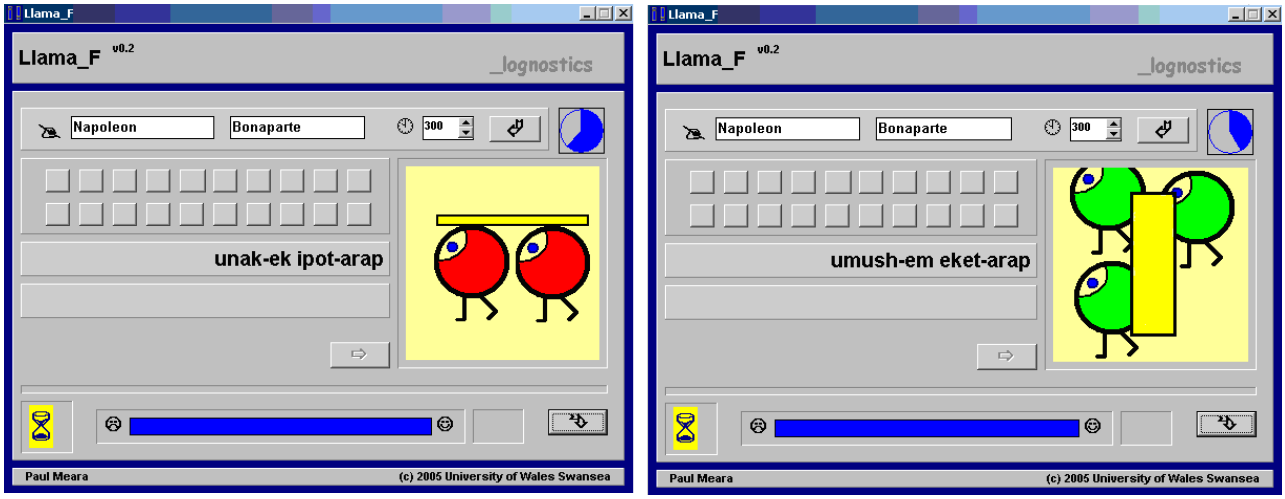
Analytic ability is the other cognitive factor, together with attention control, to be tested in the present study in order to observe its possible connection with CS use and development. As explained in section 2.2.2.2, this factor has been measured in prior research by means of other tests like the Words in Sentences MLAT subtest or grammar judgment tests (GJT), for example in DeKeyser (2000). In this research, this cognitive ability was defined as the ability to infer grammatical rules, as tested by means of the LLAMA F software (Meara, 2005).

This test is part of the LLAMA test battery, which includes four tests that measure different components of linguistic aptitude: LLAMA B is a vocabulary learning task, LLAMA D is a test of phonetic memory, LLAMA E is a test of sound-symbol correspondence and LLAMA F tests grammatical inferencing. The test used in the present study, LLAMA F, is an updated version of the Lat_C test, which was part of a previous language aptitude test battery. According to the authors, the older Lat_C had proven to be satisfactory in its measurement of analytic ability, but the LLAMA battery is an L1-independent aptitude test set, and the grammar inferencing test was adapted to work on L1 speakers of any language. Despite the creators' warning to use all of their instruments with caution, the LLAMA tests have already been implemented in previous research (Artieda, 2013) and tested for internal consistency with positive results (Granena, 2013), which is why one of the parts of this battery of tests, LLAMA

F, was selected for this study. Moreover, it is a computer-administered test and it takes ten minutes to perform, which means that, apart from being trustworthy in predicting analytic ability applied to language learning, it is also a practical instrument.

The test presents the participants with a series of pictures, each with a short phrase that describes the corresponding picture in an unknown language. By observing the different pictures and descriptions, test-takers are to infer different grammar features and are given limited time, five minutes, to infer the rules to construct short phrases in this language. The phrases refer to the shape, number and colour of the elements in the picture and their position relative to the yellow bar (see examples below in figure 4.4). Test-takers click on the square buttons on the software interface to see each picture-phrase combination as many times as they want within the time limit, and they have the possibility to take notes on a piece of paper so they do not have to memorise anything.

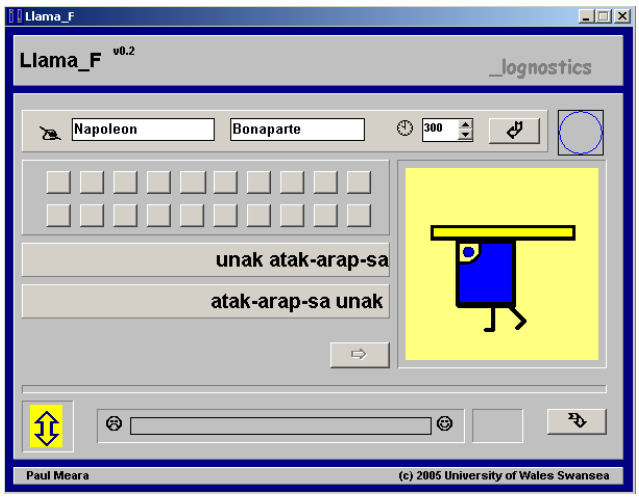
Figure 4.4: Screenshots of LLAMA F learning items



After the learning time, the programme provides a 20-item test, each item consisting of a picture and two possible short phrases to describe it (see screenshot below in figure 4.5), from which the test-taker has to select one. The programme

produces a different sound depending on whether each answer is correct or incorrect. The results of the tests are expressed as a percentage of correct answers at the end.

Figure 4.5: Screenshot of LLAMA F test item



4.2.6. Data collection

As mentioned above, data were collected at the beginning and at the end of each programme (September and January). In both cases the data collection was carried out in individual sessions via Skype. This decision was made in light of the fact that SA participants would be from different home universities and going to different host universities abroad, so face-to-face data collection would have been logistically impossible. Besides, if recruitment had been limited to a single home or host university, ideally the University of Barcelona as the AH group, probably fewer participants would have been recruited. Therefore, this seemed the most effective data collection system for the SA group to avoid losing participants for practical reasons. This way, participants could arrange an appointment with the researcher at any time within a three-week period in September and a similar period in January, and they could take the tests anywhere that suited them. Skype also allowed the researcher to monitor every task in the same way as face-to-face data collection, i.e. to give all necessary instructions in real time and to answer any doubts if instructions were not clear, as well

as to interact with the participant to carry out the CS elicitation task and audio record it. The necessary materials were sent to the participants via e-mail a few minutes before every data collection, so they would download them onto their own computers and later send back the results at the end of the session. Since this system was to be implemented with SA participants, the same procedure was followed with AH participants to control for differences in data collection conditions.

The data collection sessions were 35 minutes to one hour long, depending on the participant and technical issues. Also, the distribution of tests was the same for both data collection rounds, with the only exception being the trail-making test for attention control. The following table represents the tests administered in every session:

Table 4.7: Distribution of tests per data collection session

	T1	**T2**
2012-2013	Linguistic background and LLS questionnaire, CS task, XYlex and LLAMA F (approx. 45 minutes)	Language use and LLS questionnaire, CS task, XYlex and trail-making test (approx. 45 minutes)
2013-2014	Linguistic background and LLS questionnaire, CS task, XYlex, LLAMA F and trail-making test (approx. 55 minutes)	Language use and LLS questionnaire, CS task and XYlex (approx. 35 minutes)

In the 2012-2013 round, tests that had to be taken only once per participant and that measured presumably stable factors (trail-making and LLAMA F) were distributed between September and January sessions to keep them equal in duration. However, after this first round, high mortality (33%) was observed and participants seemed to have difficulties to find time during the January exam season to schedule the T2

session. As a consequence, it was decided that in the 2013-2014 data collection round as many tests as possible would be administered in the September session, when participants seemed to have more free time, and so the January session would not take them as long and it would be easier for them to schedule. Also, if participants failed to take the tests in T2, information about both stable factors and the initial levels of other factors could still be used in further studies.

4.3. CS effectiveness analysis

For practical reasons, prior to the analysis of CS effectiveness, the material collected by means of the CS elicitation tasks was reduced to five T1 and five T2 item descriptions, as predicted in the procedure and result report of the CS task pilot study in section 4.2.1. A preliminary CS identification round was carried out on the 22 descriptions produced by ten of the participants (five AH and five SA participants) and the items that elicited a higher number of CS instances were selected for analysis, namely the eye, the house, the tent, the elephant and the teapot from T1; and the target, the glass, the present, the camera and the telephone from T2. Therefore, the analysed corpus consisted of 650 item descriptions, five per participant and data collection. The decision to collect an extensive corpus and then only analyse the descriptions that encompassed the highest quantity of CS instances made it possible to focus the analysis on the most productive items. At the same time, it minimised the risk of not finding collaborators for the rating approach explained below, since the time needed from the raters would be less than half of what it would have taken them to rate 22 descriptions of 65 participants (a total of 1430 descriptions).

The chosen corpus was subjected to two different analysis approaches: the mini-Delphi CS scale (Montero, 2011; Montero et al., 2013) and raters' individual description assessment. The intention behind this double approach was to triangulate

the observation of the inherently subjective matter under study, i.e. the effectiveness of CS use. On the one side, the mini-Delphi scale offered a *systematic* quantification of how successful the participants were in trying to get their messages through based exclusively on the type and frequency of CS they used regardless of context. On the other side, raters would complement the scale with context-dependent evaluation, even if possibly contaminated by other factors, such as general proficiency. Each procedure is thoroughly described in subsequent subsections.

4.3.1. The mini-Delphi CS scale

The scale was developed as part of a previous study (Montero, 2011; Montero et al., 2013) as a tool to operationalise CS effectiveness. It consists of a list of CS types, based on Dörnyei and Scott's (1997) CS inventory (see section 2.1.3 and appendix A), and a numeric value or points assigned to each CS. This inventory was chosen to develop the scale because it was the most comprehensive of those reviewed, as it is actually a compilation of previous taxonomies (Bialystok 1983; Dörnyei and Scott 1995; Faerch and Kasper 1983; Paribakht 1985; Poulisse 1993; Tarone 1980; Willems 1987). It comprises 33 CS types plus variants of some of them, with descriptions and examples of each one. The following paragraphs present how the mini-Delphi scale was created, describe the instrument in detail and explain how it was implemented in the present study.

As explained by Montero et al. (2013), the instrument was the result of applying a 'mini-Delphi' process (Berquez, Cook, Millard & Jarvis, 2011). The Delphi method is a procedure to consult experts' opinions and usually find consensus on a certain issue within their domain, commonly used as a forecasting method (hence the name 'Delphi'). The 'mini-Delphi' is the face-to-face version of that procedure, in which the experts to be queried are gathered together and first asked open questions to discuss, usually in several rounds, so that opinions may evolve by building on other expert opinions (Cuhls, 2001). In a second and final stage of the process, a Likert scale is

developed based on the common ground found during the first discussion rounds, and administered to the same panel of experts. In the case of Montero et al. (2013), this procedure was adapted to fit the needs of the study: there was one single round where a panel of three experts (all of them Applied Linguistics researchers and native ESL teachers) was presented with the already finalized Likert scale, on which the different CS types with definitions and examples were the items of the Likert-scale questionnaire, but the experts were requested to reach a consensus to complete it together. This way, the opportunity to build their opinion on other expert views and to obtain a final expert answer (which is the nature of the Delphi method) was still preserved.

During the mini-Delphi session, the panel of experts received information about CS, definition and examples, and about the study and the purpose of developing the scale as a systematic measuring instrument for CS effectiveness in monologue narration with visual support (cartoon strips), which was the CS elicitation task administered in Montero et al. (2013). They were also offered the opportunity to ask any questions about the questionnaire, the Likert scale or any other related issue, since they were experts in language acquisition and communication but not necessarily in communication strategies. The panel was presented with a selection of CS types from Dörnyei and Scott's (1997) inventory (first column in table 4.8; the selection criteria are explained in subsequent paragraphs), with a definition and an example to illustrate each of them. Then the experts were asked to discuss and reach a consensus to assign a numeric value to each CS type based on how effective they agreed it generally was in monologue narration with visual support: 2 points for 'effective' CS, i.e. strategies that succeed completely in getting the message through; 1 point for 'partially effective' CS, i.e. strategies that present drawbacks for complete message transmission; and 0 points for 'ineffective' CS, i.e. strategies that do not contribute to communication or even hinder it. The table below presents the result of the mini-Delphi session. The examples correspond to transcription extracts of the cartoon strip stories.

Table 4.8: CS effectiveness mini-Delphi scale (Montero et al., 2013: 7-8)

Strategy	**Description**	**Example**	**Effectiveness**
1. Message abandonment	Leaving a message unfinished because of some language difficulty	'and the hole is very hmm is very hmm is very I don't know and the fourth…' [moving on to the next picture in the strip]	0
2. Message reduction (topic avoidance)	Reducing the message by avoiding language structures or topics considered problematic languagewise or by leaving out some intended elements for a lack of linguistic resources	'they put bread and and hmm well they put the lunch' [the speaker tries to name all lunch components but reduces the message when faced with lack of vocabulary]	1
3. Message replacement	Substituting the original message with a new one because of not feeling capable of executing it	[false start][5]	1 (unified: 2)
4. Circumlocution (paraphrase)	Exemplifying, illustrating or describing the properties of the target object or action	'the ball like it comes up because of the water, it doesn't sink, comes up' [float]	2
5. Approximation	Using a single alternative lexical item, such as a superordinate or a related term, which shares semantic features with the target word or structure	'the lunch that was in the in the *bag* so to say' [basket]	1
6. Use of all-purpose words	Extending a general, 'empty' lexical item to contexts where specific words are lacking	'he gets *something* with water' [bowl]	2 (if used just once)
7. Word coinage	Creating a non-existing L2 word by applying a supposed L2 rule to an existing L2 word	'the snake scaries the boys' [applying 3rd person –s to the adjective 'scary']	2
8. Restructuring	Abandoning the execution of a verbal plan because of language difficulties, leaving the utterance unfinished, and communicating the intended message according to an alternative plan	[false start]	2 (unified)
9. Literal translation (transfer)	Translating literally a lexical item, an idiom, a compound word or structure from L1 to L2	'the boys… prepare some *dinner* to go to the mountain' [lunch in Catalan is *dinar*]	1
10. Foreignising	Using a L1 word by adjusting it to L2 phonology (i.e. with L2 pronunciation) and/or morphology	'in a little *collin*' [hill in Spanish is *colina*]	1
11. Code switching (language switch)	Including L1 words or chunks with L1 pronunciation in L2 speech	'the dog hmm is looking is looking the *cistell*…' [basket in Catalan is *cistell*]	0

[5] See explanation of CS 3, 8 and 16a below in this subsection.

12. Use of similar-sounding words	Compensating for a lexical item whose form the speaker is unsure of with a word (either existing or non-existing) which sounds more or less like the target item	'one boy can't *arrive* [pronounced as /arrif/] to the ball'	0
13. Mumbling	Swallowing or muttering inaudibly a word (or part of a word) whose correct form the speaker is uncertain about	'they *boke xxx* water in the hole'	0
14. Omission	Leaving a gap when not knowing a word and carrying on as if it had been said	'but he can't the ball' [get back, reach]	0
16a. Self-repair	Making self-initiated corrections in one's own speech	[false start]	2 (unified)
18. Over-explicitness (waffling)	Using more words to achieve a particular communicative goal than what is considered normal in similar L1 situations	'some people, some young, some not that young, are they young? I don't know, there's four people'	1
19. Mime	Describing whole concepts nonverbally, or accompanying a verbal strategy with a visual illustration	'the dog is trying to get some food of the the *this*.'	2
23. Verbal strategy markers	Using verbal marking phrases before or after a strategy to signal that the word or structure does not carry the intended meaning perfectly in the L2 code	'the ball hmm go up *I don't know how to say it,* is like swimming'	2
24a. Direct appeal for help	Turning to the interlocutor for assistance by asking an explicit question concerning a gap in one's L2 knowledge	'the ball go to one *agujero, ¿cómo se dice agujero en inglés?*' [hole, how do you say hole in English?]	L1 (SP/CAT): 0
		Participant: 'Sam opens the *what's the name?*' Researcher: 'basket'	L2 (ENG): 2
24b. Indirect appeal for help	Trying to elicit help from the interlocutor indirectly by expressing lack of an L2 item either verbally or nonverbally	'they boys are very *preocupados?*' [worried?]	L1: 0
		Participant: 'the boys want to *hmm I don't know hmm*' Researcher: 'take it'	L2: 1

Each CS occurrence can thus be coded for effectiveness by means of the mini-Delphi scale, and an effectiveness total represented by a numeric value can be calculated for each participant at each testing time. This effectiveness score is presented as a percentage of the maximum number of points the speaker would be

awarded if all the CS they used in a specific speech production were ‘effective’ (2 points each), i.e. the amount of CS found in the speech would be multiplied by 2 and the total points this participant actually got according to the scale would constitute a percentage of the latter result. For instance, participant A used 4 strategies: 1 circumlocution (2 points), 2 approximations (1 point each) and 1 omission (0 points), which add up to 4 points total. Since 4 CS were used, the participant could have been awarded a maximum of 8 points; 4 points = 50% of 8 potential points; effectiveness score for participant A is 50%.

As can be observed on the table, the original taxonomy from Dörnyei and Scott (1997; appendix A) underwent several modifications for the purpose of Montero et al.’s study. First of all, the original model included both direct and interaction strategies. Since the study analysed monologue productions, although they did include help requests, most interaction strategies were discarded from the selection presented to the panel. Only direct strategies and help-request interaction strategies were considered. Also, some of the direct strategies from the original taxonomy, although they were identified in the speech productions, were disregarded. This is the case with “processing time pressure” strategies, commented on in section 2.1.1. Examples of such CS are ‘retrieval’ (“in an attempt to retrieve a lexical item, saying a series of incomplete or wrong forms or structures before reaching it”) and ‘use of fillers’ (“using gambits to fill pauses, to stall, and to gain time in order to keep the communication channel open”, represented on the transcriptions as ‘hmm’ or ‘ehh’). These strategies contribute to keeping the communication channel open, but they do not seem to contribute to, or interfere with, message fulfilment. In fact, they have been previously considered as ‘production strategies’ rather than ‘communication strategies’ (Tarone, 1980), since they are not used for meaning negotiation.

Secondly, what appeared on the audiorecordings and transcriptions as a false start for a sentence could be interpreted as both a ‘message replacement’, ‘re-structuring’ and ‘self-repair’ according to Dörnyei and Scott’s (1997) classification, and

there was no information about the speaker's intention to tell which strategy was being used in each case. For instance, '…but hmm *a snake* hmm they found a snake so…' could be interpreted as either 're-structuring' if the syntactical function of 'a snake' was being switched from subject to object, or as 'message replacement' because the speaker may have found some difficulty with the message whose subject was 'a snake' and decided to move on to another message. Since it was impossible to tell them apart, these three CS were coded as one single strategy. It must be noted though that during the mini-Delphi process they were presented separately and two of them, 're-structuring' and 'self-repair', were judged as 'effective' by the panel (2 points), while 'message replacement' was deemed 'partially effective' (1 point). Considering the need to code them together as one single strategy due to lack of information to differentiate them, and since two out of three were regarded as 'effective' by the mini-Delphi panel and the third one not as completely ineffective, the final unified strategy was considered as effective for statistical analysis purposes.

Finally, both 'direct appeal for help' and 'indirect appeal for help' were duplicated to distinguish when such requests were uttered in the speaker's L1 or in the L2 (English). As a consequence, these two strategies became four: 'direct appeal for help: L1', 'direct appeal for help: L2', 'indirect appeal for help: L1' and 'indirect appeal for help: L2'. This decision was made on the grounds that the mini-Delphi panel observed that the effectiveness of requesting help, either directly or indirectly, varied depending on the language in which the appeal was expressed.

The mini-Delphi scale was originally intended to be implemented in *monologue* productions with visual support, as was the case of the study described in Montero et al. (2013). In fact, the mini-Delphi panel was presented only with the selection of CS previously identified in that study's corpus. The participants in Montero et al. were not supposed to ask questions during the task, just produce a monologue telling the story reflected on a comic strip. However, as the previous paragraph implies, some of them did try to interact with the researcher by asking for help with the vocabulary they were

missing. Therefore, as mentioned before, the scale did not include most interaction strategies considered in Dörnyei and Scott's inventory, with the only exception of direct and indirect appeals for help.

In the present study, since participants were expected to interact to some extent with the interlocutor, instances of more interaction strategies could have occurred. If that had been the case, the same mini-Delphi procedure would have been carried out on those strategies in order to measure their effectiveness: a panel of experts would have been gathered and consulted about the effectiveness of those strategies that were not included in Montero et al. (2013). However, instances of such strategies (e.g. asking for repetition, asking for confirmation, interpretive summary and comprehension check) were not found in the final corpus for analysis of the present study, so there was no need to extend the scale. The chances for this kind of interaction may have been reduced because the task description was modified from Khan's (2010) procedure, which established that the task would be carried out by two participants at a time. In this case, the researcher would only prompt the participant to cover all the details in the description instead of having the participant listen and understand their interlocutors' descriptions as well. The role of the researcher was in fact to formulate questions that avoided the difficult vocabulary needed to describe the picture. As a consequence, participants did not need to resort to interaction strategies to check their own understanding of the interlocutor's message (named 'other-performance problem-related strategies' in Dörnyei & Scott, 1995) and not much room for negotiation was left. They could have used 'own-performance problem-related strategies' (comprehension check, own-accuracy check) but they did not.

In case CS types not included in the mini-Delphi scale were to be found in the corpus under study, Dörnyei and Scott's (1997) taxonomy was used to identify CS in the transcriptions of T1 and T2 oral CS elicitation tasks. It must be noted that CS type identification in this study was based on evidence found within the transcriptions and taking into account the materials and description of the task. Similarly to Montero et

al.'s study, participants were not questioned as to their intentions when replacing the needed vocabulary with some other compensating expression. The use of CS is, therefore, *interpreted* from the data. In the following table, some illustrations of this interpretation are provided:

Table 4.9: Examples of CS identification

Transcription extract	**Identified CS**
*INV: how is it locked? *1111: because it is it is i don't know what to say it it is ehh like envolted ehh with a verge I don't know what what it means hmm it seems to be or also can could be a grass a a grass so so big	Topic avoidance: 'fence / gate'
*1118: […] there is also a...hmm a well a device to expel the the the smoke of the kitchen i don't know how it's said	Circumlocution: 'chimney'
*1217: […] ah it's doing with the trump the... I suppose is the trump I'm not sure it's like going up	Foreignising: from *trompa* in Spanish, instead of 'trunk'
2110: […] it has big ears and four legs a little tail and a trumpet	Approximation: 'trumpet' for 'trunk'[6]
*1107: the tent […] it has these strings subjecting the tent	Literal translation: from *subjectar* in Catalan, instead of 'holding'

[6] 'Trumpet' is an existing word in English that refers to an object that shares features with the intended concept: both trumpet and trunk refer to a sort of pipe through which air runs producing a certain sound. Meanwhile, in the case of 'trump', there is no such semantic connection to be found. Therefore, it is interpreted that the participant in the previous instance was just adapting the Spanish word for 'trunk' phonetically to make it sound like an English word (foreignising).

*1118: […] the tent is ehh fixed to the ground with...with six ehh I think they are hmm well fixments	Coinage: for ‘pins’
*2210: number ten ehh first of all we can see an elephant which which has his...I don't know the word but is is positioned in the left side	Omission: ‘trunk’

Due to the interpretive nature of the CS identification procedure described above, an inter-rater reliability check was carried out: 15.4% of the transcriptions, which amounted to 100 transcriptions (10 per item), were double-coded by an American ESL teacher, highly proficient in Spanish and Catalan. The collaboration of someone with deep linguistic awareness and knowledge of the main languages involved was important for this procedure, since they were expected to tell when a non-existing word was based on the L1 (foreignising) and when on the L2 (coinage), and to identify literal translations. The result of this inter-rater reliability check was 80% agreement, which is a commonly accepted percentage in the field (Marques & McCall, 2005).

After identification of all CS instances in the selected corpus, the scale was implemented as explained above over all the CS used by each participant during each testing time, resulting in a numeric value, the CS effectiveness score explained above, so each participant would have a T1 and a T2 effectiveness score with which to carry out statistical analyses in order to answer the research questions of the present study. For example, if participant A used 2 approximations to describe the eye, 3 circumlocutions and 1 omission for the house, 1 mumbling and 2 literal translations for the tent, 1 similar-sounding word for the elephant and 2 circumlocutions for the teapot, i.e. 14 effectiveness points out of 24 potential points (12 CS used multiplied by 2 maximum points each), their effectiveness score for T1 would be 58.3%.

One possible drawback of implementing the mini-Delphi CS scale to assess effective use of CS is that, since it quantifies effectiveness systematically based on the

number and type of CS used by speaker, it disregards the context-dependent nature of CS effectiveness. As discussed in chapter 2, it is assumed that a certain CS may be effective in some communicative contexts but not in others. For example, abandoning a message halfway, while totally ineffective in most communicative situations, might prove effective in cases where the interlocutor has enough previous knowledge to understand the intended message. Being aware of the possibility that applying the scale would offer a somewhat partial perspective of participants' use of CS, a second assessment approach was implemented to triangulate this observation and complement the information provided by the mini-Delphi instrument. The following subsection describes this parallel analysis method.

4.3.2. Raters

The second analysis approach to quantify CS effectiveness in the CS elicitation task consisted in having three raters assess every item description from each participant individually. The individual rating approach has been implemented before in other studies on the use and effectiveness of CS, such as Haastrup and Phillipson's (1983). By means of this method, parallel to the mini-Delphi scale implementation, whatever role context may play in CS effectiveness, a role that is not reflected in the mini-Delphi score, would be accounted for. Raters were to be briefed on the general goal of the study and the concept of CS, although they would not be instructed in detail about specific CS types. The purpose of this methodological approach was to get the raters to assess each description as a whole, i.e. not every strategy individually but as part of a context. Nonetheless, raters were expected to base their judgment as much as possible on the intelligibility of each production, on whether or not they could understand the description despite the lack of vocabulary of the participants, and not to penalise participants in their evaluation for lacking the appropriate vocabulary.

For the purposes of the in situ approach to CS effectiveness evaluation, the collaboration of three Applied Linguistics postgraduate students was needed. All three

of them were English L1 speakers with no knowledge of Spanish or Catalan. Apart from the introduction to the study and the explanation of what purpose they were expected to fulfil, raters were informed about the specific goals of the oral task, all the instructions the participants had received (see task description in subsection 4.2.1), and they were asked to rate on a 3-point scale how effective each description was (1 = not effective, i.e. a standard English speaker would not understand; 2 = partially effective; 3 = effective, i.e. a standard speaker would fully understand). See instructions provided to the raters, apart from oral briefing and in situ assistance throughout the session, in appendix J.

All three collaborators rated the whole reduced corpus, five T1 and five T2 descriptions performed by every participant (a total of 650 valid descriptions, plus 30 descriptions from participants who joined the study in the second data collection year but then failed to participate in T2). The assessment was carried out on the transcriptions of the oral tasks in order to prevent background noise, bad sound quality or accents from interfering with their concentration on the strategies. For each description performed by each participant, the mean of the three values awarded by the raters was calculated. The scores used for statistical analysis in this study are the means of the five descriptions from each data collection, so for each participant there is a T1 score and a T2 score according to this analysis approach. The table below shows an example to illustrate this calculation of both scores. Under each item described, the scores awarded by the three raters are expressed with the mean of those three values underneath them.

Table 4.10: Example of raters' score calculation

	Eye			**House**			**Tent**			**Elephant**			**Teapot**			**T1 score**
Raters	2	2	3	1	1	2	2	1	2	1	1	1	1	2	1	
M	2.3			1.3			1.7			1			1.3			1.52

	Target			Glass			Present			Camera			Phone			T2 score
Raters	3	2	3	2	2	2	3	2	2	2	2	1	3	3	3	
M	2.7			2			2.3			1.7			3			2.34

Raters were also asked to perform a think-aloud protocol to express the type of strategies that, in their opinion, usually contributed to effective or ineffective communication. In fact, they did this twice: 30 minutes after the beginning of the assessment task, which was audio recorded, and towards the end, when they wrote down a few lines explaining their position. The idea was to ensure as much as possible that they remained focused on the effect of CS on the descriptions and that they had been consistent in their judgments, especially considering they did their assessment in two separate sessions: they assessed 75% of the corpus in November 2013 and the remaining 25% in February 2014, after the last data collection and oral task transcription processes. Before the second session, they had to listen to the comments they had made during the first one. It took each one of them around eight hours in total to assess the whole corpus, six for the first session and two for the second one.

Through the think-aloud recordings it can be appreciated that the opinions and examples provided by the raters show that they agree with the mini-Delphi scale in general. Even though they were not offered any information on the scale, when asked what made a description effective or not effective, the raters' observations referred directly to certain CS types and compared them. For example, in the following extract we can find a distinction between circumlocution (effective CS type according to the mini-Delphi scale) and approximation (partially effective):

> RATER 1: …for example the chimney…I'm trying to make a distinction. Obviously if someone says 'there's something on the roof and there's smoke coming out', I can understand what that means. But if someone says 'there's a fire', I don't think that's clear enough.

RESEARCHER: There could actually be a fire on roof.

RATER 1: Yeah, yeah, it's not specific enough.

And here follows a distinction between circumlocution and message abandonment (not effective according to the scale) and code-switching (also not effective):

RATER 2: … they'll often say 'the thing that smoke comes out' […] I don't think that hinders intelligibility at all. It's completely fine. Even when they give the example of 'something around the garden' I think they say ehm 'to help their privacy' I think that it still communicates even if there's lack of vocabulary […]

RESEARCHER: And what kind of thing hinders communication or gets you confused or…?

RATER 2: I think that it's partly motivation like some of the students they just give up. Some of them clearly try to communicate their ideas while others are 'I don't…'. They just abandon. […] and obviously code-switching, when they are using Spanish words or Catalan words.

RESEARCHER: Yeah usually what they are saying is 'I don't know how to say this in English'.

RATER 2: Yeah, that's… For me, I sort of prefer the students who don't refer back to their L1 when they… it feels really…kind of puts me off when I see it. I have to read the whole thing again.

This analysis approach presents the advantage of evaluating CS use in context, as opposed to the mini-Delphi scale. However, the fact that raters were exposed to the whole transcripts of speech production (not only to the CS utterances) and that they spent hours doing the assessing task may have entailed certain drawbacks. Some of

the comments provided by the raters in the audiorecorded think-aloud protocol and their final written reflections could hint at the collateral effects of this type of assessment of CS effectiveness. A possible consequence of being exposed to information beyond that which they were supposed to evaluate and being asked to provide a holistic score on each description was that they might have, consciously or unconsciously, taken other aspects of speech production into consideration, apart from whether they could understand the message despite lexical shortcomings. In fact, one of the raters mentioned that lack of organisation of the information within each description as a factor in his rating and that he deemed shorter more direct descriptions as more effective.

This comment may reflect as well that the specific conditions of the CS elicitation task were not taken into account at all times (maybe due to exhaustion from assessing for hours). The task involved participants describing each picture in enough detail so that their interlocutor, *who was looking at a similar set of pictures*, could decide if each item was identical or slightly different to hers. Another rater also shared with the researcher that some descriptions were not specific enough. When reminded of the parallel set of pictures that the interlocutor was looking at while receiving the descriptions, the rater admitted she thought she had to be able to imagine the picture, which would require speakers to be more specific. Other than these isolated incidences, raters' comments dealt for the most part with different instances of CS use they paid special attention to, either because they compensated for lack of vocabulary or because they hindered communication.

After both analysis approaches were implemented, some individual descriptions were discarded: those where no CS instances had been identified when following the mini-Delphi method (n=54, 8.3%) and those where raters had proven to be in extreme disagreement (n=63, 9.7%), i.e. descriptions that had been awarded three points by one of the raters and one point by another rater. Five descriptions fulfilled both requirements for exclusion, so the final number of discarded descriptions was 112

(17.2%). As a result, the analysis of 538 descriptions according to both approaches was subjected to statistical tests in order to answer the research questions proposed.

4.4. Summary

This chapter has first described the participants recruited for the present study. Participants in both context groups (SA n=33; AH n=32) chose to participate voluntarily through different recruitment processes. They were all Spanish undergraduate learners of English at the intermediate to upper-intermediate level, though some of them also spoke Catalan or Galician. SA students spent four months in an English-speaking country as part of a university exchange programme, while AH participants took 13.5 hours a week of university courses in English in Spain. The information collected in reference to the linguistic experience of the participants in each context revealed that SA participants perceived a higher degree of improvement of their linguistic skills as compared to their AH counterparts. Both groups had had similar habits of reading and watching TV products in English, but the SA students took part in more organised leisure activities in L2 and had more relationships in which communication was held in English in both their personal and academic or professional environments than the AH participants.

Second, the chapter has presented the instruments and procedure implemented during data collection for the present study and it has explained the criteria applied in every decision made throughout the process. CS were assessed both at the beginning (T1) and at the end (T2) of the SA or AH programme by means of Khan's (2010) information gap picture description task and a second similar set developed to that end. CS instances were identified on the transcripts of this task using Dörnyei and Scott's (1997) CS inventory. Information about LLS use was collected with Tragant et al.'s reduced Likert-scale questionnaire. Proficiency was measured as vocabulary size by

administering the X/Y_lex test, both in T1 and T2. The Trail-Making Test for attention control, particularly task shifting, planning and inhibition, was adapted to computer logistics and analytic ability was assessed as grammatical inferencing by means of the LLAMA F test.

Finally, the double analysis and operationalisation method applied to measure CS effectiveness have been presented and justified. On the one hand, the mini-Delphi scale (Montero et al., 2013) provided a completely CS-focused systematic assessment, though it lacks consideration of the communicative context. On the other hand, three raters assessed communicative effectiveness of each individual speech production, so context was considered, though other factors may have affected the focus of assessment in some cases. The double approach was intended to provide a more complete perspective on CS effectiveness by including two methods that compensate for each other's possible drawbacks. The following chapter covers the procedures related to the statistical analysis of the data.

5. Results

This chapter includes the results of all the statistical tests performed on the data collected. It is organised in three different sections: 'descriptive statistics', where a global overview of the data obtained is provided; 'preliminary analysis', where the triangulation between both CS effectiveness analytical approaches and comparability between context groups are observed; and 'inferential statistics', where the specific tests to answer each of the questions are presented.

5.1. Descriptive statistics and normality tests

This section presents descriptive information, i.e. maximum (Max.), minimum (Min.), mean (M) and standard deviation (SD), for each learning context group (AH and SA) and for the total sample in all the variables involved in the present study. Each variable is presented in a different subsection within this section.

With regard to normality assessment, considering the size of the total sample (65) and of each context group (AH: 32; SA: 33) and based on the central limit theorem (CLT), all variables could be considered normally distributed. CLT establishes that, with 30 or more subjects, the sampling distribution of the mean will approach normality (Reid, 2014: 229) and therefore it would be safe to use parametric tests for all the research questions of the present study. However, in order to be as cautious as possible, the Shapiro-Wilk normality test was consulted to decide whether a parametric or a non-parametric test should be implemented. Therefore, the results of the Shapiro-Wilk test for normal distribution are also included, together with the descriptive data of all variables (S-W on the tables). Besides, in cases where the Shapiro-Wilk result was p=.04, which is very close to the generally accepted p value (≥.05), skewness and

kurtosis *z*-scores were taken into account, i.e. if both *z*-scores proved to be <±1.96, the distribution was deemed to be normal (Corder & Foreman, 2011).

5.1.1. CS effectiveness

CS were analysed based on a CS elicitation task (picture description with interlocutor; see subsection 4.2.1) administered in both T1 and T2 and their effectiveness was assessed by means of a double approach, the raters' assessment and the mini-Delphi scale (presented in section 4.3). Considering the double analysis of CS effectiveness, all statistical tests were performed twice, once per approach. The tables below include descriptive information on the results of both T1 and T2 CS effectiveness tasks and on the progress shown between both data collections for each context group and for the sample as a whole. The descriptive information according to each method is presented in a separate table.

The progress values indicated on both tables, which represent the gains in CS effectiveness of each context group and the total sample during the four-month period, were calculated by regressing T2 scores on T1 scores and saving the residuals of this regression. This is a measure of gains that controls for pre-test scores. Since one of the main goals of the present study was to compare the development of CS effectiveness in AH and SA learning contexts, it was deemed highly important to measure CS effectiveness gains while controlling for initial scores. The residuals of the regression to account for progress between T1 and T2 in each context group were then used to perform further inferential analyses (see 5.3) and compare the possible effects of both learning contexts on the development of effective CS use.

Table 5.1 presents descriptive information on the scores obtained by means of the raters' analysis approach. Raters assessed individual descriptions on a three-point Likert scale. The mean of all three raters' evaluations per description was calculated, as was the mean of all five descriptions from each data collection, in order to obtain a single raters' score per participant and data collection (see table 10). Therefore, the

possible range of results was 1-3, 1 being non-effective communication of the message (item description) and 3 representing completely successful transmission of the intended message. The table reflects a slight initial advantage of the SA group over the AH group and a more noticeable difference between groups in the final score. In fact, the progress shown by AH learners seems to be negative, i.e. according to this approach, their performance was poorer in the T2 CS elicitation task than in the T1 task. This was not the case of SA participants, who did display a modest but positive CS development over the four-month period. Only T1 and progress data distribution will be relevant to further statistical tests, and all of them proved normal distribution according to the Shapiro-Wilk test (p>.05).

Table 5.1: CS effectiveness scores according to raters' assessment

		Min.	**Max.**	**M /3**	**SD**	**S-W**
AH n=32	T1	1.4	2.7	2.07	0.31	p=.90
	T2	1.57	2.43	2.06	0.26	p=.07
	Progress	-0.61	0.41	-0.07	0.26	p=.90
SA n=33	T1	1.3	2.6	2.11	0.30	p=.39
	T2	1.4	2.7	2.26	0.27	p=.03
	Progress	-0.65	0.63	0.06	0.25	p=.05
Total sample n=65	T1	1.3	2.7	2.09	0.30	p=.48
	T2	1.4	2.7	2.13	0.27	p=.02
	Progress	-0.65	0.63	-0.00	0.26	p=.26

The table below (5.2) includes the information related to the mini-Delphi scores. The numeric value extracted from the mini-Delphi scale system, based on the number and type of CS used during the description task, was a percentage of effectiveness out of the total potential effectiveness score, as explained in the previous chapter. Therefore, the possible range of scores is 0-100. As can be observed in the table, the

AH group's initial (T1) score mean was slightly higher than the SA group's, but their T2 score means were closer to each other. It should also be noted that both AH and SA groups showed certain progress between data collections. Again, only T1 and progress values distribution need to be considered for further statistical tests in the present study and the Shapiro-Wilk test showed normal distribution (p>.05) in all of them.

Table 5.2: CS effectiveness scores according to mini-Delphi scale

		Min.	**Max.**	**M /100**	**SD**	**S-W**
AH n=32	T1	25	85.7	60.67	14.42	p=.33
	T2	37.5	100	64.84	16.54	p=.04
	progress	-28.3	33.6	0.28	15.56	p=.11
SA n=33	T1	22.7	94.4	58.97	15.92	p=.98
	T2	31.3	90	64.66	14.41	p=.8
	progress	-32.42	28.74	0.40	15.12	p=.93
Total sample n=65	T1	22.7	94.4	59.81	15.10	p=.59
	T2	31.3	100	64.75	15.38	p=.15
	progress	-32.42	33.6	0.34	15.21	p=.35

5.1.2. Proficiency level (vocabulary size)

Participants' proficiency level was measured by means of a vocabulary test, the X/Y lex test. Even though vocabulary size offers only a partial perspective on proficiency level, this measurement was considered appropriate in the present study since it had proven to correlate with other aspects of proficiency (see subsection 4.2.3) and it would allow for the creation of groups according to proficiency level (understood here as vocabulary size) for the purposes of inferential analyses. The result of this test provided an estimated amount of known words out of 10,000 (5,000 per part of the test, X and Y).

The table in this subsection presents the descriptive data relevant to the results of the X/Y_lex test in T1 and T2, plus the progress shown between data collection times. This progress was calculated in the same way as progress in CS effectiveness, i.e. regressing T2 scores on T1 scores and saving the residuals. This information should contribute to contrast progress in CS effectiveness and progress in proficiency as measured by vocabulary size. The data on the table below suggest that, similarly to CS effectiveness as assessed by the raters, SA participants presented a slight initial advantage over AH students in terms of vocabulary size, while the difference between T2 scores is more noticeable. This is due to the fact that AH learners showed negative progress during the testing period, whereas SA students displayed a positive evolution in this respect. According to the Shapiro-Wilk test, the scores proved to be normally distributed in all cases.

Table 5.3: X/Y lex scores

		Min.	**Max.**	**M /10000**	**SD**	**S-W**
AH n=32	T1	3350	7600	5751.56	967.50	*p*=.89
	T2	3800	7650	6068.75	1105.98	*p*=.11
	Progress	-2481	1130	-290.29	833.95	*p*=.51
SA n=33	T1	3900	8150	5875.89	1150.28	*p*=.62
	T2	4650	8800	6753.03	1095.37	*p*=.74
	Progress	-769	1915	298.43	713.27	*p*=.29
Total sample n=65	T1	3350	8150	5814.62	1057.77	*p*=.93
	T2	3800	8800	6416.15	1145.10	*p*=.87
	Progress	-2481	1915	8.60	824.15	*p*=.71

5.1.3. Language learning strategies

Language learning strategies (LLS) were assessed through a questionnaire in which participants had to state how frequently they resorted to each of the 17 learning strategies proposed on a six-point Likert scale, 1 being ‘never or hardly ever’ and 6 meaning ‘always or almost always’. The strategies included on the questionnaire referred to intentional attitudes towards the development of the participants’ general level of English, vocabulary learning, studying grammar, reading and writing. Participants took this questionnaire in both T1 and T2. Only T1 answers were necessary to answer the research questions proposed, and these are presented in more detail in tables 5.4 (frequency of use reported for each individual LLS) and 5.5 (presentation of LLS use distinguishing between skills oriented and language study strategies). However, an overview of the evolution observed over the testing period is also offered in table 5.6.

The following table presents the frequency with which participants used each of the individual language learning strategies according to their answers on the T1 questionnaire. Descriptive information is offered for each LLS per context group (AH and SA) and for the total sample. In order to fit all this information into a single table, LLS are represented by numbers in the order in which they appear on the questionnaire (appendix H) but the skill they contribute to is indicated (general English learning, vocabulary, grammar, reading and writing). Minimum and maximum values have also been discarded from the table: considering in most cases the minimum value is 1 and the maximum value is 6, including them would have overloaded the table without adding much information.

Table 5.4: T1 language learning strategies

Skill	LLS	AH (n=30)			SA (n=28)			Total sample (n=58)		
		M	**SD**	**S-W**	**M**	**SD**	**S-W**	**M**	**SD**	**S-W**
Gen.	LLS 1	3.3	1.7	*p*=.01	2.9	1.5	*p*=.01	3.1	1.6	*p*=.00
	LLS 2	5.3	0.9	*p*=.00	5.1	0.8	*p*=.00	5.2	0.9	p=.00
	LLS 3	5.7	0.6	*p*=.00	5.5	0.7	*p*=.00	5.6	0.6	*p*=.00
Voc.	LLS 4	2.9	1.5	*p*=.02	2.6	1.5	*p*=.00	2.75	1.5	*p*=.00
	LLS 5	3.3	1.7	*p*=.01	2.7	1.5	*p*=.00	3.00	1.6	*p*=.00
	LLS 6	3.3	1.6	*p*=.02	2.2	1.5	*p*=.00	2.76	1.6	*p*=.00
	LLS 7	3.1	1.5	*p*=.01	3.3	1.8	*p*=.00	3.17	1.6	*p*=.00
Gram.	LLS 8	3.5	1.7	*p*=.02	3.4	1.8	*p*=.01	3.45	1.7	*p*=.00
	LLS 9	3.6	1.7	*p*=.01	3.3	1.7	*p*=.01	3.5	1.7	*p*=.00
Read	LLS 10	4.3	1.3	*p*=.01	4.1	1.5	*p*=.01	4.2	1.4	*p*=.00
	LLS 11	4.1	1.6	*p*=.00	5.1	1.1	*p*=.00	4.6	1.5	*p*=.00
	LLS 12	5	1.2	*p*=.00	5.4	0.9	*p*=.00	5.2	1.1	*p*=.00
	LLS 13	4.4	1.3	*p*=.01	4.9	1.3	*p*=.00	4.62	1.3	*p*=.00
	LLS 14	4.7	1.3	*p*=.00	5.1	0.96	*p*=.00	4.9	1.1	*p*=.00
Write	LLS 15	5	1.2	*p*=.00	5.1	1	*p*=.00	5.03	1.1	*p*=.00
	LLS 16	5.2	1.1	*p*=.00	4.6	1.3	*p*=.01	4.9	1.2	*p*=.00
	LLS 17	5.1	0.8	*p*=.00	4.4	1.5	*p*=.00	4.76	1.3	*p*=.00

The table shows that the AH group reported higher frequency of use of LLS to improve their general level of English and vocabulary (except for LLS7, “I repeat the words out loud several times”) than the SA group. With regard to grammar learning oriented LLS, the means of the frequency reports from both context groups are very close to each other: SA participants reported slightly higher use of LLS8 (“I write summaries or outlines of the structures that we are learning”) while AH did so with

LLS9 (“I review the structures mentally or out loud”). On the reading section of the questionnaire, SA students presented a higher average use of LLS in most cases with the only exception of LLS10 (“Before I start reading, I look at what the text is about”). And finally, AH participants seemed to resort to LLS to improve their writing more often than SA students in two out of the three relevant items: the AH group preferred LLS16 and 17 (“I review what I have written carefully” and “I use some book or the class notebook”) while SA were more inclined towards LLS 15 (“Before I start writing, I first think about what I want to say”).

The factor analysis that allowed for the development of this LLS questionnaire revealed two clusters of strategies within the final selection of items: seven language study strategies and ten skills oriented strategies, as mentioned in section 4.2.2. Considering the division of LLS between deep processing strategies (skills oriented) and surface strategies (language study oriented), two new numeric values were calculated per participant, i.e. one for the addition of frequencies of deep LLS they reported using and another one for the addition of frequencies of surface LLS. The table below presents the descriptive information with regard to this perspective on LLS used by the participants in the present study, which shall be involved in further statistical analyses to attempt answers to the relevant research questions. All values refer to the T1 questionnaire, administered at the beginning of the testing period. As can be inferred from the table above, AH and SA participants reported similar frequency of use regarding deep processing or skills oriented LLS. The means of the value for surface or language study oriented LLS presented a slightly bigger difference between context groups in favour of the AH students, but they were fairly close to each other as well.

Table 5.5: T1 deep and surface language learning strategies

Group	**LLS type**	**Min.**	**Max.**	**M /60 deep** **M /42 surface**	**SD**	**S-W**
AH (n=30)	Deep	28	60	48.73	6.25	*p*=.03
	Surface	12	36	22.87	6.76	*p*=.38
SA (n=28)	Deep	32	58	49.25	6.12	*p*=.11
	Surface	9	35	20.36	7.09	*p*=.55
Total sample (n=58)	Deep	28	60	48.98	6.14	*p*=.00
	Surface	9	36	21.66	6.98	*p*=.25

Finally, the following table presents global scores from both T1 and T2 questionnaires and progress shown between them, calculated though regression of T2 values on T1 values and saving the residuals. Global scores correspond to the numeric value provided by the six-point Likert scale questionnaire when adding the frequencies expressed for the 17 learning strategies by every participant at each testing time. The possible range was therefore 6-102. T1 global LLS scores were also used to answer the research questions of the present study. T2 global scores and LLS progress are included here only as additional information. According to these data, AH students reported a lower frequency of use of LLS at the end of the testing period while SA participants seemed to have increased their LLS use during the period.

Table 5.6: Language learning strategies global scores and progress

		Min.	**Max.**	**M /102**	**SD**	**S-W**
AH n=30	T1	41	86	71.16	10.60	*p*=.07
	T2	32	91	70.38	12.48	*p*=.00
	Progress	-33.4	16.3	-0.83	11.62	*p*=.00
SA n=28	T1	42	90	70.13	10.95	*p*=.79

		Min.	Max.	M /102	SD	S-W
	T2	43	90	73.03	9.45	*p*=.15
	Progress	-22.6	14.8	2.19	8.16	*p*=.20
Total sample n=58	T1	41	90	70.66	10.70	*p*=.12
	T2	32	91	71.70	11.10	*p*=.00
	Progress	-33.4	16.3	0.63	10.13	*p*=.00

5.1.4. Attention control

Attention control was measured through participants' performance on a trail-making test. More specifically, the test measured participants' attention shifting and inhibition abilities. The TMT consisted of two similar tasks: the first consisted of connecting numbers in ascending order, and the second of alternating numbers and letters in alphabetical order. The result of the trail-making test was calculated by subtracting the time spent on the first part from the time spent on the second part.

The numeric values represented in the following table refer to the seconds of difference between the two parts. In some cases this result was negative, i.e. the participant was faster on the second part of the test than on the first one. The table shows a difference of 1.64 seconds between the means of AH and SA participants. The Shapiro-Wilk test indicates distribution in groups and total sample was below the threshold of normality ($p \geq .05$).

Table 5.7: Trail-making test scores

	Min.	Max.	M	SD	S-W
AH group (n=32)	-1.03	62.69	17.90	16.75	*p*=.00
SA group (n=33)	-1.16	104.56	19.54	18.30	*p*=.00
Total (n=65)	-1.16	104.56	18.73	17.44	*p*=.00

5.1.5. Analytic ability

Analytic ability was measured by means of the LLAMA F test, which assesses skills for inference of grammar rules. Participants had to observe sentences in a new language, each sentence describing an image, and infer the underlying rules of that language. After the learning stage, there followed a 20-item test in which each item presented the participants with an image and two sentences in the new language and the test-takers had to choose which one matched the relevant image.

As a result, the LLAMA F provided the percentage of correct answers on the grammar-inference test. Therefore, the possible range was 0-100. As can be seen on the table, the mean of SA learners was 11 points higher than that of AH students. Distribution was accepted as normal in each context group but not in the total sample.

Table 5.8: LLAMA F test scores

	Min.	**Max.**	**M /100**	**SD**	**S-W**	**Skewness**	**Kurtosis**
AH group (n=32)	0	90	52.58	24.22	*p*=.11		
SA group (n=33)	10	100	63.94	22.07	*p*=.04*	*z*=-1.95	*z*=0.23
Total (n=65)	0	100	58.44	23.65	*p*=.00		

*distribution considered normal, since skewness and kurtosis *z*-scores are <±1.96 (Corder & Foreman, 2011)

5.2. Preliminary analyses

This section presents two different sets of statistical analyses performed before the inferential tests that would contribute to answer the research questions proposed in the present study. These preliminary analyses correspond to the triangulation of methodological approaches, i.e. correlation between the two CS effectiveness analysis methods implemented (see section 4.3), and to the comparability between AH and SA

students. Each set of analyses is presented in a separate subsection within this section.

5.2.1. Triangulation of methodological approaches

As mentioned in the previous chapter, two different methodological approaches were implemented in the analysis of CS effectiveness: the raters' assessment and the mini-Delphi scale. In this subsection the statistical tests performed to confirm the connection observed in the raters' think-aloud protocols are presented. Relevant descriptive statistics are presented in table 5.9. Correlations were chosen to look into this connection using two different approaches: on the one side, a correlation between participants' raters and mini-Delphi total CS effectiveness scores per data collection (n=130, since there were two raters and two mini-Delphi scores, from T1 and T2, per participant); and on the other side, another correlation between individual item descriptions' mini-Delphi and raters' scores (n=538, which is the final valid number of descriptions).

When considering the connection between scores per participant, the mini-Delphi method showed normal distribution (Shapiro-Wilk p=.23) but the raters' did not (p=.02). Consequently, a non-parametric alternative, Spearman's rho correlation, was implemented and proved a significant moderate positive correlation between the two analysis methods (p=.00, r_s=.34).

In the case of triangulation per individual item descriptions, both raters' and mini-Delphi scores failed to adjust to normal distribution (Shapiro-Wilk p=.00 in both cases). Therefore, Spearman's rho correlation was chosen again, and it indicated a significant positive though weak correlation between the two approaches (p=.00, r_s =.24).

Table 5.9: Comparison between CS effectiveness analysis approaches

	Raters					**Mini-Delphi**				
	Min.	Max.	M	SD	S-W	Min.	Max.	M	SD	S-W
Per participant (n=130)	1.3	2.7	2.11	0.29	*p*=.02	22.7	100	62.28	15.38	*p*=.23
Per individual description (n=538)	1	3	2.12	0.51	*p*=.00	0	100	62.01	28.39	*p*=.00

Based on the fact that the triangulation between methodological approaches seems to present some degree of contamination, for the correlations are both significant and positive but not strong ($r<.50$), a decision was made to perform inferential statistics to answer the research questions twice, once per analysis approach, in order to retrieve as much information as possible to investigate the assessment methods and possible connections among variables.

5.2.2. Comparability between SA and AH groups

The sample recruited for the present study was determined by the students' interest in collaborating with research and receiving information about their linguistic progress, plus a modest thank-you gift in the case of SA students or course credit in the case of AH students, which made this a quasi-experimental study. Since there was no random assignment of participants to the context groups, there was the possibility that the groups might be different in multiple aspects. This subsection presents several statistical tests that were performed in an attempt to control for some of those possible differences, considering the information gathered from the participants in both groups. The information presented here would apply to research question number 1, which is the only question in which the two groups are compared.

The most clearly noticeable difference between the groups was that the SA group included multidisciplinary students while the AH group consisted of only linguists (English Studies students), which might have entailed an advantage for the language oriented participants and therefore, the AH group. As a consequence, statistical analyses were carried out to confirm that, despite this apparent imbalance between the groups, they were suitable for comparison in the present study. Values corresponding both to stable variables (attention control and analytic ability)[7] and initial (T1) scores for variables subject to progress (raters and mini-Delphi CS effectiveness, proficiency and total LLS) were considered in this comparability check.

Regarding stable variables, attention control showed non-normal distribution in both the AH and the SA group, while analytic ability did adjust to normal distribution. Therefore, a Mann-Whitney U test was implemented to compare learning context groups in terms of attention control and an independent samples t-test to compare them in terms of analytic ability. The difference between AH (M=17.9, SD=16.75) and SA (M=19.54, SD=18.30) groups in attention control was shown to be far from significant (p=.51; Cohen's d=0.09). The test on analytic ability (AH M=52.58, SD=24.22; SA M=63.94, SD=22.07) indicated that the difference between context groups approached significance (p=.054; d=0.49). The SA group presented a higher rate for this factor.

With respect to variables subject to progress during the testing period, both T1 CS effectiveness scores, T1 proficiency (X/Y_lex score) and T1 total LLS conformed to normal distribution. Independent samples t-tests were implemented to prove that neither T1 raters score (AH M=2.07, SD=0.31; SA M=2.11, SD=0.30; p=.68; d=0.13) nor T1 mini-Delphi score (AH M=60.67, SD=14.42; SA M=58.97, SD=15.92; p=.65; d=-0.11) nor T1 proficiency (AH M=5751.56, SD=967.50; SA M=5875.89, SD=1150.28;

[7] The stability of cognitive variables involved in language aptitude has been questioned in previous research (Kormos, 2013). It is apparently possible to train certain cognitive abilities in adult speakers. However, participants in the present study were unlikely to receive such specific training during the testing period and thus attention control and analytic ability are here considered as stable variables.

p=.64; *d*=0.12) nor T1 LLS values (AH M=71.16, SD=10.6; SA M=70.13, SD=10.95; *p*=.71; *d*=-0.10) showed significant differences between learning context groups.

All things considered, regarding attention control, initial proficiency level, initial use of LLS and initial CS effectiveness according to both approaches, AH and SA groups proved not to be statistically different. There could be a difference between the groups in terms of analytic ability[8]. Groups have been compared in terms of the factors involved in the study. However, it shall be reminded that they could be different in other aspects, like motivation or attitude, which were not analysed but could have influenced their progress during the testing period. Based on the information available, the groups could be accepted as comparable in the present study.

5.3. Inferential statistics to answer the research questions

In this section the goal is to attempt to come to possible answers to the research questions stated in chapter 3 by means of relevant statistical analyses. This section is divided into five subsections, one per research question. Again, in each case, the statistical tests were performed twice, once for each methodological approach (raters' assessment and mini-Delphi score). Effect size according to Cohen's *d* in mean comparisons and *r* or r_s in the case of correlations (Pearson and Spearman, respectively) are reported in all cases to facilitate interpretation of results (Plonsky, 2012).

[8] This difference should not compromise the results in research question 1, which compares groups in terms of effectiveness progress, because, as will be shown in subsection 5.3.5.2, analytic ability did not predict progress in the SA context. Therefore the fact that SA participants showed a higher rate of this cognitive factor should not present an advantage for the SA group over the AH group.

5.3.1. Is learning context a determining factor in the development of effective communication strategies?

This question suggests a comparison in progress regarding CS effective use between participants in each of the two language learning contexts involved in the present study, AH and SA. In order to answer the question, both intra-group and inter-group comparisons were carried out. CS effectiveness progress values were normally distributed in both groups and according to both assessment approaches, which allowed for parametric tests to be implemented in all cases.

Within group comparisons were carried out by means of paired samples t-tests. As can be observed on the table below, differences within both the AH and SA groups following both assessment methods proved non-significant. It can be inferred from this that neither of the two relevant four-month context experiences favoured significant progress in CS effectiveness, though in the SA group Cohen's *d* indicated a positive effect according to the raters' score just below the medium size threshold (*d*=0.45; the effect is considered medium from 0.5).

Table 5.10: CS effectiveness within context group comparison

Context group	**Assessment method**	**T1**		**T2**		**Paired-samples t-test**
		M	**SD**	**M**	**SD**	
SA	Raters	2.11	0.30	2.26	0.27	*p*=.09; *d*=0.45
	Mini-Delphi	58.97	15.92	64.66	14.41	*p*=.17; *d*=0.25
AH	Raters	2.07	0.31	2.06	0.26	*p*=.78; *d*=-0.03
	Mini-Delphi	60.67	14.42	64.84	16.54	*p*=.16; *d*=0.26

Independent samples t-tests were performed to compare progress between the two context groups. The table below presents the descriptive data involved and the result of the test in each case. According to the raters' assessment, there was a

significant difference in CS effectiveness progress between AH and SA students in favour of the latter, with a medium positive effect of the SA context (*d*=0.51). However, following the mini-Delphi method, the test indicated otherwise. It must be noted here that CS effectiveness progress in the AH group according to the raters' assessment was negative. This entails that the significant difference found between the progress rates of the two groups is not necessarily due to a positive effect of the SA experience, but rather to a decrease in effectiveness shown by AH participants.

Table 5.11: CS effectiveness between context groups comparison

Assessment method	**AH progress**		**SA progress**		**Independent samples t-test**
	M	**SD**	**M**	**SD**	
Raters	-0.07	0.26	0.06	0.25	*p*=.03; *d*=0.51
Mini-Delphi	0.28	15.56	0.40	15.12	*p*=.97; *d*=0.01

Additionally, some of the information collected from the linguistic experience questionnaire (T2) was used to gain a more accurate understanding of the conditions in both learning contexts, i.e. how much and what kind of L2 contact the participants had had in either the AH or the SA context, and thus obtain a deeper insight into the effect of learning context on CS effectiveness development. Though based on self-reported data, it is important to explore actual L2 use because, as research on learning context has repeatedly suggested, learners do not always take advantage of the opportunities they have to practise the L2, especially in the SA context (Mitchell, Tracy-Ventura & McManus, 2015; Mitchell, Richard & Romero, 2014; Serrano et al., 2014). In the present study, however, the L2 contact outside the academic environment in the AH context was also explored, on the grounds that both the Internet and the cosmopolitan reality of cities like Barcelona, where all AH participants study, provide a wide range of options to use the L2 outside the language classroom (e.g. language exchange activities, leisure activities in English and access to all kinds of L2 written and audio-

visual materials). In the same way as there are differences with respect to L2 contact abroad, it can also be expected that AH learners might show different profiles concerning how much they use English outside the class. These differences could have an impact on the use of communication strategies, and that is why it is considered an area worth investigating.

The T2 questionnaire included a variety of items aimed at illustrating actual L2 contact in each of the learning contexts under study. Some of the questions did not contribute to distinguishing between participants in terms of how much and what kind of contact they had had with the L2, i.e. participants did not show an approximately even distribution among the possible answers. This was the case of the question on frequency of L1 contact among SA participants, which could not explain differences in CS progress since 30 out of 33 participants had daily communication in their L1 while two of them did not and one did not answer this question. Also, in the case of AH students, three out of 32 travelled abroad using the L2 to communicate during the testing period whereas the rest did not, except for two who did not answer the question. If groups were made according to the answers provided in these questions, the group of SA participants who did not communicate daily in their L1 and the group of SA participants who did travel abroad and use their L2 would be too small to run statistical analyses. A similar situation was created by the answers to the type of housing in the SA group, since 31 of them reported having lived in a hall of residence or a shared flat, i.e. most probably with other students, but no information about the language used at home was collected in that question. In contrast, other questions did elicit more varied answers among participants, namely, the ones on the number of personal and professional or academic contacts with whom the participants used the L2 regularly, whether or not they participated in organised leisure activities in English, the time spent outside the L2 country in the case of SA students, and whether or not AH students took extra classes or joined language exchange activities. Tests were performed on the

information retrieved from these questions, particularly on the effect of these variables on CS progress. The results obtained are explained in subsequent paragraphs.

For both contexts, the correlations between CS effectiveness progress and the number of personal or academic and professional acquaintances with whom learners reported to communicate in the L2 were calculated. The table below shows the descriptive statistics and normality tests on the number of personal, academic or professional and total acquaintances reported by the participants in each of the context groups and the total sample. Based on the results of the Shapiro-Wilk normality test, Spearman's rho correlations were run in all cases.

Table 5.12: Descriptive statistics of amount of L2 acquaintances

Type of contact	**AH**			**SA**			**Total sample**		
	M	**SD**	**S-W**	**M**	**SD**	**S-W**	**M**	**SD**	**S-W**
Personal	0.81	0.99	*p*=.00	2.91	1.57	*p*=.00	1.86	1.68	*p*=.00
Academic/ Profess.	0.41	0.87	*p*=.00	1.81	1.86	*p*=.00	1.11	1.60	*p*=.00
Total	1.22	1.41	*p*=.00	4.72	2.99	*p*=.00	2.97	2.92	*p*=.00

When considering the AH participants separately, none of the CS progress scores correlated significantly with any of the acquaintances types or the total (raters CS progress: total contacts p=.39, r_s=-.16, personal p=.19, r_s=-.24, academic and professional p=.55, r_s=.11; mini-Delphi CS progress: total contacts p=.90, r_s=-.02, personal p=.90, r_s=-.02, academic and professional p=.61, r_s=.09). The SA group however did reveal some significant though negative connections. The raters' CS progress score indicated a significant negative correlation with total acquaintances (p=.03; r_s=-.40) and with personal contacts (p=.02; r_s=-.42), but not with academic and professional contacts (p=.09; r_s=-.31). This result entails that the participants who reported on more L2 relationships, particularly personal relationships, showed less CS

progress than those who reported fewer contacts. In contrast, the mini-Delphi CS progress score showed a significant negative correlation with academic and professional acquaintances (p=.003; r_s=-.51), but not with personal (p=.97; r_s=.01) or total contacts (p=.07; r_s=-.32). When taking into account the total sample, the raters' assessment progress score did not show significant correlations with either personal (p=.67; r_s=-.53), academic and professional (p=.98; r_s=.00) or total acquaintances (p=.72; r_s=-.04). The mini-Delphi progress score did not correlate with personal (p=.79; r_s=-.03) or total acquaintances (p=.27; r_s=-.14), but it did show a significant though negative correlation with academic and professional contacts (p=.04; r_s=-.26), which implies that students who reported on more academic or professional contacts made less CS progress than those who reported fewer such relationships. The fact that the few significant correlations found between number of relationships and CS progress were negative should lead one to conclude that the quantity of frequent self-reported L2 interaction does not predict CS development in either context.

The question on whether participants had taken part in any organised activity other than university courses (sports, dancing lessons, theatre, cultural societies, research or work) in which they had to use the L2 did provide some insight into the matter. The table below shows the relevant descriptive statistics and the result of the independent samples t-tests implemented. It must be kept in mind here that, in the AH group, only three students participated in organised leisure activities, whereas 29 did not. In the SA group, seven participants did not participate in such activities, while 25 did. It follows that, from the total sample, 28 students reported having participated in organised leisure activities, while 36 did not. Due to the small size of some of these groups, it was expected that statistical analyses on the context groups would turn out non-significant. When considering the whole sample, from the raters' assessment perspective, there seems to be a statistically significant difference in CS effectiveness development between students who participated in organised activities and those who did not, with a medium-large positive effect of participating in organised activities

(*d*=0.71). However, according to the mini-Delphi method, the difference between these two groups of participants in terms of CS progress was not significant.

Table 5.13: Comparison between participants with or without L2 organised leisure

		Organised leisure			**No organised leisure**			**T-test results**
		M	**SD**	**S-W**	**M**	**SD**	**S-W**	
AH n=32	Raters progress	0.02	0.23	*p*=.38	-0.84	0.26	*p*=.95	*p*=.51; *d*=3.51
	Mini-Delphi progress	-1.21	18.75	*p*=.54	0.44	15.57	*p*=.19	*p*=.86; *d*=-0.10
SA n=33	Raters progress	0.11	0.19	*p*=.67	-0.03	0.35	*p*=.19	*p*=.34; *d*=0.52
	Mini-Delphi progress	0.78	14.39	*p*=.9	1.15	18.78	*p*=.40	*p*=.96; *d*=-0.02
Total n=65	Raters progress	0.10	0.19	*p*=.47	-0.07	0.28	*p*=.68	*p*=.00; *d*=0.71
	Mini-Delphi progress	0.57	14.51	*p*=.78	0.58	15.96	*p*=.54	*p*=.99; *d*=0.00

In the case of SA students, possible connections between the number of weeks spent outside the L2 country (M=2.87; SD=1.06; S-W *p*=.00) and CS progress were examined. The Spearman's rho correlation was implemented since distribution was not normal. The test showed a non-significant connection according to both analysis methods (raters' *p*=.21, r_s=0.23; Delphi *p*=.90, r_s=-.02). It follows that the amount of weeks spent outside the SA environment did not affect CS effectiveness progress in the SA group.

And finally, in the case of AH students, an independent samples t-test was performed to compare CS progress between those students who took additional

language classes outside the university or participated in language exchange activities and those who did not. As shown in table 5.14, the t-test indicated that, according to both raters' and mini-Delphi scores, participating in additional language learning-oriented activities did not have an effect on the development of CS effectiveness among AH participants. It must be highlighted once again though that the group of AH students who did take extra classes is rather small (n=9), so any potential effect of extra classes was not expected to show as significant. Therefore, the means of progress in each case and Cohen's d effect size should be observed. The raters' assessment progress scores indicates a negative mean in both groups, those with and without extra classes, though the value in the group with extra classes is higher than in the group without such classes, and Cohen's *d* indicated a medium-large positive effect (*d*=0.67). According to the mini-Delphi method, taking extra classes did not have a positive effect on CS effectiveness progress since the mean is negative (M=-4.32), while it is positive among the students who did not take extra classes (M=2.09).

Table 5.14: Comparison between AH students with or without extra classes

Assessment method	**Extra classes (n=9)**			**No extra classes (n=23)**			**T-test result**
	M	**SD**	**S-W**	**M**	**SD**	**S-W**	
CS raters progress	-0.18	0.26	*p*=.82	-0.35	0.25	*p*=.93	*p*=.17; *d*=0.67
CS mini-Delphi progress	-4.32	15.71	*p*=.63	2.09	15.46	*p*=.11	*p*=.30; *d*=-0.41

5.3.2. Are attention control and analytic ability involved in the selection of effective communication strategies?

The possible effects of attention control and analytic ability on the use of effective communication strategies were analysed separately. In both cases a double

statistical approach was performed, with the intention of obtaining a double perspective on the data collected and thus a more complete vision of the possible connections between variables. The same rationale behind the double statistical analysis approach will apply to research questions 4 and 5. In this research question, on the one side, a cluster analysis was run on the whole sample to try to identify two groups with higher and lower scores for each of the cognitive factors considered, so that the CS effectiveness scores could be compared by means of independent samples t-tests or Mann-Whitney U tests, depending on the distribution of these new groups. On the other side, correlations between each of the cognitive factors under study and CS effectiveness scores were implemented. The second approach can be performed in all cases, whereas the first one can only be implemented in cases where the cluster analysis suggests two groups of acceptable size to compare. The sample could have also been divided into two groups of the same size: 50% of participants in each group, one group with a higher rate and another with a lower rate of the relevant cognitive factor. However, in that case, any differences between groups would have been even more difficult to detect as the lowest participants in the high-rate group would have had a very similar rate to the highest-rate participants in the low-rate group. For all statistical analyses, the values corresponding to T1 CS effectiveness of the whole sample (n=65), both raters' and mini-Delphi's, were considered to control for the possible interference of the context experiences.

In the case of attention control, the cluster analysis to create groups of higher and lower skill was attempted but the result did not provide comparable groups (high attention control n=62; low attention control n=3). Therefore, only the second approach could be implemented. Considering the fact that this variable showed no normal distribution according to the Shapiro-Wilk test, Spearman's rho correlation (non-parametric) was chosen to look into a possible connection between this variable and CS effectiveness. According to both effectiveness assessment systems, there is no

correlation (raters p=.68, r_s=.05; mini-Delphi p=.48, r_s=.09) between attention control and CS effectiveness.

Regarding analytic ability, the cluster analysis provided two groups: one with participants who showed high analytic ability (n=47) and another one with low analytic ability (n=17). As table 5.15 shows, following the raters' assessment, the distribution was not normal in the group with low analytic ability, so a Mann-Whitney U test was implemented to compare means in CS effectiveness between high and low analytic ability groups, and the result showed no significant difference. According to the mini-Delphi method, the distribution was normal in both groups, and therefore an independent samples t-test was performed. The result suggested no significant difference between groups with different levels of analytic ability in terms of CS effectiveness.

Table 5.15: Comparison between groups with different level of analytic ability

Assessment method	**High analytic ability**			**Low analytic ability**			**Test**	**Result**
	M	**SD**	**S-W**	**M**	**SD**	**S-W**		
CS raters	2.10	0.31	p=.52	2.08	0.27	p=.02	Mann-Whitney U	p=.99; d=0.07
CS mini-Delphi	59.42	15.21	p=.61	61.46	15.43	p=.57	I.S. t-test	p=.64; d=-0.13

In addition, Spearman's rho correlations were run on the whole sample between analytic ability and CS effectiveness, with data from both raters and mini-Delphi. No significant correlation was found in either case (raters p=.42, r_s=.10; mini-Delphi p=.36, r_s=.12). According to all these results, there seems to be no connection between analytic ability and CS effectiveness.

5.3.3. Is there a connection between language learning strategies and the use of effective CS?

For the purposes of answering this question, since total LLS conformed to normal distribution when considering the sample as a whole, but individual LLS did not, Spearman's rho correlations were run between each LLS and T1 CS effectiveness, and a Pearson correlation was implemented between the total addition of LLS frequencies and CS effectiveness. Regarding the correlations of individual LLS with the raters' assessments, only LLS13 ("When reading in English, I try to understand the structure of the sentence") showed a significant correlation with CS effectiveness (p=.02, r_s=.29) and LLS14 ("When reading in English, I pay attention to the key words or the words that I already know") approached significance (p=.06, r_s=.24). None of the individual LLS proved to have a significant correlation with mini-Delphi CS effectiveness.

When considering LLS as the sum of frequencies of all LLS types, Pearson correlations were implemented and the results indicated that both raters' assessment of CS effectiveness (p=.30, r=.13) and the mini-Delphi approach (p=.26, r=.14) showed no significant correlation with LLS. In addition, in order to get a more fine-grained perspective on possible connections between LLS and CS effectiveness, the values for deep and surface LLS were looked into separately, and correlations were performed between each LLS group and both CS effectiveness scores. Deep LLS did not conform to normal distribution in the sample as a whole, so Spearman's rho correlations were calculated, and none of the CS scores presented a significant correlation with these LLS (raters p=.12, r_s=.21; Delphi p=.24, r_s=.15). Surface strategies did adjust to normal distribution, which led to the use of Pearson correlations with the CS scores, and both proved non-significant as well (raters p=.74, r=.04; Delphi p=.37, r=.12).

5.3.4. Is proficiency level as assessed through X/Y_lex a determining factor in the choice of effective communication strategies?

The possible connection between proficiency level, as assessed through the vocabulary size test X/Y_lex, and CS effectiveness was looked into in different ways: firstly, the role of T1 vocabulary size in T1 CS effectiveness; secondly, the connection between T1 vocabulary size and CS progress; and thirdly, the connection between vocabulary size progress and CS progress. The same double procedure as the one followed in subsection 5.3.2 was applied to this question: an attempt was made in each case to divide the sample into two groups by means of a cluster analysis according to the participants' vocabulary size so that mean-comparison tests could be performed (independent samples t-test or Mann-Whitney U). Additionally, correlations were run to gain a double perspective on the connections between variables.

Regarding the first approach, observing the role of T1 vocabulary size in T1 CS effectiveness, the clusters analysis created the following two groups: a high proficiency (according to X/Y_lex score) group (n=27) and a low proficiency group (n=38). Relevant descriptive data and test results are presented in the first rows of the table below. After confirming normality in both groups in terms of CS effectiveness according to both assessment methods, independent samples t-tests were run to compare their CS effectiveness. The comparison between the high and low proficiency groups following the raters' assessment criterion approached significance, with a medium positive effect of high proficiency level (d=0.51), while the difference between groups in mini-Delphi effectiveness turned out not to be significant. Similarly, when running Pearson correlations between T1 X/Y_lex scores and both CS effectiveness values, the raters' assessment approached significance and presented a weak positive correlation with vocabulary size (p=.058, r=.24), whereas the mini-Delphi score showed a clearly not significant correlation (p=.15, r=.18).

Table 5.16: Comparison between groups with different level of proficiency (X/Y_lex score)

Variable	**High X/Y_lex score**			**Low X/Y_lex score**			**Test**	**Result**
	M	**SD**	**S-W**	**M**	**SD**	**S-W**		
T1 CS raters	2.18	0.26	*p*=.35	2.03	0.32	*p*=.92	I.S. t-test	*p*=.054; *d*= 0.51
T1 mini-Delphi	61.28	16.87	*p*=.61	58.76	13.85	*p*=.74	I.S. t-test	*p*=.51; *d*=0.16
CS progress raters	0.00	0.27	*p*=.02	-0.01	0.26	*p*=.99	Mann-Whitney U	*p*=.55; *d*=0.05
CS progress mini-Delphi	3.66	16.03	*p*=.82	-2.01	14.35	*p*=.51	I.S. t-test	*p*=.14; *d*=0.37

The last two rows of the table above show the descriptive data and test results relevant for the second approach, the connection between T1 vocabulary size and CS progress. The same high proficiency and low proficiency groups were used to confirm a possible effect of vocabulary size on the progress made during the four-month period in terms of CS effectiveness. The distribution of the high proficiency group in terms of CS progress according to the raters' method proved not to be normal. As a consequence, a Mann-Whitney U test was chosen to compare it with the low proficiency group, and it showed no significant difference between groups. The distribution in both groups in terms of CS progress following the mini-Delphi method, however, was normal, which allowed for the implementation of independent samples t-test. This test also indicated no significant difference in CS progress between groups with different initial proficiency levels. Effect size was small-medium (*d*=0.37), but progress in the low proficiency group was negative, which may have amplified the effect of high proficiency. In very much the same way, Pearson correlations showed a non-significant connection

between these variables when considering the whole sample (raters *p*=.29, *r*=.13; mini-Delphi *p*=.93, *r*=.01).

And finally, the cluster analysis on vocabulary size progress provided very unequal groups (more proficiency progress n=59; less proficiency progress n=6), so only Pearson correlations were run between this variable and both CS effectiveness progress scores. According to both the raters' assessment (*p*=.10, *r*=.20) and the mini-Delphi method (*p*=.76, *r*=.04), the connection between these variables was not significant.

5.3.5. Can we define an ideal student profile for each learning context regarding development of CS effectiveness in terms of attention control, analytic ability, LLS and proficiency level?

As in research questions 2 and 4, two different approaches were taken to attempt an answer to this question. First, the connection between each of the possible affecting factors (proficiency level as measured by vocabulary size, LLS, attention control and analytic ability) and CS effectiveness progress was analysed in each context group (AH and SA) by means of correlations. Complementarily, multiple linear regressions provided an overview of how the different factors under study may have influenced CS effectiveness progress in each learning context.

The multiple linear regression is a parametric test which attempts to produce an equation to predict the value of the dependent variable from the values of the independent variables, i.e. to model the relationship between two or more explanatory variables and a response variable. There seems to be no non-parametric alternative to this test. In the present study, one of the variables involved in this question (attention control) did not conform to normal distribution according to the Shapiro-Wilk test, but based on the central limit theorem (CLT), since the sample for each context group included more than 30 subjects, the distribution was accepted as normal (Reid, 2013:

229). If the CLT is applied, and in the absence of any non-parametric alternative, multiple linear regressions can be implemented for this purpose.

5.3.5.1. Factors affecting CS progress in AH context

In the AH group, the distribution of CS effectiveness progress according to both raters and mini-Delphi, T1 proficiency, T1 total LLS and analytic ability proved normal, while the distribution of attention control did not. Pearson correlations were therefore performed between proficiency, LLS and analytic ability, and CS progress, while Spearman's rho correlations were implemented to observe the connection between attention control and CS progress.

All correlations between the two measures of CS effectiveness progress and proficiency (raters p=.91, r=-.02; mini-Delphi p=.88, r=-.03), LLS (raters p=.33, r=-.18; mini-Delphi p=.61, r=.09) and attention control (raters p=.92, r_s=.02; mini-Delphi p=.25, r_s=-.21) proved non-significant. In the case of analytic ability, the results indicated a significant moderate positive correlation (p=.01; r=.48) with progress in raters' assessment of effectiveness; however, the correlation with mini-Delphi effectiveness progress was also non-significant (p=.50, r=.13). When separating LLS into deep and surface strategies, both Spearman's rho for deep LLS (raters p=.72, r_s=.07; Delphi p=.17, r_s=.25) and Pearson correlations for surface LLS (raters p=.07, r=-.33; Delphi p=.69, r=-.07) turned out non-significant. Therefore, the only indicator of an ideal student who would benefit especially from the AH context in terms of CS effectiveness development, according to the raters assessment method, would be one with high analytic ability.

Additionally, a stepwise multiple linear regression was implemented to obtain a global perspective as to how the different factors may impact CS effectiveness development. The stepwise approach produces a model to predict the value of the dependent variable including only the most relevant independent variables, so it either

produces a model with statistical significance or it does not provide any model. A model was produced to predict raters' assessment CS progress in the AH context (adjusted R^2=.21, *p*=.01, F=8.90). This model only included the analytic ability variable and rejected all other independent variables, which matches the correlations' results. In conclusion, analytic ability seems to explain 21.4% (B=0.00, β=0.49) of CS progress according to the raters' method. However, the stepwise multiple regression to predict mini-Delphi progress score produced no model, as the correlations had already indicated. It follows that none of the factors under study can predict progress in the AH context when measured by the mini-Delphi approach. Similar results were obtained when considering deep and surface LLS separately as two different independent variables: a model which included only analytic ability was produced for the raters' score, while no model resulted from the regression for the Delphi CS score.

5.3.5.2. Factors affecting CS progress in SA context

In the SA group, similarly to the AH group, the distribution of CS effectiveness progress as assessed both by raters and mini-Delphi, T1 proficiency, T1 total LLS and analytic ability proved normal, while the distribution of attention control did not. Consequently, Pearson correlations were chosen to look into the connection between T1 proficiency, T1 LLS and analytic ability, and CS progress; while Spearman's rho correlations were performed between attention control and CS progress.

According to both CS analysis approaches, correlations between CS progress and proficiency (raters *p*=.17, *r*=.24; mini-Delphi *p*=.81, *r*=.04), and CS progress and analytic ability (raters *p*=.23, *r*=.21; mini-Delphi *p*=.38, *r*=.16) turned out to be non-significant. CS progress according to raters' assessment also showed a non-significant correlation with attention control (*p*=.20, r_s=-.23) and total LLS (*p*=.06, *r*=-.35), though the latter approached significance. The mini-Delphi measure of CS effectiveness progress showed a significant correlation with both LLS (*p*=.01, *r*=-.46) and attention

control (p=.03, r_s=-.38). In both cases the correlation was moderate and negative. This entails that participants who reported less overall use of LLS and participants with higher attention control (i.e. those cases where the difference of time between the first and second part of the trail-making test was lower) developed their use of effective CS in the SA context more than those with lower attention control and those who reported more frequency of use of LLS. When regarding LLS separately as deep and surface strategies, Pearson correlations indicated that there was no significant connection between surface LLS and either CS assessment method (raters p=.13, r=-.28; Delphi p=.15, r=-.27). Deep LLS did not show a significant correlation with the CS raters' score (p=.30, r_s=-.20), but they did present a strong negative Spearman's rho correlation with the CS Delphi score (p=.00; r_s=-.54).

The stepwise multiple linear regression produced one prediction model per each CS progress analysis method in the SA context. In the case of raters' assessment (adjusted R^2=.26, p=.00, F=11.12), the model included only attention control as a prediction variable (B=-0.01, β=-0.53) and rejected all other variables. Similarly, the regression on mini-Delphi progress (adjusted R^2=.31, p=.00, F=12.62) also relied only on attention control (B=-0.44, β=-0.56) and excluded all other factors. According to these results, attention control can predict between 25.9% and 31.1% of CS progress according to both CS analysis methodological approaches. It is worth mentioning that the negative B and β values in both models imply that students with high attention control (lower attention control score indicates higher rate of this cognitive factor) were the ones that made the most progress according to both raters' and mini-Delphi methods.

Multiple linear regressions were implemented, again considering deep and surface LLS separately, and they produced no model for raters' CS progress, but they did for Delphi CS progress. This model (adjusted R^2=.23, p=.01, F=7.95) included deep LLS (and not attention control) as an explanatory variable (B=-1.10, β=-0.48), which predicted 23.4% of Delphi CS progress. In this case, the negative B and β values

indicate that students who reported using deep LLS *less frequently* were the ones who made the most progress in CS effectiveness according to the mini-Delphi approach.

It must be noted here that the fact that regressions included attention control as a predicting factor in raters CS progress and excluded total LLS in mini-Delphi CS progress contradicts the correlations' results. One possible explanation for this, in the case of attention control, is that non-parametric correlations (Spearman's rho) were implemented, while regressions are parametric tests. They were implemented despite including one not normally distributed independent variable, since there is no non-parametric alternative to regressions. This explanation would not apply, though, in the case of LLS, since both Pearson correlations and regression are parametric tests. Results thus must be interpreted cautiously.

5.4. Summary

This chapter has presented all the descriptive information on the data collected and statistical analyses performed in order to triangulate CS effectiveness analysis approaches, to confirm comparability between context groups and to answer the research questions proposed in the present study. The results obtained seem to indicate that the CS effectiveness analysis methods implemented in the present study do share a common core, since they presented a significant positive correlation, though not a strong one. It was also ensured that the AH and SA groups were not statistically different according to the factors measured in T1. Analytic ability did approach significance, but in all other aspects the groups seem comparable.

Inferential statistics showed that none of the context groups made significant progress in terms of CS effectiveness during the four-month programme, yet the effect of the SA context on CS effectiveness progress according to the raters' assessment was practically medium size (d=0.45). Differences in CS progress between groups

were significant only according to the raters' approach, although it may have been due to a decrease in CS effectiveness in the AH group, not only to the progress made in the SA group. Some of the information collected in the language experience questionnaire in T2 was analysed as well to try in order to find differences in language contact during the four-month programme connected to differences in CS progress. Correlations between CS progress and number of relationships in which communication was held in English were either non-significant or negative, which implies that a higher amount of L2 acquaintances did not contribute to CS progress in either learning context. In contrast, participating in organised leisure activities did show a medium-large positive effect on CS progress according to the raters' assessment approach. In the case of SA students the amount of time spent outside the L2 country did not affect CS progress. In the case of AH students, participating in additional L2 classes or language-exchange activities did not result in significant differences, but it could have had some effect according to the raters' assessment, perhaps statistically undetected because the sample was too small.

Regarding cognitive factors, neither attention control nor analytic ability seems to be connected to effective CS use (T1 effectiveness scores, to control for potential learning context effect). Similarly, with reference to the use of LLS, neither the addition of frequencies of all LLS types, nor either of the groups of deep-processing LLS or surface LLS showed to correlate significantly with either CS progress score. The only significant or almost significant connections found were two LLS types, "When reading in English, I try to understand the structure of the sentence" and "When reading in English, I pay attention to the key words or the words that I already know", which correlated with the raters' score. Finally, T1 proficiency level seemed to have a medium positive effect that approached significance on effective CS use according to the raters' assessment. It did not, however, show any significant effect on CS progress, nor was there any significant correlation to be found between proficiency progress and CS progress.

To finish with, tests were run to identify interactions between each learning context and the other factors. In the AH group, analytic ability seemed to be able to predict 21.4% of CS effectiveness progress according to the raters' approach, while the other factors could not. From the mini-Delphi perspective, none of the factors under study was found to predict CS progress. In the SA group, although results were not consistent across all tests, attention control was found to predict between 25.9% and 31.1% of CS progress according to both CS analysis approaches. Also, students who reported using deep LLS less frequently made more CS progress according to the mini-Delphi approach. The following chapter will be devoted to the interpretation and discussion of the results of all the presented statistical analyses.

6. Discussion

This chapter is devoted to a discussion of the results obtained from the data analysis and statistical tests performed and described in the previous chapter. The goal here is to find possible interpretations or explanations for these results in order to contribute to answering the research questions proposed in chapter 3. To that end, both the relevant theoretical framework from earlier research and empirical data will be drawn upon, and specific aspects of the methodology implemented (participants' characteristics, instruments and procedure) will be re-examined.

The discussion chapter is structured in five main sections, according to the five different research questions that make up the present study. Each of these sections will summarise the statistical analyses performed and results obtained on the data to answer the relevant question, contrast these results with previous research findings and venture an answer to the research question based on all the information available. A "0" section on validity and reliability issues has been included before those main sections with the intention of explaining certain methodological aspects of the study, particularly the reliability of the CS elicitation task and the double analysis approach, which will affect the interpretation of results addressed in all the subsequent research question sections.

6.0. Validity and reliability issues

This section is devoted to the discussion of two aspects of the methodology implemented in the present study, namely the reliability of the data obtained from the CS elicitation task and the validity of the CS effectiveness double analysis approach. The discussion of these aspects has been kept separate from and offered in advance

of the discussion of the results of each individual research question since it will influence the interpretation of results in several questions. In other words, the arguments developed in the following paragraphs shall contribute to the attempts to provide answers to the original research questions in the subsequent sections of this chapter.

The first methodological aspect under discussion is the reliability of the CS elicitation task. The picture description information gap task was adopted from Khan's (2010) doctoral dissertation and expanded to fit the needs of the design of the present study, i.e. a second set of pictures with similar characteristics was produced (one to be administered in T1 data collection and the other one in T2). Both of them were piloted in order to make sure that they would indeed elicit the use of CS in upper-intermediate L2 learners because the pictures included elements whose lexical labels learners at this level generally lack in their vocabulary repertoire. Each set consisted of eleven different pictures for the participant to describe. After all data were collected, and with the purpose of facilitating the application of the raters' assessment analysis approach and focusing on the most productive material, five pictures from each set (the five pictures that elicited the highest number of CS instances) were selected as the objects of the CS effectiveness analysis. As a result, the data under analysis would consist of five T1 and five T2 speech productions (picture descriptions) per participant, upon which both CS effectiveness analysis methods and statistical analyses would be implemented.

The decision to reduce the amount of data collected and the decision as to the number of raters for the raters' assessment approach were made not only out of convenience (amount of time and effort requested from the available collaborators), but also because they seem to be supported by the results in Schoonen (2005). This study estimated the generalisability of writing scores depending on the number of tasks and raters, based on the generalisability theory or "G theory". The results showed that writing scores, intended to represent participants' writing overall proficiency, were

influenced by facets of the assessment such as number of tasks and of raters, and also by rating procedure (holistic or analytic) and rating focus (content and organisation or language use). As a consequence, the minimum number of tasks and raters required for the results to be generalisable (reliable, i.e. Gc=.80) varies depending on the rating focus and rating procedure followed. It requires fewer tasks and raters to reach the accepted generalisability level of .80 in rating language use than in rating content and organization, and in holistic rating than in analytic rating.

In the case of the present study, raters performed a holistic assessment of CS effectiveness, i.e. of how easily they could understand the elements of the picture the participants were referring to in their descriptions. This construct can be understood as *language use* if we consider that control of grammatical and morphological rules in the L2 is needed to implement CS such as circumlocution, word coinage, restructuring or foreignising, while it may overlap with the *content* focus in cases where other CS are used, such as message abandonment or reduction, approximation or omission. Each five picture set can be understood as five different tasks since, even though the instructions for each one of them were the same, the topic was different in each case. According to Schoonen's results, in the case of holistic rating of *language use*, a combination of four tasks and three raters would be enough to ensure generalisability of results (Gc=.82), whereas in the case of holistic rating of *content and organisation*, seven tasks would be needed if three raters are available (Gc=.82). The combination of five tasks and three raters for holistic rating of content and organisation results in a generalisability coefficient of .76, or .04 below the generally accepted coefficient. However, *organisation* is not supposed to be a factor in rating CS effectiveness as it has been described in this dissertation, so the *content and organisation* parameters apply only partially to the present study, while the *language use* parameters apply completely. As a result, the combination of five tasks and three raters for a rating focus that combines both language use and content following a holistic procedure was deemed sufficiently reliable.

Regarding the other methodological issue, the validity of the CS effectiveness analysis by means of a double approach, the results of the correlations implemented to represent the triangulation of the Mini-Delphi scale and the raters' assessments are here under discussion. The double methodological approach was implemented in an attempt to gather as much information as possible about the use of CS in the corpus collected, since each of the approaches considered seemed to present certain drawbacks that would be complemented by the other approach. On the one side, the mini-Delphi scale, an inventory of CS with a numeric value per CS type to assess how effective they are as agreed upon by a panel of experts, is a systematic approach to CS effectiveness testing. It cannot only be implemented in other speech productions of similar characteristics (referential communication as represented by picture description), but it also undoubtedly focuses on CS instances, performing an analytic rating procedure (each CS instance contributes individually to the final score), albeit without regard to the factor of context-dependence inherent to CS use.

On the other side, the raters' assessment method consisted of three collaborators assessing the communicative effectiveness of each item description (i.e. how well they could understand it) provided by each participant in the present study. This approach does observe the effectiveness of CS implementation in context through a holistic rating procedure (one score per picture description), which should make it a reasonably representative measurement of the construct. However, minor drawbacks to this method might rely on the fact that raters were given the whole speech production to decide on a holistic score to represent all the CS used in each description and may also be affected by the amount of hours spent doing the task. The raters were briefed on the conditions of the CS elicitation task and on the existence of CS, but not in too much detail, and they were specifically instructed to focus on whether the speakers managed to convey meaning in referring to the different elements in the pictures. Even so, the think-aloud protocols revealed that one of them seemed to place part of his attention on how the speakers organised the information in the descriptions,

which is not related to the effectiveness of the CS implemented. Also, another of the raters seemed to have forgotten about some of the conditions in which the CS task was carried out, particularly, the visual support for the interlocutor who held a set of pictures very similar or identical to the ones being described. The in situ assessment of CS use granted that context would be taken into account, and most of the comments retrieved from the think-aloud protocols do refer to the use of different CS, yet this approach could have had as a consequence a certain contamination of the rating, due to occasional (conscious or unconscious) disruptions of the raters' focus of attention.

After observing the raters' comments retrieved from a think-aloud protocol performed during the rating sessions, and in view of the fact that they seemed to label as effective the same CS types that had been granted the "effective" label by the mini-Delphi panel when creating the relevant scale, correlations were run between the two scores per participant and data collection, and per individual description. The results showed significant positive correlations in both cases but slightly different Spearman's rho values. Based again on Schoonen's results (2005), the correlation on the participants' score per data collection (i.e. the average of the scores in five subtasks) was deemed more reliable than the correlation on individual descriptions: both the mini-Delphi scale score and the raters' assessment per participant and data collection represented a combination of five tasks instead of just one, so they should prove to have higher reliability. The result of this correlation, though significant, was of .34 according to Spearman's rho, which represents a weak-moderate correlation, not a strong one. It is here interpreted that both approaches do share a common core with regard to the relevance of the selection of certain CS types in the communicative effectiveness of speech. Nonetheless, the observation of CS use in context, employed by previous studies on CS effectiveness (Haastrup & Phillipson, 1983; Littlemore, 2003; Paribakht, 1984), is trusted as the most reliable measurement. This last statement would make this correlation a possible validity test for the mini-Delphi scale,

albeit one with an unclear result due to the fact that the correlation was significant and positive but weak-moderate.

The following sections will deal with the interpretation of the results relevant to each research question. The results provided by both the mini-Delphi and the raters will be discussed, since as described in chapter 5, in most cases they do provide the same answer to the research questions (questions regarding learning context, attention control, analytic ability and LLS). In the cases in which the results differ across analytical approaches, the answer as to CS effectiveness provided by the raters' assessment will be accepted as the most reliable one. The mini-Delphi score will provide, in turn, information about the selection of CS made by the speakers, i.e. which CS types they resorted to more often.

6.1. Is learning context a determining factor in the development of effective CS?

This research question was intended to shed light on the effect of the L2 learning context (SA and AH) on the development of CS, as measured by learners' success in solving communication break-downs and comparing progress made by participants in each of the contexts involved in this study. With this purpose in mind, both within group and between groups comparisons were performed, i.e. on the one hand, tests were done to determine whether significant progress had been made in each context group; and on the other hand, groups were compared in terms of how much progress they had made, to see if there were significant differences between them and, therefore, to contribute to proving the hypothesis regarding a significant effect of learning context on the development of effective CS. Additionally, further tests were run on some of the information collected in the T2 questionnaire to get a clearer picture of each learning context and see if the language contact reported had

influenced progress in either context. Before comparing progress between context groups, statistical analyses were performed to confirm that the groups were not too different at the beginning of the testing period. The results indicated that there were no significant differences between groups in terms of attention control, initial proficiency level, initial use of LLS and initial CS effectiveness according to both approaches, so they were suitable for comparison. The only aspect that approached significance was analytic ability, which was higher in the SA group[9]. It shall be remembered that each context experience lasted for four months (one academic semester) and that participants were tested on their CS effectiveness both at the beginning (T1) and at the end (T2) of their corresponding academic programmes, so progress could be measured.

The statistical analyses performed to answer this question revealed, regarding the comparison within groups, that neither learning context, SA nor AH, significantly led to improvement in CS effectiveness; in other words, there were no significant differences between T1 and T2 scores in any of the learning contexts involved. Moreover, strictly respecting the .05 threshold of significance, results were consistent across analysis methods: both raters' assessment and the mini-Delphi score resulted in the same answer from the tests performed to test within group CS effectiveness progress. It shall be mentioned, though, that the difference between T1 and T2 in the SA group according to the raters' assessment came closer to the significance threshold than the difference in the AH group. This marginally significant effect (.09) was of almost medium size according to Cohen's *d* (=0.45). From the mini-Delphi perspective, differences were non-significant in both SA and AH groups.

As for the comparison between context groups in terms of progress in CS effectiveness, according to the raters' assessment, the test implemented showed a statistically significant difference between CS effectiveness progress (i.e. differences

[9] As was shown in section 5.3.5.2, analytic ability does not seem to play a role in the development of effective CS in the SA context. This would entail that this possible difference between groups should not compromise the results in this research question.

shown between T1 and T2 CS elicitation tasks) in AH participants and in SA students, apparently favouring the latter, with a medium size effect. A factor to take into account upon observing and interpreting these results is that the progress in CS effectiveness among AH participants was negative. This fact implies that the significant difference found between both groups does not necessarily represent evidence for a medium-size positive effect of the SA experience, but rather that, since the AH participants performed more poorly in terms of CS effectiveness during the T2 CS elicitation task than in the T1 task, the difference in progress between groups was larger. The test on the mini-Delphi score, though, indicated otherwise (i.e. no significant difference between groups in CS progress), as could be predicted by observing that the small effect in both intra-context comparisons was practically the same size and positive in both cases (SA *d*= 0.25; AH *d*=0.26). This can be interpreted as an indication that during the testing period participants in both groups slightly increased their use of CS like circumlocution, word coinage, restructuring or self-repair (effective), and/or they decreased their message abandonments, omissions, code-switching and use of similar-sounding words (non-effective).

Further tests were performed on some of the information collected from the questionnaires about the amount and type of language contact participants had during the four-month experience. This was done in order to illustrate actual L2 use during the testing period, as previous learning context research has suggested that not all learners take advantage of the L2 practice opportunities granted in SA (Mitchell et al., 2014). L2 contact in the AH group was also analysed since, living in Barcelona and having Internet access, AH students could easily have benefitted from extracurricular L2 contact as well. The questions referring to the number of L2 acquaintances, participation in L2 organised leisure activities, time spent outside the L2 country in the SA group and participation in language learning oriented activities in the AH group were considered.

The results as to these aspects of learning context indicated that the number of acquaintances did not correlate positively with CS progress according to both effectiveness analysis approaches. In contrast, participating in organised leisure activities did seem to have a positive medium-large effect on CS progress in agreement with the raters' assessment of the sample as a whole. The number of weeks spent outside the English-speaking environment in the SA group did not influence this group's CS progress. The effect of extra classes in the AH group did not show as significant, maybe due to the small size of one of the groups (only nine students took extra classes). However, according to the raters' assessment, the effect of taking extra classes was medium-large and positive, whereas it showed as small-medium and negative on the mini-Delphi effectiveness progress. In short, based on this information, it seems like organised activities, leisure and maybe extra classes as well have a greater role in the development of effective CS than the exact time spent in the SA context or the amount of regular relationships in the L2. The fact that activities are organised and probably led by some kind of instructor or organiser may ensure that communication is held in the L2 and that the activity takes place regularly, which does not apply to contact with friends, though it could apply to professional contact.

As seen in subsection 2.2.1, a few authors, DeKeyser (1991), Lafford (2004), Rubio (2007) and Montero et al. (2013) among them, had already studied the effect of L2 learning context on changes in CS use patterns, particularly the effect of SA programmes, before the present study was carried out. However, the findings provided in those studies seemed to contradict each other to some extent. Lafford and Rubio reported on a lower frequency of use of L1-based CS among SA students as compared to AH students, whereas DeKeyser maintained, after analysing CS use among SA and AH participants, that the differences found were not influenced by learning context, while findings in Montero et al. seemed to hint that learning context did have an effect on children's development of CS effectiveness and decrease in frequency of use of L1-based CS but not on that of adults.

The results obtained in the present study seem to align with DeKeyser's (1991) and Montero et al.'s findings (2013). The sample participants included adult L2 learners, and the statistical analyses in fact showed that neither of the learning contexts under study, SA nor AH, contributed, at least on their own (see research question number 5, in section 6.5), to the participants' development of effective CS use if the p=.05 threshold is respected strictly. Two observations shall be made here to more accurately weigh the importance of the current findings coinciding with previous studies, particularly with Montero et al. First, the CS elicitation task implemented in each study was different (comic strip narration monologue in Montero et al. and controlled interactive picture description in the present study); and second, even though both studies shared the application of the mini-Delphi scale for CS effectiveness, the present study has added a second analysis approach, which has provided additional evidence on the same side of the discussion. The fact that different research designs, different tests and analysis approaches produce the same results seems to contribute to a corroboration of the position that SA and AH learning contexts do not affect CS effectiveness progress, at least during stays of up to four months.

This last statement should, however, be interpreted cautiously. The design of the present study does also differ from the studies on the other side of the question (Lafford, 2004, and Rubio, 2007) with regard to other factors. All the studies reviewed here sampled undergraduate students (except for Montero et al., which additionally sampled children) and compared the effect of SA programmes to that of AH courses on the development of CS. However, two main differences can be observed across their designs and participants recruited. First, the participants' L1-L2 combination: both Lafford and Rubio sampled L1 English L2 Spanish learners (although so did DeKeyser with opposite results), while Montero et al. and the present study recruited participants with the inverse language combination, L1 Spanish L2 English speakers. And second, the length of stay in each of the learning contexts considered was not the same across all the studies: while both Lafford and DeKeyser reported on analysis of the effect of a

semester of either SA or AH experience, Montero et al. for their part observed possible changes in both contexts after three months, Rubio in turn left this factor unspecified ("an extended period of time", p. 49), and the present study focuses on developments produced during a four-month SA or AH experience (an academic semester). As seen in subsection 4.1.3, 22 out of the 33 participants in the SA context reported having spent three or four weeks outside the L2 country during the testing period, but the tests performed on this information revealed that this did not seem to affect CS progress in this group.

Such factors as the language combination and length of stay in either context could have led to different results, and therefore different conclusions, in each piece of research. Indeed, the difference between T1 and T2 raters' score in the SA group in the present study did present an almost medium size effect, so maybe a longer stay would have resulted in significant progress in the SA group as measured by the raters' assessment. It must be noted though that DeKeyser and Lafford shared these characteristics (both sampled L1 English L2 Spanish participants spending a semester in SA or AH contexts), and their findings still contradict each other, although this may well be due to the fact that they implemented different analysis approaches.

In conclusion, according to the data collected in the present study, which is based on controlled interactive picture description, and the double analysis approach performed, and in agreement with other previous findings in the CS literature (Montero et al., 2013), SA and AH L2 learning contexts do not seem to have a significant effect on the development of effective CS in Spanish adult learners of English during a four-month programme. Differences found when comparing these results with other published studies may stem from a variety of factors. Some indications have been observed that the SA context might have a significant effect over a longer period of time and that participating in organised activities may foster CS development to a greater extent than non-organised L2 contact. Further findings in this regard, i.e.

factors that may interact with learning context in fostering the development of effective CS use, will be dealt with in section 6.5.

6.2. Are attention control and analytic ability involved in the selection of effective CS?

The second question in this piece of research was formulated to address certain individual differences, i.e. language aptitude or cognitive abilities, in L2 speakers that may have a connection with whether or not learners successfully solve their communication break-downs by means of certain CS. More specifically, two factors of language aptitude are looked into here: attention control and analytic ability. In both cases, previously validated tests (the Trail-Making Test for attention control and particularly attention shifting, and the LLAMA F to measure analytic ability as grammatical inference ability) were administered to the participants in the present study in order to obtain a numeric value to represent each cognitive ability to allow statistical analyses to be performed. In fact, two statistical approaches were implemented for each construct: first, a cluster analysis on the whole sample was conducted in an attempt to identify two groups (one with a higher score for the relevant ability and another one with a lower score), which, if successful, would allow for a comparison of CS effectiveness scores by means of independent samples t-tests or Mann-Whitney U tests, depending on the distribution of those new groups; and second, running correlations between each of the cognitive abilities and both CS effectiveness scores.

In the case of attention control, only the second approach could be carried out since the cluster analysis did not provide groups of sufficient size to run further statistical analyses. Therefore, only correlations with both CS effectiveness scores, the raters' assessment and the mini-Delphi result, were performed, and they indicated that there is no significant connection between attention control and the use of effective CS

in L2 communication (r_s value was below .10 in both analysis approaches). And regarding analytic ability, both statistical approaches were fully implemented. Similarly to the case of attention control, all tests performed (between group comparisons and also correlations, with both CS effectiveness scores available) pointed towards an absence of any connection between analytic ability and effective use of CS. The effect size in both analysis methods did not reach the small effect threshold of d=0.20, and correlations were both r_s=.10.

The possibility that there might be a connection between L2 speakers' cognitive abilities, such as attention control and analytic ability, and certain patterns in the use of CS, as reviewed in subsection 2.2.3, was primarily based on a theoretical framework. Specifically, theoretical explanations of the mental processes involved in the selection of CS to solve break-downs in L2 communication, such as those published by Bialystok (1990), Dörnyei & Kormos (1998), Kasper & Kellerman (1997) and Shatz (1978), led to the hypothesis that attention control and analytic ability could be among those factors involved. To the researcher's knowledge, no previous research had provided specific empirical evidence on the influence of attention control in the selection of CS. The question of whether this cognitive ability played a part in the selection of L2 CS was formulated on the grounds that several CS researchers had implied the existence of such a role of attention control or processing control in the use of CS.

Bialystok (1990) claimed that the key component of processing control and control-based strategy use was selective attention (attention shifting). She did not link attention control to CS effectiveness, just to certain patterns of CS selection, but not necessarily to the effect of such selection. In addition, from Dörnyei (2005) and Dörnyei and Kormos (1998) a link could be hypothesised between the nature of CS implementation as a problem-solving task and attention control, as it is a cognitive ability involved in problem-solving activities in general and in speech monitoring in particular. In this case, the connection between attention control and CS effectiveness was inferred from the fact that the researchers consider CS implementation to be a

problem-solving mechanism and believe that attention control is partly responsible for that function. In turn, Shatz (1978) also saw CS use as a problem-solving activity and claimed that success in solving such problems could be predicted by the speakers' information processing capacity, and similarly Kasper and Kellerman (1997) identified a high degree of processing control as the key to choosing effective CS. They pointed directly towards "information processing capacity" or "high degree of processing control" as a predictor of success in applying CS to solve communicative problems, particularly in referential communication, according to Shatz. In great contrast to these statements, the empirical data collected in the present study, which is apparently the only empirical evidence available, seems to indicate that there is no such correlation between a high degree of attention control, particularly attention shifting, and effective use of CS in L2 communication, as measured by two different CS effectiveness analysis approaches.

And as for analytic ability, only Littlemore had published findings that showed with empirical data that L2 speakers with an analytic (or ectenic) cognitive profile matched a certain pattern in the selection of CS (2001) and, moreover, that they proved more effective in the implementation of CS as compared to learners with a holistic or synoptic cognitive style (2003). In addition to this empirical evidence, there was also a theoretical basis for the linkage of analytic ability and effective use of CS. Grañena (2013) claimed that analytic ability is, as is the case of attention control, involved in problem solving and Carroll (1973) had long before stated that this cognitive ability was responsible for inferring grammatical rules and semantic relationships. And more precisely within the field of CS research, both Bialystok (1990) and Kasper and Kellerman (1997) posited that L2 speakers had to analyse and understand the conceptual features of the intended referent in order to attempt to communicate with the interlocutor, despite lexical shortcomings, by implementing such CS types as circumlocution or word coinage (both deemed as fully effective on the mini-Delphi scale, see subsection 4.3.1). All in all, both the theoretical framework and the empirical

evidence available seemed to point towards this connection between analytic ability and CS effectiveness.

However, the results obtained in the present study seem to indicate otherwise: two different statistical approaches and two different CS effectiveness measures showed that there is no significant correlation between a high rate of analytic ability and the selection of effective CS in L2 communication. As mentioned above, the only other empirical studies available for comparison in this regard are Littlemore's (2001, 2003). In these studies, the analytic nature of participants' cognitive style was measured by means of Riding's (1991) Cognitive Styles Analysis, which included a non-language based set of tasks to test analytic ability. The fact that the test was different to the one implemented in this piece of research (LLAMA F, which measured grammatical inference ability based on an unknown language) may have led to the difference in the results across studies. Otherwise, the analysis of CS effectiveness (understood as ease of comprehension) in Littlemore (2003) and the raters' assessment in the present study seem to present a very similar design, the only differences being the number of raters (three raters in this case and two in Littlemore's), and the assessment scales implemented (a three-point effectiveness scale in this case and a five-point scale in Littlemore's).

To sum up, the empirical evidence produced in the present study indicates that there is apparently no connection between attention control and analytic ability on the one hand and CS effectiveness on the other. These results are not in line with the theoretical assumptions that analytic ability and attention control might be among the key factors causing L2 learners to effectively implement CS to bridge potential gaps in referential communication as a way to compensate for their limitations in lexical knowledge in the target language. In the case of analytic ability, the results obtained also contradicted the empirical evidence previously offered by Littlemore. Differences across studies might arise from variations in the instruments implemented to measure analytic ability or analytic cognitive profile.

6.3. Is there a connection between language learning strategies and the use of effective CS?

This research question was aimed at exploring the possibility of a connection between CS effectiveness in L2 communication, analysed using two different approaches as explained in section 4.3, and LLS, as measured by a self-report questionnaire on specific actions or attitudes towards language learning. The Likert-scale questionnaire implemented in this study, which had been previously reduced from a longer questionnaire and validated (Tragant et al, 2013), included 17 LLS oriented toward improving the learners' general level of English, vocabulary range, grammar knowledge, and reading and writing skills. Each of these 17 items was also found to fall under one of two possible categories: 'skills-based deep processing strategies' and 'language study strategies' (also known as 'surface strategies'). With all this information at hand in the present study, different correlations were carried out between both measures of CS effectiveness and three different perspectives adopted to observe LLS use, namely, each LLS type individually, deep and surface LLS separately, and the addition of the frequency of use of all LLS types, all of them as reported by the L2 speakers themselves.

Statistical results with regard to the possible correlation between LLS use and CS effectiveness were negative overall. The tests implemented showed that none of the individual LLS seemed to have a significant correlation with the mini-Delphi score, and only two LLS types indicated a significant (LLS13, "When reading in English, I try to understand the structure of the sentence") or approaching significant (LLS14, "When reading in English, I pay attention to the key words or the words that I already know") correlation with the score obtained from the raters' assessment. In turn, the correlation between the addition of frequencies of all LLS types and both CS effectiveness scores

proved to be non-significant (r_s below .15 in both cases), and so did the groups of LLS, identified by the questionnaire's authors as 'deep' and 'surface' strategies, when correlated separately with each of the effectiveness values. The strongest correlation found, apart from the two LLS types, was r_s=.20 in the case of deep LLS and the raters' score. All in all, only a few isolated connections were revealed in these data with reference to this research question.

The hypothetical link between the selection and frequency of use of certain LLS or the overall frequency of LLS use and effective use of CS in L2 communication was based on previous literature that had either grouped LLS and CS together as a single object of study, (LePichon et al., 2010; O'Malley & Chamot, 1990; Oxford, 1990) or had found the two concepts to overlap to some extent (Khan, 2010; Maleki, 2007), since the implementation of CS may grant the L2 learner the opportunity to maintain longer conversations and thus be exposed to more L2 input and multiply chances to improve language learning. Additionally, this research question was also inspired by the fact that both CS and LLS represent strategic behaviour in L2 learners, which could have been a reflection of somehow related underlying cognitive processes in the selection of both types of L2 strategies, as Ying-Chun (2009) and the previous research question ('Are attention control and analytic ability involved in the selection of effective CS?') intended to imply.

According to the literature reviewed in subsection 2.2.3, the correlation between the use of LLS and that of CS had not yet been examined before the present study. There are, as a consequence, no other findings or methodologies to directly compare these results with. There are, however, arguments that work to support the findings produced in this piece of research and also limitations to consider when interpreting the statistical results. In order to back up such findings, the methodology implemented is endorsed by the use of a Tragant et al.'s validated questionnaire and a triangulated analysis of CS effectiveness. Except for the case of two individual LLS (both addressed at improving reading skills), both effectiveness scores led to the same answer to this

research question: there seems to be no correlation between LLS and effective CS use. Nonetheless, a methodological factor must be taken into account when resorting to self-report questionnaires as data collection instruments, i.e. there is always the risk that learners (presumably unconsciously) will attempt to reflect in their answers the projection of the language learners they consider they are supposed to be, rather than the learners they actually are. The use of self-report questionnaires is obviously more cost-effective than actually observing LLS use in practice and it is therefore broadly implemented, but it does present this drawback of offering a potentially biased report on reality. In addition to this limitation, as mentioned in section 4.2.2, the questionnaire did not include LLS specifically oriented towards oral skills (speaking and listening), yet it was expected to represent strategic behaviour and attitude towards language learning. However, it might be the case that learners' attitudes towards improving their oral skills did not match their interest in written skills, and as a consequence the LLS tested did not correlate with CS use.

To conclude the answer to this research question, according to the data collected in this piece of research, there seems to be no significant correlation between the frequency of use of LLS in general, or between skills-based or language study LLS in particular, and the implementation of tools that contribute to the bridging of communication gaps in L2 communication. One LLS type did clearly correlate with CS effectiveness according to the raters' assessment score, although not with the mini-Delphi score: "When reading in English, I try to understand the structure of the sentence". However, the fact that this is the only significant correlation among all correlations run (with overall LLS use, deep LLS, surface LLS and 17 individual items) leads to the conclusion that there is no relevant connection between these two variables. This final statement relies on the use of a type of measurement instrument which is broadly accepted in the field (Cohen & Chi, 2002; Oxford, 1990; Pintrich et al., 1991; Tseng et al., 2006), a self-report questionnaire, and a double analysis approach of CS effectiveness.

6.4. Is proficiency level a determining factor in the choice of effective CS?

As was already stated in sections 2.2.4 and 3.4, the initial version of the design of the present study did not include L2 proficiency level as a factor to be tested with regard to its influence on the selection of CS, given the fact that this seemed to be one of the few connections already mostly agreed upon in the CS research community: as L2 learners progress in their learning process, they are supposed to develop more effective CS (Haastrup & Phillipson, 1983) and decrease the frequency of their attempts to solve communication breakdowns by means of L1-based CS (Liskin-Gasparro, 1996). A measure of proficiency was made a part of the original procedure design of the study with the intention of contributing to the greatest possible comparability among participants. However, since the data on proficiency level of the sample in the present study would be available, the opportunity was taken to further contribute to this hypothesis. Actually, the CS elicitation task chosen to be administered in this piece of research (Khan, 2010) had previously provided findings that contradicted this generally accepted connection between proficiency and CS. Here, however, the approach taken to analyse CS use was different to Khan's: the effect of CS in L2 communication and not the quantity of compensation strategies used was measured.

Supported by Milton's (2013) conclusions on the correlation between vocabulary size and overall language proficiency, proficiency level was measured in terms of vocabulary size, both at the beginning (T1) and at the end (T2) of the relevant testing period. Then, the vocabulary test results were statistically connected to CS use and development from three different perspectives: the role of T1 vocabulary size in T1 CS effectiveness, the role of T1 vocabulary size and CS progress between T1 and T2 shown by the whole sample (both SA and AH students), and correlation between

progress in vocabulary size and progress in CS effectiveness. In all three cases, the same approach as the one adopted with the research question on attention control and analytic ability was followed: creating groups of different proficiency levels by means of a clusters analysis so as to implement independent samples t-tests if the resulting groups were of comparable size, and also running correlations in all cases considered.

The first perspective adopted to look into the connection between the relevant variables implied that the difference between high and low T1 proficiency groups was not significant following the mini-Delphi method, but it did approach significance according to the score provided by the raters' assessment, with a medium positive effect. In other words, participants with higher proficiency level showed a more effective use of CS than those with a lower level. Similar results were reached through the implementation of Pearson correlations (raters' score r=.24). From the second perspective, the tests showed no significant connection between T1 proficiency level and any of the measures of CS effectiveness progress. According to the mini-Delphi perspective, the effect size of high proficiency was small-medium (d=0.37), but this size may have been enlarged by the fact that progress in the low proficiency group was negative. In other words, the high proficiency group seemed to have increased to some extent the use of CS like circumlocution, word coinage or restructuring or decreased frequency of CS such as message abandonment or omission, whereas the low proficiency group apparently evolved in the opposite direction. In contrast, the correlation performed did not show this tendency (r=.01). And finally, no significant correlation was found between proficiency progress and CS progress either, though progress according to the raters' score came closer to the significance threshold, with an r=.20 correlation.

As mentioned before, the level of L2 proficiency had previously been argued and empirically shown to correlate with the development of effective use of CS (Haastrup & Phillipson, 1983) and decrease in L1-based CS (Liskin-Gasparro, 1996), with only minor exceptions. Haastrup and Phillipson concluded that their results

represented a continuum, which seemed to run parallel to progress in L2 proficiency and ranged from mostly L1-based and less effective CS to interlanguage-based and more effective CS. Pointing in a similar direction, Liskin-Gasparro's findings revealed that the level of L2 proficiency and frequency of use of L1-based CS were inversely proportional. As for the minor exceptions mentioned, Liskin-Gasparro clarified that beginners and near-native speakers were to be excluded from such a continuum since the former cannot implement CS due to their limited linguistic resources and near-natives did not need CS any more than a native speaker would. Poulisse and Schils (1989) only observed quantitative and not qualitative differences among proficiency levels in terms of CS. Finally, one of the three CS elicitation tasks administered in Khan (2010), i.e. the information gap picture description task used in the present study, failed to prove such an inverse correlation between proficiency level and need for compensation strategies, while the other two tasks did prove it.

The findings provided in the present study with regard to the connection between L2 proficiency level and the use and development of CS seem to indicate an almost non-existent link between these variables if the .05 threshold of significance is strictly respected. Nonetheless, there are a couple of factors to take into account when interpreting such results. First, the lack of a significant connection between either T1 proficiency level or proficiency progress and CS effectiveness progress might result from the length of the testing period, i.e. four months, which could be too short for changes in proficiency level to be reflected in the vocabulary test administered. Indeed, progress made in terms of CS effectiveness during that time did not show as significant either, as commented on in sections 5.3.1 and 6.1.

And second, the connection between T1 proficiency and T1 CS effectiveness, which would be the one perspective that can be compared to previous findings in relation to this research question, produced two different results depending on the CS analysis approach taken. This is one of the few cases throughout this study where it happens that the two analysis approaches produce different results when attempting to

answer one of the research questions. In this case, the mini-Delphi score proved to have a clearly non-significant connection with T1 proficiency, while the raters' score closely approached significance (p=.054), which would corroborate the mostly agreed upon continuum of proficiency and CS use. As already argued in section 6.0, in case of disagreement between the results provided by the two analysis approaches, the raters' assessment would be preferred to answer the research questions. It follows that the correlation between T1 proficiency and T1 CS effectiveness that approached significance should be considered a more reliable result in this case.

All in all, the findings reported in this piece of research seem to align with the previously accepted connection between L2 proficiency level and the selection of appropriate CS to bridge gaps in L2 communication, as measured by means of the raters' assessment method. The results produced here contributed to the discussion by showing that the parallel progress of proficiency and CS effectiveness reported by previous authors (Haastrup & Phillipson, 1983; Liskin-Gasparro, 1996) was not observed over a period of four months when considering a sample that encompassed both SA and AH participants. Either the fact that participants were divided into two different learning contexts (though according to the first research question, section 6.1, it had no significant effect), or the length of the testing period, might have played a role in this respect. There is also the possibility that the proficiency measure implemented, the vocabulary size test, was not sensitive enough to reflect progress in proficiency over such a short period of time. Besides, the test provides only a partial measure of proficiency, though one expected to correlate with general proficiency according to Milton's conclusions (2013), whereas aspects of proficiency other than vocabulary size might present a stronger connection with CS effectiveness. For example, grammar control could play a greater role in implementing circumlocution, and mastery of L2 morphology rules may prove more useful to perform word coinage than knowledge of existing vocabulary.

6.5. Can we define an ideal student profile for each learning context regarding development of CS effectiveness in terms of attention control, analytic ability, LLS and proficiency level?

The fifth and final research question of this study combines all the variables involved, placing particular attention on the learning context as a potential affecting factor in CS effectiveness development. Indeed, the goal of this question was to look into possible interactions between language learning contexts and all other potential factors (attention control, analytic ability, initial proficiency level and initial LLS) that may influence the development of effective CS in each context, as measured by progress made by learners in terms of CS effectiveness during the testing period. The idea in formulating this question was that if one or several factors were revealed to predict CS development in one context but not in the other, it would imply that these characteristics at least partially define an ideal student profile for the given context. With this goal in mind, multiple linear regressions were run with all factors involved on each of the context groups separately. Additionally, correlations were implemented between each individual factor and progress made by participants in each context group. In addition, the effect of LLS was tested twice: once as a total of the addition of frequencies of all LLS and again as two separate variables, deep LLS and surface LLS. This entailed not only running correlations between deep LLS and surface LLS and each of the CS progress scores, but also running multiple linear regressions twice, once with the total of LLS and again with both deep and surface variables.

As stated in section 3.5, to the researcher's knowledge, there is no precedent of any empirical study on the effect of a possible interaction between learning context and any of the other factors on the development of effective CS. Indeed, findings with regard to the effect of context on the development of effective CS contradict each other (see Lafford, 2004, and Rubio, 2007, vs. DeKeyser, 1991, and Montero et al., 2013).

The rationale behind this question was that contradictions found among different studies could arise from individual differences among learners affecting the extent to which they improve their CS use in each of the learning contexts, in a similar way to the interactions found between learning context and cognitive abilities with an effect on other aspects of SLA (Segalowitz & Freed, 2004; Sunderman & Kroll, 2009; Tokowicz et al., 2004).

6.5.1. Factors affecting the development of effective CS in the AH context.

As outlined in section 3.5, there is no previous research as to which of the factors considered in the present study might interact with AH L2 learning regarding CS development. The only factor that seemed to be linked to this learning context according to the literature reviewed was analytic ability as measured by means of the LLAMA F test, which Grañena (2013) had found to predict aptitude for *explicit* learning. The AH students in the present study received L2 input mostly in a classroom environment, but not explicit CS training, and not all university courses entailed explicit language instruction. Therefore, the connection with regard to CS effectiveness development was not so straightforward in this case.

After running the relevant statistical tests, the following results were obtained: both the regression and correlation approaches indicated that, according to the mini-Delphi method, none of the factors could predict development of effective CS in the AH context. The strongest connection found from this approach was r_s=-.21 with attention control. However, both statistical approaches seemed to reveal that, based on the raters' assessment, high analytic ability does significantly predict to some extent (21.4% according to the regression; r=.48 according to the correlation) successful evolution of CS use. When testing LLS as two separate variables, deep and surface strategies, similar results were obtained: the mini-Delphi score did not seem to interact with any of the other factors, whereas the raters' assessment score did interact with analytic ability.

Since there are no previous comparable studies to contrast these results with, only the implemented methodology is subject to discussion at this point. The tests administered to measure the possible affecting factors (Trail-Making Test, LLAMA F, X/Y_lex and Tragant et al.'s LLS questionnaire) had been previously validated as measurement instruments for their relevant constructs. Additional information regarding L2 contact during the testing period revealed that personal, academic or professional L2 contacts outside class hours did not affect either measure of CS effectiveness progress. Organised leisure activities showed a positive effect from the raters' perspective, but only three students in this group took part in such activities. It is true that only one of the analysis approaches adopted, the raters' assessment, revealed a positive correlation between analytic ability and CS development in the AH context, while the other one, the mini-Delphi scale, did not. However, according to the arguments presented in section 6.0, it is the raters' assessment result that should be regarded as more reliable. In turn, the mini-Delphi correlation results revealed that it might have been the students with higher attention control that showed a stronger (though non-significant) tendency towards solving their communication problems by resorting to CS like circumlocution or word coinage or by avoiding message abandonment or omission.

In conclusion, the present study may have provided, to the researcher's knowledge, the first piece of empirical evidence of L2 learners with high analytic ability benefitting from the AH learning environment in terms of development of CS effectiveness. It bears mentioning here again that these results apply to a learning period of four months in learners with intermediate to upper-intermediate level (according to the certificates they had been awarded before the testing period, see section 4.1.2).

6.5.2. Factors affecting the development of effective CS in the SA context.

In the case of the SA context, previous literature on links between this context and other possible factors in terms of development of CS effectiveness was scarce too. According to the results in Segalowitz and Freed (2004), Sunderman and Kroll (2009) and Tokowicz et al. (2004), there seems to exist an interaction between working memory and some of its individual components (attention control) on the one hand, and the SA learning context that shapes the development of several aspects of SLA (oral skills in general, oral fluency, lexical production, and type of translation errors) on the other. In the case of the present study, it was therefore foreseen that attention control, which is a component of working memory, could play a role in L2 learners developing CS effectiveness in the SA context. Moreover, DeKeyser's (2010) findings regarding the interaction between SA and initial proficiency level in accuracy gains led to the speculation that a similar interaction would result in gains in CS effectiveness in the present study as well.

Statistical results regarding this research question turned out to provide somewhat contradictory answers, especially from the perspective of the mini-Delphi scale method. When applying scores obtained by means of the mini-Delphi approach, significant correlations were found in the SA group between progress in CS effectiveness and attention control (the higher the attention control score, the more CS progress), T1 total LLS, and specifically deep LLS (in both cases, the more LLS use reported, the less CS progress). In other words, participants who showed a *higher* attention control rate and/or reported using deep LLS or LLS in general *less* frequently presented a greater difference in CS effectiveness according to the mini-Delphi score between T1 and T2, so they must have increased their use of circumlocution and restructuring and reduced their use of omission and similar-sounding words to compensate for their lexical shortcomings. When implementing the multiple linear regression with the main factors involved (attention control, analytic ability, initial proficiency level and T1 total LLS), attention control was the only factor found to predict

CS effectiveness progress in the SA context, not total or deep LLS. Attention control was shown to explain 31% of progress from the mini-Delphi perspective. However, when running the regression separating LLS into deep and surface strategies, the model produced was different: it included deep LLS as a predictor of negative CS progress, but it excluded attention control in this case. This result would entail, as mentioned above, that participants who reported using deep LLS more frequently showed less CS progress according to the mini-Delphi scale: they did not increase their use of generally effective CS such as restructuring or self-repair or decrease their code-switching or omissions.

Meanwhile, from the raters' assessment perspective, results were more consistent across statistical tests. Correlations between the raters' scores and all other factors proved non-significant, and this was also the case when separating LLS into deep and surface strategies. Weak non-significant connections showed with T1 proficiency (r=.24), analytic ability (r=.21), attention control (r_s=-.23) and total LLS (r=-.35, which approached significance). These results would suggest a possible tendency of participants with higher initial proficiency levels, higher analytic ability, higher attention control and/or lower frequency of use of total LLS towards making more progress in effectiveness. The multiple linear regression in turn deemed attention control to be a significant predictor of progress in CS effectiveness in the SA context. Based on the model produced by the regression, attention control should explain 25.9% of such progress. However, when running the regression separating LLS into the two possible categories, no model was produced, which means that no factor was found to predict development in CS effectiveness significantly.

Similarly to the AH context, the lack of previous studies looking into this interaction leaves the researcher to discuss only the methodology implemented. Again, previously validated tests were administered to measure the possible affecting factors (Trail-Making Test, LLAMA F, X/Y_lex and Tragant et al.'s LLS questionnaire). The information collected regarding L2 contact during the testing period revealed that the

time spent outside the L2 country did not affect either measure of CS effectiveness progress. Personal, academic or professional L2 contacts outside class hours showed non-significant or negative correlations, which is interpreted to imply that they do not explain CS progress. Organised leisure activities, reported by 28 out of the 33 students in this group, showed a positive medium size effect from the raters' perspective. As in the case of the AH context, the CS effectiveness analysis approaches produced some different results with regard to the possible interactions with an effect on CS development, though they coincided in others. Both the mini-Delphi and the raters' method pointed towards attention control as a significant explanatory variable in at least one of the two regressions implemented in each case, the regression that included the main variables in the study. The results of the parallel regression that separated LLS into deep and surface strategies turned out different across analysis methods: the mini-Delphi method hinted at a possible interaction between the SA context and deep LLS, which indicated a negative effect of employing such strategies, whereas the raters produced no significant results to represent interactions between SA context and other factors. These results were consistent with correlations, with two exceptions: raters' score did not correlate significantly with attention control while it did on the regression, and the negative correlation between the mini-Delphi score and total LLS did not show on the regression.

All in all, results were not consistent across the different tests performed. Nonetheless, indications of a possible interaction between attention control and SA context were observed in the present study by means of the multiple linear regression. This interaction would entail that participants in the SA context with a higher rate of attention control should make greater progress in CS effectiveness, both as observed by raters' in context and as measured by their increased use of CS like circumlocution or word coinage and decrease in CS such as message abandonment or appeals for help in the L1. The lack of consistency across results could be due to having performed a multiple linear regression, a parametric test, on a set of variables among which one

(attention control) was not normally distributed, yet this decision was supported by the central limit theorem, since groups included more than 30 subjects. Another possible explanation to the somewhat contradictory results could be the effect of organised leisure activities, which is the only language contact factor that revealed itself as affecting CS progress, though this would only apply to the raters' results. Further research on this point would be needed before drawing conclusions with regard to the effect of the interaction between SA context and other factors on the development of L2 CS effectiveness.

The answer to the last research question, as suggested by the results set out above, seems to be that AH students require high analytic ability to improve their CS effectiveness, while SA learners might benefit from high attention control, though not from using deep LLS frequently, in developing effective use of CS over four months. In reference to the interaction between analytic ability with the AH context but not with SA, venturing an explanation becomes somewhat challenging. Analytic ability is said to be connected to explicit learning, which characterises the AH context in comparison to the SA context, but students in the AH context presumably did not receive explicit training in CS use. The explanation behind this interaction could be that analytic ability is particularly related to grammar learning and it may have been the case that learners with higher analytic ability improved their level of grammar through explicit teaching in the AH classroom environment. This progress, which was not reflected on the correlations between proficiency and CS progress because proficiency was measured as vocabulary size and not grammar knowledge, may have resulted into better development of effective CS since grammar control is necessary to implement certain CS types like circumlocution.

With regard to the possible interaction between SA and attention control, it could be explained by a combination of two facts. Firstly, attention control is involved in selecting relevant information and, consequently, in retrieving appropriate resources to

solve problems, e.g. communication breakdowns. Secondly, SA learners were exposed to a wide variety of communicative situations and therefore varied communicative problems brought about by the SA experience, in which they were able to practice and thus improve their communicative problem-solving skills. Those with higher attention control could have taken the most advantage of this practice since they were cognitively more predisposed to solve problems. Besides, attention control showed as a significant predictor of CS effectiveness also from the mini-Delphi perspective, according to some of the tests performed. This would entail that participants with a high rate of attention control increased their circumlocutions and word coinages and/or decreased their use of omission and code-switching. Attention control ability in an SA context might foster this particular selection of CS types. Exposure to real communication problems must have taught learners with stronger perceptive abilities that resorting to their L1 or omitting the intended concept generally hinders communication with L2 native speakers who might not speak their L1, whereas describing the concept or creating a new word based on the L2 does contribute to communication. This learning process might take longer than four months for those with a lower attention control rate. In the AH context, exposure to communicative problems must have been less varied and less frequent. This might explain the fact that attention control did not show as a significant predictor of CS effectiveness development in that case: selecting resources to solve communicative problems had not been practised enough in AH courses.

Finally, in order to interpret the results regarding the possible negative interaction between deep LLS and SA, it is worth keeping in mind that deep LLS encompass strategies oriented towards reading, writing and some of the improving general level of English. It may well be the case that the learners who reported devoting particular effort to written skills did not necessarily expend as much energy on their oral skills, which were not explicitly represented in the LLS questionnaire. The result shall not be interpreted to mean that deep LLS have a negative effect in CS

development in the SA context, but rather that other LLS that are more related to oral production might be more helpful at the time of developing effective CS.

6.6. Summary

This chapter has attempted to answer the research questions proposed in chapter 3, based on all the information available: previous research, statistical results and the validity and reliability issues discussed in 6.0. Regarding research question number 1 ('Is learning context a determining factor in the development of effective CS?'), strictly speaking, the statistical analyses indicated that neither of the learning contexts under study had a significant effect on CS effectiveness development during a four-month stay. However, it was argued that the marginally significant effect shown in the SA group in the raters' score could suggest that longer stays in this learning context might result in significant progress, in line with previous studies like Lafford's (2004). The mini-Delphi progress showed a similar evolution of CS use in the two groups, revealing that both SA and AH participants must have used in T2 more CS such as circumlocution or restructuring and less frequently resorted to omission and code-switching as compared to T1, though this evolution did not show as significant (small effect). In reference to the positive effect on CS progress of organised leisure activities, engaged in by most of the participants (79%) in this context group, the structure and regular schedule of such activities could have favoured more continuous L2 contact and thus better development of communicative effectiveness.

Research questions 2 ('Are attention control and analytic ability involved in the selection of effective CS?') and 3 ('Is there a connection between language learning strategies and the use of effective CS?') received clearly negative answers from the statistical analyses. Attention control and LLS had not been, to the researcher's knowledge, studied before as possible factors affecting effective CS use. The

hypotheses were based solely on theoretical connections posited by previous research, so the results and the design of the present study could not be compared to other similar studies. Only the case of analytic ability could be compared to Littlemore's study (2003), which indicated that learners with an analytic cognitive profile proved to be more effective in their use of CS according to an analysis method very similar to the raters' assessment in the present study. It was speculated that the differences in results between the two studies could arise from the measurement tool of analytic ability implemented.

Research question number 4 ('Is proficiency level a determining factor in the choice of effective CS?') was divided into three subquestions: the connections between T1 proficiency level and T1 CS effectiveness, between T1 proficiency level and CS progress, and between proficiency progress and CS progress. The result of the first connection, which indicated that more proficient participants used CS more effectively according to the raters' assessment, was in line with most previous research on this matter. The second and third connections did not show any significant results. Only T1 proficiency seemed to have a small-medium effect on CS progress according to the mini-Delphi approach, yet it was argued that this effect had probably been amplified due to the negative progress in the low proficiency group. Nonetheless, it suggested a slight increase in successful CS (word coinage, self-repair) and decrease in interfering CS (omission, L1-based strategies) in the high proficiency group and the opposite tendency in the low proficiency group. The length of stay or the partial measure of proficiency could have reduced the chances of finding more significant results for this question. The proficiency test may not be sensitive enough to detect progress after only four months, or it could be that other aspects of proficiency such as grammar or morphology have a stronger connection with CS effectiveness than vocabulary size.

Research question number 5 ('Can we define an ideal student profile for each learning context regarding development of CS effectiveness in terms of attention control, analytic ability, LLS and proficiency level?'), similarly to questions 2 and 3,

provided some presumably unprecedented answers. In the AH context, analytic ability revealed itself as the best predictor of CS effectiveness progress from the raters' perspective. The mini-Delphi pointed towards a possible though statistically non-significant role of attention control in CS progress as measured using the criteria of the mini-Delphi scale. The results could have also been influenced by the fact that nine of the participants in this group attended extra L2 classes or language-exchange activities, since it seems to have had a medium-large effect on their progress (non-significant probably due to the small sample size). However, the role of analytic ability was consistent across both statistical approaches implemented (correlations and regression) and the raters' method was deemed probably more reliable according to the discussion in 6.0. An explanation behind this interaction between analytic ability and AH was ventured: learners with high analytic ability may have benefitted from explicit grammar teaching and this progress in grammar level could have resulted in more effective implementation of CS like circumlocution.

In the SA context, the statistical results were less consistent. Regressions with the main variables (attention control, analytic ability, total LLS and proficiency) on both effectiveness scores pointed towards attention control as the only significant explanatory variable. It was speculated that this possible interaction may be explained by learners with stronger perceptive abilities benefitting faster (in four months) from the communicative problem-solving practice offered them by the circumstances of the SA context, i.e. real and frequent communication problems. Meanwhile, the role of attention control in the AH context was not as salient because exposure to communication problems must have been more limited in a classroom environment. Attention control seems to be the clearest predictor of CS progress in the SA context, together with organised leisure activities, which might have differentiated the 79% of participants who engaged in them from the 21% who did not. However, the other tests implemented did not reveal significant connections, or in fact indicated a negative effect of deep LLS from the mini-Delphi perspective. It was argued that the unclear results

could result from the use of multiple linear regression despite not all variables being normally distributed.

The double methodological approach to CS effectiveness analysis was implemented on the grounds that each of the methods presented advantages that complemented the other's drawbacks: the mini-Delphi scale provided systematicity and clear focus on CS, while the raters were better able to consider the context. However, the latter seemed to comply with Schoonen's (2005) criteria for generalisability and had been employed before by other researchers (Haastrup & Phillipson, 1983; Littlemore, 2003; Paribakht, 1984) to measure CS effectiveness, so it was deemed to be the most reliable score in case of discrepancy between results in their answers to the research questions. The mini-Delphi results, in turn, provided more information on the type of CS used or avoided by the participants in each case.

7. Conclusion

The goal of this dissertation was to contribute to the disentanglement of the issue of which factors or combinations of factors are involved in the use and development of effective CS in L2 speakers. To that end, a series of potentially affecting factors, i.e. factors whose influence on CS use had previously been suggested but was as yet unclear, were identified in the literature and used as a starting point to formulate the research questions in the present study. The factors chosen were learning context (SA vs. AH), attention control, analytic ability, LLS and proficiency level. Particular emphasis was placed on exploring the effect of learning context by analysing possible interactions between each learning context and all the other factors under study. These factors were tested on a sample of 65 adult Spanish presumably intermediate to upper-intermediate learners of English as an L2. Out of these participants, 33 were to participate in a SA programme for four months, while the remaining 32 would take around 15 hours of university courses taught in English in Spanish universities (AH group) during the same period of time. In order to measure the effect of learning context on the development of effective CS, the participants' use of CS in English L2 communication was tested both at the beginning (T1) and at the end (T2) of their respective programme. CS effectiveness in both T1 and T2 was analysed by means of two different approaches: raters' assessment and the application of the mini-Delphi scale.

The results obtained with regard to the individual connections between the factors under study and CS effectiveness use or development indicated that, according to both analysis approaches implemented, neither of the learning contexts studied showed a statistically significant effect on the development of effective CS over a period of four months. Similarly, attention control, analytic ability and LLS showed a lack of connection with effective CS use when tested on their own. As for proficiency

level, statistical tests based on the two analysis methods revealed no statistically significant connection between proficiency level and CS effectiveness progress or between proficiency progress and CS effectiveness progress over a period of four months. The connection between initial proficiency level and initial CS effectiveness, in which the possible effect of learning context should be controlled for, did approach significance in accordance with the raters' assessment approach, but it did not from the mini-Delphi method's perspective.

In reference to possible combinations of factors that may have an influence on the development of CS effectiveness in L2 speakers, the statistical results displayed greater variety when comparing the two analysis approaches. Specifically, the combination of potentially affecting factors was tested in each of the learning contexts under study in order to try to depict, at least partially, a learner profile that would benefit the most from each context in terms of CS effectiveness development after only four months. In the case of the AH context, the results indicated that, according to the mini-Delphi approach, none of the factors measured helps to predict CS effectiveness development, whereas following the raters' assessment method, analytic ability does seem to be a predictor of CS progress in this context over a period of four months. This interaction between analytic ability and AH may stem from learners with high analytic ability benefitting from explicit grammar teaching and thus implementing certain CS like circumlocution, which require grammar control, more effectively.

Regarding the SA context, from the mini-Delphi perspective, the tests performed pointed towards attention control and deep-processing LLS as possible predicting factors of CS effectiveness progress, although the results were not consistent across tests (each of the factors correlated individually with CS progress in the SA context, but the multiple linear regressions implemented selected either one or the other, not a combination of both). It bears repeating here that low frequency of deep-processing LLS was found to predict positive progress in CS mini-Delphi effectiveness, so engaging in the deep LLS included in the relevant questionnaire

would not favour CS progress, based on this result. The statistical tests for the data from the raters' assessment approach did not identify any predictors of CS progress during the four-month SA experience, with the exception of one of the multiple linear regressions, which pointed towards attention control as a predicting factor of CS effectiveness development, in line with some of the results obtained from tests based on the mini-Delphi approach. The possible interaction between SA and attention control was explained by the combination of stronger perceptive abilities and the frequent exposure to communication problems associated with the SA circumstances. Learners with a higher attention control rate benefitted more from this problem-solving practice than students with lower attention control, for whom significant progress might require longer L2 exposure.

These results must be explained in light of different circumstances of the study, like the characteristics of the sample and certain limitations in terms of methodological design. As previously stated, the participants recruited were all Spanish undergraduate students speaking English as an L2 at the intermediate to upper-intermediate level, with the possible exception of three participants who had been awarded certificates of advanced (C1 of the CEFR) or near-native (C2) level of proficiency, but whose results in the vocabulary size test (here used as proficiency test) did not identify them as outliers. Therefore, the findings reported in this dissertation may not apply to different L1-L2 language combinations, age populations or L2 proficiency levels.

All participants volunteered to participate in the present study through different recruitment processes, depending on the context in which they were going to continue their L2 learning process during the testing period (e-mailed call for participants via International Relations Offices of all state-funded universities in Spain, or optional activity in an Applied Linguistics course at the University of Barcelona). This implies that random assignment to the learning context groups was not possible, which made this a quasi-experimental study, and that the participants' motivation to participate might have been different in each context group or even in each individual case

(learning about their linguistic profile and progress during the SA experience or AH academic semester, a desire to collaborate with SLA research, or course credit in the case of AH participants). Although they were not asked to express their motivation, it could have influenced their interest in the study and the effort they expended in completing the different tests and tasks. However, other characteristics of the participants recruited were tested in both groups (see subsection 5.2.2), and statistical tests revealed that the two groups were comparable (statistically not different) in terms of attention control, initial proficiency level, initial use of LLS and initial CS effectiveness according to both analysis approaches, although they may differ in terms of analytic ability.

Another important aspect to consider in reference to the participants recruited is the size of the sample (total n=65; SA n=33; AH n=32). Based on the article on statistical power analysis by Cohen (1992), depending on the statistical tests implemented, the sample size and the significance criterion (α=.05 in the present study), different effect sizes can be expected to reveal themselves as statistically significant. In other words, the larger the sample, the more powerful the tests are to identify smaller effects, but the minimum sample size for each effect size (small, medium or large) depends on the test used. The sample recruited for the present study is a relatively small one, and therefore only large effect size should show as significant ($p \leq .05$) on the statistical tests chosen, according to Cohen. It follows that a possible explanation for the lack of statistically significant results in some of the connections tested might result from the sample size and not necessarily from a lack of any connection at all.

When observing the size of *d* and *r* values in all statistical analyses performed, it can be inferred that this could be the case of most connections tested in research question number 1 for the raters' progress assessment: the effect of SA context (d=0.45), of organised leisure on both SA (d=0.52) and AH (d=3.51), the time spent outside the L2 country (r_s=.23) or the effect of extra classes (d=0.67). The same

situation can be observed in the connections between analytic ability, LLS and proficiency on the one hand, and CS use on the other, and similarly in the correlations and regressions run within each learning context group to answer question 5. According to Cohen, they might in fact affect CS use and development but present a small or medium effect size which does not result in statistical significance with a sample size of n=65 (or 32 and 33 in some cases) using the tests implemented (t-tests, correlations and multiple linear regressions). In contrast, the cases that presented effect sizes of $d<0.2$ or $r<.1$ could be deemed more clearly not connected. This would apply, for example, to the correlation between attention control and CS use (raters $r_s=.051$; Delphi $r_s=.089$). In any case, the combination of context and some of these factors (including attention control) did show a certain interaction with a positive effect on CS development, so some degree of connection between these factors and CS cannot be completely ruled out.

Yet another limitation of the data collection methodology stems from the selection of instruments to measure the different constructs involved in the study, specifically, the fact that each of the potential affecting factors was measured by means of only one instrument. Though all of them (the Trail-Making Test, the LLAMA F, the LLS reduced questionnaire and the X/Y_lex) had been previously validated, performing several measures of each construct could have increased the reliability of the data collected. Also, particularly in the case of proficiency level, the instrument used was a measure of L2 vocabulary size, which is a partial measure rather than a comprehensive proficiency test. This could be regarded as a limitation as well, though it is emphasised that the selection of this measurement instrument for this construct complies with Milton's (2013) review of three different studies that showed that vocabulary size correlates with other more comprehensive measures of proficiency level.

As mentioned in section 4.3.1, the analysis of CS is based on the researcher's *interpretation* of the transcripts. Possible evidence of CS use was identified and

classified according to Dörnyei and Scott's (1997) inventory, definitions and examples, taking into account the materials and instructions of the task. This interpretation was later supported by an inter-rater reliability check that showed 80% agreement. A more precise CS analysis would have required asking the speaker about their intention in producing their message in a certain way for every instance of possible CS use, through think-aloud or stimulated recall protocols. Also, contrasting the performance on the task in English to the same task in their L1 would add information as to where linguistic idiosyncrasy ends and resources to compensate for lack of L2 vocabulary begin. However, either procedure would have entailed requesting more time from volunteering participants and thus risking less participation or higher mortality.

The results regarding the effect of learning context on the development of CS effectiveness might be explained by the length of stay in either context, which was of four months. A previous study of the SA effect on CS progress (though not specifically CS effectiveness but rather reduction of L1-based CS) conducted by Lafford (2004) did find significant differences between the SA and AH contexts over a period of six months. Taking into account the results from both studies, it can be hypothesised that a period of four months in either SA or AH may not have a significant effect on the development of effective CS but that a longer period might. Actually, the difference between the T1 and T2 raters' scores in the SA group was practically medium size. Besides, the interaction of L2 learning context with some of the other factors under study (analytic ability in the case of AH context and attention control and maybe deep-processing LLS in the case of SA) did appear to show an effect on the development of effective CS. This effect seemed to be positive in the case of AH plus high analytic ability, and in that of SA plus high attention control, but negative in the case of SA plus high frequency of use of deep-processing LLS. It is worth pointing out again here that the use of LLS was based on participants' self-reports, though, so results regarding this factor should be taken cautiously. Also, based on the information collected from the participants about L2 contact, it was observed that organised leisure activities could

have been one of the factors affecting progress in the SA context, whereas the amount of time spent outside the L2 country in the SA group and the number of L2-speaking acquaintances in both groups did not seem to influence progress.

Despite the limitations mentioned, this dissertation has made a series of original contributions to the literature on L2 CS. It has presented what to the researcher's knowledge is the first study on the connection between attention control and LLS on the one hand, and CS use on the other, and also on possible interactions between L2 learning context (both SA and AH) and other factors with an effect on the development of effective CS. In addition to that, the present study has brought further empirical evidence on some understudied areas, such as the effect of learning context on CS development, an area that had been investigated by few previous researchers and with contradictory results (Lafford, 2004, and Rubio, 2007, vs. DeKeyser, 1991, and Montero et al., 2013), and the connection between analytic ability and effective CS use, apparently examined only by Littlemore (2003) prior to this study.

Future research on factors affecting the use and development of effective CS could test the effect of attention control, analytic ability and LLS on larger samples of participants in order to detect potential small or medium effects or to rule out this connection if no effect size reveals itself as statistically significant in a large sample. As previously mentioned, reliability can also be enhanced by administering several tests to measure each of the factors involved. In addition, future studies should consider testing the effect of different lengths of stay in the SA and AH contexts, as well as other learning contexts (e.g. immersion programmes, computer-assisted language learning or CALL), so that conclusions can be drawn as to the point in either learning experience at which different types of language learners (according to different combinations of cognitive abilities, proficiency levels, LLS use or other individual differences) start to significantly benefit in terms of development of CS effectiveness. In all cases the same effectiveness measure should be implemented so that comparisons between different learning conditions and affecting factors can be established.

With regard to the analysis of CS effectiveness, a more sophisticated systematic measurement instrument could be developed. If the required resources became available, a large varied corpus of L2 speech production could be collected, and then all CS instances on the corpus would have to be identified and raters would be asked to assess the effectiveness of each CS instance. With this information at hand, a numeric value could be assigned to each CS type based on the proportion of instances that were deemed effective, partially effective or non-effective, instead of the rather simplistic three-point scale used to create the mini-Delphi instrument. The new CS effectiveness scale that would result from performing this process would still need to be used cautiously because the fact that an instrument is used systematically implies that it disregards the communicative context of whatever speech production it is applied to. However, a more fine-grained analysis than that provided by the mini-Delphi scale would be possible, and the new tool could be used in further studies on CS effectiveness at least as a preliminary analysis that could reduce the cost of in-situ raters' assessment.

The latter idea is one of the possible further uses of the data collected in the present study. Moreover, other connections could be tested on the same data. For example, a contribution could be made to the discussion on the effect of gender on the use of CS or potential differences in CS use between adult monolinguals and bilinguals could be looked into. According to Ting and Kho (2009) and Zeynep (1997), differences between men and women were found in terms of CS preferences, although the interlocutor's gender influenced both male and female speakers' CS use as well. Also, although investigating strategic behaviour in general and not CS in particular, LePichon et al. (2010) observed that monolingual children who had learned an additional language in a formal context used more varied and diverse strategies than children who had acquired two languages in a non-formal context before the age of 4 ('simultaneous bilingual children'). In the present study, both the male sample (n=18) and the monolingual sample (n=14) are rather small for the purposes of powerful

statistical analyses, as was explained before when referring to Cohen's power analysis (1992), but exploratory studies could be carried out. Additionally, connections between the affecting factors tested and the use of specific CS types and L1 or L2-based strategies could be analysed so that the resulting findings would be easier to compare with previous studies on the matter (DeKeyser, 1991; Lafford, 2004; Liskin-Gasparro, 1996; Paribakht, 1985; Rubio, 2007).

The contention here is that the potential results from all the studies proposed could have relevant implications for L2 teaching and learning. Student exchange programmes should be interested in findings on the optimal length of stay for all learners to make significant improvement in their communicative competence, information that would help them to make more cost-effective decisions. Furthermore, language teaching institutions could benefit from empirical data as to which context and accompanying conditions would foster the development of effective CS in L2 learners, and in which cases. This would help reveal the types of learners and environmental conditions that would be ideal for the explicit or implicit teaching of these strategies, and the occasions when they can be expected to develop spontaneously, which would allow L2 teachers to create the appropriate classroom conditions for all kinds of learners to develop effective CS. For example, based on the interactions suggested in the present study, learners with higher attention control could improve their CS use through implicit teaching, like with role-play activities representing real communication problems, in line with the moderate view of CS instruction supported by Kellerman (1991). Also, although not specifically explored in this dissertation, learners with higher analytic ability might benefit more from the explicit CS training proposed by Dörnyei (1995). In addition to that, according to the findings reported in this dissertation, teaching programmes should consider integrating organised leisure activities, where real and frequent L2 communication is held and therefore communication problems occur, as a way to promote CS effectiveness development in L2 learners. As stated at the beginning of this dissertation and previously by Bialystok (1990), CS are inherent to

communication in general and most especially to L2 communication. Thus, facilitating their development in L2 learners may well be understood as laying some of the most important paving stones on their path towards autonomous L2 use.

REFERENCES

Anderson, S.L. (2012) Differential gains in oral proficiency during study abroad: The role of language learning aptitudes. Unpublished doctoral dissertation, Georgetown University.

Artieda, G. (2013) Individual differences in adult learners of English as a foreign language at two levels of proficiency. Unpublished doctoral dissertation, University of Barcelona.

Baddeley, A. (1996) Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, 49A(1), 5-28.

Baddeley, A.D. (2003) Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189-208.

Bahrani, T., & Tam, S.S. (2012) Exposure to audiovisual programs as sources of authentic language input and second language acquisition in informal settings. *Southern African Linguistics and Applied Language Studies*, 30(3), 347-359.

Berquez, A.E., Cook, F.M., Millard, S.K., & Jarvis, E. (2011) The stammering information programme: a delphi study. *Journal of Fluency Disorders* 36, no. 3: 206–21.

Berry-Bravo, J. (1993) Teaching the Art of Circumlocution. *Hispania*, 76(2), 371-377.

Bialystok, E. (1983) Some factors in the selection and implementation of communication strategies. In C. Faerch and G. Kasper (Eds.) *Strategies in Interlanguage Communication*. Harlow: Longman.

Bialystok, E. (1990) *Communication Strategies: a Psychological Analysis of Second Language Use*. Oxford: Blackwell.

Bialystok, E. (2010) Global-local and trail-making tasks by monolingual and bilingual children: beyond inhibition. *Developmental Psychology*, 46(1), 93-105.

Bialystok, E., & Frohlich, M. (1980). Oral communication strategies for lexical difficulties*. Interlanguage Studies Bulletin, 5* (1), 3-30.

Bialystok, E., & Martin, M.M. (2004) Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, 7(3), 325-339.

Brown, H. (1987) *Principles of Language Learning and Teaching*. Englewood Cliffs, N.J.: Prentice-Hall.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In: C. Jack, Richards & Richard W. Schmidt (Eds.), *Language and Communication*. London: Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1* (1), 1-47.

Carroll, J.B. (1973) Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics*, 2, 5-14.

Carroll, J.B. (1981) Twenty-five years of research in foreign language aptitude. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.

Celce-Murcia, M., Dornyei, Z., & Thurrell, S. (1995).Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-3.

Cohen A.D. & Chi, J.C. (2002) Language strategy use inventory and index. In R.M. Paige, A.D. Cohen, B. Kappler, J.C. Chi & J.P. Lassegard (Eds.), *Maximizing study abroad*. Minneapolis: Center for Advanced Research for Language Acquisition, University of Minnesota.

Cohen, A.D. (1998). *Strategies in Learning and Using a Second Language*. Harlow: Longman.

Cohen, J. (1992) A Power Primer. *Psychological Bulletin*, 112(1), 155-159.

Corder, G.W., & Foreman, D.I. (2011) *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. Hoboken, NJ: Wiley.

Corder, S. (1981) *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Corder, S. (1983) Strategies of communication. In C. Faerch, & G. Kasper (Eds.), *Strategies in Interlanguage Communication*. Harlow: Longman.

Council of Europe (2001) *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge, Cambridge University Press.

Cox, J. (2013) Bilingualism, aging, and instructional conditions in non-primary language development. Unpublished doctoral dissertation, Georgetown University.

Cubillos, J., Chieffo, L., & Fan, C. (2008) The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, 41(1), 157-185.

Cuhls, K. (2001). Foresight with delphi surveys in Japan. *Technology Analysis & Strategic Management, 13*(4), 555.

De Leeuw, E., & Bogulski, C. (2015) L2 pronunciation proficiency, language use and age of acquisition as predictors of executive control in bilinguals. Paper presented at the *18th International Congress of Phonetic Sciences,* Glasgow, August 10-14, 2015.

Dechert, H. (1983) How a story is done in a second language. In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. Harlow: Longman.

DeKeyser, R. (1986) From learning to acquisition? Foreign language development in a US classroom and during a semester abroad. Unpublished doctoral dissertation. Stanford University.

DeKeyser, R. (1991) Foreign language development during a semester abroad. In B.F. Freed (Ed.), *Foreign Language Acquisition Research and the Classroom*, 104–19. Lexington, MA: D.C. Heath.

DeKeyser, R. (2000) The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533.

DeKeyser, R. (2010) Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, 43(1), 80-92.

DeKeyser, R. (2014) Research on language development during study abroad: Methodological considerations and future perspectives. In C. Pérez-Vidal (Ed.) *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam/Philadelphia: John Benjamins.

DeKeyser, R., & Koeth, J. (2011) Cognitive aptitudes for L2 learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning, Vol. II*. New York: Routledge.

Dörnyei, Z. (1995). On the Teachability of Communication Strategies. *TESOL QUARTERLY*, *29*, 55-85.

Dörnyei, Z. (2005) *The Psychology of the Language Learner. Individual Differences in Second Language Acquisition*. Mahwah, NJ: Erlbaum.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, *20,* 349-385.

Dörnyei, Z., & Scott, M. (1995) Communication strategies: an empirical analysis with retrospection. In J. Turley and K. Lusby (Eds.), *Selected Papers from the Proceedings of the 21st Annual Symposium of the Deserted Language and Linguistics Society*, 155–168. Provo, UT: Brigham Yount University.

Dörnyei, Z., & Scott, M. (1997) Communication strategies in a second language: definitions and taxonomies. Review article. *Language Learning* 47, no. 1: 173–210.

Ellis, R. (1997) *Second Language Acquisition. Oxford Introductions to Language Study*. Oxford, New York: Oxford University Press.

Faerch, C., & Kasper, G. (1983) On identifying communication strategies in interlanguage production. In *Strategies in Interlanguage Communication*, ed. C. Faerch and G. Kasper, 210–38. Harlow: Longman.

Freed, B.F., N. Segalowitz and D. Dewey. (2004) Context of learning and second language fluency in French. *Studies in Second Language Acquisition* 26: 275–301.

García-Amaya, L., & Darcy, I. (2013) Attention control in study abroad context: Longitudinal data from L2 learners of Spanish. Paper presented at the AAAL conference in Dallas, March 16-19, 2013.

Ghout-Khenoune, L. (2012) The effects of task type on learners' use of communication strategies. *Procedia – Social and Behavioral Sciences*, 69, 770-779.

Gobel, P., & Kano, M. (2014) Implementing a year-long reading while listening program for Japanese University EFL students. *Computer Assisted Language Learning*, 27(4), 279-293.

Goo, J. (2012) Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34(3), 445-474.

Grañena, G. (2006) Age, proficiency level and interactional skills: evidence from breakdowns in production. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning*, 183–207. Clevedon, UK: Multilingual Matters.

Grañena, G. (2013). Cognitive aptitudes for L2 learning and the LLAMA Language Aptitude Test: What aptitude does the LLAMA measure? In G. Granena & M.H. Long (Eds.) *Sensitive periods, language aptitude and ultimate L2 attainment*. John Benjamins.

Grey, S., Cox, J.G., Serafini, E.J., & Sanz, C. (2015) The role of individual differences in the study abroad context: cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, 99(1), 137-157.

Guo, J. (2011) Empirical Studies on L2 Communication Strategies over Four Decades: Looking Back and Ahead. *Chinese Journal of Applied Linguistics (Quarterly)*, 34(4), 89-106.

Haastrup, K., & Phillipson, R. (1983) Achievement strategies in learner/native speaker interaction. In C. Faerch and G. Kasper (Eds.), *Strategies in Interlanguage Communication*, 140–58. Harlow: Longman.

Harding, E. (1983) Compensation strategies. Occasional Paper No. 9. Dublin: Trinity College. Centre for Language and Communication Studies.

Hsiao, T.-Y., & Oxford, R.L. (2002) Comparing theories of language learning strategies: A confirmatory factor analysis. *Modern Language Journal*, 86(3), 368-383.

Isaacs, T. & Trofimovich, P. (2011) Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics* 32, 113-140.

Johnson, J.S., & Newport, E.L. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.

Juan-Garau, M., & Pérez-Vidal, C. (2007). The effect of context and contact on oral performance in students who go on a stay abroad*. Vigo International Journal of Applied Linguistics*, 4, 117-134.

Kasper, G., & Kellerman, E. (Eds.) (1997). *Communication strategies: Psycholinguistic and sociolinguistic perspectives.* New York: Addison Wesley Longman.

Kellerman, E. (1991) Compensatory strategies in second language research: A critique, a revision, and some (non-) implications for the classroom. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharmwood Smith, & M. Swain (Eds.), *Foreign/second language pedagogy research: A commemorative volume for Claus Faerch*. Clevedon, UK: Multilingual Matters.

Kellerman, E., Ammerlaan, T., Bongaerts, T., & Poulisse, N. (1990) System and hierarchy in L2 compensatory strategies. In R.C. Scarcella, E.S. Andersen & S.D. Krashen (Eds.), *Developing communicative competence in a second language*. New York: Newbury House.

Khan, S. (2010) Strategies and spoken production on three oral communication tasks: A study of high and low proficiency EFL learners. Unpublished doctoral dissertation, Universidad Autónoma de Barcelona.

Kongsom, T. (2009) The effects of teaching Communications strategies to Thai learners of English. Unpublished doctoral dissertation, University of Southampton.

Kormos, J. (2013) New conceptualizations of language aptitude in second language attainment. In G. Grañena, & M.H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*. Language Learning and Language Teaching, Vol. 35. Amsterdam: John Benjamins.

Lafford, B. (2004) The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition* 26: 201–225.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.

LePichon, E., De Swart, H., Vorstman, J. & Van den Bergh, H. (2010) Influence of the context of learning a language on the strategic competence of children. *International Journal of Bilingualism* 14 (4), 447.

Levelt, W.J.M. (1989) *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, S. (2013) The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97(3), 634-654.

Liskin-Gasparro, J. (1996) Circumlocution, communication strategies, and the ACTFL proficiency guidelines: an analysis of student discourse. *Foreign Language Annals* 29, no. 3: 317–330.

Littlemore, J. (2001) An empirical study of the relationship between cognitive style and the use of communication strategy. *Applied Linguistics* 22 (2), 241-265.

Littlemore, J. (2003) The communicative effectiveness of different types of communication strategy. *System* 31, 331-347.

Llanes, A. (2010). Children and adults learning English in a study abroad context. Unpublished doctoral dissertation, Universitat de Barcelona.

Llanes, A. (2011) The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*. DOI: 10.1080/14790718.2010.550297.

Llanes, A. (2012) The short- and long- term effects of a short study abroad experience: the case of children. *System*, 40, 179–90.

Llanes, A., & Muñoz, C. (2009). A short stay abroad: does it make a difference? *System* 37, no. 3: 353–65.

Llanes, A., & Muñoz, C. (2013). Age effects in a study abroad context: children and adults studying abroad and at home. *Language Learning* 63: 63–90.

Llanes, A., Tragant, E. & Serrano, R. (2011) The role of individual differences in a study abroad experience: the case of Erasmus students. *International Journal of Multilingualism*. DOI: 10.1080/14790718.2011.620614.

Maleki, A. (2007) Teachability of communication strategies: An Iranian experience. *System*, 35, 583-594.

Marques, J.F., & McCall, C. (2005) The Application of Interrater Reliability as a Solidification Instrument in a Phenomenological Study. *The Qualitative Report*, 10(3), 439-462.

Meara, P. (2006) X_lex manual. *Lognostics* [Online] Retrieved from: http://www.lognostics.co.uk/tools/

Meara, P. & Miralpeix, I. (2006) *Y_Lex: the Swansea Advanced Vocabulary Levels Test*. *v2.05*. Swansea: Lognostics.

Meara, P. (2005) *X_Lex: the Swansea Vocabulary Levels Test*. *v2.05*. Swansea: Lognostics.

Milton, J. (2013) Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis.* Eurosla Monographs Series, 2.

Milton, J., & Hopkins, N. (2006) Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners. *The Canadian Modern Language Review*, 63(1), 127-147.

Milton, J., Wade, J., & Hopkins, N. (2010) Aural word recognition and oral competence in a foreign language, In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning*. Bristol: Multilingual Matters.

Mitchell, R., Richard, L., & Romero, P. (2014) Social networks, target language interaction and second language acquisition during the year abroad: a longitudinal study. ESRC End of Award Report, RES-062-23-2996. Swindon: ESRC.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2015) Placement type and language learning during residence abroad. In (Eds.) *Social interaction, identity and language learning during residence abroad*. Eurosla Monograph Series, 4.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., & Howerter, A. (2000) The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.

Montero, L. (2011) The role of L2 learning context and age in the use of communication strategies. Unpublished master thesis. University of Barcelona.

Montero, L., Serrano, R., & Llanes, À. (2013) The influence of learning context and age on the use of L2 communication strategies. *The Language Learning Journal*. DOI:10.1080/09571736.2013.853824.

Nakatani, Y. (2010) Identifying strategies that facilitate EFL learners' oral communication: A classroom study using multiple data collection procedures. *The Modern Language Journal*, 94(1), 116-136.

Nation, P., & Beglar, D. (2007) A vocabulary size test. *The Language Teacher*, 31(7), 9-13.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29(4), 557-581.

O'Malley, J.M., & Chamot, A.U. (1990) *Learning strategies in second language acquisition*. New York: Cambridge University Press.

Oliver, R. (1998) Negotiation of meaning in child interactions. *The Modern Language Journal* 82, no. 3: 372–86.

Oweis, T. (2013) A literature review on communication strategies in language learning. *European Scientific Journal*, vol. 9, No. 26, 245-267.

Oxford, R.L. (1990). *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House.

Oxford, R.L. (1999) Learning strategies. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics*. Oxford: Elsevier.

Oxford, R.L. (2001) Language learning strategies. In R. Carter & D. Nunan (Eds.) *Teaching English to Speakers of Other Languages*. Cambridge: Cambridge University Press.

Paribakht, T. (1986) On the pedagogical relevance of strategic competence. *TESL Canada Journal*, 3, 53-66.

Paribakht, T. (1984) The relationship between the use of communication strategies and aspects of target language proficiencies - A study of ESL students. Retrieved from: http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED250911

Paribakht, T. (1985) Strategic competence and language proficiency. *Applied Linguistics* 6: 132–46.

Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1991) *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: NCRIPTAL, School of Education, University of Michigan.

Plonsky, L. (2012) Effect size. In P. Robinson (Ed.) *Routledge encyclopedia of second language acquisition*. New York: Routledge.

Poulisse, N. (1987) Problems and solutions in the classification of compensatory strategies. *Second Language Research*, 2, 141-153.

Poulisse, N. (1990). *The Use of Compensatory Strategies by Dutch Learners of English*. Dordrecht: Foris Publications.

Poulisse, N. (1997). Compensatory strategies and the principles of clarity and economy. In G. Kasper & E. Kellerman (Eds.), *Communication strategies: Psycholinguistic and Sociolinguistic Perspectives* (pp. 49-64). London: Longman.

Poulisse, N. (1993) A theoretical account of lexical communication strategies. In R. Schreuder and B. Weltens (Eds.), *The Bilingual Lexicon*, 157–89. Amsterdam: Benjamins.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*(1), 36-57.

Poulisse, N., & Schils, E. (1989) The influence of task- and proficiency-related factors on the use of communication strategies: a quantitative analysis. L*anguage Learning* 39: 15–48.

Poulisse, N., Bongaerts, T. & Kellerman, E. (1984) On the use of compensatory strategies in second language performance. *Interlanguage Studies Bulletin*, 8, 70-105.

Rababah, G. (2002). “Second Language Communication Strategies: Definitions, Taxonomies, Data Elicitation Methodology and Teachability Issues. A Review Article.” ERIC ED472698.v.

Rababah, G., & Bulut, D. (2007). Compensatory strategies in Arabic as a second language. *Poznań Studies in Contemporary Linguistics, 43* (2), 83–106.

Raupach, M. (1983) Analysis and evaluation of communication strategies. In C.Faerch & G. Kasper (Eds.) *Strategies in interlanguage communication*. Harlow: Longman.

Reid, H. (2014) *Introduction to Statistics: Fundamental Concepts and Procedures of Data Analysis*. Los Angeles: Sage.

Riding, R. (1991) *Cognitive Styles Analysis*. Birmingham: Learning and Training Technology.

Roehr, K. (2007) Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, 29(2), 173-199.

Rubio, F. (2007) The use of communicative strategies among advanced speakers of Spanish. [El uso de estrategias comunicativas entre hablantes avanzados de español]. *Círculo de Lingüística Aplicada a la Comunicación* 29: 44–62.

Sanz, C. (2014) Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA Project, an appraisal. In C. Pérez-Vidal (Ed.) *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam/Philadelphia: John Benjamins.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, 54(3), 525-582.

Savignon, S.J. (1990) Communicative language teaching: Definitions and directions. In J.E. Alatis (Ed.), *Georgetown University Round Table on Language and Linguistics*. Washington, DC: Georgetown University Press.

Savignon, S.J. (1972) *Communicative competence: An experiment in foreign language teaching.* Philadelphia: The Center for Curriculum Development.

Savignon, S.J. (1983) *Communicative competence: Theory and classroom practice*. Reading, MA: Addison-Wesley.

Schoonen, R. (2005) Generalizability of writing scores: an application of structural equation modeling. *Language Testing*, 22(1), 1-30.

Schoonen, R. (2010) The development of lexical proficiency: knowledge and skill. Paper presented at *Symposium on Approaches to the Lexicon*, Copenhaguen, December 6-8, 2010.

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173_199.

Segalowitz, N., & Freinkiel-Fishman, S. (2005) Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, 33(4), 644-653.

Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Interdisciplinary Journal of Study Abroad*, 10, 1_18.

Selinker, L. (1972). Interlanguage. *IRAL*, *10*, 209-231.

Serrano, R., Tragant, E., & Llanes, A. (2014) Summer English courses abroad vs. “at home”. *ELT Journal*, 68, 397-409.

Shatz, M. (1978) The relation between cognitive processes and the development of communication skills. In C. B. Keasy (Ed.), *Nebraska Symposium on Motivation*, vol. 25. Lincoln: University of Nebraska Press.

Skehan, P. (1998) *A cognitive approach to language learning*. Oxford: Oxford University Press.

St. Clair-Thompson, H.L., & Gathercole, S.E. (2006) Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59(4), 745-759.

Staehr, L.S. (2008) Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.

Stern, H. (1983) *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.

Storch, N., & Tapper, J. (2009) The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, 8(3), 207-223.

Stuss, D.T., Shallice, T., Alexander, M.P., & Picton, T.W. (1995) A multidisciplinary approach to anterior attentional functions. In J. Grafman, K.J. holyoak, & F. Boller (Eds.), *Structure and functions of the human prefrontal cortex (Annals of the New York Academy of Sciences*, Vol. 769, 191-211). New York: New York Academy of Sciences.

Sunderman, G., & Kroll, J. (2009). When study abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, 30, 1-21.

Taguchi, N. (2008). Cognition, language contact, and the development of pragmatic comprehension in a study-abroad context. *Language Learning*, 58(1), 33-71.

Tarone, E. (1977). Conscious communication strategies in interlanguage: A progress report. In H. Brown, C. Yario & R. Crymes (Eds.), *On TESOL '77* (pp.194-203). Washington DC.: TESOL .

Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL Quarterly,15* (1), 285-295.

Tarone, E. 1980. Communication strategies, foreigner talk, and repair in interlanguage. *Language Learning* 30: 417–28.

Tercanlioglu, L. (2004) Achievement goal theory: A perspective on foreign language learners' motivation. *TESL Canada Journal / Revue TESL du Canada*, 21(2), 34-49.

Ting, S.-H-, & Kho, T.-P. (2009) Gender and communication strategy use in learning English as a second language. *ITL – International Journal of Applied Linguistics*, 157, 93-108.

Tokowicz, N., Michael, E.B., & Kroll, J.F. (2004) The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, 7(3), 255-272.

Towell, R. (1987) A discussion of the psycholinguistic bases for communicative language teaching in a foreign language teaching situation. *British Journal of Teaching English*, 26, 90-99.

Tragant, E., & Victori, M. (2012) Language learning strategies, course grades and age in EFL secondary school learners. *Language Awareness*, 21(3), 293-308.

Tragant, E., Thompson, M.S., & Victori, M. (2013) Understanding foreign language learning strategies: A validation study. *System*, 41, 95-108.

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006) A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78-102.

Varadi, T. (1983) Strategies of target language learner communication: message adjustment. In C. Faerch & G. Kasper (Eds.), *Strategies in Interlanguage Communication*. Harlow: Longman.

Wanger, J. (1983). An analysis of interlangauge communication in instructions. In C. Faerch & G. Kasper (Eds.), *Strategies in Interlanguage Communication*. Harlow: Longman.

Wayne, R. (2006) Measuring grammatical complexity: The Gordian Knot. *Language Testing*, 23(4), 497-519.

Willems, G.M. (1987) Communication strategies and their significance in foreign language teaching. *System* 15, no. 3: 351–64.

Wong, L.L.C., & Nunan, D. (2011) The learning styles and strategies of effective language learners. *System*, 39, 144-163.

Yilmaz, Y. (2013) Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, 34(3), 344-368.

Ying-Chun, L. (2009) Language learning strategy use and English proficicency of univeresity freshmen in Taiwan. *TESOL Quarterly,* 43(2), 255-280.

Yule, G., & Tarone, E. (1990) Eliciting the performance of strategic competence. In R.C. Scarcella, E.S. Andersen & S.D. Krashen (Eds.), *Developing communicative competence in a second language*. New York: Newbury House.

Zeynep, K. (1997) The role of gender on communication strategy use. Paper presented at the *Annual Meeting of the Teachers of English to Speakers of Other Languages* (Orlando, FL).

Zied, K.M., Phillipe, A., Pinon, K., Havet-Thomassin, V., Ghislaine, A., Roy, A.R., & Le Gall, D. (2004) Bilingualism and adult differences in inhibitory mechanisms: Evidence from a bilingual stroop task. *Brain and Cognition*, 54(3), 254-256.

APPENDICES

APPENDIX A: CS inventory (Dörnyei & Scott, 1997: 188-192)

Inventory of Strategic Language Devices with Descriptions/Definitions, Examples (Based on Dörnyei & Scott, 1995a,1995b), and Indications Whether They Were Included in Any Other Taxonomies (T=Tarone, 1977; F&K=Færch & Kasper, 1983b; B=Bialystok, 1983; P=Paribakht, 1985; W=Willems, 1987; N=Nijmegen Group)

STRATEGY	DESCRIPTION	EXAMPLE	OTHER TAXONOMIES
1. Message abandonment	Leaving a message unfinished because of some language difficulty.	*It is a person er... who is responsible for a a house, for the block of house... I don't know... [laughter]*	T, F&K, W
2. Message reduction (topic avoidance)	Reducing the message by avoiding certain language structures or topics considered problematic languagewise or by leaving out some intended elements for a lack of linguistic resources.	[Retrospective comment by the speaker:] *I was looking for "satisfied with a good job, pleasantly tired," and so on, but instead I accepted less.*	T, F&K, W
3. Message replacement	Substituting the original message with a new one because of not feeling capable of executing it.	[Retrospective comment after saying that the pipe was broken *in the middle* instead of "the screw thread was broken":] *I didn't know "screw thread" and well, I had to say something.*	F&K, W
4. Circumlocution (paraphrase)	Exemplifying, illustrating or describing the properties of the target object or action.	*it becomes water* instead of "melt"	T, F&K, W, P; B: "description"; N: appr. "analytic strategies"
5. Approximation	Using a single alternative lexical item, such as a superordinate or a related term, which shares semantic features with the target word or structure.	*plate* instead of "bowl"	T, W; B and P: "semantic contiguity"; F&K: "generalization"; N: appr. "holistic str."
6. Use of all-purpose words	Extending a general, "empty" lexical item to contexts where specific words are lacking.	The overuse of *thing, stuff, make, do,* as well as words like *thingie, what-do-you-call-it*; e.g.: *I can't can't work until you repair my ... thing.*	W: "smurfing"
7. Word-coinage	Creating a non-existing L2 word by applying a supposed L2 rule to an existing L2 word.	[Retrospective comment after using *dejunktion* and *unjunktion* for "street clearing":] *I think I approached it in a very scientific way: from 'junk' I formed a noun and I tried to add the negative prefix "de-"; to "unjunk" is to 'clear the junk' and "unjunktion" is 'street clearing'.*	T, F&K, B, W; N: appr. "morphological creativity"
8. Restructuring	Abandoning the execution of a verbal plan because of language difficulties, leaving the utterance unfinished, and communicating the intended message according to an alternative plan.	On Mickey's face we can see the... so he's he's he's wondering.	F&K; W: under "self-repair"
9. Literal translation (transfer)	Translating literally a lexical item, an idiom, a compound word or structure from L1/L3 to L2.	I'd made a big fault [translated from French]	T, W, N; F&K: under "interlingual transfer"; P and B: "transliteration"
10. Foreignizing	Using a L1/L3 word by adjusting it to L2 phonology (i.e., with a L2 pronunciation) and/or morphology.	reparate for "repair" [adjusting the German word 'reparieren']	B, W; F&K: under "interlingual transfer"; N: under "transfer"
11. Code switching (language switch)	Including L1/L3 words with L1/L3 pronunciation in L2 speech; this may involve stretches of discourse ranging from single words to whole chunks and even complete turns.	Using the Latin *ferrum* for "iron".	T, F&K, B, W; N: under "transfer"
12. Use of similar-sounding words[1]	Compensating for a lexical item whose form the speaker is unsure of with a word (either existing or non-existing) which sounds more or less like the target item.	[Retrospective comment explaining why the speaker used *cap* instead of "pan":] *Because it was similar to the word which I wanted to say: "pan".*	
13. Mumbling[1]	Swallowing or muttering inaudibly a word (or part of a word) whose correct form the speaker is uncertain about.	And uh well Mickey Mouse looks surprise or sort of XXX [the 'sort of' marker indicates that the unintelligible part is not just a mere recording failure but a strategy].	
14. Omission[1]	Leaving a gap when not knowing a word and carrying on as if it had been said.	*then... er... the sun is is... hm sun is... and the Mickey Mouse....* [Retrospective comment: *I didn't know what 'shine' was.*]	
15. Retrieval	In an attempt to retrieve a lexical item saying a series of incomplete or wrong forms or structures before reaching the optimal form.	*It's brake er... it's broken broked broke.*	F&K

Inventory of Strategic Language Devices with Descriptions/Definitions, Examples (Based on Dörnyei & Scott, 1995a,1995b), and Indications Whether They Were Included in Any Other Taxonomies (T=Tarone, 1977; F&K=Færch & Kasper, 1983b; B=Bialystok, 1983; P=Paribakht, 1985; W=Willems, 1987; N=Nijmegen Group)

STRATEGY	DESCRIPTION	EXAMPLE	OTHER TAXONOMIES
16a. Self-repair	Making self-initiated corrections in one's own speech.	*then the sun shines and the weather get be... gets better.*	W
16b. Other-repair	Correcting something in the interlocutor's speech.	Speaker:... *because our tip went wrong...* [...] Interlocutor: *Oh, you mean the tap.* S: *Tap, tap...*	
17. Self-rephrasing[2]	Repeating a term, but not quite as it is, but by adding something or using paraphrase.	*I don't know the material...what it's made of...*	(Tarone & Yule, 1987)
18. Over-explicitness (waffling)[3]	Using more words to achieve a particular communicative goal than what is considered normal in similar L1 situations.	(This CS was not included in Dörnyei & Scott's , 1995a, 1995b, taxonomy)	(Tarone & Yule, 1987)
19. Mime (nonlinguistic/ paralinguistic strategies)	Describing whole concepts nonverbally, or accompanying a verbal strategy with a visual illustration.	[Retrospective comment:] *I was miming here, to put it out in front of the house, because I couldn't remember the word.*	T, F&K, B, P, W; N: under either "analytic" or "holistic strategies"
20. Use of fillers[4]	Using gambits to fill pauses, to stall, and to gain time in order to keep the communication channel open and maintain discourse at times of difficulty.	Examples range from very short structures such as *well*; *you know*; *actually*; *okay*, to longer phrases such as *this is rather difficult to explain*; *well, actually, it's a good question.*	
21a. Self-repetition[5]	Repeating a word or a string of words immediately after they were said.	[Retrospective comment:] *I wanted to say that it was made of concrete but I didn't know 'concrete' and this is why "which was made, which was made" was said twice.*	(Tarone & Yule, 1987)
21b. Other-repetition	Repeating something the interlocutor said to gain time.	Interlocutor: *And could you tell me the diameter of the pipe? The diameter.* Speaker: *The diameter? It's about er... maybe er... five centimeters.*	
22. Feigning understanding[6]	Making an attempt to carry on the conversation in spite of not understanding something by pretending to understand.	Interlocutor: *Do you have the rubber washer?* Speaker: *The rubber washer? ... No I don't.* [Retrospective comment: *I didn't know the meaning of the word, and finally I managed to say I had no such thing.*]	
23. Verbal strategy markers[7]	Using verbal marking phrases before or after a strategy to signal that the word or structure does not carry the intended meaning perfectly in the L2 code.	E.g.: (strategy markers in bold): (a) marking a circumlocution: *On the next picture...* **I don't really know what's it called in English**... *it's uh this kind of bird that... that can be found in a clock that strikes out or [laughs] comes out when the clock strikes*; (b) marking approximations: *it's some er... it's* **some kind of** *er... paper*; (c) marking foreignizing: ... *a panel [with an English accent],* **I don't know whether there's a name in English or not** *[laughter] just it's a panel flat*; (d) marking literal translation: *it's er... a smaller medium flat and in,* **we call them** *blockhouse, but it's not it's not made of blocks*; (e) marking code switching: *the bird from the clocks come out and say "kakukk"* **or I don't know what**; see also the example for message abandonment.	
24a. Direct appeal for help	Turning to the interlocutor for assistance by asking an explicit question concerning a gap in one's L2 knowledge.	*it's a kind of old clock so when it strucks er... I don't know, one, two, or three 'clock then a bird is coming out. What's the name?*	T, F&K, W
24b. Indirect appeal for help	Trying to elicit help from the interlocutor indirectly by expressing lack of a needed L2 item either verbally or nonverbally.	*I don't know the name...* [rising intonation, pause, eye contact]	T, F&K, W
25. Asking for repetition	Requesting repetition when not hearing or understanding something properly.	*Pardon? What?*	
26. Asking for clarification	Requesting explanation of an unfamiliar meaning structure.	*What do you mean?*, *You saw what?* Also 'question repeats,' that is, echoing a word or a structure with a question intonation.	W
27. Asking for confirmation	Requesting confirmation that one heard or understood something correctly.	Repeating the trigger in a 'question repeat' or asking a full question, such as *You said...?*, *You mean...?*, *Do you mean...?*	W
28. Guessing	Guessing is similar to a confirmation request but the latter implies a greater degree of certainty regarding the key word, whereas guessing involves real indecision.	E.g.: *Oh. It is then not the washing machine. Is it a sink?*	

Inventory of Strategic Language Devices with Descriptions/Definitions, Examples (Based on Dörnyei & Scott, 1995a,1995b), and Indications Whether They Were Included in Any Other Taxonomies (T=Tarone, 1977; F&K=Færch & Kasper, 1983b; B=Bialystok, 1983; P=Paribakht, 1985; W=Willems, 1987; N=Nijmegen Group)

STRATEGY	DESCRIPTION	EXAMPLE	OTHER TAXONOMIES
29. Expressing non-understanding	Expressing that one did not understand something properly either verbally or nonverbally.	Interlocutor: *What is the diameter of the pipe?* Speaker: *The diameter?* I: *The diameter.* S: *I don't know this thing.* I: *How wide is the pipe?* Also, puzzled facial expressions, frowns and various types of mime and gestures.	
30. Interpretive summary	Extended paraphrase of the interlocutor's message to check that the speaker has understood correctly.	*So the pipe is broken, basically, and you don't know what to do with it, right?*	W
31. Comprehension check	Asking questions to check that the interlocutor can follow you.	*And what is the diameter of the pipe? The diameter. Do you know what the diameter is?*	W
32. Own-accuracy check	Checking that what you said was correct by asking a concrete question or repeating a word with a question intonation.	*I can see a huge snow... snowman? snowman in the garden.*	
33a. Response: repeat	Repeating the original trigger or the suggested corrected form (after an other-repair).	See the example of other-repair.	
33b. Response: repair	Providing other-initiated self-repair.	Speaker: *The water was not able to get up and I...* Interlocutor: *Get up? Where?* S: *Get down.*	
33c. Response: rephrase	Rephrasing the trigger.	Interlocutor: *And do you happen to know if you have the rubber washer?* Speaker: *Pardon?* I: *The rubber washer... it's the thing which is in the pipe.*	
33d. Response: expand	Putting the problem word/issue into a larger context.	Interlocutor: *Do you know maybe er what the diameter of the pipe is?* Speaker: *Pardon?* I: *Diameter, this is er maybe you learnt mathematics and you sign er with th this part of things.*	
33e. Response: confirm	Confirming what the interlocutor has said or suggested.	Interlocutor: *Uh, you mean under the sink, the pipe? For the...* Speaker: *Yes. Yes.*	
33f. Response: reject	Rejecting what the interlocutor has said or suggested without offering an alternative solution.	Interlocutor: *Is it plastic?* Speaker: *No.*	

APPENDIX B: Original and translated AH initial questionnaire (without LLS)

Apellidos y nombre ……. Edad…….

1. Estudios (España)…….Curso…..

2. Lengua(s) materna(s)…….

3. Haz una valoración de tu nivel de inglés según este baremo:

(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado

listening	**speaking**	**reading**	**writing**

4. ¿Has estado en algún país extranjero en que te comunicaras en inglés durante dos semanas o más? Indica cada estancia y especifica lugar, actividades realizadas (curso intensivo, turismo, estancia con familia nativa, prácticas en empresa, trabajo), duración y tu edad al inicio.

5. ¿Has asistido a academias u otro tipo de clases extraescolares en España? Por favor, indica duración, tipo de curso (intensivo o extensivo) y edad de inicio.

6. ¿Sueles ver películas, televisión o leer libros en inglés? ¿Con qué frecuencia? ¿Desde hace cuánto tiempo?

7. ¿Has tenido algún otro tipo de contacto con la lengua inglesa (amigos, familiares, parejas)?

8. ¿Tienes algún título de inglés? En caso afirmativo, especifica cuál(es)

9. Además de tu lengua materna y el inglés, ¿qué otras lenguas conoces? Haz una valoración general de tu nivel:
(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado
Especifica brevemente dónde / cómo las has aprendido (colegio, universidad, academia, viajes, estancias largas, relaciones personales, etc.)

Lengua	**Nivel**	**Contexto de aprendizaje**

Full name ……. Age…….

1. Degree……..Year …..

2. First language(s) ……

3. Estimate your level of English using the following levels:

(1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced

listening	**speaking**	**reading**	**writing**

4. Have you had any experiences abroad that lasted for two weeks or longer during which you communicated in English? Please indicate each stay and specify place, purpose of the trip (intensive course, tourism, stay with native family, internship, work), duration and your age at onset.

5. Have you attended English extracurricular classes in Spain? Please specify duration, type of course (intensive or extensive) and your age at onset.

6. Do you usually watch films or television, or read books in English? How often? How long have you been doing it?

7. Have you had any other contact with the English language (friends, family, partners)?

8. Have you been awarded any English language certificate? Please specify which one.

9. Apart from your first language and English, what other languages do you speak? Please estimate your level: (1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced
Briefly specify where/how you learned them (at school, university, language school, trips, personal relationships).

Language	**Level**	**Learning context**

APPENDIX C: Original and translated SA initial questionnaire (without LLS)

Apellidos y nombre ……. Edad…….

1. Estudios (España)……Curso……

2. Fecha de llegada a la universidad de acogida ……..……..

3. Fecha aproximada de finalización de la estancia……..……..

4. Lengua(s) materna(s) ……

5. Haz una valoración de tu nivel de inglés según este baremo:

(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado

listening	**speaking**	**reading**	**writing**

6. ¿Has estado en algún país extranjero en que te comunicaras en inglés durante dos semanas o más? Indica cada estancia y especifica lugar, actividades realizadas (curso intensivo, turismo, estancia con familia nativa, prácticas en empresa, trabajo), duración y tu edad al inicio.

7. ¿Has asistido a academias u otro tipo de clases extraescolares en España? Por favor, indica duración, tipo de curso (intensivo o extensivo) y edad de inicio.

8. ¿Sueles ver películas, televisión o leer libros en inglés? ¿Con qué frecuencia? ¿Desde hace cuánto tiempo?

9. ¿Has tenido algún otro tipo de contacto con la lengua inglesa (amigos, familiares, parejas)?

10. ¿Tienes algún título de inglés? En caso afirmativo, especifica cuál(es)

11. Además de tu lengua materna y el inglés, ¿qué otras lenguas conoces? Haz una valoración general de tu nivel:
(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado
Especifica brevemente dónde / cómo las has aprendido (colegio, universidad, academia, viajes, estancias largas, relaciones personales, etc.)

Lengua	**Nivel**	**Contexto de aprendizaje**

Full name ……. Age…….

1. Degree……..Year …..

2. Date of arrival to the host university……

3. Estimated date of return……

4. First language(s) ……..

5. Estimate your level of English using the following levels:

(1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced

listening	**speaking**	**reading**	**writing**

6. Have you had any experiences abroad that lasted for two weeks or longer during which you communicated in English? Please indicate each stay and specify place, purpose of the trip (intensive course, tourism, stay with native family, internship, work), duration and your age at onset.

7. Have you attended English extracurricular classes in Spain? Please specify duration, type of course (intensive or extensive) and your age at onset.

8. Do you usually watch films or television, or read books in English? How often? How long have you been doing it?

9. Have you had any other contact with the English language (friends, family, partners)?

10. Have you been awarded any English language certificate? Please specify which one.

11. Apart from your first language and English, what other languages do you speak? Please estimate your level: (1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced
Briefly specify where/how you learned them (at school, university, language school, trips, personal relationships).

Language	**Level**	**Learning context**

APPENDIX D: Original and translated AH final questionnaire (without LLS)

Apellidos y nombre …….

1. Haz una valoración de tu nivel de inglés actual según este baremo:

(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado

listening	**speaking**	**reading**	**writing**

2. Durante estos cuatro meses, ¿has realizado algún viaje a países de habla inglesa o donde hayas utilizado el inglés para comunicarte? ¿A dónde? ¿Durante cuánto tiempo? ¿Qué actividad realizaste durante la estancia?

3. En este tiempo ¿has asistido a academias u otro tipo de clases extracurriculares? Por favor, indica duración y tipo de curso (intensivo o extensivo).

4. ¿Aproximadamente cuántas películas en inglés (en casa o en el cine) has visto durante este tiempo?

5. ¿Has leído o estás leyendo algún libro en inglés en estos meses? Sí/No. En caso afirmativo, ¿Cuántos?

6. ¿Con qué frecuencia sueles leer algún periódico en inglés (impreso o bien on-line)?

- ☐ Cada día o prácticamente cada día
- ☐ Entre tres y cinco días a la semana
- ☐ Uno o dos días a la semana
- ☐ Casi nunca o nunca

7. ¿Qué promedio de tiempo A LA SEMANA sueles dedicar a mirar la televisión (o productos de televisión por Internet) en inglés?

……..

8. ¿Cuántas horas de clase en inglés has tenido a la semana este semestre? (incluidas asignaturas de lengua inglesa de la universidad)

9. ¿Has realizado alguna actividad extra (teatro, deporte, etc…) en inglés? En caso afirmativo, ¿cuál/cuáles?

10. *Durante estos 4 meses, fuera de la universidad, en tu entorno personal, ¿te has comunicado con alguien en inglés en tu vida diaria? Indica la siguiente información sobre un máximo de 4 personas con las que tienes más relación en inglés:*

NOMBRE	Frecuencia con que ves a esta persona	Nacionalidad

11. *Cuando has estado* **en la universidad** *(y en el trabajo si estás trabajando), ¿te has comunicado con alguien en inglés, fuera del horario de clase? Indica la siguiente información sobre un máximo de 4 personas con las que tienes más relación en inglés:*

NOMBRE	Frecuencia con que ves a esta persona	Nacionalidad

Full name …….

1. Estimate your level of English using the following levels:

(1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced

listening	**speaking**	**reading**	**writing**

2. During these four months, have you traveled to English-speaking countries or done any other trips in which you have communicated in English? Where? For how long? What did you do during your stay?

3. During these four months, have you attended extracurricular language classes? Please specify duration and type of course (intensive or extensive).

4. How many films in English have you watched during this time?

5. Have you read during this time or are you reading any book in English? How many?

6. How often do you read the newspaper in English?

- ☐ Every day
- ☐ 3-5 days a week
- ☐ 1-2 days a week
- ☐ Never or hardly ever

7. How many times a week do you usually watch television (or online television products) in English? ……..

8. How many hours of classes taught in English have you had per week this semester? (including English courses at the university)

9. Have you participated in any other activity (theatre, sports, etc.) in English? Which one(s)?

10. During these four months, outside the university, in your personal environment, have you used English to communicate with anyone regularly? Specify the information requested about up to four people with whom you communicate in English the most:

NAME	How often do you see them?	Nationality

11. When at the university (and at work if you have been working), have you used English to communicate with anyone outside class hours? Specify the information requested about up to four people with whom you communicate in English the most:

NAME	How often do you see them?	Nationality

APPENDIX E: Original and translated SA final questionnaire (without LLS)

Apellidos y nombre …….

1. Haz una valoración de tu nivel de inglés actual según este baremo:

(1) elemental, (2) intermedio bajo, (3) intermedio, (4) intermedio alto, (5) avanzado

listening	**speaking**	**reading**	**writing**

2. ¿Desde que llegaste al país de acogida en septiembre, aproximadamente cuantos días has estado fuera del Reino Unido/Irlanda/Estados Unidos?

3. ¿Qué promedio de tiempo DIARIO estimas que has destinado a la comunicación vía correo electrónico, teléfono, chats, Skype, etc… con amigos y/o familia con la que has utilizado tu lengua materna?

4. ¿Dónde te has alojado este semestre?
   - Residencia
   - Piso compartido
   - Familia

5. ¿Aproximadamente cuántas películas en inglés (en casa o en el cine) has visto durante este tiempo?

6. ¿Has leído o estás leyendo algún libro en inglés en estos meses? Sí/No. En caso afirmativo, ¿Cuántos?

7. ¿Con qué frecuencia sueles leer algún periódico en inglés (impreso o bien on-line)?
   - ☐ Cada día o prácticamente cada día
   - ☐ Entre tres y cinco días a la semana
   - ☐ Uno o dos días a la semana
   - ☐ Casi nunca o nunca

8. ¿Qué promedio de tiempo A LA SEMANA sueles dedicar a ver la televisión (o productos de televisión por Internet) en inglés?
……..

9. ¿Cuántas horas de clase en inglés has tenido a la semana este semestre? (incluidas asignaturas de lengua inglesa de la universidad)

10. ¿Has realizado alguna actividad extra (teatro, deporte, etc…) en inglés? En caso afirmativo, ¿cuál/cuáles?

11. *Fuera de la universidad, en tu **entorno personal** (en casa, con los amigos, con pareja), ¿te comunicas con alguien en inglés en tu vida diaria? Indica la siguiente información sobre un máximo de 4 personas con las que tienes más relación en inglés:*

	Frecuencia con que ves a esta	Nacionalidad

NOMBRE	persona	

12. *Cuando estás* **en la universidad** *(y en el trabajo si estás trabajando), ¿te comunicas con alguien en inglés, fuera del horario de clase? Indica la siguiente información sobre un máximo de 4 personas con las que tienes más relación en inglés:*

NOMBRE	Frecuencia con que ves a esta persona	Nacionalidad

13. Intenta valorar en porcentajes aproximados la exposición a y el uso activo de las diferentes lenguas (ej.: inglés 50%, lengua materna 40%, otras 10%)

	Exposición a las lenguas (recepción)	**Uso activo de las lenguas (producción)**
Inglés		
Lengua(s) materna(s)		
Otras		

Full name …….

1. Estimate your level of English using the following levels:

(1) elementary, (2) lower-intermediate, (3) intermediate, (4) upper-intermediate, (5) advanced

listening	**speaking**	**reading**	**writing**

2. Since you arrived in September to the host country, how many days have you spent outside the UK/Ireland/United States?

3. How much time do you estimate you have spent DAILY communicating with friends and/or family per e-mail, telephone, chats, Skype, etc. in your first language?

4. Where have you been living during this semester?

- Hall of residence
- Shared flat
- Host family

5. How many films in English have you watched during this time?

6. Have you read during this time or are you reading any book in English? How many?

7. How often do you read the newspaper in English?

- ☐ Every day
- ☐ 3-5 days a week
- ☐ 1-2 days a week
- ☐ Never or hardly ever

8. How many times a week do you usually watch television (or online television products) in English?
……..

9. How many hours of classes taught in English have you had per week this semester? (including English courses at the university)

10. Have you participated in any other activity (theatre, sports, etc.) in English? Which one(s)?

11. During these four months, outside the university, in your personal environment (at home, with friends, with partners), have you used English to communicate with anyone regularly? Specify the information requested about up to four people with whom you communicate in English the most:

NAME	How often do you see them?	Nationality

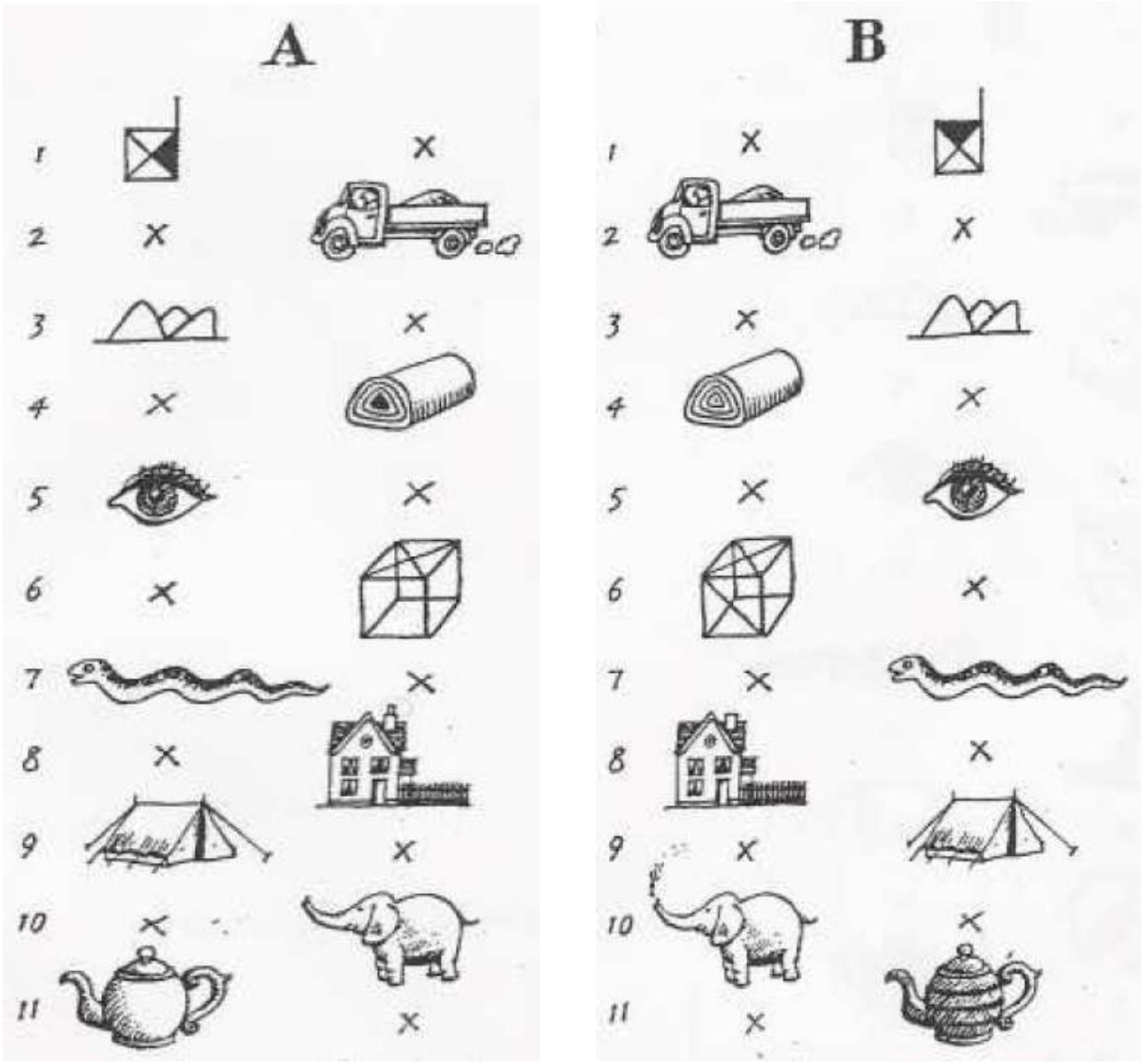
12. When at the university (and at work if you have been working), have you used English to communicate with anyone <u>outside class hours</u>? Specify the information requested about up to four people with whom you communicate in English the most:

NAME	How often do you see them?	Nationality

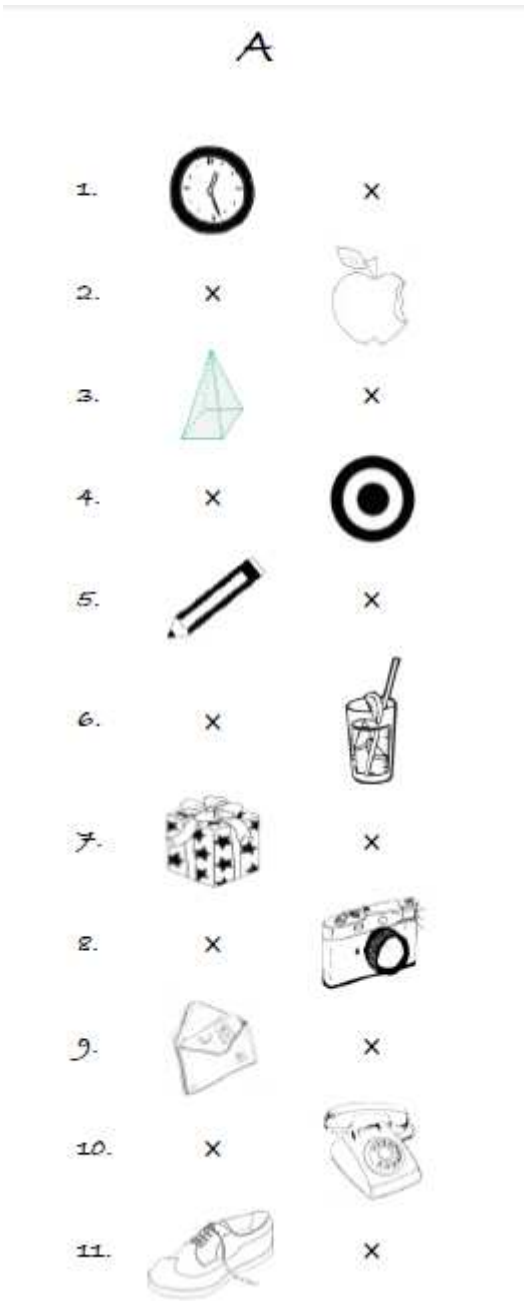
13. Estimate the percentage of time spent exposed to and using actively the different languages you have been using (e.g. English 50%, first language 40%, others 10%).

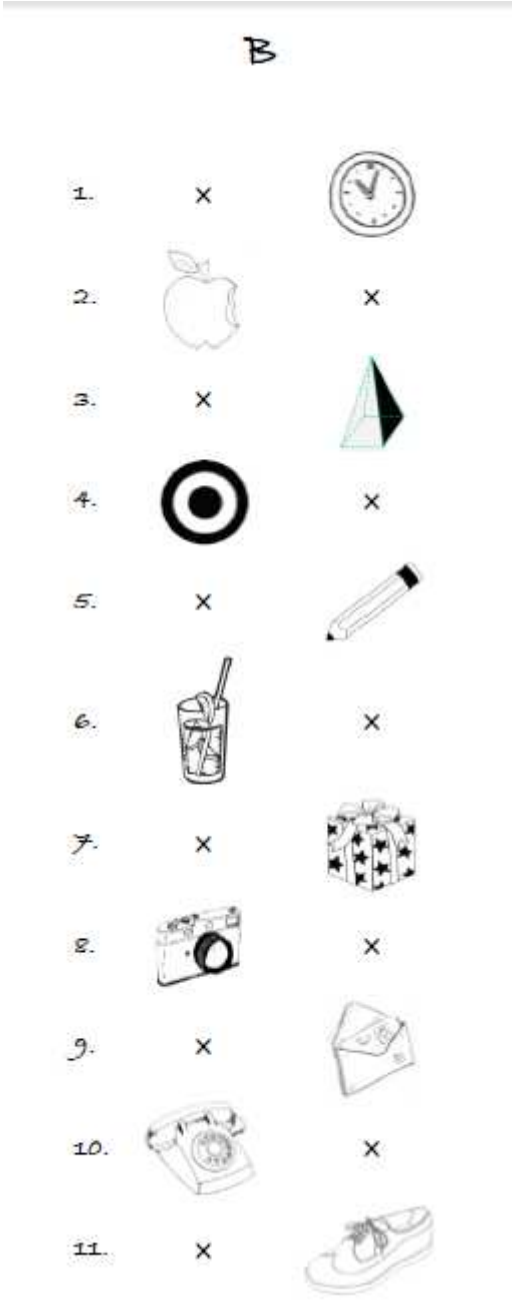
	Exposure to languages (reception)	**Active use of languages (production)**
English		
First language(s)		
Others		

APPENDIX F: T1 CS elicitation task (Khan, 2010)



APPENDIX G: T2 CS elicitation task





APPENDIX H: Original and translated LLS questionnaire (Tragant et al., 2013)

A continuación, tendrás que indicar con qué frecuencia utilizas diferentes medios para aprender inglés. Cada persona aprende de un modo diferente y hay muchos otros medios aparte de los que se mencionan aquí, por lo que nos gustaría que fueras completamente sincero/a en tus respuestas. En cada caso, indica la frecuencia marcando con una X según el siguiente baremo:

(1) Nunca o casi nunca. (2) Muy pocas veces. (3) Pocas veces.

(4) Algunas veces. (5) A menudo. (6) Siempre o casi siempre.

	1	**2**	**3**	**4**	**5**	**6**
Para mejorar mi nivel de inglés, repaso lo que he visto en clase de inglés o me pongo a prueba en casa.						
Cuando veo textos cortos en inglés, intento averiguar qué significan.						
Cuando oigo a alguien hablar inglés, hago un esfuerzo por ver lo que puedo entender.						
Para aprender vocabulario en inglés, hago listas de palabras y me las aprendo.						
Escribo las palabras nuevas tal como suenan o hago alguna anotación para recordar cómo se pronuncian.						
Escribo las palabras nuevas junto a una frase de ejemplo.						
Repito las palabras nuevas en voz alta varias veces.						
Para aprender gramática, escribo resúmenes o esquemas de las estructuras que estoy aprendiendo.						
Repaso las estructuras gramaticales mentalmente o en voz alta.						
Para leer en inglés, antes de empezar miro de qué trata el texto.						
Relaciono lo que sé sobre el tema con lo que dice el texto.						
Presto atención a las palabras que aparecen junto a aquellas que no entiendo en el texto.						
Intento entender la estructura gramatical de la frase.						
Presto atención a las palabras claves del texto o a las palabras que ya conozco.						
Para escribir en inglés, antes de empezar pienso en lo que quiero decir.						
Después, repaso lo que he escrito cuidadosamente.						
Mientras escribo en inglés, si quiero usar palabras, expresiones o formas gramaticales que no conozco o no recuerdo en el momento, utilizo algún libro o el cuaderno de clase.						

Please indicate below how often you use different means to learn English. Every person learns in a different way and there are many other ways to learn apart from those listed here, so we would like you to be completely honest with your answers. In each case, indicate the frequency of use with an "X" according to the following options:

(1) Never or hardly ever (2) Very rarely. (3) Rarely.
(4) Sometimes. (5) Often. (6) Always or almost always.

	1	2	3	4	5	6
To improve my English, I review what we have done in class or I test myself on my own.						
When I see short texts in English, I try to figure out what they mean.						
When I hear someone speak in English, I make an effort to see what I can understand.						
At home I make lists of words and study them.						
I write down the word as it rounds or make some sort of annotation to remember its pronunciation.						
I write down the word together with an example sentence.						
I repeat the word out loud several times.						
To improve my grammar, I write summaries or outlines of the structures that we are learning in class.						
I review the structures mentally or out loud.						
Before I start reading, I look at what the text is about.						
I relate what I know about the topic with what the text is about.						
I pay attention to the words that appear next to what I do not understand in a text.						
I try to understand the structure of the sentence.						
I pay attention to the key words or the words that I already know.						
Before I start writing, I first think about what I want to say.						
I carefully review what I have written.						
While writing in English, if I want to use words, expressions or grammatical forms that I do not know or do not remember at that moment, I use some book or the class notebook.						

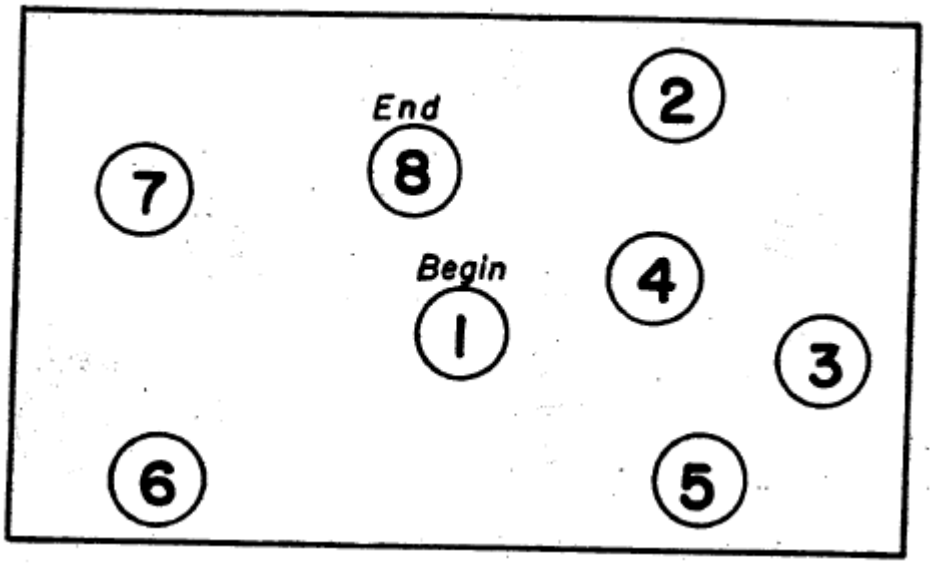
APPENDIX I: Trail-Making Test (Isaacs & Trofimovich, 2011)

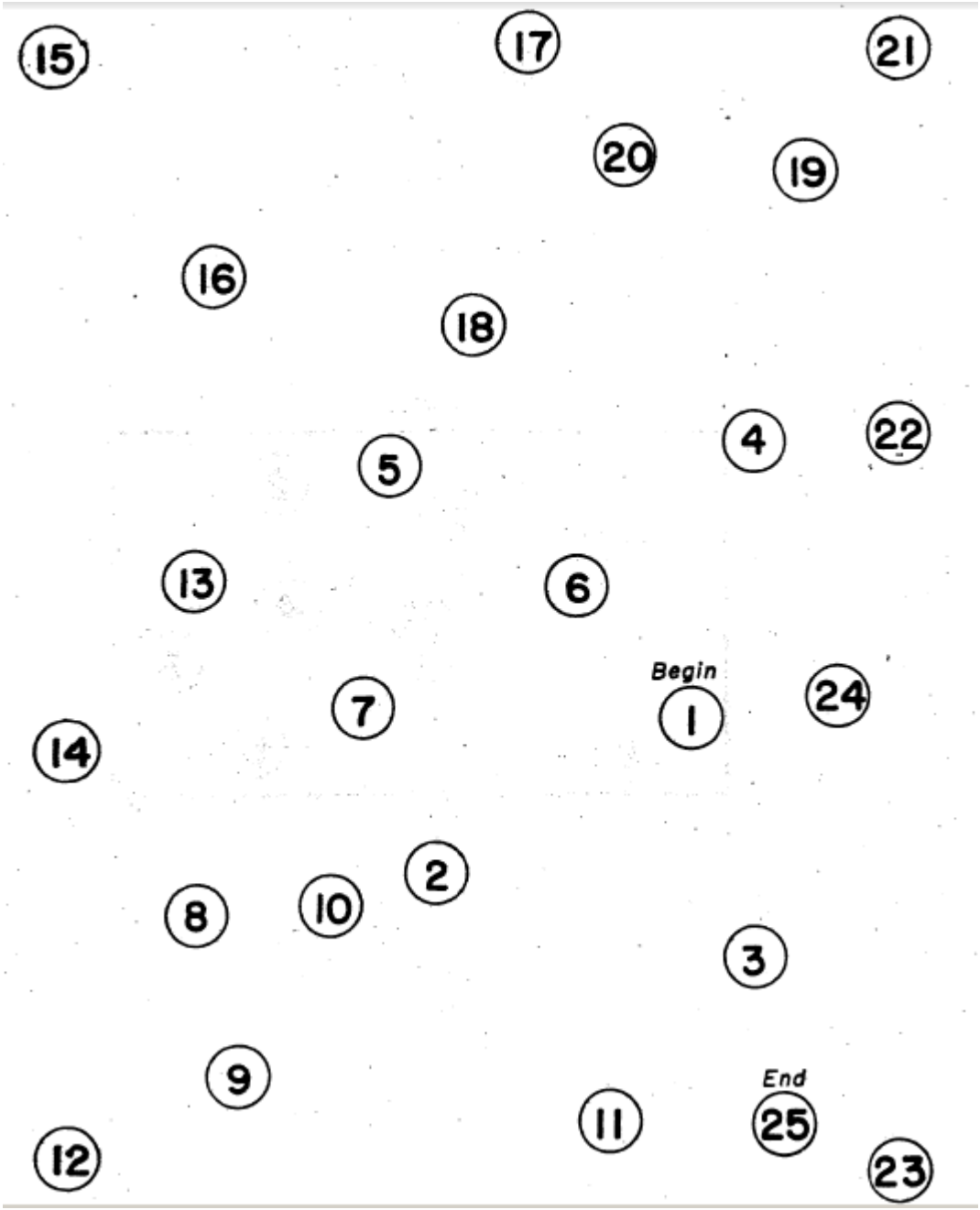
NAME:
DATE:

TRAIL MAKING

Part A TIME:

SAMPLE




15
17
21
20
19
16
18
4
22
5
13
6
Begin
24
7
1
14
2
8
10
3
9
End
11
25
12
23

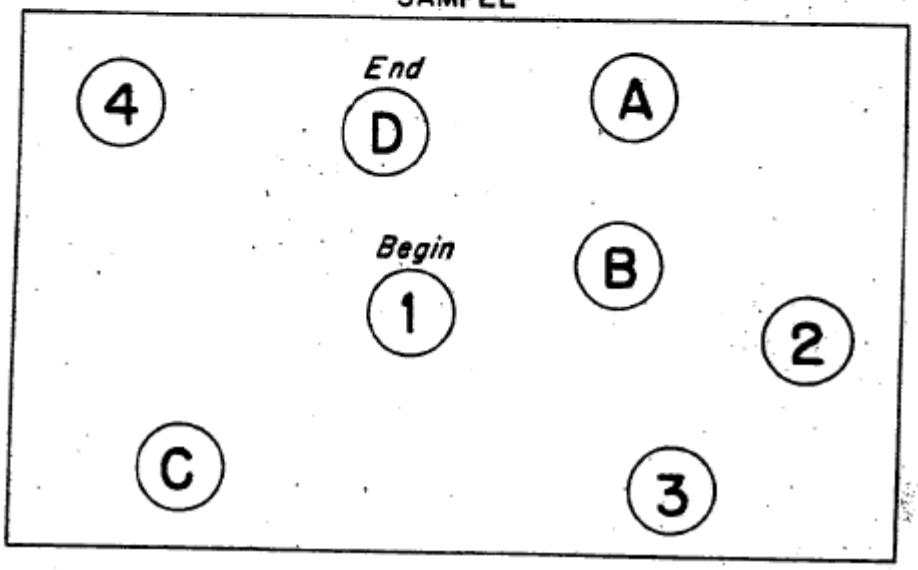
NAME:
DATE:

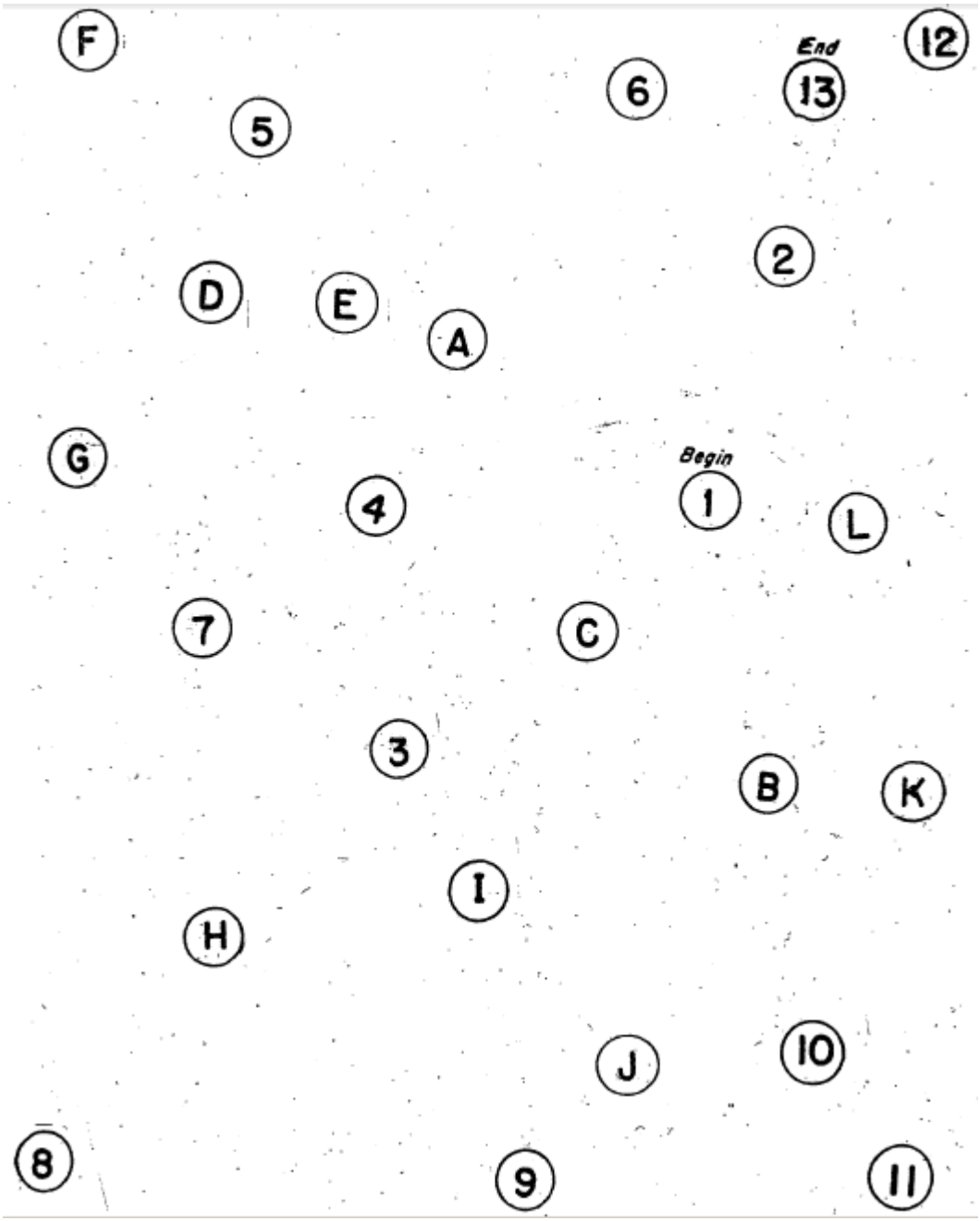
TRAIL MAKING

Part B

TIME:

SAMPLE




F
End
12
6
13
5
2
D
E
A
G
Begin
4
1
L
7
C
3
B
K
I
H
J
10
8
9
11

APPENDIX J: Instructions for raters

As a rater for this study, you will have to assess the communicative effectiveness of the participants when describing a series of pictures (i.e. how successful are they in getting their message through by referring to the different elements in each picture?). You have to express your evaluation with points, 1 point being the lowest mark (non-effective), 3 points being the highest (fully effective). It is important that you use the whole scale. Do not be afraid of using 1 when the participant does not get their message through in a generally intelligible way (for native or near-native speakers of English without any knowledge of Spanish or Catalan).

TASK FOR THE PARTICIPANTS:

Participants were asked to describe these pictures in full detail with information like identification of the object, shape, colours, position, components and activity (in the case of animate objects). Their descriptions were supposed to be detailed enough for the interviewer to create a mental image of what the participant was seeing and compare it to another set of pictures (some identical, some slightly different to the participants').

IMPORTANT CONSIDERATIONS IN RATING:

- Considering it is communicative effectiveness and not vocabulary range or grammatical correctness what we are looking into, participants should not be punished for not using the most accurate words or structures as long as they succeed in conveying meaning.
- In some cases participants over-interpret the picture or mix up left and right or do similar things that are not caused by lack of linguistic resources. They should not be punished for that either. Examples: the elephant is eating something or greeting its family; the elephant is fat/young/big/small…; in the windows of the house there are people looking out; there are X (number other than 6) ropes attaching the tent to the ground; the eye looks sad.

- The reason why there are several raters doing the same task is because independent opinions are necessary for the purpose of the study. Therefore, you should not comment on any aspect of the task with the other raters until after it is completely finished (February 2014).

Please do not hesitate to take as many breaks as you may need, to ask the researcher about anything related to the task and to share your thoughts on the task with her.

Thank you very much for your collaboration.

Pictures being described










Main conventions:

esp@ / cat@ = Spanish / Catalan word … = silence

[MP] [ei] [trIangle]… = mispronounced

in the le- in the right = unfinished word (left) xxx = unintelligible

[laughs] [cough] [interruptions]

NOTE: if a word is pronounced exactly as another existing English word, phonetic transcription might not be provided, just the pronounced word. Ex. “The elephant has a short tile” (“tail” was intended).