




Facultade de Filoloxía

Traballo de
Fin de Grao

FIRST LANGUAGE
ACQUISITION
CORPORA IN
ENGLISH AND
THEIR INFLUENCE
ON OTHER
LANGUAGES

Pablo Figueiredo Palacios


Milagros Fernández Pérez

Xullo 2016



Traballo de Fin de Grao presentado na Facultade de Filoloxía da Universidade de Santiago de Compostela para a obtención do Grao en Lingua e Literatura Inglesas

INTRODUCTION	1
1. CORPUS LINGUISTICS	3
1.1. Brief historical overview of Corpus Linguistics.....	3
1.2. Corpus Linguistics: theory or methodology?.....	4
1.3. Corpus-based and corpus-driven approaches	6
2. CHILD LANGUAGE.....	8
2.1. Language acquisition: three approaches	9
2.2. Child language and Cognitive Linguistics.....	11
2.3. Stages of language development.....	13
3. CHILDES: Child Language Data Exchange System	15
3.1. The database	16
3.2. The CHAT system	16
3.3. The CLAN programs	17
3.4. Sociolinguistics and CHILDES	17
4. SELECTION OF CORPORA	19
4.1. Koiné corpus	19
4.2. Matching Koiné with the other corpora.....	20
4.2.1. Belfast corpus	22
4.2.2. Wells corpus	23
4.2.3. Gathercole corpus	23
4.2.4. Gleason corpus.....	23
5. ANNOTATION IN OUR SELECTION OF CORPORA	24
5.1. Differences between Spanish and English.....	24
5.2. Dependent tiers used in the English corpora	26
5.3. Dependent tiers used in the Koiné corpus	26
5.4. Belfast corpus	27
5.4.1. Grammatical relations (%gra) and Morphemic semantics (%mor).....	27
5.4.2. Actions of the speaker (%act) and General comments (%com)	28
5.5. Wells corpus	29
5.5.1. Intonation (%int).....	29
5.5.2. Phonological processes (%xpho).....	30
5.6. Gathercole corpus	32
5.6.1. Errors (%err).....	32
5.6.2. Phonological processes (%xpho).....	35

5.7. Gleason corpus.....	37
5.7.1. Addressee (% add).....	37
5.7.2. Paralinguistic behaviors (% par).....	38
5.8. Koiné corpus	39
5.8.1 Developmental profile of a child	39
5.8.1.1. Phonetic-phonological component	40
5.8.1.2. Syntactic component	40
5.8.1.3. Lexical component	41
5.8.1.4. Pragmatic component.....	41
5.8.1.5. Evaluation.....	43
6. CONCLUSIONS.....	44
7. BIBLIOGRAPHIC REFERENCES	47
ANNEX 1.....	50
ANNEX 2.....	53

INTRODUCTION

This *Traballo de Fin de Grao* was initially conceived as an exploration on how corpora of language acquisition in English influenced (or not) corpora in other languages, namely in Spanish. We had to make some slight changes during the course of our investigation due to circumstances that we shall mention hereafter.

The paper can be seen as having two differentiated parts. The first one provides the background for the topics that we want to address, which can be grouped in two disciplines of the linguistic studies: Corpus Linguistics and Psycholinguistics (more specifically, first language acquisition). The second part presents an analysis of language acquisition corpora in English. The corpora are first analyzed and then confronted with one in Spanish. Chapter 3 functions as a pivot for the two parts. It brings together Corpus Linguistics and first language acquisition in Brian MacWhinney and Catherine Snow's CHILDES project.

This first part comprises chapters 1 and 2; chapter 1 contains the basics on Corpus Linguistics and the main lines of work that can be seen from a historical perspective. Chapter 2 deals with Child Language – acquisition theories, Cognitive Linguistics, and language development. Then, as we said, Corpus Linguistics and Language Acquisition meet in chapter 3, where we present the CHILDES project.

Then in the second part, chapter 4 deals with the corpora that we selected, and answers a series of basic questions: which corpora we will work with, why we chose them, and what characteristics they have. Chapter 5, which is the longest one, presents the analysis and the comparison of the corpora. However we considered relevant to explain firstly some of the issues that allow carrying out the comparison, such as linguistic differences of English and Spanish, formal features of the corpora, and so on. After that there is the analysis of each separate English corpus, and the characteristics that make them different from each other. The slight changes that we have previously mentioned concern the depth of the analysis that we wanted to do when we started working on the paper. The corpora in English (their annotations, in fact) did not allow us to make a comparison as detailed as we would have liked to.

The last section of chapter 5 contains the analysis of the Koiné corpus. Because of its particularities, it allowed us to prepare a developmental profile of one of the children of this corpus.

Chapter 6 brings together the conclusions of the whole paper, situating it in the context of our undergraduate studies, as well as in the context of the Linguistic Studies. Additionally, it contains several of the conclusions we arrived to in each of the chapters. A commentary on the path we followed when conceiving this paper is also present, and it is useful to see the relation (or relations) between the different sections, and how everything is ultimately connected.

Finally, we included in the annexes some charts that we elaborated and helped us to analyze the corpora, and the transcripts we used for the developmental profile. The charts were useful for chapters 4 and 5, and the transcripts for the last subsections of chapter 5.

1. CORPUS LINGUISTICS

In the first place, we shall explain what a corpus is. In its etymological sense, “corpus” means “body.” It was a term initially used to refer to collections of commentaries of the Bible. However, when used in Linguistics, a corpus used to be something slightly different: A collection of linguistic data produced in real-life situations, collected in order to serve a purpose – the study of language in use. This is the traditional definition of a corpus (rather simple, as we can see), but in the latest years this definition has been refined. As Hunston points out, a corpus is “a collection of texts¹ (or parts of texts) that are stored and accessed electronically” (Hunston 2002: 2). This distinction is crucial, and it is what confers Corpus Linguistics (CL from now on) its versatility. It is impossible for us now to think of CL without immediately associating it with computers.

Furthermore, it is important to point out that a corpus, if we want to draw significant conclusions from it, has to be carefully planned; it has to be designed according to the objectives of the investigations we want to conduct. The design of a corpus involves a series of factors, ranging from the environment where the recordings will take place, to the representativeness of the sample, or the socio-economic background of the people recorded; however, these are not the only ones, there are many others.

1.1. Brief historical overview of Corpus Linguistics

If we want to find the antecedents of modern-day corpus, it is mandatory for us to mention the first electronic corpus ever compiled: the Brown University Standard Corpus of Present-day American English (or plainly Brown Corpus), which was created in the 1960s. It was designed by investigators Henry Kučera and W. Nelson Francis at Brown University, and it contained roughly 1 million words. At the time it was seen as a waste of time and financial resources, due partially to the rising popularity of Chomsky’s nativism, which is diametrically opposed to the empirical approach that the use of corpora brought about.

However, acquisition corpora existed long before the appearance of computers; they were corpora in the strict sense of “collections of text”, that is, in the traditional sense.

¹ The “texts” that Hunston mentions include both discursive and textual data. This means that they can be drawn from actual spoken utterances (discourse) or from written texts.

Charles Darwin, for instance, kept regular, systematic diaries of the speech productions of his children. Even in the early stages of the discipline, in the 20th century, the methodology was rather cumbersome – notepads and analogical tape recorders were used, which were later substituted by digital recording systems, and in some cases, even video recording. With the rise of computers, thanks to specialized software, the once tedious task of listening and transcribing those recordings has become much quicker and easier to perform. It is also worth mentioning the pioneering work of Roger Brown, *A First Language: The Early Stages* (1973), which follows longitudinally the language acquisition process of three children, – Adam, Eve, and Sarah – namely, the semantic and grammatical aspects of their development. In Spanish we find an interesting study, conducted by Fuensanta Hernández Pina, entitled *Teorías psicosociolingüísticas y su aplicación a la adquisición del español como lengua materna* (1984) in which Hernández Pina describes the process of acquisition of her own son, as well as verifying the validity of the existing theories.

1.2. Corpus Linguistics: theory or methodology?

When we define CL, it is mandatory to mention a classic long-standing debate that has occupied the scientific community: is CL a linguistic theory of its own or is it a methodological approach to Linguistics? There are extensive lists of scholarly works dealing with this issue, but for the sake of our investigation we have to consider CL a methodology. The reasoning behind this assertion can be summarized in a few lines: CL provides us with the resources to investigate, on the one hand, the structural components of language –Phonology, Syntax, Morphology, and so on; and on the other hand, the branches of Linguistics – Sociolinguistics, Psycholinguistics, Historical Linguistics, etc. What this means is that in addition to us being able to investigate the *parole*, we can also obtain information about the *langue*. In this regard, Berber Sardinha says that

a Linguística de Corpus é uma perspectiva, isto é, uma maneira de se chegar à linguagem, e faz alusão ao conceito de teoria linguística como janela que molda como enxergamos a linguagem. Dessa forma, segundo Hoey, a Linguística de Corpus não seria apenas um instrumental, mais sim uma *abordagem*. (Berber Sardinha 2004: 37)

Moreover, CL has proven to be a very useful tool for the development of materials related to second-language teaching, clinical linguistics, forensic linguistics, and many

other linguistic-related fields. CL can be used to research, analyze and obtain new knowledge about virtually every language-related field. This is its major asset, the enormous researching capacity with which we are provided.

This brings us to the very origins of our discipline, in a time when there was a strong disbelief about the necessity of compiling linguistic corpora, especially by Noam Chomsky and those who shared his views and theories. One of the main arguments against CL was that it dealt with language from a statistical or probabilistic point of view; Chomsky believed that the true object of study of Linguistics was the *langue*, that is, all the possible sequences in language, and not the *parole*, the actual language in use. From his point of view, the only way to access the linguistic knowledge was through introspection, and not through the use of corpora.

We have to take into account that when this initial criticism took place, CL was still in its very early stages, computer sciences were not as developed as they are today, and the researchers' work was harder and they were more prone to making mistakes. On the other hand, these initial corpora were infinitely smaller than the ones that exist today, which are really significant samples of the language. Linguists today are very careful when designing their corpora; now we know that one of the most important prerequisites when creating a corpus is the criteria used to include texts in the corpus. A badly designed corpus can lead to highly inaccurate or incorrect results, so this particular point should never be ignored.

As time has passed and CL has developed, Chomsky's criticism has been refuted. Michael A. K. Halliday has worked extensively on this, concluding that "there is no discontinuity when we rewrite frequency as probability" (Halliday & James 1993: 66). And this means that "the relative frequency of occurrence observed in the corpus, therefore, can be equated with the instantiation of probability in the grammar" (Tognini-Bonelli 2001: 75).

Furthermore, the bias that Chomsky referred to, and so greatly feared, can be compensated if the corpus is correctly designed. The fact remains that in certain aspects of language a statistic approach is not incorrect, as Guillermo Rojo points out: "las estructuras lingüísticas (en fonología, morfología y sintaxis) no son infinitas y, como

veremos en un punto posterior, presentan características estadísticas realmente interesantes” (Rojo 2002: 3).

The initial skepticism has been resolved and now practically all the departments of Linguistics include CL in their agendas.

1.3. Corpus-based and corpus-driven approaches

We should bear in mind that even when CL has been accepted as a perfectly valid methodological framework an internal division appears. This division concerns the ways in which corpora can be approached, the ways in which they are used, and the ways data are extracted from them. Two approaches can be differentiated: corpus-based and corpus-driven. Even though both consider corpora as their source of linguistic information, they also differ greatly. We shall see what these differences are.

On one hand, the corpus-based approach uses corpora as a means to corroborate a series of already formulated linguistic theories, or ideas. In some cases, also, this approach could be used to clarify certain points of those theories which need further refining. Corpus data would serve as the basis to back the ideas of the linguist: “the corpus is considered useful because, on occasions, it indicates where minor corrections and adjustments can be made to the model adopted and, of course, it can be also valuable as a source of quantitative evidence” (Tognini-Bonelli 2001: 66).

On the other hand, a corpus-driven approach would not consider corpora as a way to provide more evidence for existing theories, but rather to extract new linguistic information. So if we follow this approach, the theoretical conclusions that can be drawn ought to adjust strictly to the data available, not the opposite: “the theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus. [...] The evidence that comes to light has to be either rejected by argument or respected – it cannot be ignored” (Tognini-Bonelli 2001: 84).

Thus, in such cases as child grammar – emergent grammars – or those atypical grammars of patients with some sort of language pathology, it is impossible not to turn to empirical data, to data from corpora. Both in language acquisition and in clinical linguistics we ought to work with real cases. Even if an intuition-based study were acceptable in some very particular cases (cf. Itkonen 2005), it is a totally inadequate

view when we want to study atypical or emergent language. The only acceptable, possible, scientific option is to work with corpora. However, authors such as Fillmore believe that the two positions are correct, and in fact, they need each other. He confronts “corpus linguists” with “armchair linguists,” and says he does not think that

there are any corpora, however large, that contain information about all of the areas of English lexicon and grammar that [he wants] to explore; all that [he has] seen are inadequate. The second observation is that every corpus that [he has] had a chance to examine, however small, has taught [him] facts that [he] couldn't imagine finding out about in any other way. [His] conclusion is that the two kinds of linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body. (Fillmore 1992: 35)

In spite of Fillmore's assertions, which were made almost 25 years ago, corpus-driven investigation is the one that has to be done; we, as researchers, have a need for a documentary base sufficiently large that it allows us to draw significant conclusions – and we are still far from that goal. Michael Tomasello and other authors (developmental psychologists and usage-based linguists, chiefly) propose working with Dense Databases (DDBs) that will allow future investigators to conduct appropriate research. It is precisely the corpus-driven approach the one that enables the use of DDBs and dense sampling².

² Usually, longitudinal corpora record 1-2% of the child's speech productions, whereas dense corpora capture around 7-15% (Lieven & Behrens 2012).

2. CHILD LANGUAGE

From the moment they are born, children start communicating. Their first manifestations of communicative behavior are eye contact with their parents, babbling, pointing, and so on. As a matter of fact, a lack of these behaviors can be red flags for neurodevelopmental disorders such as autism (Fernández Pérez 2006). Specifically, pointing is the only way in which human babies communicate that is different from other animal species. Pointing creates a joint context in which the infant and the parent are involved. This is Tomasello's "joint attentional frame" or "common conceptual frame" (Tomasello 2009: 70). The joint attentional frame is highly important for the subsequent cognitive development of the child; they are prelinguistic conceptualizations that will later structure complex (linguistic) utterances.

Child language, then, is an ability that appears around one year of age, and goes on until the early teens, with the acquisition of some pragmatic uses and complex constructions. And of course, lexical learning continues through our whole lives. So child language is a developing system which has to be studied from a new perspective, different from that of the adult grammar, paying attention to its own peculiar constructions, also different from those of the final stages of acquisition:

En cualquiera de las vertientes lo prioritario son los DATOS DE ADQUISICIÓN. Son necesarias pues teorías instrumentales que faciliten el acceso a los materiales que se observan o con los que se experimenta, y que canalicen las descripciones necesarias en cada momento. Teorías analíticas suficientemente flexibles y versátiles que puedan dar cabida a materiales lingüísticos infantiles sin acomodarlos o someterlos al modelo adulto. (Fernández Pérez 2003: 274)

An approach based on concepts and categories that apply to the adult grammar would not take into the account the whole range of peculiarities and genuine features that characterize child language and make it unique. Child grammar is, therefore, an emergent grammar, so it cannot be studied with the categories of adult grammar in mind:

The emergentist approach to language acquisition views language as a structure arising from interacting constraints, much as the shape of the coastline arises from pressures exerted by ocean currents, underlying geology, weather patterns, and human construction. (MacWhinney 1998: 201)

But first, we shall see what have been the different perspectives concerning language acquisition.

2.1. Language acquisition: three approaches

Traditionally, the different visions on how children acquire language can be grouped around three categories or schools of thought: behaviorism, nativism, and interactionism.

Behaviorism, where the figure of B. F. Skinner stands out, defends that language is not different from any other human behavior, and its acquisition can be explained in terms of imitation and reinforcement learning, as a stimulus-response mechanism; the internal mental processes are completely dismissed: behaviorism only accounts for observable phenomena. According to this, the process of language learning is understood as one of training, rather than one of maturation.

Nativism was postulated by Noam Chomsky in the late 1950s, when he wrote a demolishing review of B.F. Skinner's *Verbal Behavior* (1957)³. In his review, Chomsky pointed out a major weakness of Skinner's approach: the so-called 'poverty of stimulus.' It is not possible for children to learn the grammar of a language through imitation, because the data they are exposed to in the early stages of their lives is very limited –and inadequate also (cf. Chomsky 1965). The 'poverty of stimulus' would become the fundamental cornerstone of Chomsky's nativist approach; given that children cannot learn a language imitating the language of their parents, there has to be an innate mechanism – the so-called *Language Acquisition Device*, or LAD – that allows them to acquire the grammar of the language they are exposed to.

The third approach to language acquisition is interactionism. Language is just one of the many cognitive processes of the human brain, and has to be studied accordingly. Many factors come into play in language acquisition, and so it is necessary for us to study the process taking all of them into account. Psychologists and psycholinguists, such as Vygotsky, Piaget, or Bruner, just to name a few, have developed their particular views within the framework of interactionism. Piaget and Vygotsky, specifically, were true pioneers when they developed their approaches in the early decades of the 20th century.

³ Chomsky, Noam (1959): "A Review of B.F. Skinner's *Verbal Behavior*," in *Language*, 35, no. 1, pp. 26-58.

It was Piaget who realized that child language could not be studied by comparing it with adult grammar, but instead what really was worthy of studying were the peculiarities of those emergent grammars. Piaget, said Vygotsky, “concentrated on the distinctive characteristics of child thought, on what the child has rather on what the child lacks. Through this positive approach he demonstrated that the difference between child and adult thinking was qualitative rather than quantitative” (Vygotsky 1934: 9).

More than eighty years later the assertion still stands; in language acquisition we study what the child has, and what processes gradually emerge. It is completely impossible to conduct a research today without this premise in mind.

Vygotsky’s social interactionist perspective shares some features with the rest of the approaches we have just mentioned, but it does not stop there, it goes a step further and provides an explanation to the dark areas of the other perspectives. Vygotsky is without a doubt one of the most important figures of the last century, despite being wrongfully unknown in Europe and the United States until the translation of *Thought and Language* in 1962. Psychology, Linguistics, Pedagogy, etc. would have been greatly benefited from Vygotsky’s knowledge had his work been known while he was still alive. Down below we will explain what these connections and clarifications are.

Both behaviorism and social interactionism consider that experience plays an important part in language acquisition; the crucial difference here is that the latter introduces mental functions, which behaviorism disregarded, as they were not directly observable.

Nativism and social interactionism agree that language is a human-specific ability that sets us apart from other animals, but this ability “is not that of generating universal grammatical structures, but rather of representing meaning shared with others through the use of conventional signs (whatever they may be)” (Bernicot 2014: 155).

And finally, both Piaget’s cognitivism and Vygotsky’s social interactionism assign a fundamental importance to cognitive processes, but social interactionism considers that “these cognitive processes are based on content-filled structures and not on structures whose elements are defined by extremely general characteristics” (Bernicot 2014: 155-156).

It is now sufficiently clear the lucidity of Vygotsky's vision, which brings together in one approach the various existing conceptions on language acquisition – even in some cases being ahead of his time. This is the strong point of the vygotskian perspective, the capacity to combine coherently all those theoretical approaches, with the addition that he never departed from empirical evidence.

2.2. Child language and Cognitive Linguistics

In the latest decades a new multidisciplinary trend of thought has taken over the task of studying child language. Linguists, psycholinguists, neurolinguists, or even Artificial Intelligence experts have occupied themselves with what is known as *Cognitive Linguistics* (Fernández Pérez 2007: 8-9). In this framework, we will follow Michael Tomasello's usage-based theory of language acquisition, which postulates that human brains have two sets of cognitive skills, not specifically "designed" for language acquisition but for other more general abilities, that will provide the child with the necessary resources for acquiring grammar. These two sets of abilities are *intention-reading* and *pattern-finding*:

'Intention-reading' is what children must do to discern the goals or intentions of mature speakers when they use linguistic conventions to achieve social ends, and thereby to learn these conventions from them culturally. [...] 'Pattern-finding' is what children must do to go productively beyond the individual utterances they hear people using around them to create abstract linguistic schemas or constructions. (Tomasello 2009: 69-70)

In the light of these facts, we should abandon the old idea that "syntax is grammar" and approach child language from another perspective, namely, from a communicative point of view. This is the point of view of Pragmatics. As a matter of fact, Tomasello's is very much a pragmatic approach to language acquisition. What he calls "mutually understood context" and "joint attentional frames" can be summarized in one word: *deixis*. Both Tomasello (2009) and Levinson (2003) agree when they say that deixis and the ability to "read" indexicality, are skills exclusive to human beings, for primates lack this set of cognitive abilities. Deixis, we must conclude, is the key to the first steps of language acquisition.

The main goal of children is to get their messages across, so what we need to do, if we want to study child language, is to value their communicative effectiveness, to see how the emergent grammar evolves into what will later be their adult grammar, and what processes (cognitive, social...) come into play⁴. The only way to do this is to resort to the data; we have to study child language directly from the communicative exchanges of children with their parents, with other children, with their teachers, and so on. This is where Corpus Linguistics shall help us with our endeavor. By recording and transcribing children we can access to the singularities and peculiarities of child language, and we will be able to understand better how language works.

However, this is not only a matter of common interest for linguists and Linguistics; if we are able to define a certain way, or certain patterns in which children learn how to use language, we can use this knowledge to help people – not only children, but adults as well – with cognitive or developmental issues: for instance, individuals with autism, aphasia, speech disorders, etc. Psychologists, neurologists, and speech therapists benefit from the study of language acquisition: this is yet another instance of the multidisciplinary nature of this field and the necessity to have a deeper understanding of the workings of the human mind.

As we have seen, corpus data are useful in acquisition for a number of reasons. Among the most important ones there is the use of data to draw linguistic profiles, a methodology initiated by the eminent British linguist David Crystal. In his 1981 book *Clinical Linguistics*, Crystal designed these charts and tables through which linguistic behaviors are evaluated. By having a standard of what is considered to be the normal development of language, it is possible to see if the process of language acquisition in children follows what is to be expected. In Crystal's own words:

A profile, as its name suggests, is no more than a first approximation to an accurate description; but it does at least imply that the salient, identifying features of a problem area have been isolated. To be useful, a profile of linguistic behaviour needs to be discriminating – to indicate the main differences between normal and abnormal, and to identify different categories of abnormality. The most useful profiles are those which

⁴ This does not mean that the rest of the components of grammar – phonology, morphology, syntax – are not important or should not be studied. In fact, the linguistic profiles that psycholinguists and speech pathologists use to evaluate child language and peculiar grammars (those of aphasic patients, for instance) take into account all components of grammar.

are based on an acquisitional dimension, because they can then be used simultaneously for assessment and remediation: by showing where a patient *is* on a profile chart, we can see immediately where he ought to be, and perhaps see paths which would enable us to get him there. (Crystal 1981: 22)

However, a profile does not have to be one of an aphasic patient, or a child with developmental issues; it can be done merely to measure the evolution of a child's language acquisition. It would be enough to select a sample of texts – transcripts – belonging to different chronological moments of the development of the child and see how he or she is doing, what processes are present, how he or she overcomes them.

2.3. Stages of language development

The process of language acquisition is a marvelous one. Children start their linguistic utterances as soon as 9 months old and by the time they are three they can already communicate a great deal of things. Not all children acquire language at the same speed, but we can roughly establish some stages in the process of acquisition, stages that characterize their degree of linguistic development. It is typically assumed that there are four main stages, or periods, which would be: (1) a prelinguistic stage⁵, from ages 0 to 1 or 1.5 years old, in which at first the child just babbles, and later on starts producing utterances similar to those of their linguistic environment (cf. Vihman 2014); (2) a holophrastic stage, from 9 or 10 months to 1.5 or 2 years old, in which the children start using “holophrases” or single words conveying complex meanings; (3) a syntactic stage, from 2 to 5 years old, in which children start using two-, three-, and four-element clauses, progressively; and finally (4) an advanced stage, from 5 years old onwards, where the use of complex syntactical structures and pragmatic competence are its defining aspects (Fernández Pérez 1999). This is a rather simple classification, and it can be further developed and subdivided; also, depending on the author the number of stages and sub-stages, or the ages of each stage can vary.

Crystal and Fletcher proposed a classification with seven stages, which fits almost perfectly the characterization we have just mentioned. They also said that

⁵ We understand the term “prelinguistic” as previous to the utterance of language-specific constructions, not as “non-linguistic;” children show communicative and interactive (therefore, linguistic) behaviors before they produce actual verbal emissions (cf. Fernández Pérez 2006).

a synthesis of the descriptive findings of the language acquisition literature provides a postulated set of age-related stages of syntactic development. Ages are averages, which will ultimately need to be refined with reference to socioeconomic, sex, and other well-known variables. (Crystal & Fletcher 1981: 170-171)

This means that they recognize the possible variation of ages depending on each and every individual child, but the theoretical foundations for the study of child language are still very solid. Nevertheless, we have to admit that what really interests us is not the existence of these stages, but the fact that every child goes through them in different chronological moments (cf. Fernández Pérez 2006). We need to talk about *development* rather than mere *acquisition*: the process is one of cognitive maturation, of cognitive development.

The difference between the number of phases, depending on the author(s) we read has a justification, precisely, on the *processes*. Depending on the relevance given to one or other of these processes, a different number of stages will be established. For example, if the process of emission of holophrases is considered to be rather important, a holophrastic stage, or period, will be established. If the process of predication combining three elements is thought to be more important than the predication with two (or four, or *n*) elements, more weight will be given to this stage over another with, let us say, two, or four elements (cf. Brown 1973). The stages of language acquisition, then, are subordinate to the processes considered and the hierarchy set between them. As we explained in the previous section, the way to evaluate in which stage the child is in – or which processes he or she is dealing with – in a precise moment, is by using an evolutionary linguistic profile.

3. CHILDES: Child Language Data Exchange System

If we want to do serious, meaningful, scientific research on how humans acquire language, we need to face it from an evolutionary point of view. The only scientifically valid approach to child language is to consider it as an emergent process and work directly with empirical data. The current status of CL and Language Acquisition is a direct consequence of these two disciplines running in parallel in the last decades, since progress in one of them has affected the other one directly. We cannot explain CL isolating it from Language Acquisition and vice versa, so in order to conduct a research in Language Acquisition it is mandatory to have access to corpus data. CL suffered in its first years a tremendously strong opposition from prestigious linguists who considered that it was a waste of time, resources and effort. Over the years CL has proven the exact opposite: that some aspects of language are only accessible through the analysis of corpora and its scientific validity is undeniable today.

It is well known that child language is an evolving system, a developmental process characterized by a grammar that differs in several aspects from adult grammar; child language has peculiar, genuine features of its own. All these characteristics point towards a different linguistic evaluation than that used in adult speech, in fact, those genuine features and uses are the very core of the acquisition studies. This is a task that would not be possible to carry out without real data, without transcripts of verbal interactions between children in an environment as natural as the circumstances allow it; this is a task that has to be carried out with the help of CL.

In our research, we are going to make use of the project that Brian MacWhinney and Catherine Snow, from the Carnegie Mellon University, started over thirty years ago: the CHILDES project (<http://childes.psy.cmu.edu/>). CHILDES stands for Child Language Data Exchange System, and its main goal is to allow researchers to access and to share child language transcript data –although data from bilingual individuals and from aphasic patients are also available. This is a later addition; CHILDES has been recently made a part of TalkBank, which is a larger corpus where the bilingual, aphasia, and second-language learning directories are stored. Another peculiarity of this project that differentiates it from others is that

in contrast, to adult corpora such as the British National Corpus (BNC) or the International Corpus of English (ICE), the CHILDES database does not have a pre-determined structure or a specific design. It includes data from many different research projects that researchers have made available to the scientific community. By contrast, adult corpora are designed based on a predetermined set of criteria. (Diessel 2008: 6)

All the data from CHILDES are publicly available. Over the years it has proved to be one of the most resourceful projects of its kind.

The CHILDES project consists of three main parts, which are described below: the database, the CHAT system and the CLAN programs.

3.1. The database

Firstly, the database is composed by transcriptions of conversations of children and adults; these have been incorporated to the system by researchers from all over the world that wanted to carry out their own investigations. It is an enormous repository of corpora with data from a great number of languages. The CHILDES database organizes the transcriptions in the following categories: North American English, British English, Bilingual Acquisition, Language Disorders, Narratives, Germanic Languages, Romance Languages, Slavic Languages, East Asian Languages, Celtic Languages, Other Languages, and PhonBank.

The CHILDES database is quite large; it contains 300 million characters, approximately 300MB of transcribed content. In some cases there are also audio and/or video files linked to the transcripts that, thanks to one of the CLAN programs, can be played as the transcript automatically scrolls down the computer screen.

3.2. The CHAT system

The second part of the project is the CHAT system, or Codes for the Human Analysis of Transcripts. It is the standardized format in which the transcriptions are computerized, and not only provides a standard notation for transcribing human speech, but also allows researchers to create new tags and codes that adjust to the typological particularities and variations of the different languages, a fact that contributes to the versatility of the whole project.

The transcript files consist of a header, where information about the participants is provided, along with the situational context, the date of the recording, and so on. Then the actual transcript is divided into lines: the main line and the dependent lines. In the main line is where the utterances of the participants are transcribed; the dependent lines contain the annotations made by the researchers, with information concerning the linguistic aspects of the utterances (Diessel 2008: 8).

However, this is also a double-edged weapon. The standardized format is quite useful if we want to compare several corpora of the same language, for instance, for we know that the tiers used are going to be the same, to a certain extent – each researcher might introduce new tiers according to the nature of their projects, but the main tiers are probably going to be the same. Although, when we want to compare different corpora in different languages we are certainly going to find a larger variety of tiers. This is the case especially if they belong to different linguistic families, which can sometimes difficult a comparison between corpora. We will elaborate on this in later sections of this paper.

3.3. The CLAN programs

The last part of CHILDES is the CLAN programs (Computerized Language ANalysis), a set of computer programs whose function is the analysis of language and the linking of transcripts to audio or video recordings. These programs were designed to analyze the data that are transcribed in the CHAT format, but ASCII files of any type can be analyzed with CLAN. The kinds of analyses that can be carried out using CLAN are “frequency counts, word searches, co-occurrence analyses, MLU counts, interactional analyses, text changes, and so on” (MacWhinney 1995: 130).

In the present paper we are only going to be using CLAN in a rather simple way, so we will not go into further detail explaining all the tools and possibilities that the programs offer.

3.4. Sociolinguistics and CHILDES

One aspect about CHILDES that we find highly relevant to our study is the treatment of English compared to the rest of the languages. The centrality of English in MacWhinney’s project is worth mentioning. We have to accept that English has been

for quite some time a sort of *lingua franca* of the Western world, as it once were Latin, French, or Spanish. There is no difference in the scientific community, as one might expect; English is, by far, the most used language in scientific research, papers, lectures, etc. When we started this particular study we expected to find a different situation, with more linguistic diversity in the corpora, but this was not the case. Not only is English indexed in separate directories –instead of finding it among the rest of the Germanic languages– but also there are separate folders for American and for British English. The rest of the geographical varieties of English are simply not there. Moreover, when MacWhinney summarizes the contents of the database, he establishes the following classification: “The corpora are divided into five major directories: English data, non-English data, narrative data, data from clinical populations, and data from bilingualism and second-language acquisition” (MacWhinney 2000).

We should not overlook that the English language makes up half of the total contributions of the database, but is this a justification of MacWhinney’s classification or is it yet another instance of the dominance of English in the scientific community (and in the world in general)?

The justification that we can find for this is that Brian MacWhinney’s background is not exactly that of a linguist; instead, he is a psychologist who has developed his professional career studying the psychological and neurological bases of the language. Taking into consideration his main researching interests, Linguistics –as linguists, not psychologists understand it– takes a second place, so to speak. Hence, the more “linguistic” factors of Linguistics, if we may put it thus, are not covered. We would like to think that a linguist would have taken the sociolinguistic approach into account and would have made things in a different way.

4. SELECTION OF CORPORA

As a comparative study, it is a point of paramount importance that we establish certain criteria in order to carry out our task in the most efficient and rigorous manner. The object of our comparison is English and Spanish child language corpora, and given that we are most familiar with Spanish, our corpus of reference shall be in this language. We have also worked in the past with one particular corpus included in the CHILDES project, which is the Koiné corpus, so we believe that this corpus shall be the starting point for our corpus selection. The Koiné corpus was compiled by the research group *Koiné*, directed by Professor Dr Milagros Fernández Pérez, from the University of Santiago de Compostela. The investigations conducted in the last years by *Koiné* have been dealing with the study of the so-called “peculiar language” and the value of communicative efficacy.

4.1. Koiné corpus

The Koiné corpus totals 71 Spanish-speaking children (although some of them were bilingual and also spoke Galician), aged 18-53 months, and recorded in the kindergartens they habitually attended. Five kindergartens collaborated, in different provinces of Galicia: *A Milagrosa* in Lugo, *Elfos* in A Estrada, and *Breogán, Santa Susana* and *Vite* in Santiago de Compostela. The recording sessions usually involved 4 or 5 children and one researcher, who tried to interact with them as little as possible, in order to let the children spontaneously interact with each other.

We need also to mention that Koiné features a set of new tiers created according to the rules of the CHAT system, in order to adapt the preexisting tiers (which are designed to best fit the typological characteristics of the English language) to the linguistic particularities of Galician and Spanish. As it is noted in the database manual:

Regarding the process of coding, it has been introduced the line %xmor, adapted to the peculiarities of the languages presented in this corpus -Spanish and Galician-. So far, this process has enabled us to record the properties of all child verbal constructions in a formal way. In addition, the annotation of line %xpar, which includes peculiar usages of children, is complete (MacWhinney 2000).

Another difference between Koiné and the rest of the corpora concerns the theoretical approach to which the researchers working with this corpus have subscribed.

MacWhinney suggests in the CHAT manual that “errors” should be annotated under the dependent line %err. But having reached this point, we need to make a huge distinction: MacWhinney does not differentiate between actual errors (honest mistakes, such as using one word instead of another one inadvertently, or mispronouncing a word) and the genuine processes of child language, which are the very essence of an emergent grammar. These processes are the most valuable resources the investigators have, and are of the utmost importance when evaluating the linguistic development of a child. What MacWhinney is doing here is focusing on language as a product rather than as a process.

Some approaches to language acquisition focus on the product – the end state to be achieved – rather than on the process. This distinction tends to capture one difference between linguistic and psycholinguistic approaches to acquisition. Linguists tend to focus on the product, for instance, what a relative clause looks like, laid out on the table for analysis. In contrast, the psycholinguist is more concerned with when the speaker needs a relative clause, how he accesses the pertinent structure, the phrases, words, syllables, and sounds, and then produces the utterance itself piece by piece. This has led to differences in emphasis, with linguistic approaches focussing [sic] more on the adultlike nature of children’s knowledge while psychological ones have focussed [sic] more on the changes that occur during development. (Clark 2009: 10)

Koiné, on the other hand, deals with those peculiar uses of child language. Dr Fernández Pérez and her collaborators paid close attention to these features and so their corpus is annotated differently to the other ones –regular and peculiar uses are labeled separately. In the next section there is a relation of all the dependent tiers that appear in all our corpora, with their respective descriptions.

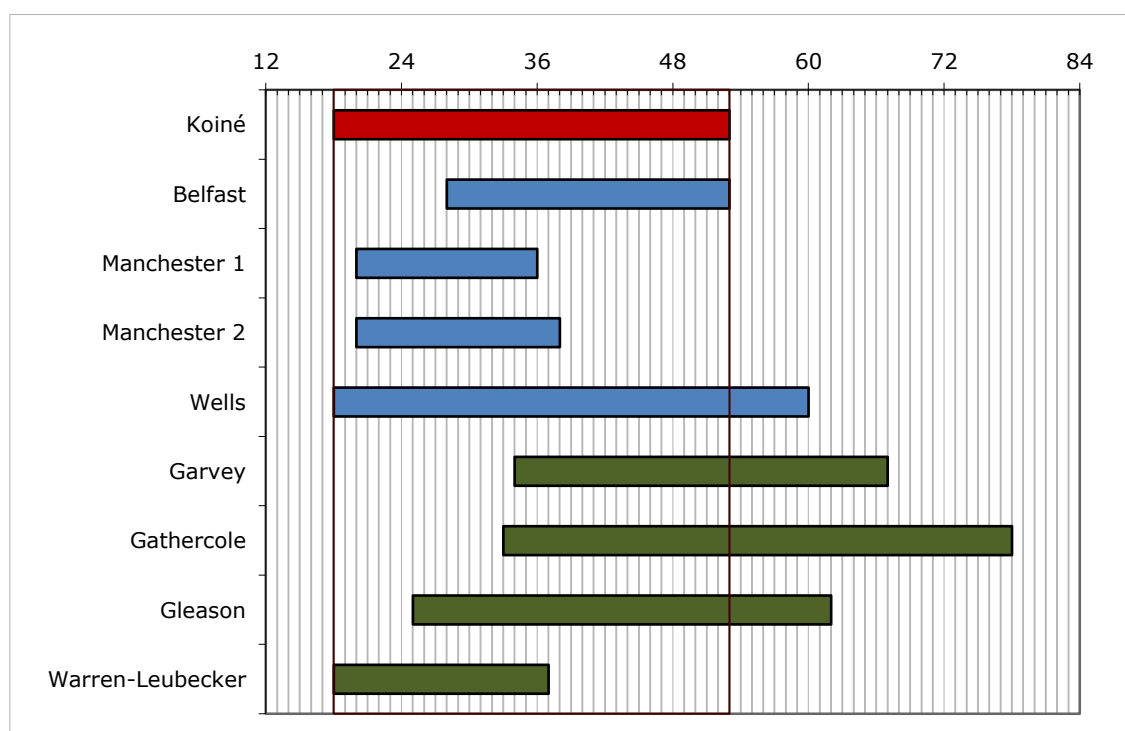
4.2. Matching Koiné with the other corpora

We have already mentioned the characteristics that we tried to find in the English directories of corpora contained in CHILDES; with an especial emphasis on the age range, but also the fact that the children are interacting with other children, with very few adult interventions. We were aware when we began this study that it was very difficult (or nearly impossible, we must say) to find one corpus that met these requirements completely, so we were looking for corpora with characteristics as close to these as we could find. In our initial selection we found eight English corpora that we

thought had an age range very similar to that of the Koiné corpus, which were the Belfast, the Manchester1, the Manchester2, the Wells, the Garvey, the Gathercole, the Gleason, and the Warren-Leubecker. In closer observations, we decided to discard both Manchester corpora, the Gathercole and the Warren-Leubecker, and keep the remaining corpora. Thus, there is a balance between the two English varieties found in CHILDES; two of them belong to the British English directory and the other two to the American English directory of the database.

The main reason for choosing these corpora is, as we have previously mentioned, the age range of the children, but there are other factors. For instance, the fact that some of them featured a non-standard dialect of English – namely the Belfast corpus and the Wells corpus – and the regional linguistic features are taken into account. We mention this because in the Koiné corpus we can also find a certain degree of intertwining or overlap between Spanish and Galician, and often some peculiar constructions or uses belonging to the Galician geographic variety of Spanish.

In the chart below we have, on the horizontal axis, the age in months, and on the vertical axis, the corpora we initially considered. The Koiné corpus appears, as a reference, in the first place in red. Then, in blue, we have the British English corpora, and finally, in green, the American English corpora. As we can see, some of the corpora that we discarded have an age range included within that of Koiné; however, we shall explain



the motivations that made us dismiss them.

The main reason for not considering working with the Manchester1 corpus was that it comprised a very short age range, compared with the rest of the corpora, so we thought it best not to have this corpus in our study. As for the Manchester2, the description in the CHILDES database manual did not include but two words: “four children”, and in comments it said “Dense sampling from four British children.” In view of the above, we believed that it would be more accurate working with other corpora.

Having selected our British English corpora, we proceeded to select two American English ones, in order to have a balance between the two varieties. Firstly, because of its reduced age range, we discarded the Warren-Leubecker corpus. Secondly, out of the other three, the one that had the fewest recordings per child was the Garvey corpus – two recordings per child, whereas the other two have at least 3 or 4 – so that is why it was discarded.

Now that we have explained our reasons for discarding those corpora, let us proceed to briefly explain the characteristics of the corpora we actually selected:

4.2.1. Belfast corpus

Directed by Alison Henry, from the School of Behavioural and Social Sciences of the University of Ulster (Northern Ireland). The participants were eight upper-working-class children: 4 boys and 4 girls, aged 2;0 to 4;6. The recording sessions took place at the children’s home once a month for 1 hour. These were parent-child interactions, with one investigator being present, although at times other children (friends of the target child) were present, and sometimes the recording sessions took place at the investigator’s house, with the investigator’s children being present.

It might seem contradictory that we discarded some other corpora precisely because they were adult-children interactions, and in fact this is the case in this particular corpus. But, as we have mentioned before, the age range of this corpus, as well as the non-standard variety of English, make this corpus interesting for our study. The goal of this study was “to establish how children treated variable forms in the input, that is, cases where there was more than one possible realization for a form or construction” (<http://childes.psy.cmu.edu/manuals/03english-uk.pdf> p.2).

4.2.2. Wells corpus

Directed by Gordon Wells, from the Ontario Institute for Studies in Education (OISE) of the University of Toronto (Canada). The participants are 32 British children, 16 boys and 16 girls, aged 1;6 to 5;0. The recording sessions took place mainly in the child's home, but eventually further recordings were made at the Research Unit of the university, in order to administer some tests to the children. The investigators placed a radio microphone under the children's top garment, linked to a recorder programmed to record 24 samples of 90 seconds each, in approximately 20-minute intervals between 9 a.m. and 6 p.m. However, this favoring of spontaneity in the children's interactions complicated the process of taking on-the-spot notes about the context. To compensate this, the investigators would play the recordings to the parents at the end of the day so they could try to remember these contextual details (location, participants, activity...) in as much detail as possible. The goal of this study was "to provide a normative survey of British children growing up in an urban environment" (<http://chilides.psy.cmu.edu/manuals/03english-uk.pdf> p. 39).

4.2.3. Gathercole corpus

Compiled by Virginia C. Gathercole, from the School of Psychology of the University of Wales. The participants were 14 children, 8 boys and 6 girls, aged 2;9 to 6;6. The recording sessions took place at school, when the children were having lunch. There was an investigator present when the recordings were being made.

4.2.4. Gleason corpus

Compiled by Jean Berko Gleason, from the Department of Psychology of the Boston University. The participants are 24 children, 12 boys and 12 girls. There were two types of recording sessions: in the laboratory with the child's mother, in the laboratory with the child's father, and at home at dinnertime, where both parents were present. They were 30-minute sessions; audio and video were recorded in the laboratory, whereas at home only audio was recorded. The goal of this project was to study the acquisition of communicative competence.

5. ANNOTATION IN OUR SELECTION OF CORPORA

Having given an account of our theoretical and methodological basis, we shall go into the analysis of the contents of our selection of corpora. From the moment this paper was conceived, the Koiné corpus occupied also a place in the spotlight for a series of reasons that appeared to us as natural and logical. Its closeness to us, geographical and academic, but also linguistic, has been fundamental in situating this corpus in a privileged place and in considering it as the reference when comparing the rest of the corpora in English.

We started our analysis by elaborating a series of tables where the data were arranged in order to proceed in a clear, orderly way (they are included in Annex 1). We made one table for each corpus, in which we present the names of the children, the number of recordings they appeared in, their age in the first and last recordings, and the dependent tiers used in the transcripts. By doing this, we can see very graphically what tiers were used in each corpus. This is relevant to our study because we can see the priorities of the researchers, in terms of which features of the language they paid attention to and to which they did not.

5.1. Differences between Spanish and English

It goes without saying that the two languages we want to compare, English and Spanish, differ in some aspects. The most obvious one may be that, even though they are not typologically different languages strict sensu, they do have features that grammatically separate the two languages – word order rigidity, or a different degree of verb morphology, for instance (see the chart below).

	SPANISH	ENGLISH
Word order	SVO, but highly flexible.	SVO, but word order distinguishes subject, object, indirect object, etc.
Gender	Yes: masculine/feminine	Few traces of gender marking.
Verb morphology	Verb affixes mark number, person, tense.	Very limited verb morphology.
Phonology	20 consonants, 5 vowels, 5 diphthongs	24 consonants, 12 vowels, 8 diphthongs.

These are just a few of the many differences between both languages; those who appear as the most prominent ones to the naked eye, so to speak, and are, we believe, the ones that will substantially affect the kind of annotations of the corpora. We have seen how the annotations for these two languages, phylogenetically very close, present certain difficulties. Should we be conducting some research with two or more languages belonging to different linguistic families, we assume that it would be an infinitely more complicated task, for we would have to develop a new set of tiers which would fit the features of the languages at issue.

Take a nominative-accusative language and an ergative-absolutive one; or an agglutinative, and an analytical language. In such cases, and despite the possibility of creating new tiers using the CHAT format, the researchers' task will multiply enormously. It is a complicated matter, and it is not our intention to deal with it in the present paper. In this line of work we follow Maya Hickmann's classification of the most relevant typological differences between languages. Hickmann suggests a ternary division, classifying languages depending on: (1) their morphological systems, (2) their general orientation towards subject versus topics, and (3) their reliance on processes of lexicalization versus grammaticalization.

- (1) Morphological systems. One major source of variation across languages stems from the extent to which they provide morphological markings and from the properties of these morphological systems. (Hickmann 2002: 50)
- (2) Orientation to subjects of topics. Languages have also been categorised as being subject-oriented vs. topic-oriented. This dimension is partly related to morphological properties, which contribute to the marking of subjecthood. Subjecthood can be described as a bundle of features, some of which are distributional and others morphological. (Hickmann 2002: 53)
- (3) Lexicalisation and grammaticalisation. A third general dimension differentiating languages concerns the ways in which systems encode information through grammatical or lexical means [...]. Some languages mark distinctions in their verbal morphology, while others express equivalent or related distinctions by other lexicalised, semi-lexicalised, or periphrastic means. (Hickmann 2002: 55)

5.2. Dependent tiers used in the English corpora

- **%act:** Actions of the speaker or the listener.
- **%add:** Addressee (who talks to whom).
- **%alt:** Alternative possible transcription.
- **%com:** General purpose comment tier.
- **%err:** Errors.
 - **\$PHO:** \$VOW, \$CON, \$CC, \$SYL...
 - **\$MOR:** \$PRE, \$SUF, \$NFX, \$NFL...
 - **\$SYN:** \$ACC, \$STR, \$NFL...
- **%exp:** Specifying the deictic identity of objects or individuals.
- **%gra:** Grammatical relations.
- **%gpx:** Gestural and proxemic material.
- **%int:** Intonations.
- **%mor:** Morphemic semantics.
- **%par:** Paralinguistic behaviors.
- **%pho:** Describing phonological phenomena.
- **%sit:** Situational information relevant only to the utterance.
- **%spa:** Speech act coding.

5.3. Dependent tiers⁶ used in the Koiné corpus

- **%par:** Paralinguistic behaviors. This tier is only used to encode peculiar uses.
 - Phonology: **\$PHO**.
 - Morphology: **\$NOU, \$VER, \$WFO**.
 - Syntax: **\$CON, \$AGR, \$PRE, \$ELL**.
- **%pho:** Used only to encode peculiar uses.
 - Phonetics and Phonology: number of syllables, stress, tonicity, syllable position, process, etc.
 - **%mor:** All uses are encoded.
 - Morphology and Syntax I (verbal constructions)
 - Morphology and Syntax II (nominal constructions)
 - **%pra:** All uses are encoded.

⁶ For a more detailed explanation of these tiers, see Fernández Pérez 2014: 28-29.

5.4. Belfast corpus

In this corpus we find eight children: Barbara, Conor, Courtney, David, John, Michelle, Rachel, and Stuart. Four of them appeared in 14 recordings, one in 11, one in 9 and two in 7 recordings.

The dependent tiers that appear the most are %gra and %mor, which are present in every utterance produced, either by children or by adults. Then %exp appears 632 times, followed by %sit (275 times) and %add (227 times). The presence of the tiers %act and %com is very limited; the former appears 23 times and the latter only once.

5.4.1. Grammatical relations (%gra) and Morphemic semantics (%mor)

The style of comments made under these lines look like this:

- David (2;1), file david03.cha:

```
*GRA: but I wear this cleaning round the house and  
      then when I come out I change .  
%mor: conj|but pro:sub|I v|wear det|this part|clean-  
      PRESP adv|round art|the n|house coord|and  
      adv:tem|then conj|when pro:sub|I v|come  
      adv|out pro:sub|I v|change .  
%gra: 1|3|LINK 2|3|SUBJ 3|0|ROOT 4|5|DET 5|3|OBJ  
      6|8|JCT 7|8|MOD 8|5|OBJ 9|13|LINK 10|13|JCT  
      11|13|LINK 12|13|SUBJ 13|8|CJCT 14|13|JCT  
      15|16|SUBJ 16|13|COMP 17|3|PUNCT
```

- Rachel (2;9), file rachel5.cha:

```
*CHI: lookin(g) in the in the wha(t) you call (th)at  
      mummie ?  
%mor: part|look-PRESP prep|in art|the adv|in art|the  
      pro:wh|what pro|you v|call det|that  
      n|mum&dadj-DIM ?  
%gra: 1|0|INCROOT 2|1|JCT 3|2|POBJ 4|5|JCT 5|3|CMOD  
      6|8|LINK 7|8|SUBJ 8|5|COMP 9|10|DET 10|8|OBJ  
      11|1|PUNCT
```

It can now be stated that these kinds of annotations were not made manually by one of the researchers, but rather using a syntactic parser and a morphological tagger. These are very useful tools used in Natural Language Processing (NLP). What a parser does is to segment and assign a grammatical structure to a given clause. A morphological tagger, as its name suggests, recognizes the words and assigns morphological labels.

These are just two of the very many tools available in the field of Computational Linguistics.

However, the abundance of these tiers is not surprising, considering that the objective of this corpus was to study structure variation. However, in our opinion, the use of the morphological annotations may be somewhat excessive, for English morphology is not a very complicated one – if we compare it, for instance, with that of the romance languages.

5.4.2. Actions of the speaker (%act) and General comments (%com)

- Barbara (2;7), file barb04.cha:

```
*MOT: will we put it over here will it do there ?  
*CHI: uhhuh .  
%act: starts cooking with her toy kitchen .  
*MOT: well here everything is falling off it !
```

- Barbara (3;2), file barb08.cha:

```
*CHI: that's a broken one .  
%com: hiccups in the middle of broken
```

- Barbara (3;2), file barb08.cha:

```
*INV: I was expecting to see you_all pale and +...  
*MOT: no I look healthy enough but +...  
*INV: and awful looking you know .  
*CHI: &ah !  
%act: panting about to sneeze .
```

- Courtney (3;7), file court03.cha:

```
*CHI: that goes (.) where's this goes ?  
%act: doing jigsaw  
*INV: that's right .
```

- John (3;6), file john02.cha:

```
*INV: mhm that looks like good fun !  
*CHI: O .  
%act: playing .  
*CHI: see all them there things !
```

5.5. Wells corpus

This is a substantially larger corpus, in which 32 children were recorded. Most of them appear in 10 or 9 recordings, although there are a few that only appear in 8.

In this corpus we find a slightly bigger variation in the tiers; even though the pragmatic component is still the most annotated (addressee, grammatical relations, morphology, and paralinguistic behaviors), intonation (%int) and phonological processes (%xpho) are also considered here.

5.5.1. Intonation (%int)

This is the only corpus where intonations are labeled; let us take a look at some examples to see how it was done:

In the case of Abigail, only two comments accompany the intonation tier: “clipped” or “sustained.”

- In the file abigai03.cha, the child is 1;8.27:

```
*CHI: wassat [: what's that] .  
%int: clipped
```

- In the file abigai04.cha, she is 2;0.01:

```
*CHI: bike Mummy (40.) .  
%int: sustained
```

It is different in the case of Benjamin because the comments made under the %int line are “lengthened final consonant” and “high pitch.”

- In the file benjam02.cha, he is 1;5.21:

```
*CHI: home .  
%int: lengthened final consonant  
*CHI: home (2.) .  
%int: lengthened final consonant
```

- In the file benjam08.cha, he is 2;11.29:

```
*CHI: xxx car .  
%int: high pitch
```


5.5.2. Phonological processes (%xpho)

A number of phonological processes appear in this corpus under the tier %xpho, although these are not the only ones; some of them are directly transcribed in the main line, instead of appearing in a dependent line. We can see an instance of this in the file benjam03.cha:

```
*CHI:      mind@c [= mine] .
```

There are numerous examples of this unorthodox way of transcribing in the whole corpus. We will not be paying attention to these, as it is much more complicated finding them than it is to find the correctly annotated cases of phonological processes.

Abigail, a 3 years and 3 months old girl produced the following utterance, transcribed in the file abigail09.cha:

```
*CHI:  yes yyy !  
%xpho: j3 j3 j3...../j3 j3 j3/
```

The structure of the syllable is affected: there is a deletion of the final consonant; the CVC syllable is then reduced to CV. The existence of this process is very common until around age 3;0, but as it is well known, the ages may vary among children. Since Abigail is 3;3, it is not strange that she has not eliminated this process yet.

Benjamin, when he was 1;5 produced the following utterances in the file benjam02.cha

```
*CHI:  hats [<] .  
%xpho: &ts...../æts/
```

```
*CHI:  her [/] her .  
%xpho: erh...../er/
```

We cannot tell if this is a phonological process or merely a pronunciation influenced by the regional accent. Benjamin is a child from Bristol, and the Bristolian accent is part of what is known as the West Country English. These linguistic varieties and accents include a certain particularities, among which we can find initial h-dropping.

In the file benjam03.cha, Benjamin is now 1;8 and retains a process which usually lasts until around 5;0 years old. It is the substitution process of a glide (in this case /j/) for the liquid /l/:

*CHI: <&l yellow> [>] .
 %xpho: lElow...../'lɛlow/

In the next file, benjam04.cha, Benjamin's sister Nicola (whose age is unknown) produces the word "peacock" in a peculiar way:

*NIC: peacock (2.) .
 %xpho: pig&wg...../'pigæwg/

The processes we can recognize here are: a voicing of the unvoiced stops (/k/ > /g/) and a fronting of the second vowel, so the open back vowel /ɒ/ is realized as a near-open front /æ/. Also, there appears a semivowel /w/, forming an unusual diphthong.

Gary, 2;0, turns the initial nasal /m/ of "monkey" into the stop /p/. The process is known as denasalization, and it often turns a nasal consonant into a non-nasal consonant with similar place of articulation.

*TRA: look at that monkey xxx .
 *TRA: xxx can't xxx him (8.) .
 *CHI: no yyy .
 %xpho: p&nki...../'pænki/

In this case both consonants are bilabials, so the data agree with the theory. Considering Gary's age, and the fact that denasalization is common until around 2;5, we must conclude that in this particular process there is no deviation from the expected development.

Geoffrey, 2;2, in the file geoff04.cha, emits the utterance "bickie" (/ˈbɪki/) to refer to some biscuits. The processes we see are two: the simplification of the consonant cluster /sk/ into /k/; and the elimination of the final consonant.

*CHI: want a bickie [= biscuit] .
 *CHI: xxx .
 *CHI: bickie (7.) .
 *MOT: come on quickly .
 *RAC: xxx .
 *CHI: want a yyy (4.) .
 %xpho: bik...../'bɪk/
 *RAC: one more (3.) .
 *RAC: oh look .
 *MOT: &mm .
 *RAC: quite good writing isn't it ?
 *CHI: I want yyy (3.) .
 %xpho: bickie...../'bɪki/

The reduction of consonant clusters is a very common process until the age of 4;0, and the elimination of the final consonants is usual until the age of 3;3. Geoffrey is still younger than those ages, so it is not strange to find the mentioned phonological processes.

5.6. Gathercole corpus

There are 14 children; ten of them appear in 4 recordings, one in 3, one in 2, and two in 2 recordings. The use of tiers in this corpus is very balanced – the annotations are rather consistent in their distribution in the different recordings. Still, the pragmatic component is very present.

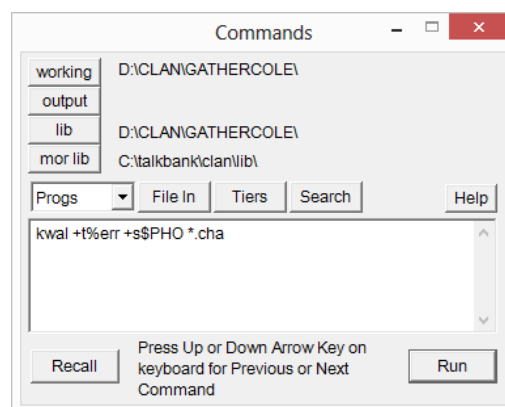
5.6.1. Errors (%err)

This is the only of our corpora which includes the dependent line %err, so we decided to search for data using CLAN. The command that we used was *kwal*, which stands for **keyword and line**. It is a rather transparent name, the use of this command allows the user to find a specified word and the line it appears in. This is how the command looks like when typed into the command box of the program:

```
kwal +t%err +s$PHO *.cha
```

and it means the following:

- **+t%err**: search only the tiers that begin with “%err”
- **+s\$PHO**: search for the keyword \$PHO (phonological errors)
- ***.cha**: the asterisk tells the program to conduct the search in all the files with the extension .cha



The results from this analysis were excessively long to be included thoroughly in this paper, so we are going to give just a brief example of what it looks like.

```
> kwal +t%err +s$PHO *.cha
```

```
-----  
*** File "01.cha": line 360. Keyword: $pho
```

```
*JEF:      we play outside to be monsters .
```

```
%err:      <6w> /manst3rz/ = /ma~t6rz/ $PHO
```

```
-----  
*** File "05.cha": line 1658. Keyword: $pho
```

```
*MIC:      macaroni [<] .
```

```
%err:      /m&k6roni/ = /m&woni/ $PHO
```

```
%err:      <3w> /snek/ = /nek/ $PHO
```

```
-----  
*** File "10.cha": line 936. Keyword: $pho
```

```
*LIL:      spaghetti .
```

```
%err:      <1w> /spAgEti/ = /pAskEti/ $PHO
```

```
-----  
*** File "13.cha": line 198. Keyword: $pho
```

```
*ERI:      it's a snake !
```

```
%err:      <3w> /snek/ = /nek/ $PHO
```

Now let us see what is happening in each individual case:

In the first example, Jeff, 2 years and 9 months old, utters the word “monsters” (adult form /'mɒnstərz/) we can see the process known as cluster reduction. The child is simplifying the consonant cluster -nst-, which results in /'mɒtərz/.

In the second example, belonging to Michael, a 3 year-old child, we can appreciate two phonological processes; the first one is a process that affects the syllable structure, a deletion of an unstressed syllable. The second one is a process of substitution, the alveolar /r/ is replaced with the semivowel /w/. The word “macaroni” (adult form /mækə'roni/), turns into /mæ'wɒni/.

In the third example, Lily (4 years and 3 months old) instead of pronouncing “spaghetti” as /spʌ'ɡeti/, she says /pʌ'sketi/; it is a case of metathesis, a sound change that alters the order of phonemes in a word.

In the last example, Eric (3 years and 3 months old) drops the first consonant of “snake” (adult form /sneik/). By this process of consonant cluster simplification we find the child form /neik/.

These examples have not been taken at random, but are representative samples of what should not be labeled as “errors” when dealing with child language. These are the normal phonological processes which children go through when they are acquiring language.

The adult generalization is, after all, the end product of protracted sifting through mountains of input data, some highly regular, some less so, some downright chaotic. Such a process of gradual refinement should NOT be expected to serve as model for the child's generalization at any specific point. But the concept of “erroneous learning” in fact leads one to make precisely this kind of dubious analogy. (Givón 1985: 122)

The lack of interest in them, or the unawareness of their existence may account for the classification of the processes as errors. They are the very essence of child language acquisition, so they ought to be recognized and studied thoroughly.

5.6.2. Phonological processes (%xpho)

However, in this corpus some phenomena appear under the annotation %xpho, which are, in fact, phonological processes. We ignore Dr Gathercole's motivations for carrying out such a division – %err and %xpho –, when the same kinds of peculiarities are annotated under different labels.

We are now going to present and analyze some of the phonological processes so we can later compare the two tiers and see if there is an actual reason for this seemingly arbitrary distinction.

Sarah (3;6) in the file 03.cha presents a reduction of a consonant cluster. Thus, /ntʃ/ becomes /un/. We can also see it as an elimination of the final consonant.

```
*SAR: I'm gonna be a yyy .  
%xpho: /griun/...../'griun/  
*SAR: I'm gonna be a Grinch .
```

There are also some transcription inconsistencies even within the same tier. In the following example, taken from file 05.cha, Michael (3;0) eliminates the unstressed syllables of the word “macaroni.” Instead of using the tiers %xpho or %err, the person who transcribed this file simply annotated it in the main line:

```
*JEF: macaroni !  
*JEF: macaroni , Bryan !  
*BRI: macaroni .  
*MIC: to many (maca)ronis .
```

And then, a few lines later, another vocal production of this child is annotated in the dependent line:

```
*VIR: it's your favorite ?  
*MIC: yyy .  
%xpho: /d&d&tS6m&woni/...../'dædætsəmæ'woni/  
*VIR: ahah .  
*MIC: macaroni ta(stes) good .
```

The lack of the adult form in the transcript makes it a bit difficult for us to know exactly what the child was saying. Our interpretation, based on the context of the conversation, is that the child said something similar to “dada’s amaroni”, whose adult equivalent would be “daddy’s macaroni.” The processes are two: gliding, which substitutes the liquid /r/ for the semivowel /w/; and metathesis.

Luke (5;2), in file 06.cha, simplifies the initial consonant cluster of “spaghetti.” There is also a voicing of the stop consonant (/p/ > /b/).

```
*RAC: for spaghetti .
*VIR: okay .
*RAC: I can hardly say it .
%par: <bef> giggles
*LUK: yyy .
%xpho: /b6gEti/...../bə'gɛti/
```

Megan (3;10) also is undergoing a process of metathesis in file 10.cha. The adult form “kitten” is realized as /'tɪkən/.

```
*VIR: kittens ?
*MEG: yeah .
*LIL: I have two of em .
*VIR: no , Megan .
*MEG: know what ?
%xpho: /tIkEn/...../'tɪkən/
```

Another instance of gliding is found in file 13.cha, when Michael (3;4) substitutes the /r/ of “Harriet” for /w/:

```
*ERI: +" Harriet (.) Harriet (.) yyy .
*MIC: xxx <my said> [//] my Daddy said yyy (.) .
%xpho: <7w> /haw6t haw6t haw6t haw6t/...../'hawət/
```

As we mentioned before, there are some cases in which we could not establish the adult equivalents of some of the child forms. Among these we find:

Saasha (5;8), in file 04.cha:

```
*SAA: are you eating your yyy already ?
%xpho: /&nd6r/...../'ændər/
```

Erin (3;4) in file 13.cha:

```
*ERN: <I want> [/] I want him to come back .
%xpho: /bajkIN/...../'bajkɪn/
*ERN: I'm afraid he's in the bath+room yyy pottie .
```

Gillian (4;9), and Luke (5;5), in file 14.cha:

```
*GIL: I know one (.) it's yyy and I don't know which
      number that is .
%xpho: /twInIs/...../'twɪnɪs/
*VIR: it's what ?
*LUK: yyy .
%xpho: /kwEFIs/...../'kwɛfɪs/
*GIL: I don't know which number yyy is , but I think
      +/.
%xpho: /twInIs/...../'twɪnɪs/
```

As we suspected, there is absolutely no reason for the distinction of %err and %xpho, all the examples we have seen of both tiers could be grouped in %xpho. We do not know the reason behind this, but it could be that the researchers took the prescriptivist approach too seriously and judged that the adult form of the grammar –grammar *as a product* – would serve as a pattern from which to evaluate children’s speech – which is a grammar still developing.

5.7. Gleason corpus

There are 24 children in the Gleason corpus. This is the corpus with the lowest number of recordings per child; twenty-two of them appear in 3 recordings, and the remaining two appear in 2 recordings.

In this corpus there is a balance between the pragmatic and the grammatical components. Morphology (%mor) and grammatical relations (%gra) are the most used tiers, closely followed by addressee (%add), paralinguistic behaviors (%par), and general purpose comment (%com).

We would like to mention that this is the only corpus that includes the audio files linked to the transcript, so that when they are played the corresponding line of the transcript is highlighted. This was achieved by using the CLAN programs, namely the Transcriber Mode.

As we have already seen the tiers %mor, %gra, and %com in previous corpora, and the variation between them is really not interesting, we will focus on some of the other dependent lines.

5.7.1. Addressee (%add)

– Father/eddie.cha:

*CHI: Daddy [<], this is way out .

*FAT: yeah .

%add: camera 1 person

*FAT: move on to the book ?

– Dinner/andy.cha:

*CHI: I'll show you which one's yours .

%add: mot

*CHI: this one's (.) yours .

*MOT: okay , you wanna sit there xxx ?
 %add: **chi**
 *CHI: no , I wanna sit in my own .
 %add: **mot**
 *CHI: zum@b , zum@b .
 %add: **self**

– Dinner/david.cha:

*CHI: I [//] you know what I ate ?
 *MOT: what did you eat ?
 %add: **chi**
 *MOT: some of my peanuts and .
 *CHI: raisins .
 %add: **mot and fat**
 *MOT: but what about the poor ducks ?
 %add: **chi**
 *MOT: didn't they get any ?
 *CHI: I [/] I [/] I gave some things to them , too .
 %add: **mot**

The addressee indication appears almost exclusively in the “Dinner” folder of this corpus, due to the presence of several speakers, and to the fact that in this case the conversations are not centered on the children; the whole family takes part in them. It was necessary to indicate who was talking to whom, otherwise the transcripts would have been too confuse to interpret correctly.

5.7.2. Paralinguistic behaviors (%par)

– Mother/ursula.cha:

*MOT: that's gonna be the tricky part , isn't it ?
 *MOT: xxx much easier +/ .
 *CHI: I don't care .
 %par: **yells**
 *MOT: 0 [=! grunts] .
 *CHI: I don't care .
 *MOT: wow !
 *MOT: I don't know , Ursula .

– Father/xavia.cha:

*FAT: that'll be forty seven dollars , please .
 *CHI: you got it !
 %par: **pretend voice**
 *FAT: &=laughs .
 *CHI: xxx here you go .

– Dinner/Isadora.cha:

*FAT: we asked her what's for supper and she said
walnuts .
*MOT: I know , I heard that .
%par: <after> laughs
*CHI: what [>] ?
*FAT: xxx [<] it's called an omelette .
*CHI: yeah , omelette .
*CHI: omelette [>] .
*MOT: <cheese omelette> [<] .
*CHI: yeah , cheese omelette .
%par: imitating

5.8. Koiné corpus

Koiné is a perfectly good example of a coherent, well-thought, and well-structured corpus. By taking a quick look at it we can see several aspects that make it stand out when compared to the English corpora. Even the description that appears in the CHILDES manual is far more extensive than its English counterparts, with several charts, diagrams, and tables listing the informants, their ages, the codes used, and so on.

This corpus has a medium degree of annotation. The ideal corpus that we want to find would not be one that has been not properly annotated, nor one that is so thoroughly annotated that it compromises too much any future research. Koiné shows a very balanced level of annotation, something which was missing in the English corpora; they annotated very thoroughly some components of the language – morphology and syntax, namely – and left other components uncovered. (cf. Fernández Pérez 2011)

The advantages of this are manifold; for instance, we can conduct some kinds of analysis that we could not in the rest of the corpora. One of them is the possibility to draw a developmental profile of the speech of a child (we already commented in §2.2 on the relevance of linguistic profiles).

5.8.1 Developmental profile of a child

The subject of this profile is Ana, a girl from the kindergarten of Vite (Santiago de Compostela). Vite is a peripheral district with a middle-class population. Ana speaks Spanish on a regular basis and Galician occasionally. In the transcripts that we will be using (which we have included in Annex 2), Ana was 2;3.21, 3;1.13, and 3;7.22. The

description of the characteristics of Ana's speech will be done according to four linguistic components: phonetic-phonological, syntactic, lexical, and pragmatic.

5.8.1.1. Phonetic-phonological component

In the first transcript Ana does not produce the rhotic [r] so she solves it by means of simplification (in syllable-initial position and in intervocalic position): [a nadá] (a nadar), [kwáto] (cuatro), [bedá] (verdad); or by substitution: [bláfo] (Breixo). There is also a process of substitution of the bilabial stop in some cases: [wénas] (buenas).

In the second transcript she is still unable to emit the rhotic, so in some cases she deletes it: [táigo] (traigo); and in others she produces a similar sound, close to an alveolar [d]. The interdental fricative [θ] is not present yet, she substitutes it by the alveolar [s]: [kalsetines] (calcetines), [sapátos] (zapatos). She is not able to emit the consonant clusters [kl], [pl], she reduces them eliminating the [l]: [páya] (playa), [tʃánkas] (chanclas).

In the third transcript she sometimes hesitates when pronouncing the rhotic [r], but it appears for the first time in some contexts. The same happens with the interdental fricative [θ]: sometimes it is present and sometimes it is not. It is relevant the appearance of a new process, a reduction of syllables in polysyllabic words: [θanórias] (zanahorias).

5.8.1.2. Syntactic component

We can see as soon as the first transcript that Ana emits utterances composed of several elements. She seems to have a certain degree of grammatical development, but these utterances are for the most part simple constructions. She does not emit complex constructions, such as coordinate or subordinate clauses. Regarding the verbal tenses, she just uses the present (simple and continuous). The use of deictics is not limited to verbal ones ("mira," "aquí," "ese"), but she also supports her utterances with body language (she shows two fingers when she says "dos").

- *ANP: yo quiero ese gato .
- *ANP: mira [?] comí las uvas .
- *ANP: <yo tengo gatos [*]> [<] .
- *ANP: <está durmiendo> [<] en la cama .

In the second and third transcripts we can appreciate an increase in the grammatical complexity of Ana's productions. The first coordinate constructions appear, sometimes introduced by the discourse marker "pero," a clear sign of her communicative disposition.

- *ANP: no porque [/] <porque (.) pongo agua> [>] en el cubo y la tiro <a la> [//] al [/] al agua !
- *ANP: pero yo tengo un [//] unas [/] unas [//] unas chancas [*] de ir <a la> [/] a la calle y de ir <a la> [/] a la piscina .
- *ANP: y <se mancó> [*] aquí na [//] na frente (.) <e se> [//] e <se mancó> [*] aquí na rodilla e a [//] e noutra rodilla e [/] e [//] <e a> [//] e na cara .

5.8.1.3. Lexical component

The lexicon that Ana uses is the expected one in her age range. It is that of the animals, family members, etc., as well as deictics such as personal pronouns: "vaca," "lata," "mama," "paiaso." Agreement is not a problem, whether it is determiner-noun (gender and number) or noun-verb (person and time): "yo quiero ese gato;" "mira un león."

In the second transcript there seems to be an increase in her lexicon and the variety of things she can name. As in the previous transcript, she names animals, family members, food, bodily parts, and so on. We can also see new vocabulary, not so related to the everyday life, for instance "playa;" "otoño;" "piedras;" "tienda." As was the case in the previous transcript, agreement is not a problem.

In the third transcript her vocabulary is more job-related, probably because the researchers are showing the children flashcards in order for them to describe: "bombero;" "medico;" "cartero;" "pescador." Adjectives, like "malo" or "mayor," and adverbs, like "más," "ahora," "después," first appear in this transcript.

5.8.1.4. Pragmatic component

Ana is a very participative girl; she is one of the most active in conversation. She very often interrupts her classmates or answers questions that are not addressed to her. She also repeats words from her classmates or the investigators.

- *PIL: cuantos tienes tú ?
 *INE: así .
 %act: enseña tres dedos
***ANP: <cuato [*]> [>] !**
 %par: cuato=cuatro \$PHO ;

- *PIL: eu creo que esto é unha porta para entrar .
 *IAN: xxx .
***ANP: porta .**
 %spa: \$IMIT

Ana participates quite frequently in the second transcript, interrupting conversations or even adding information that was not relevant to the communicative interaction. She very actively uses the discourse marker “pero.”

- *BLA: hai animais que viven debaixo dunha pedra ?
 *BRE: si .
***ANP: pero yo también (.) cogí piedras .**

- *PIL: una rana en la playa !
 *BRE: no [//] na [/] na [//] na playa non hai ranas .
 *PIL: verdad que non ?
 *BRE: non .
***ANP: pues yo sí que las ví .**

- *BRE: yo teño &do dous abuelos e [//] e dúas abuelas .
 %act: usa los dedos para enseñarme el número
 *PIL: si (.) que ben !
***ANP: yo teño cuatro abuelas .**
 %spa: \$IMIT

In the third transcript Ana still interrupts her classmates, sometimes even arguing with them. She keeps answering questions that are not directed to her. In one occasion she intervenes without being asked, contributing with irrelevant information (“Uno dos tres

cuatro”). She repeats even more what other classmates say. In one of the annotations the researchers wrote that when Ana is not able to answer the question she was asked, she describes what she sees in a picture. This shows Ana’s desire to engage in communicative exchanges, even when she has lexical limitations.

- *PIL: e como se chama ese señor que rega o fogo?
%add: BRE
***ANP: bombero.**
%com: contesta aunque la pregunta no va dirigida a ella

- *PIL: que é eso?
*BRE: unha mosca.
***ANP: unha mosca.**
%spa: \$IMIT

5.8.1.5. Evaluation

Ana is a child with a high degree of communicative value. Considering her evolution in the different linguistic components and her age (2;3-3;7), we could say that she is in the syntactic stage of language development. This stage consists of

el uso del llamado «lenguaje telegráfico», con la utilización de estructuras y combinaciones reducidas para conseguir significados complejos. Posteriormente, las construcciones se amoldan a generalizaciones y analogías diversas, con una gradación tal que desde oraciones simples [...] se pasa a oraciones incrustadas, hasta llegar a oraciones coordinadas y a relaciones proposicionales complejas. (Fernández Pérez 1999: 159)

The phonological processes are the expected ones in her age. Ana’s descriptive abilities are remarkable, as well as her referential abilities (deictic use). It is also worth mentioning her conversational abilities and her rather high participative intentions.

6. CONCLUSIONS

The realization of this paper in the context of the *Grao en Lingua e Literatura Inglesas*, and more specifically, within the minor in Linguistics, has allowed us to demonstrate our academic trajectory and the knowledge acquired in the last four years of our undergraduate studies. Moreover, not only is it a clear demonstration of our comprehensive vision of Linguistics, but also it is representative of the methodology and researching activity required to carry out a project of this characteristics. As the contents, the schedules, and the resources of a *Grao* are limited, we feel that by working on this project we have filled those areas of knowledge that we were curious about and were not covered in class, because of the aforesaid reasons.

Throughout these pages we have been giving brief but – in our opinion – precise brushstrokes of the different areas that compose Linguistics today. We have worked with CL firstly, with a focus on both a theoretical and practical view. This starting point was the root from which then we attended the other fields: Computational Linguistics, Psycholinguistics (First-language acquisition, Clinical Linguistics, etc.), Linguistic Typology, Sociolinguistics, and many others. It goes without saying that knowledge of General Linguistics constituted the very base on which the rest of the paper was built.

Within Corpus Linguistics we have been outlining the different perspectives adopted by the different authors: CL as methodology or as linguistic theory. We have subscribed to the former in order to be coherent with our project, giving sound reasons and relevant quotations. Once this matter was settled, there appeared the consideration of corpus methodology as a background from which to corroborate pre-existing theories (the corpus-based approach), or as a motivation for research (corpus-driven).

Also, the use of the CLAN programs was an interesting experience. It introduces a new way of dealing with linguistic data, which otherwise would be impossible to process. Technology is a huge part of today's lifestyle, and of course is also present in the new ways of approaching Linguistics and other (related and non-related) fields. We are aware that we have not used the maximum of electronic resources that CHILDES offer, but it has served as a first contact with it.

Delving deeper in the corpora we arrived to first-language acquisition, where the preeminence of Linguistics was not so clear, opening the door to Psychology. The

reason is simple: it was traditionally a field dealt with by psychologists – with a strong background on Linguistics. In the last decades we have been witnesses of the growing multidisciplinary of the field, where linguists, psychologists, neurologists, speech pathologists and professionals of other types have been working alongside each other, contributing with great masses of knowledge until now undiscovered.

We have also been able to deal with some sociolinguistic issues (implicitly present in the title), when we commented on the folders of the CHILDES database. We have corroborated our initial hypothesis: English has indeed an enormous influence over the rest of the languages in the different parts of MacWhinney and Snow's project; we saw it in the folders and also in the CHAT codifications, designed to fit the linguistic characteristics of English-speaking individuals – even though the possibility of adapting the codes to fit other languages.

Once we arrived to the data of the corpora we could see that CHILDES is a project designed with a user in mind that leans more toward the profile of a psychologist rather than a linguist. Having said this, we insist on the fact that the project is not lessened in prestige, merits, or validity. Precisely, and as we have mentioned before, the versatility of CHILDES and its great potential for adaptation depending on the objectives of the research that we want to conduct is what makes CHILDES a one-of-a-kind project. We would like to insist on the differences between the English corpora and Koiné. Despite the relative ease for English-speaking researchers that CHILDES offers, we found in Koiné a more carefully annotated corpus, with more attention paid to the balance of the tiers used. The English corpora annotated to a high degree the morphosyntactic component, sometimes leaving behind the rest of the components. Koiné, on the other hand, showed a consideration for the characteristics of child language, which provides a vision of the emergent grammar as peculiar and genuine.

All things considered, and answering our research question, indeed English acquisition corpora affect other languages, but it was to a lesser extent than we had initially thought. It is true that English plays a determinant role as *lingua franca* in academia, but its influence –at least in the question at hand – is not an insurmountable obstacle. The existence of corpus such as Koiné proves that it is possible to conduct scientific research without submitting to the demands of the English language. We should always bear in mind that linguistic diversity – and in fact, all kinds of diversity – is extremely

valuable and should be treated with the utmost respect. This is the case especially in the globalized world of today, where it seems that everything, from culture, customs and traditions to language, tends to converge in a single, homogeneous culture.

Finally, we would not want to conclude this paper without addressing a personal issue. The realization of this project has been, without a doubt, a personal challenge that has allowed us to show to the possible readers, as well as to ourselves, the different abilities and capacities needed to undertake it. It has also been a process of academic and personal maturation, providing us with a vast amount of new and non-trivial knowledge.

7. BIBLIOGRAPHIC REFERENCES

- BERBER SARDINHA, TONY (2004): *Lingüística de Corpus*. São Paulo: Manole.
- BERNICOT, JOSIE (2014): "Bases of Linguistic Development," in Fäcke, Christiane (ed.): *Manual of Language Acquisition*, pp. 143-161. Berlin/Boston: De Gruyter.
- BROWN, ROBERT (1973): *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- CHOMSKY, NOAM (1959): "A Review of B.F. Skinner's *Verbal Behavior*," in *Language*, 35, no. 1, pp. 26-58.
- (1965): *Syntactic Structures*. The Hague: Mouton.
- CLARK, EVE V. (2009): *First Language Acquisition*. New York: Cambridge University Press.
- CRYSTAL, DAVID & FLETCHER, PAUL (1979): "Profile Analysis of Language Disability" in Fillmore, Charles J. et alii (eds.), *Individual Differences in Language Ability and Language Behavior*, pp. 167-188. New York, San Francisco, London: Academic Press.
- CRYSTAL, DAVID (1981): *Clinical Linguistics*. London: Edward Arnold.
- DIESEL, HOLGER (2008): "Corpus Linguistics and Language Acquisition" in Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics: An International Handbook. Volume 1*. Berlin: Mouton de Gruyter.
- FERNÁNDEZ PÉREZ, MILAGROS (1999): *Introducción a la lingüística. Dimensiones del lenguaje y vías para su estudio*. Barcelona: Ariel.
- (2003): "Dinamismo construccional en el lenguaje infantil y teoría lingüística", *ELUA*, 17 (volumen especial), pp. 273-287.
- (2006): "El lenguaje infantil. Algunos lugares comunes revisitados", en *Actas del XX Congreso de la Asociación de Jóvenes Lingüistas*, pp. 1-21 (Málaga 2005), *Interlingüística*, 16.

- (2011): “El corpus *Koiné* de habla infantil. Líneas maestras,” en Fernández Pérez (coord.): *Lingüística de Corpus y Adquisición de la Lengua*, pp. 11-36. Madrid: Arco Libros, D. L.
- (2014): “Lingüística y Déficit Comunicativos: Lingüística Clínica y Logopedia,” in Fernández Pérez, M. (coord.): *Lingüística y Déficit Comunicativos*, pp. 19-44. Madrid: Síntesis.

FILLMORE, CHARLES J. (1992): “‘Corpus linguistics’ or ‘Computer-aided armchair linguistics,’” in Svartvik, Jan (ed.): *Directions in Corpus Linguistics: proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*, pp. 35-60. Berlin: Mouton de Gruyter.

GIVÓN, TALMY (1985): “Function, Structure, and Language Acquisition,” in Slobin, Dan Isaac (ed.): *The Crosslinguistic Study of Language Acquisition, Vol. 2*, pp. 118-160. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

HALLIDAY, MICHAEL A. K. & JAMES Z. (1993): "A Quantitative Study of Polarity and Primary Tense in the English Finite Clause". In Sinclair et al. (eds.), *Techniques of Description: Spoken and Written Discourse*, pp. 32-66. London: Routledge.

HERNÁNDEZ PINA, FUENSANTA (1984): *Teorías Psicosociolingüísticas y su Aplicación a la Adquisición del Español como Lengua Materna*. Madrid: Siglo XXI de España.

HICKMANN, MAYA (2003): *Children’s Discourse: Person, Space and Time Across Languages*. Cambridge; New York: Cambridge University Press.

HUNSTON, SUSAN (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

ITKONEN, ESA (2005): “Concerning the Synthesis between Intuition-based Study of Norms and Observation-based Study of Corpora” in Haddington, Pentti et alii (eds.): *SKY Journal of Linguistics 18*, pp. 357-377. Linguistic Association of Finland.

LEVINSON, STEVEN C. (2003): “Deixis” in Horn, Laurence R. & Ward, Gregory (eds.), *The Handbook of Pragmatics*, pp. 97-121. Oxford: Blackwell.

- LIEVEN, ELENA & BEHRENS, HEIKE (2012): "Dense Sampling," in Hoff, Erika (ed.): *Research Methods in Child Language. A Practical Guide*, pp. 226-239. Malden: Wiley-Blackwell.
- MACWHINNEY, BRIAN (1995): *The CHILDES Project. Tools for Analyzing Talk*. Hillsdale, NJ; Hove, UK: Lawrence Erlbaum Associates, Publishers.
- (1998): "Models of the Emergence of Language" in *Annual Review of Psychology*, Vol. 49, pp. 199-227.
- ROJO, GUILLERMO (2002): "Sobre la Lingüística basada en el análisis de corpus" in *Jornadas sobre corpus lingüísticos*. San Sebastian: UZEI.
- SKINNER, B. F. (1957): *Verbal Behavior*. New York: Appleton-Century-Crofts, Inc.
- TOGNINI-BONELLI, ELENA (ed.) (2001): *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.
- TOMASELLO, MICHAEL (2009): "The Usage-Based Theory of Language Acquisition" in Bavin, Edith L. (ed.), *The Cambridge Handbook of Child Language*, pp. 69-88. Cambridge: Cambridge University Press.
- VIHMAN, MARILYN MAY (2014): *Phonological Development: The First Two Years*. Chichester, West Sussex: Wiley-Blackwell.
- VYGOTSKY, LEV (1979): *Thought and Language*. Cambridge: The M.I.T. Press.

ANNEX 1

	CHILD	# REC	AGE	DEPENDENT TIERS														
				%act	%add	%alt	%com	%err	%exp	%gra	%gpx	%int	%mor	%par	%xpho	%sit	%spa	
BELFAST CORPUS	Barbara	14	2;4 - 4;1	X	X		X		X	X			X			X		
	Conor	14	3;8 - 4;6		X				X	X			X				X	
	Courtney	7	3;4 - 4;0	X	X				X	X			X				X	
	David	14	2;0 - 4;2	X	X				X	X			X				X	
	John	7	3;5 - 4;4	X	X				X	X			X				X	
	Michelle	14	2;4 - 4;4		X				X	X			X					X
	Rachel	9	2;5 - 3;2	X	X				X	X			X					X
	Stuart	11	3;5 - 4;5		X				X	X			X					X

	CHILD	# REC	AGE	DEPENDENT TIERS														
				%act	%add	%alt	%com	%err	%exp	%gra	%gpx	%int	%mor	%par	%xpho	%sit	%spa	&tim
GATHERCOLE CORPUS	Michael	4	2;10 - 3;4	X	X	X	X	X	X	X	X		X	X				
	Jeff	1	2;9	X	X	X	X	X	X	X	X		X	X				
	Erin	4	2;10 - 3;4	X	X	X	X	X	X	X	X		X	X				
	Gillian	4	4;3 - 4;9	X	X	X		X	X	X	X		X	X	X			
	Erik	1	4;6	X	X	X		X	X	X	X		X	X	X			
	Luke	4	4;11 - 5;5	X	X	X		X	X	X	X		X	X	X			
	Matthew	4	4;7 - 5;1	X	X	X		X	X	X	X		X	X	X			
	Megan	4	3;6 - 4;0	X	X	X		X	X	X	X		X	X	X			
	Sarah	4	3;6 - 4;0	X	X	X		X	X	X	X		X	X	X			
	Lily	4	3;10 - 4;5	X	X	X		X	X	X	X		X	X	X			
	Brian	4	5;11 - 6;6	X	X	X		X	X	X	X		X	X	X			
	Saasha	3	5;8 - 6;2	X	X	X		X	X	X	X		X	X	X			
	Nicole	4	5;4 - 5;10	X	X	X		X	X	X	X		X	X	X			
Eric	2	3;2 - 3;3	X	X	X		X	X	X	X		X	X	X				

	CHILD	# REC	AGE	DEPENDENT TIERS														
				%act	%add	%alt	%com	%err	%exp	%gra	%gpx	%int	%mor	%par	%xpho	%sit	%spa	%tim
WELLS CORPUS	Abigail	10	1;5 - 4;8	X	X		X		X	X	X	X	X	X	X	X		
	Benjamin	10	1;5 - 5;0	X	X		X		X	X	X	X	X	X	X	X	X	
	Betty	9	1;6 - 4;11		X		X			X			X	X				
	Darren	10	1;6 - 4;10		X		X			X			X	X	X			
	Debbie	10	1;6 - 3;11				X			X			X	X				
	Ellen	10	1;5 - 4;9		X		X			X			X	X				
	Elspeth	10	1;5 - 5;0		X		X			X			X	X				
	Frances	10	1;6 - 4;10		X		X			X			X	X	X			
	Gary	10	1;6 - 4;9		X	X	X						X	X	X			
	Gavin	10	1;6 - 4;9		X		X			X			X	X				
	Geoffrey	10	1;6 - 4;11		X		X			X			X	X	X			
	Gerald	9	1;6 - 4;9		X		X			X			X	X	X			
	Harriet	10	1;6 - 4;10		X		X			X			X	X				
	Iris	10	1;6 - 4;8		X		X			X			X	X				
	Jack	10	1;5 - 4;9		X		X			X			X	X	X			
	Jason	10	1;6 - 5;0		X		X			X			X	X				
	Jonathan	10	1;6 - 4;7		X		X		X	X			X	X				
	Laura	9	1;6 - 3;6		X		X		X	X			X	X				
	Lee	8	1;5 - 3;5		X		X			X			X	X				
	Martin	9	1;5 - 3;5		X		X			X			X	X				
	Nancy	8	1;6 - 3;3		X		X			X			X	X				
	Neil	8	1;6 - 3;6		X		X			X			X	X				
	Neville	9	1;5 - 3;5		X		X			X			X	X				
	Olivia	9	1;6 - 3;5		X		X			X			X	X				
	Penny	9	1;6 - 3;5		X		X			X			X	X				
	Rosie	9	1;7 - 3;6		X		X			X			X	X				
Samantha	9	1;6 - 3;6		X		X			X			X	X					
Sean	9	1;6 - 3;6		X		X			X			X	X					
Sheila	9	1;11 - 3;6		X		X		X	X			X	X					
Simon	9	1;5 - 3;5		X		X		X	X			X	X					
Stella	9	1;6 - 3;5		X		X			X			X	X					
Tony	9	1;5 - 3;6		X		X		X	X			X	X					

	CHILD	# REC	AGE	DEPENDENT TIERS													
				%act	%add	%alt	%com	%err	%exp	%gra	%gpx	%int	%mor	%par	%xpho	%sit	%spa
GLEASON CORPUS	Andy	3	4;0 - 4;2	X	X		X			X		X	X				
	Bobby	3	4;1 - 4;4		X		X		X	X		X	X				
	Charlie	3	2;11 - 3;0	X	X		X			X		X	X				
	David	3	4;1 - 4;2		X		X			X		X	X				
	Eddie	3	4;3 - 4;4		X		X			X		X	X				
	Frank	3	5;2	X	X		X			X		X					
	Guy	3	3;0 - 3;1		X		X		X	X		X	X				
	Helen	3	4;4 - 4;11		X		X		X	X		X	X				
	Isadora	3	3;6 - 3;7	X	X		X			X	X	X	X				
	John	3	4;1 - 4;2		X		X		X	X		X	X				
	Katie	3	3;2	X	X		X		X	X		X	X				
	Laurel	3	2;11 - 3;0	X	X		X			X		X	X				
	Martin	3	2;5 - 2;6	X	X		X		X	X		X	X				
	Nanette	3	2;1 - 2;2	X	X		X		X	X		X	X				
	Olivia	3	3;2 - 3;3	X	X		X			X	X	X	X				
	Patricia	3	2;5 - 2;6	X	X		X			X		X	X				
	Richard	2	2;8 - 2;9	X						X		X					
	Susan	3	3;2 - 3;8	X	X		X		X	X		X	X				
	Theresa	3	4;0 - 4;2	X	X		X			X		X	X				
	Ursula	2	3;7							X		X	X				
Victor	3	2;3 - 2;5	X	X		X			X		X	X					
Wanda	3	3;11 - 4;0		X		X			X	X	X	X					
William	3	2;2 - 2;3	X	X		X			X		X	X					
Xavia	3	4;0 - 4;5	X	X		X			X		X	X					

ANNEX 2

@Begin
@Languages: es, gl
@Participants: PIL PilarOtero Investigator, ANC AnaCodesido
Investigator,
BRE Breixo Target_Child, ANP Ana Target_Child, IAN Ian
Target_Child, INE
Irene Target_Child
@ID: gl|koiné|PIL||female|Vite||Investigator||
@ID: gl|koiné|ANC||female|Vite||Investigator||
@ID: gl, es|koiné|BRE|2;11.20|male|Vite||Target_Child||
@ID: es|koiné|ANP|2;3.21|female|Vite||Target_Child||
@ID: es|koiné|IAN|2;11.27|male|Vite||Target_Child||
@ID: gl, es|koiné|INE|3;0.05|female|Vite||Target_Child||
@Transcriber: PIL
@Date: 14-JAN-1998
@Location: Santiago de Compostela
@Situation: escuela infantil de Vite
@Activities: juego simbólico y denominación
@Tape Location: 1.4
*INE: <este (.) yo tengo uno> [>] .
%add: PIL
%act: señala y enseña su dedo
*IAN: <yo (..) uno> [<] .
%add: PIL
%act: también enseña su dedo
*PIL: a_ver (.) e como te chamas ti ?
%add: INE
*INE: Irene .
*PIL: e ti ?
%add: IAN
IAN: Ia [] .
%par: Ia=Ian \$PHO ;
*PIL: e ti ?
%add: ANP
*ANP: Ana .
*PIL: y este homiño de verde ?
*INE: Breixo !
*ANP: <Breixo> [>] !
*IAN: <yo soy verde> [<] !
*PIL: así que (.) a quen lle trouxeron os reyes un payaso (.) a
quen ?
IAN: do [] .
%par: do=yo \$PHO ; do=a mí \$CON ;
*PIL: a ti ?
%add: IAN
*IAN: sí .
%act: asiente con la cabeza
*PIL: si (.) e como é ?
*IAN: xx .
*PIL: bueno (.) entonces esta casa de quen era ?
ANP: <a Blaxo> [] [?] .
%par: a Blaxo=de Breixo \$PHO \$CON ;
*PIL: de Breixo (.) era (.) a casa ?

*ANP: sí .
 *PIL: sí (.) e quen é este señor Breixo ?
 *IAN: yo !
 *INE: é Breixo .
 *PIL: é Breixo que está (.) que está facendo Breixo ?
 *INE: ten [/] (.) ten plantas .
 *PIL: a onde ibas con plantas ?
 %add: BRE
 *PIL: non sabes a donde ibas ?
 %act: niega con la cabeza
 INE: ibra [] [//] (.) ía á casa .
 %par: ibra=iba \$PHO ;
 *PIL: iba a dar un paseo .
 *IAN: xx casa .
 *INE: non ves aquí a casa ?
 %add: IAN
 %act: señala en el libro
 *ANC: aquí .
 %act: señala en el libro
 *INE: <un conexo> [/] un conexo !
 *INE: una flor .
 %act: señala en el libro
 *BRE: flor .
 *IAN: flor .
 *ANP: xx .
 *PIL: oh !
 *BRE: un [/] <un gato> [>] .
 *IAN: <gatos> [<] .
 INE: otro jato [] .
 %par: jato=gato \$PHO ;
 INE: mira (.) un jatiño [] (.) ves ?
 %par: jatiño=gatiño \$PHO ;
 *INE: mira (.) un [/] un [//] unha <flor (.) azul> [>] .
 %add: PIL
 %com: dice azul
 IAN: <un dos tes [] xx> [<] .
 %par: tes=tres \$PHO ;
 *PIL: unha flor azul ?
 *INE: si.
 *PIL: oh (.) unha flor azul !
 *PIL: e esto que é ?
 IAN: peda [] [//] (.) <o peda> [*] .
 %par: peda=pera \$PHO ; o peda=a pera \$PHO \$AGR ;
 *PIL: e de que color é ?
 IAN: <un pelo Ana> [] .
 %par: un pelo Ana=un pelo de Ana \$CON ;
 %act: señala el pelo de Ana
 *PIL: é unha pera para Ana ?
 IAN: sí (.) Ia [] xx .
 %par: Ia=Ian \$PHO ;
 *PIL: si ?
 *INE: yo quiero este .
 %act: señala hacia la lámina
 *ANP: yo quiero ese gato .
 %act: también señala

*PIL: quieres este gato ?
 *PIL: pero este gato é desta casiña eh !
 ANP: e [] mío .
 %par: e=es \$PHO ;
 *PIL: é o gato da casa .
 *INE: eu quero ese can .
 *PIL: esto é un can ?
 *INE: <<eu quero este> [/] (.) <eu quero este> [/] (.) eu quero
 este> [>] .
 %add: PIL
 %act: señala el que quiere
 IAN: <mía (.) mía> [/] (.) mía (.) <y [] mía (.) y [*] mía>
 [/] (.) y
 [*] mía !
 %par: y=es \$PHO ;
 %add: PIL
 %act: se señala con la mano a sí mismo
 *ANP: guau@o guau@o guau@o!
 *ANP: xx pollito [//] (.) <un pollito> [/] (.) un pollito .
 %add: PIL
 *PIL: un pollito !
 %add: ANP
 *INE: <<eu quero este> [/] (.) eu quero este> [<] .
 %add: PIL
 %act: señala el que quiere
 *PIL: pero que é esto ?
 *IAN: <xx> [>] .
 *INE: <un can> [<] .
 *PIL: seguro ?
 *IAN: no (.) mi xx +/ .
 *INE: yo quiero esto .
 IAN: no (.) <mi [] tere [*] ezo [*]> [//] (.) mi [*] tere [*]
 etos [*]
 !
 %par: mi=yo \$CON ; tere=quiero \$PHO ; ezo=eso \$PHO ; etos=estos
 \$PHO ;
 *INE: yo quiero esto !
 *IAN: no .
 *PIL: cómo dices (.) Ian ?
 *IAN: xx .
 *INE: este es mí(o) (.) este é pa(ra) min !
 %act: señala la lámina
 IAN: ese <é ta [] min> [>] .
 %par: ta=para \$PHO ;
 *PIL: <pero que me> [<] dixeché (.) mi quere esto ?
 IAN: sí (..) ete [] e [*] mía sí (.) todo [*] mías .
 %par: ete=esta \$PHO \$AGR ; e=es \$PHO ; todo=todas \$AGR ;
 %act: señala en el libro
 %add: PIL
 *INE: <no ese es mío> [>] .
 %act: señala el que quiere ella
 *PIL: <todas tuyas> [<] ?
 *PIL: todas no (..) a_ver (.) uno pa(ra) ti y otro pa(ra) ella .
 %add: IAN
 *IAN: este +/ .

*IAN: estas .
*INE: yo quie +//.
*INE: yo quiero este .
%act: señala
*ANP: yo quiero o pollito !
%act: señala
%add: PIL
*PIL: yo quiero este !
*INE: yo quiero este .
%act: señala
%add: PIL
*PIL: ah (.) bueno (.) vale !
IAN: tero [] este xx .
%par: tero=quiero \$PHO ;
*PIL: pois nada (.) son os dous pa(ra) min .
*ANP: yo o pollito .
*PIL: a ver !
*ANP: <yo la vaca> [>] !
*BRE: <xx> [<] !
*PIL: tú quieres el pollito ?
%add: ANP
*ANP: sí (.) xx .
IAN: <un pedito [] (.) gua@o gua@o gua@o> [>] !
%par: pedito=perrito \$PHO ;
*PIL: y tú (.) que quieres tú ?
%add: BRE
*BRE: esto (.) <un pez> [>] +//.
*ANP: <guau@o guau@o guau@o guau@o guau@o> [<] !
*PIL: y esto qué es ?
*BRE: un pez .
*IAN: xxx .
*PIL: <un pez> [<] (..) quién tiene un pez en casa ?
*BRE: eu .
*PIL: tes un pez (.) na casa ?
%add: BRE
%act: asiente
*PIL: e donde está ?
*IAN: &mi +//.
*INE: mira (.) <un &ca> [//] un can !
*INE: <este é pequeniño (.) e este é a mamá> [>] .
%act: señala en la lámina
IAN: <xx (.) este e [] mía [*]> [<] .
%act: señala
%par: e=es \$PHO ; mía=mío \$AGR ;
*INE: este é pequeniño e este é a mamá .
*PIL: <o pequeniño e a mamá> [>] .
IAN: <e [] mía [*] (.) mamá [*]> [<] .
%add: INE
%par: e=es \$PHO ; mía=mío \$AGR ; mamá=la mamá \$PRE ;
*PIL: <senta> [>] .
%add: INE
*INE: <y este é pequeniño> [<] (.) e este é (.) bueno xx .
%act: señala
*PIL: ese é pequeniño e este é a mamá (.) si .
%add: INE

IAN: <este e [] mía [*]> [//] (.) <este e mía (.) mamá> [*] !
 %par: e=es \$PHO ; mía=mío \$AGR ; mamá=la mamá \$PRE ;
 %act: señala en la lámina
 *PIL: este é de Ian e este é de Irene .
 *PIL: donde tes o pez ?
 %add: BRE
 *IAN: yo !
 %act: se señala a sí mismo
 *INE: &te teño aquí .
 %act: señala en la lámina
 *PIL: non te acordas ?
 %add: BRE
 %act: Breixo asiente
 IAN: <el [] vaca seseta [*]> [>] .
 %par: el=la \$AGR ; seseta=es ésta \$PHO \$CON ;
 %act: señala la vaca
 *PIL: <e donde está> [<] ?
 %add: BRE
 *BRE: na casa (.) xx .
 %add: PIL
 *PIL: si (.) pero dónde está ?
 *PIL: está na habitación (.) na cociña ?
 %add: BRE
 *INE: que te [?] (.) na cociña ?
 %com: no sé si dice te o ten
 %exp: dice cosiña
 *PIL: a ver onde ten o pez Breixo (.) donde terá o peixe ?
 *ANP: aquí !
 *ANP: una vaca !
 *PIL: non o sabemos (.) uy e esto que é ?
 %act: paso página
 *INE: xx una vaca !
 *BRE: <a vaca> [>] .
 *IAN: <esto é a vaca> [<] !
 *PIL: si non te sentas ben non seguimos lendo o libro .
 %act: cierro el libro
 *INE: mira (.) &a (.) unha pequeniña e a mamá !
 %act: señala la lámina
 *PIL: outra vez (.) unha pequeniña e a mamá !
 *IAN: xx mamá .
 %act: señala la lámina
 *INE: &un +/.
 *PIL: <e como fan as vacas> [>] ?
 *BRE: u@o !
 *INE: <<no (.) no> [/] (.) no (.) esa é a pequeniña mamá> [<] !
 *ANP: u@o u@o u@o !
 *IAN: <u@o u@o u@o u@o> [>] !
 *INE: <esa é a mamá> [<] !
 *PIL: uuuuu !
 %com: todos hacen u@o u@o u@o !
 *IAN: <é malo> [/] é malo .
 *INE: eso que é ?
 %act: señala en la lámina
 *PIL: son malas as vacas ?
 *PIL: <no> [>] !

IAN: <e [] nalas [*]> [?] [<] .
 %par: e=son \$AGR ; nalas=malas \$PHO ;
 *INE: eh son bue(nas) .
 %com: ese eh no es una interjección sino que está usado como
 apoyatura
 *INE: e esto que é ?
 ANP: <on güenas> [] .
 %par: on güenas=son buenas \$PHO ;
 *PIL: a_ver (.) que será esto ?
 IAN: <se güenas> [] [>] .
 %par: se güenas=son buenas \$PHO ;
 INE: <un carabajo []> [<] .
 %par: carabajo=escarabajo \$PHO ;
 %act: PIL
 ANP: <o güenas> [] [>] .
 %par: o güenas=son buenas \$PHO ;
 *PIL: <un que> [<] ?
 %act: INE
 INE: un carabajo [] (..) verdad ?
 %par: carabajo=escarabajo \$PHO ;
 *PIL: un escarabajo .
 *IAN: <e ti o xxx> [>] .
 %act: señala en el libro la regadera
 *PIL: <mira como o sabe (.) ves> [<] ?
 %add: ANC
 *PIL: e entón pa(ra) que me preguntas (.) si xa o sabes ?
 *INE: si que é .
 *PIL: e esto que é ?
 *INE: un +/.
 IAN: <a pantas> [] .
 %par: a pantas=para las plantas \$PHO \$PRE \$CON ;
 *INE: <eso é un cubo (..) y un paxaro (.) y una mazán (.) mira
 [/] (.)
 mira (.) <un perrito> [//] (.) un perrito de ahí> [>] !
 %act: señala
 %com: dice masán
 *ANP: una [//] (.) una lata !
 *PIL: que é ?
 %add: ANP
 *ANP: es una lata .
 *PIL: <que será (.) Breixo (.) tu sabes o que é> [<] ?
 %add: BRE
 %act: Breixo niega con la cabeza
 IAN: <e sa (.) e sal> [] !
 %par: e sa e sal=el agua sale por ahí \$PHO \$ELL ;
 *ANP: es una lata !
 IAN: <esto se sal> [] .
 %par: esto se sal=esto sale \$PHO \$CON ;
 %act: señala la regadera
 *PIL: pois mira (.) unha lata pode ser (.) un caldeiro tamén
 (..) pero
 (.) e para as plantas tamén (.) eu creo que é unha regadeira eh
 !
 *PIL: e esto ?
 *INE: un paxaro .

*PIL: <un paxaro> [>] !
 IAN: <un paiaso [] pío@o pío@o pío@o> [<] .
 %act: con la mano imita el pico de los pájaros
 %par: paiaso=paxaro \$LEX ;
 *INE: pío@o pío@o pío@o !
 *PIL: quen ten un paxaro na casa ?
 *IAN: allí !
 %act: señala al cielo
 *PIL: naide (.) ten un paxaro na casa ?
 IAN: <tá [] allí> [/] (.) <tá [*] allí> [//] (.) tá [*] alí !
 %par: tá=está \$PHO ;
 *INE: yo tengo un paxa(ro) (.) eu (.) o paxaro ten Ana (..) <na casa> [>]
 .
 *PIL: <que Ana> [<] ?
 *PIL: quen é Ana ?
 INE: aesa [] !
 %par: aesa=esa \$PHO ;
 %act: señala a Ana
 *PIL: tes un paxariño na casa ?
 %act: ANP
 ANP: no xx a [] casa .
 %par: a=en \$CON ;
 *INE: ti tes esta vaca ?
 %add: ANP
 %act: señala en la lámina
 *IAN: <no (.) yo (..) este yo> [>] !
 %add: INE
 %act: se señala a sí mismo
 *ANP: <xx tiene xx mamá> [<] !
 *INE: xx .
 IAN: <este yo mamá> [] [>] !
 %par: este yo mamá=este es de mi mamá \$CON ;
 %act: señala en la lámina
 *PIL: <qué tiene mamá> [<] [/] (.) qué tiene mamá ?
 *ANP: <lalalá la> [//] (.) <lalalá la mamá (..) lalalá la mamá> [/] (.) lalalá la mamá .
 *INE: este co +/-.
 IAN: <este é mía> [] xx a mi xx +/-.
 %par: este e mía=esta es mía \$PHO \$AGR ;
 %act: señala en la lámina
 *PIL: oh (.) y esto qué es ?
 ANP: una galina [] !
 %par: galina=gallina \$PHO ;
 IAN: co@o co@o (.) lila [] (.) me@o .
 %par: lila=gallina \$PHO ;
 %com: en una lámina aparece una gallina y en la de al lado una oveja
 *ANP: xx .
 INE: unha galina [] .
 %par: galina=galiña \$PHO ;
 %com: dice jaliña
 *ANP: un pollito !
 INE: unha galina [] .

%par: galina=galiña ;
%com: dice jalina
*PIL: e que (.) que fan as galiñas (.) que fan ?
%com: digo jaliñas
***ANP: pío@o pío@o pío@o pío@o !**
*BRE: co@o co@o co@o co@o !
*INE: co@o co@o có@o !
*IAN: xx me@o !
%act: señala la oveja
*INE: a ovella fai +/.
*PIL: y qué podemos comer (.) que ponen las gallinas (..) qué podemos comer ?
*BRE: huevo !
*PIL: un huevo (.) qué rico (.) um !
*PIL: a quién le gustan los huevos con patatas fritas ?
*IAN: <yo> [>] !
*BRE: <yo> [<] !
*PIL: qué rico (.) qué hambre tengo !
*PIL: y esto ?
*IAN: un mé@f !
*BRE: unha ovella !
*INE: mira que hay ahí !
%act: señala la portada del libro
*PIL: a ver (.) qué hay aquí ?
***ANP: yaya [?] !**
*INE: patitos (.) alí .
*PIL: patitos ?
*PIL: ay sí (.) patitos !
*PIL: que están haciendo ?
*INE: están nadando .
*IAN: ir nadando .
%com: tanto Irene como Ian pronuncian casi una r en nadando es una d muy relajada
*PIL: y quién sabe xx ?
*IAN: nadar y nadar .
*PIL: quién sabe nadar ?
%com: todos se ponen a gritar y no puedo transcribir nada
INE: eu sabo [] nadar (.) vale ?
%par: sabo=sei \$VER ;
%com: sabo es una sobrerregularización y en nadar vuelve a usar una d muy relajada
*PIL: tu sabes nadar (..) e quen +/.
*BRE: <eu tam(én)> [//] yo también !
%com: cambio de código en el mismo turno
*IAN: yo xx a palas !
*BRE: eu tamén !
*INE: <un perrito (.) mira un perrito> [>] !
%act: señala la lámina
*PIL: <eu tamén> [<] !
*PIL: quién máis sabe nadar ?
*PIL: <todos sabemos nadar> [>] ?
*INE: <mira un perrito (.) mira [/] mira> [<] !

%act: señala el libro
*PIL: a_ver (.) e a que estilo sabes nadar ti ?
%add: INE
INE: así xxx sabo [] nadar .
%par: sabo=sei \$VER ;
%add: PIL
%act: mueve los brazos como si nadase a braza
*PIL: ah (.) bueno !
*ANC: xx Pili (.) sabe nadar así .
%com: no transcribo los comentarios entre ANC y PIL
*PIL: qué era esto (..) Ana ?
%add: ANP
*IAN: <<un cocós@f> [/] (..) un cocós@f xx pa(ra) mi papá (.)
un> [>] +/.
%act: señala la lámina
*PIL: <bueno (.) ahora xa é un cocós> [<] .
ANP: <<a nadá []> [/?] (.) a nadá [*]> [>] !
%par: nadá=nadar \$PHO ;
%add: IAN y BRE
*INE: <<mira ovos> [/] (.) <mira ovos> [//] (..) ovos> [<] !
%act: señala en la lámina
*PIL: home (.) estos son ovos (.) mira (.) os ovos (.) que non
nos
dábamos conta que estaban aquí !
%add: IAN
IAN: <ma cumelos> [] !
%par: ma cumelos=para comelos \$PHO ;
*PIL: e esto que é ?
*PIL: esto si que non sei o que é .
*ANP: &s uvas .
*BRE: uvas .
*PIL: ay (.) son uvas !
IAN: <a mi> [] !
%par: a mi=mías \$CON ;
%act: se señala a sí mismo
INE: uvas asules [] .
%par: asules=azules \$PHO ;
*PIL: e cando se (.) si (.) a verdad é que uvas azules son un
pouco raras
(.) non é ?
*INE: si .
*INE: é meu eh@i !
%com: deixo de transcribir un fragmento porque no se entiendo
nada
*PIL: quen comeu as uvas ?
*IAN: no !
*PIL: non se comen as uvas ?
%add: IAN
*IAN: no .
%add: PIL
*PIL: como que no ?
%add: IAN
*IAN: <e si picas> [?] .
*ANP: mira [?] comí las uvas .
%add: PIL

*PIL: tu comiche as uvas ?
 %add: ANP
 *ANP: xx .
 IAN: xx vuvas [?] [] (..) <a min tamé(n)> [//] (..) yo [//]
 (.) <yo
 tamén ua [*] (.) Ana> [*] .
 %par: vuvas=uvas \$PHO ; ua=uvas SPHO ; yo tamén ua Ana=yo
 también comí
 uvas como Ana \$ELL ;
 %add: PIL
 ANP: <a pa la comía mía> [] .
 %par: a pa la comía mía=son para mi comida \$PHO \$PRE \$CON \$ELL;
 IAN: <a mía [/] mía> [] !
 %par: a mía=son mías \$PHO \$CON ;
 *ANP: xxx .
 *INE: que hai ahí?
 *PIL: vamos a_ver outra cousa (.) a_ver que aparece por aquí
 a_ver .
 *ANC: Ian cantos anos tes (.) Ian (.) cantos anos tes ?
 *IAN: dos .
 %act: enseña dos dedos
 *ANC: así dos (.) e cantos anos ten Ana ?
 *IAN: tes .
 %act: vuelve a poner dos dedos
 *INE: yo tengo dos .
 *ANC: dos .
 ANP: gato [] .
 %par: gato=cuatro \$PHO ;
 %act: enseña las dos manos
 *INE: <no (.) yo tengo> [>] &a (.) mira [//] mira que tengo .
 ANP: <yo tengo gatos []> [<] .
 *PIL: cuantos tienes tú ?
 *INE: así .
 %act: enseña tres dedos
 ANP: <cuato []> [>] !
 %par: cuato=cuatro \$PHO ;
 *ANC: <así tres> [<] (.) tres ?
 *ANP: muchos !
 *ANC: tres (.) bien !
 *PIL: si .
 *ANC: y Breixo cantos anos ten ?
 *BRE: así .
 %act: enseña dos dedos
 *ANC: dos !
 *PIL: non se pode pelear un (.) eh !
 *INE: unha casa (.) xxx hai ahi unha casa !
 %act: señala el dibujo
 *PIL: a_ver a_ver (.) uy !
 *ANP: <sí xxx una casa> [>] .
 *PIL: <mira (.) ahora unha casa> [<] !
 *IAN: yo casa !
 *INE: <que hai aqui> [/] que hai aqui ?
 *PIL: unha ventana .
 *INE: e que hai ahi adentro?
 *PIL: a_ver contamo ti (.) que hai ahi dentro .

*IAN: ventana !
 *INE: que hai ahí ?
 *PIL: a_ver (.) vamos a imaxinar (.) que hai ahí dentro ?
 BRE: un tabús [] .
 *PIL: que hai ?
 BRE: un tobús [] .
 *PIL: ah pero eso é despóis (.) pero que hai aquí dentro da casa ?
 *ANP: vacas !
 *PIL: vacas ?
 *IAN: xxx .
 *INE: que hai ahí ?
 *PIL: hai unha familia vivindo ahí .
 *INE: hai unha familia .
 *BRE: vou nese autobús .
 %act: señala el dibujo
 IAN: un tobús [] !
 %par: tobús=autobús \$PHO ;
 *PIL: quen ven en autobús ó cole ?
 IAN: do [] !
 %par: do=yo \$PHO ;
 *BRE: eu !
 *ANP: <xxx> [>] .
 IAN: <yo cole> [] [<] .
 %par: yo cole=yo voy al cole \$ELL ;
 *BRE: i eu .
 *BRE: <i eu> [//] i eu tamén .
 *PIL: gústache ir en autobús (.) si ?
 *INE: a mí me gusta ir en el autobús .
 IAN: ei@i mía [] <e poche mía> [*] [//] <e poche mía> [*] xxx !
 %par: mía=mira \$PHO ; e poche mía=este coche es mío \$PHO \$AGR \$PRE \$CON ;
 %act: sinala a lámina dos coches
 *INE: mira (.) coches !
 *PIL: que son ?
 *INE: <coches> [>] !
 *BRE: <coches> [<] !
 BRE: xxx cho(cou) que eses chocou [] !
 %par: chocou=chocaron \$VER ;
 *PIL: de que color son (.) a_ver ?
 IAN: ete [] e [*] mía [*] .
 %par: ete=este \$PHO ; e=es \$PHO ; mía=mío \$AGR ;
 BRE: é mía [] .
 %par: mía=mío \$AGR ;
 %spa: \$IMIT
 *PIL: de que color son (.) a_ver quen o sabe ?
 ANP: a [] mío .
 %par: a=es \$PHO ;
 IAN: <no (.) mía []> [>] !
 %par: mía=mío \$AGR ;
 *INE: <ese é verde (.) ese é azul e ese> [<] +//.
 *PIL: y este ?
 *BRE: eh (.) rojo .
 *INE: rojo .
 *INE: rojo .

*BRE: rojo .
*BRE: esto é un oso !
*PIL: que está facendo ?
BRE: dumindo [] .
%par: dumindo=durmindo \$PHO ;
INE: está dumindo [] .
%par: dumindo=durmindo \$PHO ;
*PIL: a_ver este osiño (.) como se chama ?
*ANP: xxx .
*PIL: como se chama ?
*ANP: lo (.) osito .
*PIL: osito (.) a_ver (.) e o osito que está facendo ?
IAN: <mío Pe(pe)> [] [//] Pepe .
%par: mío Pepe=el mío se llama Pepe \$ELL ;
*PIL: <chamámoslle Pepe> [>] ?
*ANP: <está durmiendo> [<] en la cama .
*PIL: está durmindo na cama (.) está soñando .
*IAN: sí (.) Pepe estás !
*PIL: e esta quen é ?
IAN: un [] nena .
%par: un=unha \$AGR ;
*PIL: que ten a nena ?
*IAN: esa .
*PIL: que ten ?
*IAN: esa .
*ANP: unha xxx .
*INE: é [//] é a pequeniña .
*PIL: pero como se chama ?
*INE: ela .
*ANP: <está xxx> [>] .
*PIL: <unha munequiña ten> [<] .
*ANP: mira (.) <un paiaso> [>] !
*BRE: <un paiaso> [<] !
*IAN: un payaso !
*PIL: quen tiña un payaso ?
%com: hay un fragmento que no transcribo por su poca relevancia comunicativa
*PIL: non me enterei (.) que é ?
*INE: <hay dibuj(os)> [//] (.) ahí hay dibujos .
*PIL: aquí hay dibujos ?
*ANP: <aquí hay dibujos> [>] .
%spa: \$IMIT
*PIL: <pero como se chaman> [<] ?
*INE: ella [?] .
*IAN: no !
*ANP: aquí hay vacas .
*INE: se llama ella [?] .
*ANP: también +//.
*IAN: vacas !
*PIL: que dibuxos hai aquí ?
IAN: do [] !
%par: do=yo \$PHO ;
*ANP: dos .
%act: saca dos dedos
*PIL: a_ver (.) que dibuxos vedes na tele ?

IAN: do [] !
 %par: do=yo \$PHO ;
 *INE: yo veo (.) vacas !
 *PIL: tu ves vacas (.) pero como se chaman <os dibuxos> [>] ?
 IAN: <<do> [<] (.) cedos> [] !
 %par: do cedos=yo cerdos \$PHO ;
 *INE: se llaman (.) xxx .
 IAN: Ana (.) bubu@b (.) Brexo [] (.) me@o .
 %par: Brexo=Breixo \$PHO ;
 *PIL: pois eu veía Heidi .
 *ANP: amimi@b (.) <papa@b> [>] .
 %com: no sé lo que quiere decir
 IAN: <a [] más> [<] [//] ahora más .
 %par: a=ahora \$PHO ;
 ANP: <eso [] e [*] de vedá [*]> [/?] .
 %par: eso=esto \$PHO \$PRE ; e=es \$PHO ; vedá=verdad \$PHO ;
 %act: enseña su reloj
 *PIL: quen ve o xabarán club ?
 *INE: yo veo el xabarán club !
 *PIL: si?
 *BRE: <i eu &ta> [//] i eu tamén !
 *PIL: si (.) mentras merendades (.) si ?
 *INE: eso de ahí é un boli .
 %act: señala la lámina
 *PIL: que estades comendo o bocadillo e vendo a tele á vez ?
 *INE: si .
 *PIL: si ?
 *INE: si .
 *PIL: si .
 *PIL: quen fai o bocadillo ?
 *INE: no [/] no (.) agora eu quero boli .
 %act: toca la lámina
 *PIL: tu queres ver o chisme enteiro (.) a_ver vemos a seguinte logo .
 *BRE: un <avión> [>] !
 *IAN: <xxx> [<] !
 *BRE: avión !
 *IAN: sí (.) pa [/] pa xxx !
 %exp: no se entiendo lo que dice
 *PIL: a_ver e que terá por aquí ?
 *BRE: <xxx> [>] .
 *IAN: <www> [<] !
 %exp: no se entiendo lo que dice
 *PIL: eu creo que esto é unha porta para entrar .
 *IAN: xxx .
 *ANP: porta .
 %spa: \$IMIT
 *PIL: e quen entrará por alí dentro ?
 *ANP: yo .
 *IAN: xxx .
 *INE: no (.) <esta non é a porta> [>] .
 *ANP: <un [/] un [//] un pujama@c (.) un &pu &jama> [<] .
 *PIL: e cal é logo ?
 *ANP: <é un p> [//] (.) <un pujama@c> [>] .

*INE: <é [//] (.) era> [<] [//] (.) este (.) era un (.) &pa [//]
&pa [//]
patito .
%act: señala la lámina
*INE: si (.) así .
%act: hace un dibujito en la lámina
ANP: no [//] no (.) e [] mío !
%par: e=es \$PHO ;
%act: Ian y Ana se pelean
*PIL: entonces onde está a porta ?
IAN: <aquí (.) porta> [] .
%par: aquí porta=aquí está a porta \$ELL ;
*PIL: esa (.) é a porta (.) é a porta (.) esa ?
*IAN: www .
%exp: no se entiende
*ANP: mira (.) un león !
%act: señala la nueva lámina
*PIL: un león (.) <pero qué es> [>] (.) esto ?
*IAN: <sí (.) agrrrrr@o> [<] !
*INE: é un león !
*PIL: pero todo esto qué es ?
*INE: esto .
*PIL: qué son ?
*BRE: pesetas !
*PIL: pesetas (.) e que se pode comprar coas pesetas ?
*INE: se xxx (.) las pesetas se pode comprar (.) cabalos !
*PIL: cabalos e que mais se pode comprar ?
IAN: dos talalos [] .
%par: talalos=cabalos \$PHO ;
*BRE: e comi [//] e comida !
*IAN: www .
%exp: no se entiende
INE: mira que tá [] fasendo [*] .
%par: tá=está \$PHO ; fasendo=facendo \$PHO ;
%add: PIL
%act: señala a Ian
*PIL: a_ver que é esto ?
BRE: camareros [] !
%par: camareros=caramelos \$PHO ;
@End

@Begin
@Languages: es, gl
@Participants: PIL Pili Investigator, BLA Blanca Investigator,
ANP Ana
Target_Child, BRE Breixo Target_Child
@ID: gl, es|koiné|PIL||female|Vite||Investigator||
@ID: gl, es|koiné|BLA||female|Vite||Investigator||
@ID: es|koiné|ANP|3;1.13|female|Vite||Target_Child||
@ID: gl, es|koiné|BRE|3;9.12|male|Vite||Target_Child||
@Transcriber: PIL
@Date: 06-NOV-1998
@Location: Santiago de Compostela
@Situation: escuela infantil
@Tape Location: 2.3
ANP: mira que taigo [] yo .
%par: taigo=traigo \$PHO;
%add: PIL
%act: levanta el pie y enseña los calcetines
*PIL: que es eso ?
ANP: calsetines [] .
%par: calsetines=calcetines \$PHO;
*BLA: calcetines !
*PIL: e quen che comprou esos calcetines ?
*ANP: mamá .
*PIL: sí (.) que chulis !
*PIL: a_ver os de Breixo .
*BRE: mira (.) aquí tengo estos calcetines .
%add: PIL
%act: también levanta los pantalones y el pie y muestra sus
calcetines
*PIL: e de que color son ?
%com: BRE se pone a mirar sus calcetines y tarda bastante en
contestar
*PIL: eu creo que son (.) azules non ?
*BRE: non .
*PIL: non ?
*BRE: é azul [/] azul (.) e escuro .
*BLA: azul escuro (.) claro (..) si (.) en cambio o pantalón de
que color
é ?
BRE: rojo [] .
%par: rojo=rojo \$PHO;
%com: parece que el sonido se acerca mas a la r francesa
*BLA: rojo (.) si .
*PIL: bueno (.) pero si Breixo e Ana xa saben todos os colores
(.) eh !
*BLA: sabedes todos os colores (.) si ?
*BRE: esto é blanco .
*BLA: si (.) bueno (.) un pouco gris (.) eh .
*BRE: este é azul .
*BLA: e este é azul (.) si .
*BRE: esto tamén blanco .
*BLA: ese tamén blanco .
*PIL: e donde estuvo Breixo este vrao (.) a_ver (.) cóntame (.)
que

fixeches ?

BRE: e estaba (.) na paya [] .

%par: paya=playa \$PHO;

*PIL: fuche á playa (.) e sabes nadar ?

%act: BRE asiente

BRE: si (.) <con fotador []> [>] [//] con flotador .

%par: fotador=flotador \$PHO;

%act: imita la forma del flotador con las manos

*ANP: <y yo también> [<] !

*PIL: ah (.) con flotador !

*ANP: y yo también !

*PIL: tú también con flotador ?

ANP: sí (.) sin gaja [] !

%par: gaja=gafas \$PHO ;

*BLA: sin gafas (.) ah !

*PIL: e sin flotador ?

%add: ANP

BRE: <yo [/] yo nado con gajas [] de nadar e e con flotador de nadar> [>].

%par: gajas=gafas \$PHO;

%add: PIL

ANP: <yo fui sin [/] sin (.) fui [//] <fui sin> [/] (.) fui sin tenis (.) fui sólo con [/] con (.) con zapatos []> [<] .

%par: zapatos=zapatos \$PHO;

%add: BLA

*BLA: con zapatos (.) si ?

ANP: <con cal> [//] con sin calsetines [] .

%par: calsetines=calcetines \$PHO;

*BLA: claro (.) te bañas con zapatos pero sin calcetines .

*BLA: e pos aletas ou non ?

*ANP: no !

*ANP: no porque [/] <porque (.) pongo agua> [>] en el cubo y la tiro <a la> [//] al [/] al agua !

*BRE: <pero yo sí> [<] !

%add: BLA

*BLA: ah sí (.) las tiras al agua ?

*BLA: moi ben !

*ANP: y salpico .

*BLA: oh (.) a quen salpicas ?

*ANP: al agua !

*BLA: a quen ?

*ANP: al [//] al agua .

*BLA: al agua la salpicas ?

*BLA: mui ben !

*PIL: e algun colleu un peixe ?

*BRE: si .

*PIL: pescache un peixe ?

%add: BRE

*BRE: <si> [<] .

%act: BRE asiente de nuevo

*BLA: <grande ou pequeno> [>] ?

*PIL: como era ?

*BRE: pequeno .
 *PIL: así (.) ou así ?
 %act: amosolle diferentes tamaños
 *BRE: no mira (.) así .
 %act: amosame o tamaño
 *PIL: así .
 *BRE: así .
 *PIL: e que fixo o peixe (.) onde o metiches (..) onde o gardache ?
 *BRE: eh@i (.) no (.) no@i [/] no@i .
 *PIL: non colliches ningún peixe (.) eh ?
 *BRE: si@i !
 *PIL: collichelo ?
 *BLA: si (.) pero despóis +/.
 *PIL: pero collichelo na praia ?
 *BRE: si .
 *PIL: entonces collíchelo co caldeiro ?
 *BRE: claro .
 *BLA: e logo botáchelo no mar para que seguise crecendo (.) verdade ?
 *PIL: ou leváchelo pa casa ?
 BRE: <lo levei> [] pa [*] casa .
 %par: lo levei=leveino \$CON ; pa=para \$PHO ;
 *BLA: si ?
 *PIL: e onde o metiches ?
 BRE: na naverera [] .
 %par: naverera=nevera \$PHO;
 *PIL: na nevera !
 *BLA: na nevera !
 *PIL: e despóis comíchelo ?
 *BRE: si .
 ANP: pero yo (.) yo salté en la paya [] !
 %par: paya=playa \$PHO;
 *BLA: saltaches na praia (.) que ben (.) moi alto (.) como saltabas (.) así (.) a ver como fas ?
 *ANP: así (.) y [/] y tiraba más [/] más agua .
 *BLA: tirabas más agua (.) bueno qué gamberra !
 *ANP: al [/] al a los papás .
 *BLA: a los papás (.) si (.) que sorte (.) e os papás se enfadaban ou non ?
 BRE: tiene unhas cousiñas negra [] ahí .
 %par: negra=negras \$AGR;
 %com: PIL le suena los mocos a ANP
 *PIL: quien ?
 *BRE: tiene unhas cousiñas (..) ela .
 *PIL: si ?
 *BLA: tes unhas cousiñas +//.
 *BRE: ++ ela .
 %add: señala a ANP
 *BLA: ahí ten pelotas (.) verdade (.) na piscina (.) para xogar os nenos (.) pero vos xa non estades nesta clase (.) a que no ?

ANP: pero yo tengo un [//] unas [/] unas [//] unas chancas []
de ir <a
la> [/] a la calle y de ir <a la> [/] a la piscina .
%par: chancas=chanclas \$PHO ;
*BLA: tienes ganas (.) de ir á praia e á piscina (.) pero agora
é inverno
(.) agora xa non se vai á praia .
*ANP: no .
*BLA: non (.) <só se vai no verán> [>] .
*BRE: <porque fai frío> [<] .
*BLA: fai frío .
*BRE: e sol .
*ANP: <y caen las hojas del otoño> [>] .
*BRE: <ahí hai sol> [<] .
*BLA: hoy hay sol ?
*BRE: mira .
%act: asiente y señala la ventana
*ANP: y caen las hojas del otoño .
*PIL: e hoxe fai frío ou non fai frío ?
*BLA: caen as follas no outono (.) claro que sí (.) e logo na
primavera
voven saír verdade ?
*BRE: <as [//] as flores> [>] .
*ANP: <sí xxx> [<] primavera .
*BLA: as flores .
*BLA: tes ganas de que sexa primaveira ?
*ANP: no (.) me gusta la primavera .
BRE: pero a mí me gusta a &pr [//] a [//] a paya [] .
%par: paya=playa \$PHO ;
*BLA: si ?
*BRE: a playa .
*PIL: a playa .
*BRE: si .
*PIL: por que che gusta ?
*BRE: poque [//] poque hai animais .
*PIL: e que animais hai ?
*BRE: hai animais [/] animais [//] animais de [/] de que
viven <nunha
pedra> [//] &vi debaixo dunha pedra .
*BLA: hai animais que viven debaixo dunha pedra ?
*BRE: si .
*ANP: pero yo también (.) cogí piedras .
*BLA: tú también cogiste piedras ?
*BRE: pero <debajo> [>] +/.
*ANP: <muchas> [<] !
*BLA: muchas ?
*ANP: yo sola [?] .
*BRE: se [//] se colles muchas pedras hai un nicho [?] .
*ANP: pero yo no (.) yo tenía un bicho en el patio [?] .
*PIL: que bicho é ese ?
ANP: la rana [] !
%par: rana=rana \$PHO ;
%com: pronuncia casi d
*PIL: una rana en la playa !
*BRE: no [//] na [/] na [//] na playa non hai ranas .

*PIL: verdad que non ?
 *BRE: non .
 *ANP: pues yo sí que las ví .
 *PIL: que hai na playa ?
 *BRE: peces .
 *PIL: e que máis ?
 *BRE: e [/] (.) e [//] e &ta (.) e +/.
 *BLA: hai area ?
 *BRE: +, e area .
 %spa: \$IMIT
 *BLA: e o mar .
 ANP: yo también xxx de la [//] la paya [] .
 %par: paya=playa \$PHO ;
 ANP: yo también fui a la paya [] .
 %par: paya=playa \$PHO ;
 *PIL: también ?
 ANP: también xxx cuando fui a la paya [] .
 %par: paya=playa \$PHO ;
 *PIL: no ?
 *PIL: e fuchedes á casa dos abuelos ?
 *ANP: sí .
 *BRE: yo sí pero &ha [//] pero hai (.) yo teño +/.
 *ANP: <yo sí> [/] <yo sí> [/] yo sí !
 *BRE: ti no !
 *ANP: sí !
 *BRE: no !
 *ANP: que sí !
 *BRE: yo teño &do dous abuelos e [//] e dúas abuelas .
 %act: usa los dedos para enseñarme el número
 *PIL: si (.) que ben !
 *ANP: yo teño cuatro abuelas .
 %spa: \$IMIT
 *BLA: tes catro abuelas (.) que sorte !
 *BRE: y el se &man (.) el (.) o abuelo Juan se (.) &cho (.) veu pola
 carretera xxx un camión e [/] e [//] e o coche rompeu todo .
 *BLA: e pasoulle algo o abuelo Juan ?
 BRE: le [/] le cayó unha pea [] encima .
 %par: pea=pedra \$PHO ;
 *BLA: si ?
 ANP: y <se mancó> [] aquí na [//] na frente (.) <e se> [//] e
 <se
 mancó> [*] aquí na rodilla e a [//] e noutra rodilla e [/] e
 [//] <e
 a> [//] e na cara .
 %par: mancó=mancouse \$CON ;
 %act: señala las partes del cuerpo
 *BLA: e agora xa está ben ?
 *BRE: no (.) estano curando no [/] no [//] no hospital .
 *BLA: está no hospital ?
 *BRE: si .
 *BLA: e ti fúchelo ver ?
 *BRE: si (.) pero [/] (.) pero non podían subir os nenos .
 *BLA: claro (.) non podían .
 *BRE: non .

*ANP: pero yo (.) yo <también lo que saltaba> [?] !
*PIL: Breixo quedou igual de resignado ca mín (.) a que si ?
BRE: pero os [/] os [//] os nenos non saltan no hestatal []
que están
as mulleres enfermas .
%par: hestatal=hospital \$PHO ;
%add: ANP
*ANP: pero yo lo ví .
*BRE: que sí !
*ANP: que yo xxx .
*BRE: eu sí (.) yo sí +/.
%add: ANP
%com: cambio de código en el mismo turno
*ANP: y yo también !
*BRE: ti no !
*ANP: sí !
*BRE: que no !
*ANP: que sí !
*BRE: que no !
*PIL: bueno +/.
*ANP: que sí !
*PIL: a_ver (.) ti fuche e ela tamén (..) ó mellor foi outro día
(.) ou
outro sitio +/.
*BRE: no [//] (.) no que eu [//] yo [/] yo [//] yo lo ví (.)
<hoxe> [>] .
*ANP: <yo fui a una> [<] tienda .
*PIL: ay (.) tu fuche a unha tienda (.) claro (.) e el +/.
*ANP: yo fui a muchas .
*PIL: fuiste fuche a muitas tiendas (.) pero el non estaba
falando das
tiendas (.) a que non Breixo ?
*PIL: claro (.) e que ela confundiuse .
%act: BRE niega con la cabeza
*BRE: <&esta (.) estaba ahi (.) ela estaba falando> [>] dos
pistales [*]
.
%par: pistales=hospitales \$PHO ;
*ANP: <xxx> [<] .
*PIL: mui ben .
*ANP: sí .
*PIL: bueno (.) volvemos ó recreo ?
@End

@Begin
@Languages: es
@Participants: PIL Pili Investigator, ANP Ana Target_Child, BRE Breixo Target_Child
@ID: es|koiné|PIL||female|Vite||Investigator||
@ID: es|koiné|ANP|3;7.22|female|Vite||Target_Child||
@ID: es|koiné|BRE|4;3.21|male|Vite||Target_Child||
@Transcriber: CHE
@Date: 16-MAY-1999
@Location: Santiago de Compostela
@Situation: escuela infantil
@Tape Location: 3/II.4
*BRE: y es que me lo regaló mi mamá .
%com: se refiere a su jersey
*PIL: regaloucho a túa mamá ?
*BRE: está un pouco sucio aquí.
%com: fijémonos en que el niño alterna en el uso de la lengua
*PIL: está un pouco sucio aí?
*PIL: quen o ensuciou?
*PIL: a que eu adivino quen foi?
*PIL: a que fuche tu quen o ensuciache?
*PIL: seguro que estiveche xogando e ensuciáchela.
*PIL: onde estuveches estos días que non viñeches nada á gardería?
*BRE: es que (..) no me acuerdo.
%com: apenas perceptible
*PIL: estarías en monforte?
%add: BRE
*ANP: sí [!].
*PIL: na casa dos abuelos?
%add: BRE
*ANP: sí.
%com: aunque las preguntas no van dirigidas a ella la niña contesta igual
*BRE: no (.) no me acuerdo.
*PIL: e desde cando Breixo me fala en castelán?
*PIL: desde cando me falas en castelán?
%com: el niño tarda en reaccionar a las preguntas
*PIL: a_ver (.) vamos ver un conto.
*PIL: vamos ver un conto contádesme?
*PIL: cóntame Breixo unha parte e ti outra (.) pero Breixo tenma que contar en galego eh (.) senón non me vale.
*PIL: vamos contar qué é o que fai cada un destes señores.
%act: saca un libro y le enseña los dibujos
*PIL: a_ver (.) primeiro vai Breixo (.) vale Ana (.) e despois vas ti.
%act: le enseña los dibujos a Breixo
*PIL: que é este señor?
%act: señala uno de los dibujos
%add: BRE
*BRE: <un> [/] un señor que está traballando.
*PIL: e que está facendo?
*BRE: facendo unha casa.
*PIL: con que?

BRE: <con> [/] con majtillos [].
%par: majtillos=martillos \$PHO ;
*BRE: e con cemento.
*PIL: e esto que está poñendo que son?
*BRE: ladrillos.
*PIL: vale [!].
*PIL: e a_ver (.) cóntame este señor qué fai.
*BRE: o señor que está regando o fuego.
%com: el niño no contesta expresamente a la pregunta sino que dice lo que ve por eso utiliza el relativo "que"
*PIL: e como se chama ese señor que rega o fuego?
%add: BRE
*ANP: bombero.
%com: contesta aunque la pregunta no va dirigida a ella
*BRE: bombero.
*PIL: moi ben Ana (.) bombeiro!
*PIL: e este?
*ANP: cartero.
BRE: castero [].
%par: castero=cartero \$PHO ;
*PIL: así non vale (.) estaslle axudando!
%add: ANP
BRE: es el castero [].
%par: castero=cartero \$PHO ;
*PIL: e que fai este señor?
BRE: está dando cajtas [].
%par: cajtas=cartas \$PHO ;
*PIL: e onde as leva?
%com: se produce un corte en la grabación que impide escuchar la respuesta del niño
BRE: le daba cajtas [] a la gente.
%par: cajtas=cartas \$PHO ;
*PIL: ah (.) e esta señora qué é?
*BRE: un policía.
*PIL: e que fai?
BRE: <pa(ra)> [/] para los coches <y que> [/?] y <que vayan> [].
%par: y que vayan=para que sigan \$LEX ;
%com: tal y como se expresa el niño la construcción carece de sentido
*PIL: y una pregunta en qué me estás hablando?
BRE: en [] policía[!].
%par: en=de \$CON ;
*PIL: no (.) en qué me estás hablando?
*PIL: en castellano o en gallego?
*PIL: en qué me estás hablando?
%act: el niño se queda callado y agacha la cabeza
*PIL: sabes en qué?
%act: BRE dice que no con la cabeza
*PIL: di a verda(de) sábelo ou non?
*PIL: non?
*PIL: e ti sábelo Ana?
%act: los niños se quedan callados
*PIL: non teñen conciencia pragmática todavía.

%add: CHE
 *PIL: vamos a_ver (.) agora tócalle un pouco a Ana eh?
 *PIL: a_ver Ana cóntame qué é este señor?
 %act: le enseña un dibujo
 *ANP: eh@i (.) médico.
 *PIL: y por qué lo sabes?
 *ANP: porque sí.
 *PIL: y qué tiene este señor (.) a_ver cuéntame qué tiene por ahí.
 *ANP: <un> [//] uno de mirar la barriga.
 *PIL: y a quién tiene en los brazos?
 *ANP: un bebé.
 *PIL: a un bebé, ah!
 *PIL: y este señor?
 *ANP: está con sus ovejas.
 *PIL: y cómo se llaman los señores que están con sus ovejas?
 *ANP: pues (.) son los que llevan bastones.
 %com: fijémonos en que la niña cuando no sabe contestar a lo que le preguntan recurre a describir lo que ve
 *PIL: y se llaman pas(tores) +//.
 ANP: ++ (pas)tones [].
 %par: pastones=pastores \$PHO ;
 *PIL: no (.) ellos (.) cómo se llaman (.) pasto(res) +//.
 %act: la niña no sabe contestar
 *PIL: pastores!
 *PIL: es un pastor!
 *BRE: pastores.
 %com: lo que dice Breixo apenas se escucha
 *PIL: y este señor qué es?
 %act: le enseña otro dibujo
 *ANP: xxx.
 *PIL: cómo?
 *ANP: xxx.
 *PIL: tu sabes o qué é?
 %add: BRE
 *BRE: un páxaro [?].
 ANP: <dos peces> [?] (.) <no (.) tres pefes []> [?] xxx.
 %par: pefes=peces \$PHO ;
 *BRE: no (.) estos son peces (.) esto es un páxaro y un señor.
 *PIL: claro y el señor +//.
 *ANP: uno dos tres cuatro!
 *PIL: el señor con quién está?
 *PIL: con los +//.
 *BRE: ++ con los peces.
 *PIL: entonces es un pes(cador) +//.
 *BRE: ++ (pes)cado.
 *PIL: un pescador!
 *PIL: e este señor quen é?
 %add: BRE
 *ANP: pescador.
 *PIL: no este non é pescador este que é?
 %add: BRE
 *PIL: este é carniceiro (.) vamos ver outro libro que este xa non nos

gusta.
*BRE: xa casi no lo sabemos.
%com: se refiere al libro
*PIL: xa casi o sabedes?
*PIL: sabedes este?
%act: le enseña un libro nuevo
*ANP: el <loro> [?].
*PIL: a_ver (.) este es un cuento de un oso.
*BRE: de un o(so) +/.
*PIL: a_ver quén está por aí (.) contádeme.
*ANP: tú no sabes?
*BRE: un osito.
*PIL: está un osito (.) con quién está ?
*BRE: con un neno.
*PIL: <e ese neno> [>] +//.
*ANP: <con un neno> [<].
*PIL: e ese neno con quen está ?
*ANP: con un oso polar!
*PIL: quen é esta de aquí?
%add: BRE
*ANP: polar!
%com: parece que la niña quiere llamar la atención
*BRE: a mamá .
*PIL: a mamá (.) e esta sabedes quen é?
*BRE: quen?
*PIL: a irmá .
*BRE: a irmá .
%spa: \$IMIT
%com: el niño repite por lo bajo
*PIL: bueno entonces ten o seu moniño e a mamá está cuidando del
por que
cuida del (.) sabes por qué cuida del?
%act: BRE dice que sí con la cabeza
*PIL: por qué?
*BRE: <porque> [/] <porque> [/] <porque> [/] porque o neno naceu
<na> [/
na barriga da mamá .
*PIL: claro e porque é un neno moi pequeno todavía.
*PIL: eu creo que Breixo si que sabe en qué me está falando.
%add: CHE
*PIL: e quen o leva aquí agora?
%act: señala un dibujo
*ANP: un papá !
*PIL: non (.) ese non é o papá .
*BRE: a mamá .
*ANP: a mamá .
%spa: \$IMIT
*PIL: non (.) esa non é a mamá quen é esta?
*ANP: un abuelo.
*PIL: este é seu irmao (.) o maior (.) tu tes irmaos maiores?
%add: BRE
*BRE: si.
*PIL: si?
*PIL: e como se chaman?
%act: el niño se queda callado sin saber qué contestar

*PIL: tes irmaos ou non?
*BRE: sí (.) un hermano que se llama xxx +//.
%act: el niño comienza a balbucear unas palabras
*BRE: no tengo.
*PIL: non tes irmaos!
%com: riendo
*PIL: e Ana ten irmaos?
*ANP: uno más mayor xxx.
%act: indica con la mano que su hermano tiene cinco años
*PIL: cantos anos ten?
%act: ANP enseña cinco dedos para indicar la edad de su hermano
*PIL: así?
*PIL: e canto é así?
%add: ANP
*BRE: cinco es muy poco.
*PIL: non é moi grande (.) verda(de) (.) case é coma ti.
*ANP: xxx.
*PIL: cómo?
*ANP: que ahora va al cole(gio) del xxx.
*PIL: non te entendín nada (.) Ana que dixeches?
*ANP: que va al cole(gio) xxx.
*PIL: va al colegio lópez ferreiro.
ANP: zí [].
%par: zí=sí \$PHO ;
*PIL: pero hai que pronunciar mellor (.) eh (.) que xa es unha nena grande para falar así.
*PIL: a_ver vamos a_ver qué cartas nos tocan [!].
*PIL: esta vez van ser difíciles (.) eh.
*BRE: por qué?
*PIL: porque xa sodes maiores.
*PIL: a_ver (.) vaille tocar unha a un e outra a outro.
*PIL: a_ver vounas baraxar ben.
*PIL: a_ver empeza Breixo.
%act: le extiende las cartas para que el niño coja una
%com: el niño se queda mirando la carta sin decir nada
*PIL: sabes o que é?
%act: BRE dice que no con la cabeza
*PIL: a_ver (.) enséñame.
*PIL: tí sabes o que é (.) Ana?
%act: ANP dice que no moviendo la cabeza
*PIL: xa vos dixen que iban ser moi difíciles!
*ANP: una paloma!
*PIL: non (.) é algo máis grande ca unha paloma.
*PIL: unha á(guila) +//.
*BRE: ++ unha águila.
*PIL: unha águila (.) moi ben!
*PIL: a_ver qué lle toca a Ana.
%act: extiende las cartas hacia la niña que coge una carta y se la queda mirando si contestar
*PIL: sabes o que é?
*ANP: un malo [?].
*PIL: un +//.
*ANP: ++ malo [?].

*PIL: un qué?
 *ANP: un malo [?].
 *PIL: non sei o que é eso .
 *ANP: un malo [?].
 *PIL: a_ver (.) ti sabes o que é?
 %act: le enseña la carta al niño
 *BRE: <un> [/] (.) un loro.
 *PIL: é algo parecido a un loro (.) un tu(cán) +//.
 *BRE: ++ un tubo.
 *PIL: no.
 *ANP: un tubo.
 %com: repite de nuevo lo que había dicho el niño.
 *PIL: un tucán!
 *PIL: chámase un tucán estos páxaros.
 *PIL: Ana estasme ensuciando todos os pantalóns desde que cheguei.
 *PIL: e eso?
 %com: se refiere a otra carta que le entrega
 *BRE: un rinoceronte.
 *PIL: no (.) non é un rinoceronte (.) é parecido [!].
 *PIL: non ten cornos (.) entón no é un rinoceronte (.) o rinoceronte ten aquí un corno.
 %act: la entrevistadora señala el dibujo
 *PIL: ti sabes o que é Ana?
 *PIL: enséñallo a Ana.
 %act: el niño le enseña la carta a la niña.
 BRE: un cocolilo [].
 %par: cocolilo=cocodrilo \$PHO ;
 *PIL: no [!].
 *ANP: <no> [/] <no> [/] no.
 *PIL: que é?
 *ANP: un monstruo [?].
 *PIL: un hi(popótamo) +//.
 *ANP: ++ (hi)popótamo.
 *PIL: <un hipopótamo> [!] [>].
 *BRE: <un hipopótamo> [<].
 %spa: \$IMIT
 *PIL: a_ver cómo é?
 %add: BRE
 *BRE: un xx.
 %com: el niño no sabe pronunciar la palabra.
 *PIL: a_ver cómo.
 BRE: popopotopo [?] [].
 %par: popopotopo=hipopótamo \$PHO ;
 ANP: hipopótano [].
 %par: hipopótano=hipopótamo \$PHO ;
 *PIL: cómo?
 ANP: <un hipo(pótamo)> [//] un hipopótano [].
 %par: hipopótano=hipopótamo \$PHO ;
 %com: la niña le repite al niño como debe pronunciar la palabra pero ella también tropieza a la hora de hacerlo
 *PIL: a_ver ti sabes dicilo?
 *PIL: a_ver díó Breixo.

*PIL: non sabes dicilo?
%add: BRE
*ANP: no.
*PIL: si que sabe (.) oh!
*ANP: pero no quiere.
*PIL: non quere dicilo (.) bueno (.) pois entón tócalle a Ana.
%act: la entrevistadora le enseña las cartas y la niña coje una
*PIL: que é eso?
%add: ANP
*BRE: unha mosca.
*ANP: unha mosca.
%spa: \$IMIT
*PIL: unha mosca [!].
*PIL: son pesadas as moscas verda(de)?
*ANP: no.
*PIL: no?
*ANP: arañan [?].
*PIL: arañan as moscas?
*PIL: non arañan as moscas o que pasa é que molestan verdade ?
*ANP: sí y en casa [?] también.
*PIL: como están todo o día na casa?
*ANP: todavía [?] están en mi casa.
%act: la entrevistadora le da una carta al niño y él la mira
*ANP: después las otras!
*PIL: despois as outras.
*PIL: a_ver enséñame o que é.
%act: el niño le enseña la carta a la entrevistadora.
*PIL: dios que difícil [!] ti sabes o que é?
*PIL: un es(carabajo) +//.
BRE: ++ carabejo [].
%par: escarabejo=escarabajo \$PHO ;
*PIL: un escarabajo (.) muy bien Breixo!
%act: el niño se pone todo contento por haber acertado
*PIL: a_ver Ana.
*PIL: xa é a última eh Ana.
*PIL: uy <qué difícil> [>].
*ANP: <una mosca> [<].
*PIL: é parecido a unha mosca pero non é unha mosca .
*PIL: ti sabes o que é Breixo?
%act: BRE dice que no con la cabeza
*PIL: andan así por encima da auga.
*ANP: <una> [/] una mariposa.
*PIL: un parecido a unha mariposa pero non é .
*ANP: xxx mariposas.
*PIL: é unha li(bélula) +//.
*PIL: ++ libé(lula) +//.
BRE: ++ lileta [].
%par: lileta=libélula \$PHO ;
%com: el niño completó la palabra sin saber realmente de qué estaba hablando
*PIL: libélula [!] chámase (.) é mui difícil eso eh .
*PIL: e a_ver se xa terminamos (.) a_ver qué é eso Breixo .
%act: le entrega una carta
*BRE: un tigre!

*PIL: non é un tigre (.) é parecido a un tigre (.) pero non é un tigre
(.) eh .
*PIL: un le(opardo) +//.
*BRE: ++ un león pardo .
*PIL: un leopardo (.) moi ben!
*BRE: bien!
%act: la entrevistadora le enseña una carta al niño
*PIL: mira esto é un tigre esto é un tigre.
*PIL: vamos a_ver as outras cartas logo?
*BRE: vale (.) sí (.) sí!
%com: mientras la entrevistadora saca las cartas el niño muestra signos de cansancio
*PIL: cantas cousas vimos hoxe (.) dios.
*PIL: a_ver estas son do que podes comprar.
%exp: se refiere a las cartas que acaba de coger
%act: los niños se acercan para ver las cartas
*PIL: pero non vale velas (.) eh?
*BRE: sólo te estoy mirando eso.
%exp: se refiere a la tapa de una carpeta
%com: fijémonos en cómo el niño cambia de lengua
%act: los niños se acercan a mirar algo y se rien entre ellos
*PIL: vouvos dar a vosoutros.
*PIL: que vos pasa?
*PIL: de que vos rides?
*PIL: a_ver quén mo conta?
*BRE: de la caja.
*PIL: por que?
*ANP: de la caja.
BRE: la caja de las cajtas [].
%par: cajtas=cajas \$PHO ;
*PIL: que ten a caixa das cartas?
BRE: hay dibijado [?] [] ahí unas cosas xxx.
%par: dibijado=dibujado \$PHO ;
*PIL: que ten dibuxado?
*BRE: comida!
*ANP: comida!
*BRE: comida de comer.
%com: los niños realizan estos comentarios en tono jocoso
%act: la entrevistadora baraja las cartas y se las ofrece
*PIL: a_ver (.) empezamos por Ana.
%act: la niña coge una carta
ANP: zanorias [].
%par: zanorias=zanahorias \$PHO ;
*PIL: zanahorias [!] soupéchelo á primera (.) oh.
*PIL: a_ver Breixo.
%act: el niño coge una carta
*BRE: un bocadillo.
*PIL: no.
%act: la niña le mira la carta
*ANP: un bocadillo.
%spa: \$IMIT
BRE: un bocadillo <y> [/] y una tajta [] y +//.
%par: tajta=tarta \$PHO ;

*ANP: ++ <y una> [//] <y un pato [?]> [>].
 *BRE: <un xxx> [<].
 *BRE: y un pato [?].
 *PIL: pode ser pero son pasteles máis ben (.) eh.
 *PIL: a_ver.
 %act: la entrevistadora le entrega una carta a Ana
 *ANP: pastel (.) no.
 *BRE: no.
 *PIL: no.
 BRE: una madanela [].
 %par: madanela=madalena \$PHO ;
 *ANP: una madalena con leche!
 *PIL: una qué?
 *ANP: una madalena con leche!
 *PIL: e tu como dixeches unha +//.
 %add: BRE
 *ANP: ++ madalena [?] con leche.
 *BRE: madalena [?] con leche.
 %spa: \$IMIT
 %com: el niño dice la palabra con mucha dificultad
 *PIL: te voy a dar yo a ti (.) que me diste esas palabras.
 %act: el niño coge una carta y la mira
 *BRE: huevos.
 *PIL: no [!] míraos ben.
 *ANP: sí <huevos> [//] son huevos con cacas!
 *PIL: que son?
 *BRE: son huevos con cacas.
 %spa: \$IMIT
 *PIL: non mira ben a_ver qué son.
 *ANP: son huevos con cacolas.
 *PIL: que son?
 BRE: son escarabejos [].
 %par: escarabejos=escarabajos \$PHO ;
 *PIL: non (.) son me(xilóns) +//.
 *ANP: ++ son cacas!
 *PIL: <me(xilóns)> [>] +//.
 *ANP: <son cacas> [<] verdad?
 %add: BRE
 *PIL: me(xilóns) +//.
 *PIL: mexi(lóns) +//.
 *BRE: ++ (meji)llones.
 *PIL: mexilóns (.) moi ben (.) a_ver ti.
 %add: ANP
 %com: fijémonos en los cambios de lengua en el niño
 *ANP: cacas.
 *BRE: unha piña +/.
 *ANP: no cacas!
 *BRE: non unha piña!
 *PIL: a_ver vaines contar un conto Breixo.
 *BRE: yo no.
 *PIL: sabes o conto de caperucita Breixo?
 *BRE: no!
 *PIL: e ti Ana sabelo?
 %act: ANP dice que no con la cabeza
 *PIL: tampouco!

*BRE: <yo> [/] <yo tengo la pelí(cula)> [//] yo tengo el cuento de caperucita.

*PIL: en dónde?

*BRE: en casa.

*PIL: y nunca lo trajiste.

*ANP: yo sí cuando tú no viniste.

*PIL: y tú dónde estuviste también estos días que no te vi nada?

ANP: porque tú no vinistes [].

%par: vinistes=viniste \$VER ;

*PIL: y tú tampoco.

*ANP: sí [!].

*PIL: no [!].

*ANP: que sí [!].

*PIL: hubo un día que no viniste (.) que estaba esto cerrado.

*ANP: dónde?

*PIL: aquí estaba cerrado (.) no había colegio.

*PIL: te acuerdas que estaba cerrado?

*ANP: sí.

*PIL: y dónde estuviste (.) cuando +/-.

*ANP: pues en casa.

*PIL: en casa sola?

*ANP: en mi casa.

*PIL: solita?

*ANP: no!

*PIL: con quién?

*ANP: con mi padre y <con mis> [/] con mis primos xxx con mi padre con mi madre y con mi hermano.

*PIL: pero tu madre no trabaja?

*PIL: sí que trabaja tu madre no?

*ANP: sí que trabaja (.) <mi madre xxx> [//] su cole está <allí> [//] allí.

%act: la niña señala hacia la ventana

*PIL: trabaja en el cole tu madre?

*ANP: sí.

*PIL: sí?

*PIL: y tu madre en que trabaja Breixo?

*BRE: <en> [/] en casa.

*PIL: en casa?

*PIL: y tu papá ?

*ANP: también mi madre trabaja en casa <y en el cole> [>].

*BRE: <mi papá> [<] [/] mi papá <trabaja> [/] trabaja curando vacas.

*PIL: ah é verda(de) que teu papá é veterinario (.) non me acordaba eu!

*PIL: y tu papá ?

%add: ANP

ANP: el tabaja [] muy lejos.

%par: tabaja=trabaja \$PHO ;

*PIL: muy lejos muy lejos [!].

*ANP: sí +/-.

*BRE: trabaja en el cielo decías verdad?

*ANP: sí (.) hasta arriba de todo.

*PIL: arriba de todo de todo xxx.
*ANP: sí.
*PIL: <qué way> [>].
*BRE: <trabajaba> [/] trabajaba en el cielo con unas escaleras.
*ANP: sí [!].
*PIL: ostras [!] <es verdad ya no me acordaba de eso Breixo> [>].
*BRE: <<hac(ía)> [/?]> [<] <hacia> [//] <haciendo> [/] haciendo castillos
verdad?
*ANP: no pero <haciendo con el martillo> [?] con el castillo.
*PIL: es verdad que nos lo contara una vez (.) no me acordaba yo ya.
*PIL: bueno vámonos!
*ANP: no!
*BRE: sí!
*ANP: mucho tiempo.
*PIL: más tiempo quieres hechar aquí?
*ANP: sí como los otros.
*PIL: y como los otros hechamos el mismo tiempo (.) el mismito el mismito.
*PIL: lo que pasa es que hoy se te pasó volando (.) normalmente quieres irte muy rápido.
@End