This is the peer reviewd version of the followng article:

Doing interlanguage analysis in school contexts / Pallotti, G. - In: Communicative proficiency and linguistic development / I. Bartning, M. Martin & I. Vedder. - ELETTRONICO. - Cambridge : Eurosla, 2010. - ISBN 9781446669938. - pp. 159-190

Eurosla
Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

27/11/2026 15:19

(Article begins on next page)

Doing interlanguage analysis in school contexts

Gabriele Pallotti
University of Modena and Reggio Emilia

In second language acquisition (SLA) research it is well-known that interlanguages are autonomous and rule-governed linguistic systems whose grammar cannot be described simply in terms of errors and deviations from L2 norms. However, assessment practices at school, both formative and summative, heavily rely on counting errors and scoring them based on various types of 'gravity'. Carrying out a systematic interlanguage analysis as it is done in SLA research would be highly time consuming and very impractical in most teaching and testing contexts. The chapter will show some examples of interlanguage analysis in such contexts. Different stages of the research process will be focussed upon, including data collection, transcription, coding, scoring, and quantitative and qualitative analysis. If CEFR scales are to be related to acquisitional sequences in teaching and testing contexts, it is necessary to find ways in which the latter can be assessed in a reasonable amount of time and without specialized skills, while preserving the procedure's validity with the respect to current SLA theorising and methodology.

1. Introduction

One of the main aims of this volume is to find relationships between linguistic development, as described in Second Language Acquisition (henceforth, SLA) research, and the acquisition of communicative proficiency, as described in the Common European Framework of Reference (Council of Europe, 2001; henceforth CEFR). Most chapters report on studies in which various measures of linguistic performance calculated by SLA researchers are matched to assessments of communicative language proficiency made by raters experienced in language testing and assessment. The question remains whether these raters, or other practitioners, would be able to conduct the linguistic part of the analysis themselves, if this could be implemented in concrete (formative and summative) assessment practices, how it should be conducted and what type of training would be needed to allow teachers and language testers to implement interlanguage analysis in their professional domains.

The aim of this chapter is to discuss these issues, by reporting on a project aimed at bringing interlanguage analysis to school. The project involved sever-

EUROSLA MONOGRAPHS SERIES 1
Communicative proficiency and linguistic development, 159-190

al teachers at different levels, from kindergarten to middle school, as part of wider teacher-training and action-research schemes. Teachers were involved in collecting data and in their subsequent analysis based on the notion of interlanguage (henceforth, IL). These procedures were designed so that they could be implemented in everyday school settings. The main goal was formative assessment, in the more usual sense of being oriented to improving didactic strategies, but also in the sense of being part of teacher training.

The present chapter thus differs from others in the volume. No attempt will be made at relating the development of communicative language proficiency, as described in the CEFR, with the acquisition of linguistic structures. The main goal will rather be that of presenting a scheme for training teachers in analysing their pupils' interlanguage. This will lead to a more general discussion on how practitioners, like teachers and language assessors, can acquire the skills that are necessary for describing linguistic development in ways that are consistent with the notion of interlanguage and current SLA research. If SLATE's efforts are to have an impact on teaching and testing practices, it is important to address the issue of how linguistic development can be assessed in such contexts and how professionals should be trained accordingly.

The present contribution is far from offering an exhaustive and systematic answer to such crucial questions. The project reported here is in fact limited in several ways. Besides being still at a rather exploratory stage, it concerns a single language, Italian, and a very special learners' group, i.e. children aged between 5 and 12. However, it is hoped that the present discussion will stimulate a more general reflection on how interlanguage analysis can be integrated into formative and summative assessment.

2. Interlanguage description

2.1. The notion of interlanguage

The term interlanguage was first introduced by Selinker (1972), who defined it as "a separate linguistic system based on the observable output which results from a learner's attempted production of a TL [= Target Language] norm" (p. 214). This orientation derived in turn from the Error Analysis Approach (Corder, 1967), which emphasized that errors are an important window on the learner's processes and strategies and that their careful analysis is more productive, from a pedagogic and scientific point of view, than the mere counting, scoring and sanctioning of 'wrong' forms.

The relationship between errors and interlanguage development has remained intricate and several authors have felt the need, at various times, to stress that the two notions should be kept conceptually apart. Bley-Vroman

(1983) for example warned against the risk of committing a "comparative fallacy", arguing that "work on the linguistic description of learners' language can be seriously hindered or sidetracked by a concern with the target language" (p. 2). Similarly, Sorace (1996) notes that, if the aim is to reconstruct a learner's linguistic system, "the evaluation of the distance between native and non-native grammars becomes an irrelevant criterion" (p. 386).

The fact that different researchers, at different times and from different perspectives, have stressed the independence of the notion of interlanguage from those of error and conformity to L2 norms testifies that the two are often mingled. However, this is not always the case and several authors have provided descriptions of interlanguage development eschewing any reference to accuracy or errors. For example Pienemann (1998) proposes the notion of "factorization" as a way of disentangling various factors bundled together in the L2 which may lead to 'errors'. A learner may develop an interlanguage system in which just one of such factors governs a set of form-function associations, which should be described in their own right, regardless of the fact that they yield forms not allowed by L2 rules. For instance, in a fusional language like Swedish or German, adjectives may be inflected based on a variety of factors, such as gender, number, definiteness or case. A learner who associates one inflectional morpheme with just one of these factors, e.g. number, will produce many forms deviating from L2 norms, but would nonetheless follow a clear interlanguage rule.

Another approach that has coherently looked at interlanguage development in its own right is that of the Basic Variety (Klein & Perdue, 1992, 1997). These authors have described some organizing principles of utterance construction in the early stages of IL development. The principles predict the order in which constituents will appear in an utterance and they have to do with semantic notions such as 'the referent with highest control on the situation', or pragmatic-textual ones, such as 'topic' and 'focus'. Based on the extensive ESF project database (Perdue, 1993), the authors found these principles to hold independently of the L1 and the L2, thus pointing to a generalisable functional explanation of interlanguage development making no reference to errors and conformity to L2 norms.

The Basic Variety Approach, as the name itself suggests, was developed to investigate the initial stages of second language acquisition. In such cases it is more obvious that learners' productions should be analysed according to their internal logic rather than by looking at their conformity to L2 norms. Norris and Ortega (2003) suggest that this way of looking at interlanguage as an independent system is more relevant for initial stages, while other types of analysis are more appropriate for later stages. At intermediate levels quantitative measures of the spread and consistency of use of a structure may be employed, while accuracy-based measures would be more meaningful for characterizing the production of advanced learners.

2.2. Describing interlanguage development in terms of accuracy

Despite these recommendations, the description of learner language in terms of errors is still quite widespread. In SLA research many studies characterise learning over time or after an experimental treatment with accuracy measures, like the number of error-free T-Units or number of errors per 100 words (e.g. Wolfe-Quintero, Inagaki, & Kim, 1998). Teachers, too, often describe and evaluate their students' performance based on the number and type of errors made, which is also a common practice in language assessment. For example, the CEFR (Council of Europe, 2001, p. 114) provides descriptors for "grammatical accuracy", which constantly refer to the number and type of errors made, but none on "grammatical development". Examples of these descriptors are: "systematically makes basic mistakes (A2); errors occur, but it is clear what he/she is trying to express (B1); does not make mistakes which lead to misunderstanding (B2); errors are rare and difficult to spot (C1)". Clearly, since the CEFR is a language-neutral instrument, it would have been impossible to refer to the development of specific grammatical features, and a scale referring to a general notion of "accuracy" (necessarily related to errors' quantity and quality) is perhaps unavoidable. Users of the CEFR are indeed invited to consider "which grammatical elements, categories, classes, structures, processes and relations are learners, etc. equipped/required to handle" (p. 114), but nothing else is said about how interlanguage development should be conceptualised and reported. Even English Profile, a comprehensive project for relating CEFR descriptors to linguistic development in L2 English, heavily relies on counting and scoring errors (Hendriks, 2008; Williams, 2008).

There is nothing inherently wrong in reporting accuracy scores, but one should be clear about what is achieved with this type of approach. As Wolfe-Quintero et al. (1998) write: "the purpose of accuracy measures is precisely the comparison with target-like use. Whether that comparison reveals or obscures something about language development is another question" (p. 33). In other words, 'accuracy', i.e. the degree of conformity to L2 norms (expressed as the ratio of 'wrong' forms to overall production), should not be taken as a direct indicator of interlanguage development, and may actually somehow distort the picture.

The error-counting approach has some clear advantages. It is relatively easy to understand and can easily be applied by teachers, assessors and other practitioners, for whom noticing the presence of an error is more intuitive than understanding the internal logic of an interlanguage system. Judgements thus tend to be rather reliable, although various authors have reported a number of problems in the operationalization of what errors are and how they should be scored, a problem that becomes even more apparent in the case of qualitative judgements on the gravity of different types of errors (James, 1998). In any case,

when searching for correlations between descriptions of linguistic and communicative competence (as is done in most chapters in this volume), one should always bear in mind that, as regards the former, 'accuracy growth' and 'interlanguage development' do not represent the same construct.

2.3. Interlanguage analysis in research contexts

Analysing interlanguages in research contexts is often a complex and laborious process. Some authors, especially in the generative framework, have focussed exclusively on acceptability judgements or comprehension tasks, as these should better reflect the underlying linguistic competence without the many confounding factors related to on-line performance. However, in other approaches the predominant sources are production data and more specifically those coming from communicative oral tasks. The preference for oral data has been motivated by their being more spontaneous, unplanned, and thus based on implicit linguistic knowledge.

Oral data need transcribing, which is very time-consuming. For a medium-grained transcription, about 10 hours are required for one hour of data, and the figure can easily increase in the case of more accurate transcriptions, e.g. including phonological, prosodic or gestural information. Transcribed data are then usually subjected to some kind of coding in which all instances of the phenomenon under investigation are tagged, counted and classified. This leads to analysis proper, consisting in observing frequencies and relationships among categories in order to arrive at generalizable conclusions.

The aim of analysis is typically the discovery of developmental orders, for instance how a certain structure emerges and then gradually spreads out in interlanguage, or how this path is related to that of other linguistic structures. Studies may be purely descriptive or they may test specific theory-based hypotheses. In both cases, conclusions need to be reliable, hence the need for robust data sets containing large numbers of tokens of the phenomenon under investigation. If the focus is on the development of a specific feature, researchers must ensure that their data sets are sufficiently 'dense' (Pienemann, 1998), i.e. that they contain several contexts requiring the production of a given structure. Studies on developmental orders, making generalisable statements about the appearance of certain structures before others, need to set explicit acquisition criteria specifying the conditions that must be satisfied in order to conclude that a structure has been acquired (Pallotti, 2007).

Performing such analyses requires not only a substantial amount of time, but also a specific expertise, including the ability to identify variable linguistic rules and to define them in abstract terms, transcending the norms of a specific language but grounded in general principles such as those proposed by Universal Grammar or Functional-typological linguistics. All this is usually

beyond the reach of most practitioners in teaching and testing contexts. The project reported here had the aim of developing simplified and practically manageable ways of doing interlanguage analysis at school, with learners aged 5-12. The main goal was that of training teachers in new ways of looking at their pupils' linguistic productions, in order to improve their teaching and promote language development in multilingual classes. The insights gained in this research might also be generalized to other contexts, such as large-scale assessment, as will be discussed in the final section.

2.4 Assessing young learners

Assessing learners younger than 12 poses specific problems and requires methods and procedures that may differ significantly from those used with adolescents and adults. The relevant literature is relatively scant but steadily growing - see e.g. reviews by Rea-Dickins and Dixon (1997), Cameron (2001), Ioannou-Georgiou and Pavlou (2003), Hasselgreen (2005), McKay (2005, 2006), Bailey (2008), and a special issue of *Language Testing* (Rea-Dickins, 2000).

Carpenter, Fujii, and Kataoka (1995) is one of the few studies entirely devoted to the identification of suitable protocols for assessing young learners' interlanguage. The authors describe an oral interview procedure for assessing L2 learning in children aged 5-10. The procedure begins with a task requiring children to manipulate objects in order to respond to simple commands and answer simple questions requiring no or minimal verbal production. After a few minutes of conversation, an information gap task follows, requiring children to discover differences between a picture they have in front of them and one held by the interviewer. Children are then given four pictures and asked to select one that does not belong with the others, motivating their choice; this elicits the production of comparisons and academic language on similarities, differences and categorization patterns. The next task is a picture-story narrative based on *Goldilocks and the three bears* and the last task is a role play where children, using puppets, impersonate first themselves talking to another child, then a teacher talking to students, in order to elicit register variation.

I am not aware of other published sources describing systematic research procedures for a comprehensive assessment of young learners' interlanguage. Ideas and practical suggestions can be drawn from empirical studies that have used one or, occasionally, more than one elicitation tool for data collection in child SLA (see e.g. chapters in Philp, Oliver, & Mackey, 2008). Other useful resources are contributions on data elicitation techniques for SLA research in general (for a comprehensive review, Gass & Mackey, 2007), most of which can be employed, perhaps with adaptations, with young children. Alternatively, many methods for assessing monolingual children can be used for bilinguals (a useful review of methods for assessing syntactic competence can be found in

McDaniel, McKee, & Smith Cairns, 1996). Finally, a few exams now exist that are specifically designed for assessing young learners of a second or foreign language, such as the Cambridge Young Learners English Tests (see Taylor & Maycock, 2007).

For the assessment of young children in school contexts, most authors indicate the portfolio as the approach of choice (e.g. Ioannou-Geogiou & Pavlou, 2003). Valdez Pierce and O'Malley (1992) examine a number of options for conducting performance and portfolio assessment in school contexts, pointing out the strengths of these approaches and some possible problems, including the amount of time required and the difficulty to administer tasks to individual students. More recently, Hasselgreen (2003) reports on a large-scale international project for the construction of portfolio based on CEFR 'can do' statements especially geared towards a population of young adolescents aged about 13-15.

Young learners' assessment is most of the times formative, i.e. it aims at diagnosing pupils' strengths and weaknesses, their achievements, difficulties and developmental paths, in order to make teaching more effective and, possibly, make pupils (especially older ones) aware of their own learning, which is supposed to increase their motivation and autonomy. Given these premises, issues of validity do not primarily concern the representativeness of the sample or its adequacy for evaluating the skills possessed by an individual in a range of everyday situations, as in more standardized high-stakes testing, but rather have to do with the usefulness of assessment for promoting learning.

3. Interlanguage analysis in school contexts: a case study

3.1. The project

The main goal of the approach described here is to bring techniques and strategies of interlanguage analysis commonly used in SLA research into school contexts, in order to increase teachers' "diagnostic competence, i.e., the ability to interpret students' foreign language growth, to skilfully deal with assessment material and to provide students with appropriate help in response to this diagnosis" (Edelenbos & Kubanek-German, 2004, p. 259).

The main aim is thus formative assessment, giving teachers conceptual tools to understand how their students make progress in the second language in order to better assist them with well-designed and appropriately timed pedagogic activities. Most teachers in the contexts where the project was carried out lack the skills needed to interpret interlanguage development. They notice pupils' errors, interpreting them with vague and completely impressionistic opinions about their 'gravity'. At the primary and middle school levels they believe that their intervention should just consist in marking all the errors and counting

them, while in preschool the prevailing attitude is 'let nature take its course', i.e. not making any specific effort at focussing children's attention on grammatical forms. Little or no attempt is made at understanding how and why errors are produced and it is often the case that even very different error types, such as grammatical, lexical, phonological or orthographical, are bundled together and hardly ever meaningfully set apart. This confusion is aggravated when it comes to summative assessment. School teachers discuss whether students should pass or fail based on their errors, what errors can be considered to be acceptable after a certain number of months or years of exposure to the L2, or how grades should be assigned based on the number and type of errors (for a discussion on critical issues and good practices in classroom-based assessment, see Rea-Dickins, 2008).

The project presented here had the aim of training teachers in a different approach. They were asked to collect and transcribe samples of their pupils' oral productions, which departs from their usual practice of assessing and grading written texts only. Although time-consuming, the act of transcribing itself promotes closer attention to interlanguage dynamics and the realisation that children construct their own rules in creative and systematic ways, rather than just 'make mistakes'. This awareness was further stimulated in subsequent analyses of the transcribed materials, in which teachers noted down, classified and interpreted various types of linguistic behaviour, making an effort to use positive formulations of what structures are present and how they work rather than just listing errors and shortcomings. Teachers were thus asked to reproduce, in a simplified and assisted way, the methodology of SLA research projects. The aim was to promote a new attitude towards interlanguage productions, based on understanding their internal logic and systematicity, in order to assist learners in their gradual approximation to the target language. The focus was mainly on grammatical structures, but teachers were also asked to look at lexical, textual and communicative features, in order to realise that grammar is just one dimension to be considered.

3.2. Context and participants

The project has been conducted for the past three years and is still under way. It involved 10 kindergarten, 7 primary and 2 middle school classes in different parts of Northern Italy, with a total of about 40 participating teachers, who were actively involved in data collection and in the creation, selection and fine-tuning of procedures for data elicitation and analysis. Altogether, about 120 NNS children of different linguistic backgrounds and 40 NS children have been included, aged between 5 and 12. The NNS children's proficiency in Italian varied from very basic to native-like. Some of them were enrolled in childcare services in Italy very early on, even before 3, and their competence in Italian was

virtually identical to that of their monolingual peers. Many others began their exposure to Italian in preschool, after 3, and a few started at 6 or later. It is not possible to provide here a more detailed description of their linguistic levels, as the assessment of their linguistic development was one of the aims of the project and the data collected thus far do not allow sorting them into precisely defined developmental levels. Actually, 'grading' pupils and sorting them into levels was not one of the aims of the scheme, and it was in fact discouraged, as will be discussed in the final section. The primary goal was to make teachers understand the logic of interlanguage systems, and asking them to classify children into bands or levels would have distracted them from the task of interpreting their productions.

The group of teachers was heterogeneous, too. All of them were already in-service and the majority had a considerable number of years of experience. Preschool teachers in Italy do not specialise in specific curriculum areas. Primary school teachers can in principle teach all areas of the curriculum, although most of them used to specialise in two broad areas, humanities and mathematics-sciences. This has been changed by a recent reform (2009), which is promoting a return to a model where a single teacher is in charge of all subjects. Most of the primary school teachers involved taught in the humanities area, as was the case for the single middle school teacher participating in the project. None of them had previous training in interlanguage analysis or applied linguistics. They took part in the training scheme voluntarily, as part of their elective training courses or in projects for experimenting effective ways of teaching Italian to language minority children.

3.3. Collecting interlanguage samples

The first step in the training scheme consisted in making teachers aware of the concept of 'data density' (Pienemann, 1998), i.e. the fact that some communicative tasks tend to promote or require the production of certain grammatical features more than others. They needed to understand that if a structure is not produced this does not necessarily mean it should have been - in other words, 'absent' is not equivalent to 'missing'. This implies a sensitivity to the relationships between activities and linguistic structures, which are quite obvious when teachers think of grammar exercises and drills, but tend to be overlooked in the case of oral communication tasks.

Secondly, it is also important to reflect on which grammatical structures are more informative about a learner's development, i.e. what are good 'diagnostic features'. A feature with high diagnostic value is one with a relatively slow development, appearing early but continuing to be challenging even to more advanced learners. This way, a communicative task providing a number of contexts for producing that structure would be relevant for assessing learners at very

different proficiency levels, whose performance may range from not supplying the structure at all, to producing it in a few sterreotypical contexts, to using it correctly in most cases except for the most irregular / exceptional ones, to complete mastery. Based on previous research (synthesized in Giacalone Ramat, 2002), the following structures were selected as having good diagnostic value for L2 Italian.[1]

Past tense marking
A number of studies have described a developmental sequence for past tense marking in Italian interlanguage (for a review, Banfi & Bernini, 2003). The past participle marker -to emerges as the first grammatical morpheme productively attached to the verb stem, expressing a perfective meaning, typically (but not exclusively) applied to past time contexts. The auxiliaries have/be required for passato prossimo ('present perfect') are at first produced erratically, then become more consistent and their choice more target-like, with be applied to unaccusative verbs and with pre-verbal clitic object pronouns (Pallotti & Peloso, 2008). When the auxiliary be is used, the past participle must agree in person and number with its subject or object, a feature emerging relatively late in L2 Italian (Chini, 1995). Verb conjugation in the imperfect (imperfetto) always appears after the emergence of perfective marking, both in the synthetic form expressing habitual, iterative aspect (lui mangiava molto, lei leggeva tutte le sere 'he used to eat much' 'she used to read every evening') and in the compound form in conjunction with gerundive used in progressive contexts (lui stava mangiando, lei stava leggendo 'he was eating' 'she was reading').

Clitic pronouns
Italian, like other Romance languages, has a series of clitic pronouns appearing only in object positions. They vary for person, number, gender and case (direct vs. oblique). Their high frequency in the input makes them appear relatively early, often as part of unanalysed formulas, but the full system is mastered only after many years (Berretta, 1986; Giannini, 2008; Maffei, 2009).

Noun phrase agreement
In Italian all the elements of the noun phrase must agree in gender and number with the head noun, as in la casa bella ('the-f.sg. house nice-f.sg') or i

1 While choice of diagnostic structures is inherently language-specific, it is also the case that similar structures prove to have good diagnostic values in different languages. For example, agreement within and across phrases and past tense marking have been included by Bartning and Schlyter (2004) in their model of French L2 development. See also Forsberg and Bartning (this volume).

> ragazzi italiani ('the-m.pl. boys Italian-m.pl.). Some adjectives have four different endings, others only two, one for singular and one for plural, regardless of gender; there are also smaller classes with invariable adjectives and other inflectional patterns. The problem of agreement is compounded with that of noun gender assignment, which can be inferred in many cases from the noun's phonological ending, but has to be learned by heart in another substantial proportion of cases. The first signs of agreement appear very early in article-noun sequences, which might often be formulaic chunks. Agreement with other elements of the noun phrase appears later and develops gradually, with application ratios growing slowly and approaching 100% accuracy only in the most advanced learners (Chini, 1995).

It should be stressed that the choice of these tasks was partly determined by the nature of the language under investigation - other languages with different diagnostic features may require a different selection of stimuli. These tasks were found to be effective also because they could be performed - of course in quite different ways - by children of all ages and at a variety of proficiency levels, from near-beginners to highly fluent speakers, thus allowing comparisons across groups of learners.

The communicative tasks used aimed at achieving a good density of these diagnostic features, plus of course several other grammatical structures such as the marking of other verb tenses or the use of prepositions. Some tasks were selected also because they had been effectively used in previous research on children learning Italian as a second language, allowing for comparison with already existing data bases. This was the case of the picture story *Frog, where are you?* (Mayer, 1969), employed in a number of studies on Italian L1 and L2 development (e.g. Serratrice, 2007), and of the cartoon Reksio, utilized in the project *Construction du discours par des apprenants de langues, enfants et adultes* (Watorek, 2004; for L1 Italian, Giuliano, 2006). The tasks were performed by the children individually in a separate room, thus requiring the presence of a second teacher or researcher during ordinary class activities.

In the course of the three years of experimentation, and in the various locations where the training scheme was implemented, a number of different communicative tasks were piloted, some of which were also constructed by the teachers. In the following pages, only those used more consistently will be described, followed by a critical discussion about their strengths, weaknesses, possible variants and suggestions for improvement.

Free conversation

The interview begins with an ice-breaking conversation about the child, his or her family, school experiences and other similar topics. In this phase some more complex questions can be asked, eliciting decontextualised speech about family,

friends or objects having a special significance to the child. Unless the child is particularly shy or lacking linguistic resources, the interviewer should ask mostly open-ended, generic questions (like *and then?* or *What do you do every morning at school?*) or use mirroring techniques for letting the child continue without prompting him or her, i.e. by repeating parts of her previous utterances (*You said you have a little cat, how nice*) which displays active listening without leading the dialogue with questions and answers. Free conversations tend to have a relatively low data density, as the interviewer often does most of the talking by asking series of questions which may receive only minimal answers (see also Pienemann, 1998, pp. 297–303). However, in some cases they may offer relevant data for analysis and they may be used to sensitize teachers to the differences between their usual instructional conversations and tasks which allow students to perform more autonomously.

Past events narration

The introductory conversation can naturally lead to the narration of past events or states. In our protocol, this has proven to be the most effective and reliable way of eliciting past tenses. In fact, the question *What did you do last Sunday?* straightforwardly provides a number of obligatory contexts for past. The interviewer asks questions involving perfective (e.g. *what did you do last Sunday/yesterday/during the holidays?*) and imperfective aspects (*what did you use to do when you were going to the creche/nursery/when you were five years old?*). In order to elicit a variety of person markings, the conversation should not concern the child alone but also other persons such as friends or family.

A more structured variant of this task was having the children do an activity in class (e.g. preparing a fruit salad), take pictures of the various phases and using them as prompts to stimulate recall of various sub-actions with questions like *What did we do yesterday in class? Okay, we washed the fruit and then?* In this way, their productions on past tense events are more similar and can be compared.

Picture-story retelling

The child looks at pictures representing a series of events and then tells the interviewer the story. The picture-story *Frog, where are you?* (Mayer, 1969) was employed in some cases as it is clear, sufficiently long and complex to provide relatively rich data samples and can be understood by children at all ages after four. Narratives elicited with this story, especially by native or near-native speakers, tend to be rather long, which is an advantage in a research context but can pose problems in school contexts, where time for transcription is limited. For this reason another story was employed, a series of six pictures representing a father and a child going to a lake, catching a fish and taking it home; when the father is about to kill the fish with a knife the child starts crying and they go

back to set it free, but as soon as the fish is thrown into the lake it is swallowed by another bigger fish.[2] This story elicited narratives of about 100 words, which despite their brevity contained a number of interesting grammatical features, such as noun phrase agreement, clitic pronouns, person marking, prepositions. Obviously the tokens were very few, but this can also be seen as an advantage, as teachers found these stories easy to transcribe, analyse and compare. It is of course impossible to make an accurate and reliable estimate of the frequency and distribution of linguistic structures with such short texts, but they are nonetheless relevant for an analysis based on the emergence or presence of structures and for a first interpretation of their functioning.

In order to avoid use of non-verbal communication, children were told to hold the book in front of their eyes, without showing it to the interviewer or pointing to the pictures. Sometimes they would ask for a specific word, especially in the more complex Frog story. The interviewer would first prompt the child to find the solution autonomously, also using a paraphrase or a related term; if this attempt failed, the word was provided for the sake of maintaining a relaxed conversation flow.

Picture stories are effective for assessing various linguistic features, including lexical variety, syntactic complexity, expression of space and motion or participants' intentional states. As regards verb conjugation, they work well for assessing person marking, while they pose problems for time-reference evaluation. In fact, the task itself does not orient the speaker towards a particular interpretation, and narrations in the past tense - (*Once upon a time) there was a boy and he was looking at a frog* - or in the present - (*here, under my eyes) there is a boy and he is looking at a frog* - are both acceptable. In our data we observed a tendency to favour past tense narration in younger learners and present tense in older ones, with many children from 5 to 7 freely alternating present and past in the same story, thus making an analysis based on obligatory contexts impossible - in other words, with these data one can list the forms that are used, not those that are missing.

This problem might be overcome by giving interviewees an explicit prompt at the beginning, like *You must start the story with 'there was a boy who lived in a small house'*. However, these prompts don't seem to work even with adult speakers, who may end up using past tense marking in a story that is supposed to be told in the present, and vice versa (Robinson, Cadierno, & Shirai, 2009).

---

2 The story comes from the series *Vater und Sohn* by Erich Ohser (better known under the pseudonym of Plauen), published by Südverlag (www.vaterundsohn.de).

Video retelling

Following previous research by Watorek (2004), a cartoon of the series Reksio was employed to elicit narrative texts. In order to make communication more effective the interviewer was not supposed to watch the video together with the child, but left the room or attended to some other task, like writing or reading documents. With this stimulus children tended, more than with picture stories, to tell the story in the past, in the form *(In the video I've just seen) there was a boy...*, although some children also used the present tense or alternated between the two. This procedure is effective for assessing the ability to construct a coherent narrative, with appropriate reference to entities, time, space and the characters' psychological states, plus more general linguistic features like lexical choice, inflectional morphology or use of determiners and prepositions.

Spot the difference

Children were asked to describe differences between two similar pictures. This proved to be a more effective procedure than describing a single picture, which was initially piloted but eventually discarded. First of all, children found it more motivating. Furthermore, their productions tended to have more comparable sizes, as there was an optimal length given by the total number of differences to be reported, whereas free descriptions may dramatically vary in length, with some of them covering several pages of transcript. Thirdly, the task worked particularly well for eliciting complex Det-Adj-N noun phrases, because they were made communicatively necessary by the nature of the differences themselves - pictures were specifically drawn and coloured so as to contain e.g. two grey knives in one and three white knives in the other.

3.4. Transcription

Transcribing oral data was the most challenging part of the project, although it was seen as an important step for helping participants familiarise themselves with the dynamics of oral communication in the L2. Teachers were trained to use the software Soundscriber (www-personal.umich.edu/~ebreck/sscriber.html) on digital audio files and some managed to transcribe all of their data by themselves. However, most teachers' typing skills were very modest and their time for the project was limited, so that they ended up transcribing only very short texts. The six-picture story worked particularly well in this regard, as it produced only a few lines of transcript that still allowed for comparison across children and made teachers aware of the value of interlanguage analysis. This held true also for the video retelling, despite its yielding slightly longer stories. Longer tasks, like the Frog story or the description of a complex picture, often produced texts whose transcription required an amount of time beyond that which could be asked of in-service practitioners. In these cases, data were not

transcribed or they were transcribed by university students or staff hired by institutions promoting the teacher training schemes. The audio files were nonetheless archived by the schools to create portfolios for the children involved.

Transcription was kept at its simplest. All the words uttered were transcribed exactly as they were produced, including false starts, retracings and cut-off segments. Inaudible speech was represented by series of 'x', while best guesses were enclosed in parentheses. The interviewer's turns were systematically transcribed only if they contained speech, while those containing backchannels could be omitted. Pauses were marked by the symbol # or series thereof, with each token representing approximately a half second.

3.5. Analysis

Analysing data can be a very time-consuming activity, too. This is further complicated by the fact that teachers need to learn the features to focus on and appropriate ways to describe them, going beyond simple error spotting. In order to help them find regularities in interlanguage and discover its internal logic, several analytic grids were developed in the course of the project. They will be presented in the following pages, with a discussion about their merits, critical points and suggestions for further use.

The first grid (Table 1 in the Appendix) consists of a simple list of the essential areas to look at, leading to a more systematic observation than simply making disordered remarks on whatever feature meets the eye. In this table, a first basic distinction is made between communicative competence and linguistic competence. The former is further divided into efficacy (the ability to reach one's communicative goals) and fluency (the ability to do so smoothly, quickly and effortlessly). Since the main goal of the project was to focus on interlanguage, this aspect was intentionally left underspecified - further training on CEFR scales, for example, might stimulate more detailed accounts of communicative competence. The part on linguistic competence was divided into broad levels of language description, viz. noun and verb systems, syntax and the lexicon. Each of these levels contained a few sub-headings for the main categories requiring special attention - for example, in the noun system teachers were asked to systematically look at noun and adjective morphology, noun phrase structure (presence of constituents, agreement phenomena) and pronouns; in the verb system, at verb conjugation and inflection for tense, aspect and modality. They were also invited to illustrate each of their remarks with a few examples, indicating the relevant transcript line (examples of analyses based on this grid can be found in Ledda & Pallotti, 2005). It turned out that this grid worked well in training sessions, where teachers' analyses were assisted by an experienced trainer. However, when they were left on their own, most teachers

were able to write just a few minimal remarks for each heading, most of which were in the form 'uses X' or 'makes errors on Y'.

A more detailed version of the grid was thus designed including an exhaustive list of the features to be looked at (Table 2 in the Appendix). The main headings and sub-headings remain basically the same as in the previous version, with the addition of a macro-area 'textuality'. However, each area is outlined in much greater detail, with a number of characteristics to be taken into consideration and a brief introduction to the phenomena to be looked at, which served to illustrate key terms and basic processes relevant for that aspect of language. For instance, different aspects of communicative competence are described using slightly modified versions of some CEFR descriptors for Spoken fluency (Council of Europe, 2001, p. 129), Phonological control (p. 117), Qualitative aspects of spoken language use (p. 28), Sustained monologue (p. 59).[3] The section on nominal morphology recalls the basic difference between number and gender, the arbitrariness of the latter for all inanimate and most animate nouns and the existence of different inflectional classes. Turning to noun phrase construction, teachers were asked to consider various types of agreement among different constituents of the noun phrase and to look for possible systematicities and differences in singular/plural or masculine/feminine phrases, and so forth for all the other categories.

Teachers reacted positively to this second grid, as it was clearer to them what features should be focussed on and with which analytic categories. However, some of them responded to some items with a simple yes or no. For example, from the list of possible verb forms (present simple, present perfect, imperfect, subjunctive etc.) they just ticked the ones being produced. This is not wrong in itself, and the table actually invites such answers in some cases. The problem may be that in this way one might not arrive at a real understanding of an interlanguage's internal logic and rationality, which are more complex than a simple list of L2 features. In other words, if training is to be effective, these longer and more detailed checklists should not be used as inventories of items to be ticked, but should rather be seen as a memory aid for conducting a careful analysis.

---

3 CEFR scales were developed for describing adults' performance and their application to young children is not straightforward. In this project the adaptation consisted of choosing only selected descriptors from a few relevant scales, slightly modifying some of them. However, rating children on CEFR levels was not one of our goals, and the descriptors were actually given without any reference to CEFR scales and levels. For an application of CEFR scales to young learners see Papp and Salamoura (2009); see also chapters by Alanen, Huhta, and Tarnanen and Martin et al. in this volume discussing the CEFLING project on adolescents aged 12–16.

In order to stimulate a more qualitative look at interlanguage strategies, a third tool was proposed, focussing on one structure at a time (Table 3 in the Appendix). For each linguistic structure with particular diagnostic value, e.g. clitic pronouns, past tenses or noun phrase agreement, teachers were asked to select in their data examples that might help them understand how the learner is using that structure. According to emergentist / functionalist theories of language acquisition (Ellis, 2006), new rules emerge in interlanguage with a gradual spread from more prototypical cases, which may be learnt as lexicalized phrases or chunks, to slightly more abstract patterns, based on relatively simple generalizations on frequently occurring form-function mappings, to more abstract and complex patterns, incorporating a variety of features at the same time, and allowing for exceptions and irregularities (see also Martin, Mustonen, Reiman, & Seilonen, this volume). For example, verb marking for imperfective aspect will first appear on prototypical verbs encoding states, like *be* or *have,* then spread to activities and only finally to verbs expressing punctual notions, where speakers display their skill by producing unusual combinations of verb actionality and aspectual marking (Andersen & Shirai, 1996; Giacalone Ramat, 2002).

Hence, not all tokens of a grammatical structure are equal, some of them displaying (or allowing to infer) more proficiency than others. Similarly, not all errors are equivalent - some may indicate complete ignorance of the structure while others may be due to imperfect, partial or not completely automatized knowledge. This third analytical tool tries to capture this state of affairs, asking teachers to note down examples displaying complete or partial lack of knowledge, or examples indicating general knowledge but problems with irregular, unusual, complex cases, or examples where application to such cases may lead to the conclusion that the structure has been thoroughly acquired.

This approach allows one to zoom into specific structures and to assess their acquisition one by one. The same level of detail may be applied to a quantitative analysis, which then becomes very similar to those used in many SLA research projects, as teachers are requested to focus on one structure at a time and to count its various target- and non-target like forms. Table 4 in the Appendix exemplifies this type of grid with *passato prossimo* ('present perfect') in Italian. The first four lines are used to score various types of correct realizations, ranked in a tentative order of difficulty from the easiest ones, involving the unmarked participle ending *-to*, to more complex cases of participles inflected for gender and number or irregular participles (although acquisition of some of the latter may be facilitated by their high frequency). Subsequent lines classify various types of forms deviating from L2 norms, in what is evidently a blend of interlanguage-based (what forms are produced and what interlanguage rules are followed) and accuracy-based (the correspondence of these forms to L2 rules) descriptions.

Despite its seeming complexity, teachers found this table rather easy to compile, as they were guided by the pre-formulated descriptions, needing only to assign each token to one of the categories. This grid allows a fine-grained distributional analysis: rather than just counting correct and wrong forms, one can obtain a picture of what types of grammatical strategies are being employed, which may be revealing of developmental levels and acquisitional paths. Furthermore, one can obtain distribution ratios, expressed as percentages, which can be used to analyse data both longitudinally and cross-sectionally.

A problem with this procedure is that compilation can be rather time-consuming: transcripts must be searched for tokens of the target structure, which are then scored in the appropriate lines. While this scoring turned out to be rather fast, as in most cases it did not involve complex decisions or interpretations, it nonetheless required a careful look at transcripts for each diagnostic feature - hence the need to limit these to not more than two or three. A second problem is that the number of tokens for each category tends to be rather small, unless substantial data samples are collected. Quantitative results that can be obtained from this table should thus be interpreted cautiously due to their limited reliability. However, one should also bear in mind that the purpose of this analysis is to ascertain the presence, absence and logic of certain interlanguage strategies, rather than producing generalizable quantitative statements about L2 development.

4. Implications for teaching and testing

A set of procedures has been presented which were piloted to analyse children's interlanguage in a variety of school settings, as part of in-service teacher training schemes. The lessons that can be gained from this pilot study, and which may be extended to other contexts, can be grouped under two main headings - implications for teaching and implications for testing.

As regards teaching, the scheme had a positive impact on the participants' professional development. Teachers were overall satisfied, with some of them even enthusiastically reporting that this close attention to their students' productions radically changed their attitudes and practices. This had an impact on the teaching of Italian to native speakers as well, both in multilingual and in monolingual classes - analysing what they knew and what they did not, and showing that differences between them and non-native speakers were often limited to just a few areas, helped teachers reconceptualize many aspects of language education in multilingual classrooms. Traditional activities based on classifying and labelling linguistic structures gave way to more functionally-oriented ones, focussing on the areas that careful observation showed to be weaker.

This led to using an active approach to language education, which most of the time took the form of cooperative learning with mixed-level groups. Some of these innovative didactic activities were published on the Web as resources to share with colleagues (www.comune.re.it/interlingua). Thus interlanguage analysis has an important role to play as part of teacher training, both pre-service and in-service, as it leads to more learner- centred activities and to the realisation that any effective pedagogic intervention should start from understanding learners' competences, strategies and processes.

The value of interlanguage analysis for formative and diagnostic assessment is thus undeniable. In this context, issues of reliability and data robustness become less crucial - the logic of a child's interlanguage can be inferred, or at least acknowledged, even with small speech samples, and the very act of transcribing and analysing a few lines may already promote such a change of attitude. If time is at a premium, accurate analysis can be conducted only on those children who need more careful monitoring, e.g. newcomers or those with special difficulties. For these and all the others, collected data (digitised oral productions, written texts etc.) can in any case be seen as part of a portfolio documenting individual learning paths, regardless of whether and how they are accurately transcribed and systematically analysed.

A second area for which this study may be relevant is language testing. Other chapters in this volume report on research aimed at matching the development of communicative proficiency with profiles of second language acquisition. The problem arises of how such linguistic profiling can be practically incorporated into language testing.

It is unlikely that procedures like the ones discussed above may be directly used in large-scale, standardised testing. Firstly, they present all the problems (but also the strengths) associated with performance assessments, i.e. assessments in which '(a) examinees must perform tasks, (b) the tasks should be as authentic as possible, and (c) success or failure in the outcome of tasks, because they are performances, must usually be rated by qualified judges' (Norris et al., 1998, p. 8). In the case at hand, these judges should have a very special type of qualification, i.e. the ability to analyse interlanguages, which requires extensive training. Furthermore, an interlanguage analysis as is usually done by SLA researchers, or even in the simplified adaptation exposed in previous pages, takes a considerable amount of time.

If linguistic profiling is to be incorporated into large-scale language assessment, less time-consuming alternatives must be devised. A possibility would be using written data, which makes transcription unnecessary. However, the equivalence of interlanguage production in the oral and written mode has to be demonstrated, not assumed. Alternatively, raters can be instructed to assess oral language performance 'on the fly', while listening to it directly or from a record-

ing. The task can be assisted by the computer, as with the Rapid Profile software, developed by M. Pienemann and associates (http://groups.uni-paderborn.de/rapidprofile/), which provides an interface where the rater can input scores for a small set of diagnostic features. An even simpler solution would be having raters formulate holistic judgements on interlanguage structures, based on a checklist like the ones presented above but applied directly to learners' productions without transcribing them. The checklist would probably be formulated in terms of a scale, with different descriptors ordered in a series of levels, corresponding to a developmental sequence. Such rating scales could concern individual structures (e.g. tense marking, noun-phrase agreement, articles) or group several grammatical structures to provide a global picture of interlanguage development at various levels (but this rests on the assumption that it is indeed possible to identify relatively stable 'levels' of interlanguage development comprising a number of features; see Bartning & Schlyter, 2004, for such an attempt).

Whatever choice is eventually made regarding the implementation of interlanguage analysis into large-scale testing, it is essential that raters have a sound understanding of what an interlanguage is, how it works and how it should be analysed. The procedure presented here may offer some ideas for their training.

First of all, raters need to learn to separate the two areas of communicative proficiency and linguistic development. While it is true that the two dimensions are often related and grow side-by-side, there is also a considerable degree of independence, so that the two constructs are separated in most models of language proficiency. The first thing that raters need to learn is to keep the two dimensions apart, at least for analytic purposes. They should also be aware that constructs such as 'error compromising/not compromising communication' are spurious, in that they mingle a linguistic dimension (accuracy) with a communicative one (adequacy) (see also Kuiken, Vedder, & Gilabert, this volume).

A second aspect that needs to be discussed in a training scheme for raters is the difference between linguistic development and accuracy. Teachers and testers alike are frequently prone to the 'comparative fallacy', which entails describing interlanguage development in terms of errors and conformity to L2 norms. Especially in the initial-intermediate stages of acquisition, it makes little or no sense to count errors and other deviations from L2 norms, while it is more productive to recognise that an interlanguage has reached a complexity and sophistication level higher than another, something that can be quite unrelated to the number of errors produced. In order to do so, raters need to become aware of how interlanguages develop over time in order to express more complex grammatical functions, and the grids presented in this chapter may help them achieve such an awareness. The importance of transcribing oral data should not be overlooked in this respect, as it is an effective way of focussing one's attention on important details of linguistic production, including the use of prosody for marking infor-

mation structure at the phrase, sentence or text level, or the subtle interactions of phonology and morphology in word endings. Once this understanding of interlanguages as autonomous systems is firmly consolidated, one may discuss whether accuracy-based analyses could be used (though not exclusively) for characterising intermediate-advanced varieties, as Norris and Ortega (2003) suggest. A grid like the fourth presented in this chapter points to this direction, as it allows one to recognise the presence and distribution of both a variety of interlanguage forms and of structures conforming to L2 norms.

Learning to perform an adequate interlanguage analysis takes a substantial amount of time, based on the experience reported here. While most teachers involved in the project reacted very positively to the 'new' way of looking at their pupils' productions, when demonstrated by an experience trainer with an academic background, they encountered some difficulties in doing the analysis themselves. The main reason is that although most of them taught Italian as their main subject, they lacked an up-to-date and scientifically appropriate metalinguistic terminology. A relevant proportion of linguistic education in Italy deals with metalinguistic description, but this is done with a multitude of traditional categories - some of them misleading or ill-founded, some no longer in use in contemporary linguistics - while other crucial ones are missing, including *aspect, determiner, morpheme, phrase*. More importantly, traditional linguistic analysis at school consists in the mechanical application of metalinguistic labels (such as 'abstract noun', 'present perfect', 'concessive clause') to written texts, with little or no understanding of the general mechanisms responsible for the production of linguistic structures. Traditional labels may work relatively well when applied to the description of standard written Italian, but they fail when other languages or varieties are to be described. In such cases, concepts from general and functional-typological linguistics are essential, because they allow one to focus on linguistic processes, on how language works, rather than simply classifying individual items in a sentence. In other words, what teachers (and probably most raters) lack is an understanding of how their own language works, let alone others, including interlanguages. Their metalinguistic awareness is limited to a set of labels plus an ordinary native speaker's sensitivity to ungrammatical constructions, with a very limited capacity to explain why they are ungrammatical or what the logic behind grammaticality is. What needs to be stimulated is thus a different attitude towards language data, based on reasoning and understanding rather than on mere tagging and classifying.

Without such an attitude and the associated analytical competence, extensive or exclusive reliance on rating scales, even if based on SLA research findings, might prove to be limited or even misleading. In our experimentation, teachers were not given scales with level descriptors for different stages of interlanguage development. In fact, on the few occasions in which teachers received

CEFR scales for communicative competence, we noted that they were very happy with assigning learners to such prefabricated scales, which they did not take to be so different from traditional grading scales - they would say 'she is at B1' instead of 'she scored 6/10'. This however distracted them from the real objective, which was understanding students' linguistic strategies and their interlanguages' logic. Prefabricated descriptor scales might thus be used at some point in teachers' training - e.g. at the start for sensitizing them to the existence of 'typical' linguistic configurations at different developmental levels, or at the end as checklists - but their use should be limited. The same holds true for raters' training. While, for practical reasons, in their professional activity they may end up using prefabricated descriptor scales containing typical traits of different levels of interlanguage development, it is important that they reach a sound understanding of how interlanguages work in order to be able to apply such scales meaningfully.

References

Andersen, R., & Shirai, Y. (1996). The primacy of aspect in first and second language acquisition: The pidgin-creole connection. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 527–569). San Diego: Academic Press.

Bailey, A. (2008). Assessing the language of young learners. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of Language and Education: Vol. 7. Language testing and assessment* (2nd ed.) (pp. 273–284). New York: Springer.

Banfi, E., & Bernini, G. (2003). Il verbo. In A. Giacalone Ramat (Ed.), *Verso l'italiano: Percorsi e strategie di acquisizione* (pp. 70–115). Rome: Carocci.

Bartning, I., & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French Language Studies, 14,* 281–299.

Berretta, M. (1986). Per uno studio dell'apprendimento dell'italiano in contesto naturale: Il caso dei pronomi atoni. In A. Giacalone Ramat (Ed.), *L'apprendimento spontaneo di una seconda lingua* (pp. 329–352). Bologna: Il Mulino.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33,* 1–17.

Cameron, L. (2001). Assessment and language learning. In L. Cameron, *Teaching languages to young learners.* Cambridge: Cambridge University Press.

Carpenter, K., Fujii. N., & Kataoka, H. (1995). An oral interview procedure for assessing second language abilities in children. *Language Testing, 12,* 157–181

Chini, M. (1995). *Genere grammaticale e acquisizione.* Milan: Angeli.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching, 5,* 161–170.

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

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence'. *Language Testing, 21,* 259–283.

Ellis, N. (2006). Cognitive perspectives on SLA. *AILA Review, 19,* 100–121.

Gass, S., & Mackey, A. (2007). *Data elicitation for second and foreign language research.* Mahwah: Lawrence Erlbaum.

Giacalone Ramat, A. (2002). How do learners acquire the classical three categories of temporality? Evidence from L2 Italian. In R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense–aspect morphology* (pp. 221–248). Amsterdam: Benjamins.

Giannini, S. (2008). Proprietà formali e distribuzionali dei clitici in Italiano L2. In R. Lazzeroni, E. Banfi, G. Bernini, M. Chini, & G. Marotta (Eds.), *Diachronica et synchronica. Studi in onore di Anna Giacalone Ramat* (pp. 231–253). Pisa: ETS.

Giuliano, P. (2006). *Abilità narrativa ed emarginazione sociale.* Napoli: Liguori.

Hasselgreen, A. (2003). *Bergen 'Can Do' project.* Strasbourg: Council of Europe. Retrieved from http://www.ecml.at/documents/pub221E2003_Hasselgreen.pdf

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing, 22,* 337–354.

Hendriks, H. (2008). Presenting the English Profile Programme: In search of criterial features. *Research Notes, 33,* 7–10.

Ioannou-Georgiou, S., & Pavlou, P. (2003). *Assessing young learners.* Oxford: Oxford University Press.

James, C. (1998). *Errors in language learning and use: Exploring error analysis.* Harlow: Longman.

Klein, W., & Perdue, C. (1992). *Utterance structure. Developing grammars again.* Amsterdam: Benjamins.

Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research, 13,* 301–348.

Ledda, F., & Pallotti, G. (2005). L'interlingua. In G. Pallotti & AIPI (Eds.), *Insegnare e imparare l'italiano come seconda lingua.* Rome: Bonacci.

Maffei, S. (2009). Osservazioni sullo sviluppo dei pronomi personali. In M. Palermo (Ed.), *Percorsi e strategie di apprendimento dell'italiano lingua seconda* (pp. 103–119). Perugia: Guerra.

Mayer, M. (1969). *Frog, where are you?* New York: Puffin.

McDaniel, D., McKee, C., & Smith Cairns, H. (1996). *Methods for assessing children's syntax.* Cambridge: MIT Press.

McKay, P. (2005). Research into the assessment of school-age language learners. *Annual Review of Applied Linguistics,* **25***,* 243–263.

McKay, P. (2006). *Assessing young language learners.* Cambridge: Cambridge University Press.

Norris, J., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments.* Honolulu: Second Language Teaching and Curriculum Center.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 717–761).

Oxford: Blackwell.

Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics, 28,* 361–382.

Pallotti, G., & Peloso, A. (2008). Acquisition sequences and definition of linguistic categories. *The Open Applied Linguistics Journal, 1,* 77–89.

Papp, S., & Salamoura, A. (2009). An exploratory study into linking young learners' examinations to the CEFR. *Research Notes, 37,* 15–22.

Perdue, C. (1993). *Adult language acquisition: Cross-linguistic perspectives.* Cambridge: Cambridge University Press.

Philp, J., Oliver, R., & Mackey, A. (2008). *Second language acquisition and the younger learner: Child's play?* Amsterdam: Benjamins.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.

Rea-Dickins, P. (Ed.). (2000). [Thematic issue on assessing young learners]. *Language Testing, 17*(2).

Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education: Vol. 7. Language testing and assessment* (2nd ed.) (pp. 257–271). New York: Springer.

Rea-Dickins, P., & Rixon, S. (1997). The assessment of young learners of English as a foreign language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of Language and Education: Vol. 7. Language testing and assessment* (pp. 151–161). Dordrecht: Kluwer.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics, 30,* 533–554.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10,* 209–231.

Serratrice, L. (2007). Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics, 39,* 1058–1087.

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–409). San Diego: Academic Press.

Taylor, L., & Maycock, L. (Eds.). (2007). [Thematic issue on Cambridge Young Learners English (YLE) tests]. *Research Notes*, 28.

Valdez Pierce, L., & O'Malley, J. M. (1992). Performance and portfolio assessment for language minority students. *NCBE Program Information Guide Series*, 9, Spring.

Watorek, M. (2004). Construction du discours par des apprenants de langues, enfants et adultes. *Acquisition et Interaction en Langue Étrangère, 20,* 130–171.

Williams, C. (2008). Challenges to parsing English text: The language of non-native speakers. *Research Notes, 33,* 10–15.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity.* Honolulu: Second Language Teaching and Curriculum Center.

APPENDIX

Table 1. Interlanguage observation grid

Communicative competence	*Linguistic competence*			
	Noun system	**Verb system**	**Syntax**	**Lexicon**
Communicative efficacy	Noun and adjective morphology	Verb conjugation	Formulas	Variety, richness
Fluency	Noun phrase construction	Verb tense, aspect and mood	Word order in different types of constructions	Communication strategies
	Pronouns			

Table 2. List of descriptors and topics for the systematic observation of interlanguage*

What aspects are systematic? What regularities emerge? What **can** children do? This is not a list of items to be ticked with a simple yes or no, but a guide for systematic observation and analysis.

Communicative competence
Fluency
Does the learner express him- or herself easily, fluently, effortlessly?
- Can manage very short, isolated, mainly pre-packaged utterances, usually stimulated by teacher's prompts.
[...]
- Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of speech.
- Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.
[...]
- Has acquired a clear, natural, pronunciation and intonation.
Communicative efficacy
Can the learner convey ideas effectively? Achieve the goals he/she aims at? Avoid misunderstandings?
In conversations
- Can communicate with a few words and memorized patterns.
[...]
- Is entirely fluent in interaction, being able to manage it effectively.
In stories and descriptions
- Can tell a story or describe something in a simple list of points
[...]
- Can give clear, detailed descriptions of complex subjects.

>>>

* Due to space limitations, the original table (Gabriele Pallotti – Stefania Ferrari) has been abridged by omitting some descriptors of communicative competence.

Linguistic competence

NOUN SYSTEM

Noun and adjective morphology

Observe how nouns are inflected for gender (masculine and feminine) and number (singular and plural). Recall that number inflection has a meaning (it depends on the number of referents being talked about) while gender inflection is almost always arbitrary and must be learned by heart (what is masculine in the sun and feminine in the moon? In German the exact contrary is true).

Nouns ending in -e give special problems as they can be both masculine and feminine.

- Singular nouns: masculine and feminine?
- Plural nouns: masculine and feminine?
- Gender of nouns ending in -e?

Noun phrase construction

How is gender and number agreement marked? What elements - e.g. articles, demonstratives, possessives, adjectives - contribute to forming noun phrases, as in *i bambini intelligenti, le ragazze simpatiche, il cerchio giallo, la tazza rossa*?

Note agreement between article and noun (i*l bambino, i coltelli*), noun and adjective (*bambino allegro, coltelli gialli*) and article, noun and adjective (*il bambino allegro, i coltelli gialli*).

Several types of determiners exist beside the article: quantifiers (*qualche matita, molti colori*), numerals (*tre, cinque*), possessives (*il suo zaino, le loro borse*), demonstratives (*questa ragazza, quel libro*).

- Article/noun agreement
- Article noun/adjective

[...]

- Agreement in singular phrases
- Agreement in plural phrases

Are demonstratives used?

Are possessives used?

Pronouns

What pronominal forms are used? Note both free pronouns (*io, tu, lui, lei, noi*...) and clitics, which can express a direct (*me, te, lo, la, li*) or indirect object (*mi, ti, gli, le, ci, vi, gli*).

Also note if there are combined pronouns (*glielo, ce li, me la*) and clitics' position with respect to the verb (sometimes you may hear *io prendoli, voglio lo vedere*).

Finally, note clitic usage typical of substandard Italian: *a lei gli/ci dico.*

- Presence and usage of free pronouns
- Presence and usage of direct object clitics
- Presence and usage of indirect object clitics
- Combined pronouns
- Pronouns' position

>>>

VERB SYSTEM

Verb conjugation

How are different persons expressed? With one fixed form, with several forms or with the entire paradigm?

- Are verbs inflected? How?
- Some persons
- All persons (required by communicative demands)

Verb's tense, aspect and mood

How are notions of tense, aspect and mood expressed? What tenses, aspects and moods of the Italian verb system are used?

- Present
- Imperative
- Past participle
- Present perfect
- Imperfect
- Conditional
- Future
- Subjunctive
- Gerund
- *Stare* + gerund (progressive)
- Simple past

SYNTAX

Formulas

Are fixed formulas used, i.e. sentence chunks memorized as if they were a single word (e.g. *come si chiama? come stai? non ce l'ho, dammi, non lo so*)? Number, appropriateness, variety.

Negation

- No + X. (*no mangiare questo, no io così, no pane*)
- Non + X (*non mangio questo, io non faccio così, non c'è il pane*)
- Non ... mica, neanche ... (*non ha mica detto così, non ha neanche un soldo*)
- With indefinites (*niente, nessuno* ...)

Word order in different types of construction

How are sentences constructed? According to a canonical word order subject-verb-complement or with more complex orders? Observe for example:

- Post-verbal subject (*è arrivato Mario, sono caduti loro, si è spenta la luce*)
- Dislocations (*il libro non l'ho visto; non l'ho visto, il libro; a Roma ci sono già stato*)

>>>

Subordination

Are subordinate clauses produced? Which ones?

- Simpler ones (causal, temporal, final)
- More complex ones (relative, hypothetical, concessive) (if communicative situation requires them)

TEXT ORGANIZATION

How are sentences and parts of text connected?

- use of temporal (*poi, allora, dopo, mentre, alla fine*), argumentative (*però, invece, eppure*), meta-textual (*insomma, e tutto questo..., in poche parole*) connectives
- cohesion across different text parts, marked by pronouns and other pro-forms (*questo lo faccio solo la domenica*)

LEXICON

Variety, richness

Is the lexicon varied? Are terms used appropriately and precisely?

- Has a very basic repertoire of simple expressions for giving personal information and satisfying concrete needs
- Can use basic structures and memorized expressions or groups of few words to speak of him/herself or other persons, about ordinary actions, places and objects owned
- Controls sufficient language structures and lexicon to express him/herself, with some hesitations or circumlocution
- Can express him/herself clearly and briefly, but in a communicatively appropriate way, about everyday topics
- Has a rich linguistic repertoire, including a wide range of specific and appropriate terms, which can vary for style and register

Communication strategies to fill lexical gaps

Are particular communication strategies used to compensate for lack of specific terms?

- Repetition
- Reformulation/paraphrase (*la casa delle api, l'animale che salta*)
- Lexical invention (*il camionaio, matrimoniare*).
- Request for clarification/teacher's help
- Other

Table 3. Qualitative analysis of one structure (e.g. present perfect)

Examples displaying lack of knowledge	
Examples displaying difficulties and uncertainty	
Examples displaying knowledge	
Examples displaying general knowledge but problems in irregular/complex cases	
Examples displaying excellent knowledge, proficiency	
Examples of probably formulaic uses	

Table 4. Passato prossimo (present perfect)*

1) Read the transcript and underline in red all tokens of present perfect. When a form is not correct, besides underlining it add a cross next to it.
2) Use the 'score' column to mark a line every time you see the type of structure described in that line; when you reach five lines, draw a line across them to facilitate counting. At the end, you will report the sum in the 'total' column. If an example contains more than one error, score more than one line in different columns. For example, if the child says 'noi ha arrivato' you will score one on the line 'auxiliary *have* instead of *be*', one on the line 'no subject-auxiliary agreement' (noi ha) and one on the line 'no subject-participle agreement' (noi … arrivato).
   Examples should be provided for some of the errors, for structures demonstrating good knowledge, for self-corrections and for possible doubts.
   Don't score cases in which the form is repeated immediately after being uttered by another person.
   If the transcript is very long you can carry out quantitative analysis only on one part (but beginning and end should be clearly marked).

Passato prossimo	*Examples*	*Score*	*Total*
Corretti con participio –to (abbiamo mangiato, sono arrivato, è rimasto) *Correct with -to participle* ('we have eaten, I have arrived, he has remained')			
Corretti con participio –ta, -te, -ti (è tornata, siamo stati, l'ho vista) *Correct with –ta, -te, -ti participle* ('she has			
returned, we have been, I have seen her')			
Correct with irregular participle			
Correct with pluperfect			
Ausiliare + verbo non passato (ha mangia, sono torno) *Auxiliary + non past verb* (he has eat, I have return)			
*Wrong auxiliary choice (*to have *instead of* to be)			
*Wrong auxiliary choice (*to be *instead of* to have)			
Lack of auxiliary have			
Lack of auxiliary be			
No agreement between subject and past participle			
No agreement between subject and auxiliary			
Other non-standard uses(e.g. analogic constructions on irregular verbs)			
Doubtful, unclassifiable or uninterpretable cases			

* Due to space limitations, the original table has been abridged by giving the Italian original wording and examples in a few cells only.