Sharon Unsworth

Department of English

Utrecht University

Trans 10

3512JK Utrecht

The Netherlands

Tel.no. +31 30 2536014

Fax. no. +31 30 2536000

Email: sharon.unsworth@let.uu.nl

Title: Comparing child L2 development with adult L2 development: How to measure L2 proficiency

Running title: Comparing child L2 with adult L2: How to measure L2 proficiency

To appear in Gavruseva, E. and Haznedar, B. (eds.) *Current Trends in Child Second Language Acquisition* Amsterdam: John Benjamins

Comparing child L2 development with adult L2 development:

How to measure L2 proficiency·

Sharon Unsworth (Utrecht University)

Abstract

Most studies investigating the role of age in L2 acquisition compare L2 children and adults in terms of ultimate attainment (e.g. Johnson and Newport 1989). This paper addresses some of the methodological and conceptual issues which arise when, following Schwartz (1992), such a comparison is made in terms of development. The paper considers three variables which need to be carefully controlled for in any such comparison: L1 transfer, age at time of testing and L2 proficiency. A new proficiency measure, the Age-Sensitive COmposite Proficiency Score (ASCOPS), is presented. Unlike previous measures, ASCOPS can be used with both L2 children and adults, and it takes into account the intrinsic differences between these two groups relating to their age at time of testing.

1. Introduction

Most researchers would agree that there are age effects in non-native (L2) acquisition: whereas L2 children usually become nativelike, L2 adults, on the whole, fall short. The significance – theoretical or otherwise – which one attaches to this observation will of course depend on one's general approach. It has been argued that child L2 acquisition has the potential to inform both native (L1) and adult L2 acquisition (Lakshmanan 1995; Schwartz 1992; 2003; 2004). For example, Schwartz (1992) proposes that a comparison of child L2 development with adult L2 development can inform the debate on the role of domain-specific principles (UG) in adult L2 acquisition. The purpose of this paper is not to evaluate the various accounts put forward in this domain, nor to determine the extent to which the aforementioned observation holds. Instead, the focus here is methodological in nature. This paper focuses on such a child L2 ~ adult L2 comparison. Rather than specifically investigating whether there is evidence for or against Schwartz's (1992) hypothesis, the goal of the present paper is to explore some of the methodological and conceptual challenges which child L2 ~ adult L2 comparisons involve.

L2 children and adults may differ with respect to a number of variables, including age, the L1, and L2 proficiency. Controlling for L2 proficiency presents a number of problems. For example, when the L2 children are still children at the time of testing, the age differences between the two learner groups will impose certain constraints on the types of methods which can be employed. The first part of this paper consists of a critical evaluation of some of these methods, as well as a discussion of some of the factors involved in selecting an appropriate proficiency measure. The second part presents a newly developed proficiency measure, the Age-Sensitive Composite Proficiency Score (ASCOPS). Based on semi-spontaneous speech data, this measure combines complexity and accuracy scores in the domains of morphosyntax and the lexicon while taking into account some of the inherent differences between children and adults.

The paper is organised as follows. The rationale for comparing L2 children and adults is outlined in more detail in § 2. Sections 3 through 5 consider certain variables which need to be controlled for in any child L2 ~ adult L2 comparison, namely L1 transfer, age at time of testing and L2 proficiency. Section 5 reviews some of the available proficiency measures in detail and evaluates their suitability for cross-age comparisons. A selection of these measures are combined in § 6 to form the Age-Sensitive Composite Proficiency Score. Finally, an example of a child L2 ~ adult L2 study using this score is presented in § 7.

2. Why compare child L2 with adult L2?

One of the perennial debates in L2 acquisition research concerns the question of whether non-native acquirers (L2ers) make use of the same language acquisition mechanism(s) as L1 children, and more specifically, whether adult L2ers develop the same kind of linguistic knowledge that native speakers have at their disposal. Many of the studies addressing this issue compare L2 adults with L2 children to determine whether there is evidence for a critical period in L2 acquisition. In its most theory-neutral form, the critical period hypothesis, originally proposed for L1 acquisition by Penfield and Roberts (1959) and Lenneberg (1967), states that 'there is a limited developmental period during which it is possible to acquire a language, be it L1 or L2, to normal, nativelike levels' (Birdsong 1999:1). As this definition suggests, most of the comparisons of L2 children and L2 adults focus on whether these two groups reach the same level of ultimate attainment, or whether, as a result of biological (e.g. Pulvermüller and Schumann 1994) and/or cognitive (Felix 1985; Newport 1990) and/or sociological (Krashen 1982; Schumann 1975) factors, the acquisition of an L2 as an adult is fundamentally different (Bley-Vroman 1989) from the acquisition of an L2 as a child.

Probably the most cited study investigating the critical period hypothesis in L2 acquisition is Johnson and Newport (1989). This study tested L2 child and adult learners with either Chinese or Korean as L1 on a number of morphosyntactic properties of English. The authors observe a significant negative correlation between age of first exposure and task score for L2ers who were first exposed to English before age 16, whereas amongst post-pubertal L2ers, no such relationship was found. This, they claim, is evidence for a critical period in L2 acquisition. Although Johnson and Newport's investigation has since been criticised (Bialystok and Hakuta 1994; Bialystok 1997; Kellerman 1995) and in replications and reanalyses their claims have been both confirmed (e.g. DeKeyser 2000; Johnson and Newport 1991) and disconfirmed (e.g. Bialystok and Miller 1999; Birdsong and Molis 2001), most researchers would still concur that, either on the basis of this and/or comparable studies, L2 children have a long-term advantage over L2 adults.

The experimental set-up of Johnson and Newport (1989) is typical of a critical period study (see Hyltenstam and Abrahamsson 2003 for a recent overview): a group of L2ers, whose age of first exposure ranges from child to adult, are tested on a series of TL properties to determine whether they fall within native-speaker range. Subsequently, correlation co-efficients are calculated to ascertain whether there is a(n inverse) relationship between age of first exposure and ultimate attainment. In such studies, the L2 children are almost always adults at time of testing.

Another approach to age effects in L2 acquisition, frequently adopted in earlier work (see e.g. Clahsen 1986; Clahsen 1988; Dulay and Burt 1974; Ervin-Tripp 1978; Meisel, Clahsen and Pienemann 1981) but somewhat neglected more recently, is to examine the development sequences which learners take rather than the level of ultimate attainment which they reach As Schwartz (to appear:2) notes, typical critical period studies, such as those mentioned above, are 'in principle disinterested' in development. It is, however, only by examining developmental data that we can understand the processes underlying child L2 acquisition. After all, when examining child L2 *development*, the L2 children are children at time of testing.

Following a proposal put forward in earlier work (Schwartz 1992), Schwartz (to appear) argues that it is a comparison between child L2 and adult L2 *development* – rather than ultimate attainment – which will allow us to determine whether these two types of acquisition are of the same epistemological type. More specifically, she claims that by comparing developmental sequence data from these two groups, it is possible to decide between a UG-based approach to adult L2 acquisition and a general problem-solving approach. Her argument works as follows: assuming that child L2 acquisition is driven by UG (based on evidence from studies such as the Johnson and Newport study that, in terms of ultimate attainment, young L2ers are generally more successful than adult L2ers), then comparing developmental sequences of child L2ers with those of adult L2ers, while holding the L1 constant, will provide evidence for or against UG involvement in L2 acquisition.[1] If L2 children and L2 adults pass through the same developmental sequences, then this would constitute evidence for UG involvement in (adult) L2 acquisition, whereas if the L2 child and L2 adult developmental sequences pattern differently, this would provide evidence against UG constraining (adult) L2 acquisition.[2]

The purpose of this paper is not to present data for or against this approach (see e.g. Gilkerson 2005); rather, it will explore some of the methodological and conceptual issues involved in making a comparison between child L2 development and adult L2 development. As noted above, focussing on child L2 *development* requires testing L2 children when they are children. This introduces an important difference between L2 children and adults, which, as this paper will demonstrate, has far-reaching consequences for the nature of the comparison between the two groups. These two groups of learners may also differ from each other with respect to a number of other variables, including L1 knowledge and L2 proficiency level. Any study seeking to determine whether L2 children and adults behave similarly or differently in their acquisition of a given TL property – based on Schwartz's rationale or otherwise – must ensure that, if a valid comparison is to be guaranteed, these variables are carefully controlled for. The remainder of the paper considers each of these variables in turn, focussing on how to control for comparability in terms of L2 proficiency in particular.

3. L1 transfer

There is accumulating evidence that, as in adult L2 acquisition, the start of the child L2 acquisition process is characterised by L1 transfer. One of the clearest examples of this comes from Haznedar's (1997) study of a Turkish-speaking child acquiring English. During approximately the first two-and-a-half months of data collection, Haznedar observes that the child's verbal utterances are predominantly XV, that is, head-final, as in the L1, Turkish, whereas after this point, the targetlike VX order prevails. Further evidence of L1 transfer in child L2 acquisition can be found in, for example, Haberzettl (1999) and Whong-Barr and Schwartz (2002).

In a comparison of L2 children and L2 adults, it is relatively simple to control for L1 transfer, namely by holding the L1 constant across the two groups, as proposed by Schwartz (1992). In this way, the effect of the L1 is neutralised as it should be the same for both groups. Of course, given that transfer is most evident at the initial stages of development, it is also important that the L2 children and adults are also at a comparable stage in their L2 development or proficiency. How to achieve this is dealt with in § 5.

4. Age at time of testing

The age of an L2er can be calculated either for the point of first exposure to the TL and/or for the time at which data collection takes place. Child L2ers differ from adult L2ers by definition with respect to age at first exposure. When investigating L2 development, as is the case here, the two groups will also necessarily differ in terms of age at time of testing. This will have important consequences for the types of tasks which may be employed.

One early study which compared L2 children and adults developmentally is Snow and Hoefnagel-Höhle's (1982) investigation of the L2 acquisition of Dutch by English-speaking children and adults. They employed a variety of production and comprehension tasks to subjects who ranged from age 3 to adult at time of testing on three different occasions. The much-cited conclusion of this study is that the older L2ers had the advantage over the younger L2ers. In their battery of tests, however, Snow and Hoefnagel-Höhle included tasks which arguably favoured the older subjects. For example, in a sentence repetition task and a sentence translation task, subjects were asked to either imitate or translate sentences of increasing length (2 to 10 words in the repetition task) and grammatical complexity, which were read aloud to them. It is quite possible that the older subjects' superior performance on this task resulted from their relative cognitive maturity (e.g. greater working memory capacity) rather than their higher L2 proficiency.

The problems with Snow and Hoefnagel-Höhle's (1982) study highlight an important methodological point which should be taken into consideration for any study testing child and adult L2ers when the child L2ers are children at the time of testing: the chosen experimental methods should not be too cognitively challenging for the younger subjects, nor, on the other hand, should they be too easy for the older subjects. This means that in many cases, they should not involve reading or writing, because, assuming the children/adults are learning to read and write in the TL, children will often not be as advanced as adults. In addition to differences in cognitive capacities, such as working memory, the child and adult groups may also differ in the amount of metalinguistic knowledge they have at their disposal (Bialystok 1993). Older subjects, if they have had language instruction and perhaps even if they have not, will have metalinguistic knowledge which they could potentially employ in certain data collection procedures.

Striking a balance between L2 children and adults in these terms, trying to ensure that one group is not favoured over the other, is a challenge. Two possible ways in which this challenge could be countered are as follows. The first option would be to tailor a particular task to the different (cognitive) abilities of each age group, i.e. use different versions of the task with the different age groups. Although this would address the problem, using different tasks with the different age groups (at least without careful piloting and validation) could make results difficult to interpret as it introduces another variable to the comparison. Disparate outcomes for the different age groups could potentially be due to the differences in the task as well as any differences in the subjects' interlanguage grammars. Another option would be to ensure that the TL property which is being tested is sufficiently sophisticated to ensure that metalinguistic knowledge cannot be applied. If this is the case, using an 'easy task' with L2 adults will be less problematic.

These issues are explored in more detail with respect to proficiency tasks in the following section.

5. L2 proficiency

L2 children and adults may also differ in terms of their L2 proficiency. Before considering how the two groups can be compared in terms of proficiency in § 5.2, this construct is first defined in § 5.1. Subsequent sections deal with various ways of measuring proficiency (§ 5.3) and more specifically, with ways of measuring morphosyntactic complexity (§ 5.4), lexical complexity (§ 5.5) and morphosyntactic and lexical accuracy (§ 5.6).

5.1 Defining proficiency

Before considering how best to measure proficiency in L2 children and adults, it is necessary to define this construct. In L2 acquisition research, the term 'language proficiency' is employed in numerous ways. As well as being used as a global indicator of an L2er's abilities in the TL, it is also used to refer to specific aspects of linguistic competence, such as phonological, syntactic, morphological, lexical and/or discourse skills. In general, language proficiency is divided up into knowledge and some aspect of use (control / communicative competence) and it often involves one or more of the following dichotomies: productive vs. receptive, written vs. oral, communicative vs. grammatical, etc.. Defining proficiency – one of the goals of a whole field of applied linguistics (see e.g. papers in De Jong and Verhoeven 1992) – is well beyond the scope of this paper. Nevertheless, in order to be able to proceed, a concrete definition of this construct is required. Intuitively, proficiency might be defined as 'a person's overall competence and ability to perform in L2' (Thomas 1994:330, fn. 1). Such intuitive definitions, however, remain somewhat tautological (Verhoeven and Vermeer 1989:26). For present purposes, the construct of L2 proficiency is operationalised as 'the ability to produce and comprehend lexically, morphologically and syntactically complex and accurate utterances in the TL'.[3,4]

5.2 Comparing proficiency level in L2 children and L2 adults

Proficiency is more difficult a variable to control for than, for example, L1 transfer. It is, however, essential that measures are taken to ensure that any comparisons between L2 children and L2 adults are based on learners who are at approximately the same proficiency level. As Thomas (1994) notes, proficiency measures are necessary when a researcher wishes to compare different groups of L2ers in their acquisition of a given phenomenon. In order to be able to say anything meaningful about the differences or similarities between these two groups, it is imperative that the 'right' children are compared with the 'right' adults. Imagine that a group of L2 children were compared with a group of L2 adults and that with respect to the TL property in question, the adults were observed to commit a certain error which the children did not. On the basis of this observation it might be claimed that adult L2 acquisition is (fundamentally) different from child L2 acquisition. This is for example the basic line of argumentation followed by Weerman, Bisschop and Punt (2003) and Blom and Polišenská (2006) in their studies on the acquisition of Dutch adjectival and verbal inflection, respectively. Without ensuring that the L2 children and adults have approximately the same proficiency level, however, such a claim remains premature. It is possible that the lack of errors on the part of the L2 children may be because, in the sample in question, the L2 children are more proficient than the L2 adults; they may have made this error in an earlier developmental stage, but this will remain undetected unless data from L2 children at a lower level of proficiency are examined. Likewise, the claim that in a particular sample, L2 children are targetlike and L2 adults are not can only be maintained if the two groups are taken from a comparable (high) proficiency level.[5]

Comparing L2 children and L2 adults in terms of proficiency places certain restrictions on the type of measure which can be used. Firstly, because the task should not involve reading/writing, a traditional cloze test, often used as a proficiency measure for adult L2ers, is immediately ruled out. Secondly, the task should not be too 'test-like', in the sense that it should not resemble the type of language tests commonly administered in L2 classrooms. This is to avoid the application of metalinguistic/ explicit knowledge, that is, conscious knowledge of language, often in the form of grammatical rules, which is viewed as distinct from the implicit knowledge considered to be central to L2 proficiency as defined here.[6] Given that metalinguistic/explicit knowledge develops with age (Bialystok 1993; following Karmiloff-Smith 1986), this is more likely to be an issue with the older child L2ers and adult L2ers. Furthermore, as noted by Appel (1984:139), older children and adults are also likely to have more developed test-taking abilities. Hence, the use of such a test could introduce an unwanted variable (or variables), which would at least partly co-vary with age.[7]

5.3 Types of proficiency measure

In a survey of 157 articles taken from four L2 acquisition journals,[8] Thomas (1994) observes that L2 proficiency is generally measured in four different ways: (i) impressionistic judgement, (ii) institutional status, (iii) standardised test and (iv) in-house assessment instrument. This section considers to what extent each of these is a valid measure of proficiency and whether they are suitable for use with both L2 children and adults.

Impressionistic judgement, which involves ‘asserting that a learner has a given level of control over [the] L2, on the basis of the experimenter’s unsupported evaluation, or the evaluation of some other (often unspecified) person’ (Thomas 1994), has the clear disadvantage of lacking generalisability: one person’s ‘advanced’ is not another’s. As Thomas (1994:317) points out, institutional status, for example, first-year students vs. final-year students, suffers from a similar weakness: standards are determined in different ways in different institutions, and hence final-year students at one university might not be the same as final-year students at another university. Furthermore, when the pool of subjects includes both adults and children, there is often no such common denominator with which they could be compared.

Standardised tests are available for many different languages. For Dutch, the TL in focus here, these include the TAK test (*Taaltoets Alle Kinderen* ‘Language test for all children’ Verhoeven and Vermeer 2001), the ISK tests (*Internationale Schakel-Klassen* ‘International transition classes’ from CITO (Dutch testing agency)) and the NIVOR test (*Niveauvorderingentoetsen* ‘Level Assessment tests’, also from CITO). While each of these tests has the advantage of being standardised, they are each designed for different age groups: the TAK is designed for children, the ISK for adolescents and the NIVOR for adults. This is problematic given the purpose of a proficiency measure here is to find a way of comparing these groups. Likewise, in English, many standardised tests are available, but these are often for one particular age group (e.g. TOEFL is for adults only). One exception to this is the *CYCLE* (*Curtiss-Yamada Comprehensive Language Evaluation*) test (Curtiss and Yamada 1985). This test, designed for a wide range of ages and cognitive abilities, was used to compare Spanish-speaking L2 children and adults in a recent study by Gilkerson (2005) on the acquisition of particle verbs in English. When such tests are available for the TL in question, they may be preferable to the non-standardised measure discussed below. However, as noted above, they are not available for all languages.

Two points mediating against the use of standardised tests are, however, that they often focus very clearly on the application of rules and they are usually rather time-consuming to administer. Focussing on rules, for example on how to form noun plurals or inflect verbs, should be avoided in the present context as this could encourage the use of metalinguistic knowledge, especially on the part of the adults – see discussion above. Employed in the context of a child L2 ~ adult L2 comparison, proficiency measures are carried out in addition to the experimental tasks designed to tap knowledge of the TL property under investigation. This means that several tasks need to be carried out in what is often the already limited period of time available with the L2er. For this practical reason, the proficiency task should not be too time-consuming. Standardised tests are often rather lengthy.

The final proficiency measure in Thomas' list is the in-house assessment instrument, which includes tests developed privately by researchers (or teachers). Thomas (1994) writes that this type of measure has the advantage of ensuring that all subjects are tested in a uniform fashion; this means that any resulting proficiency scores are internally consistent within the sample and 'subgroups may be compared with respect to proficiency on some rational basis' (Thomas 1994:322). Given that such sub-group comparisons are the locus of interest here, an in-house assessment instrument seems to be a suitable option for child L2 ~ adult L2 comparisons. What exactly such an assessment instrument should measure will be addressed in the following sections. Prior to this, however, the details of the task itself are given.

Following Larsen-Freeman (1983) and Whong-Barr and Schwartz (2002), the data used for this measure were collected using a picture description task. Subjects were presented with sets of between four and eight pictures which depicted a series of events and their task was to describe/tell a story about what they saw. Importantly, all the actions depicted in the pictures were considered general enough so as not to (dis)favour either the child subjects or the adult subjects. They included, for example, planting flowers and watering them, digging a hole in the beach, and having an accident on a bicycle. During the task, which lasted approximately 10 minutes, subjects were encouraged to speak as much as possible, and when necessary, the experimenter(s) provided prompts and asked questions designed to elicit more data, such as '*And what happened next?*'. The data were transcribed in CHAT format using the CLAN programme available via CHILDES (MacWhinney 2000) by one of the experimenters who was present at the time of recording (either the present author or a native-speaker student assistant) and they were checked by either the other experimenter or another native-speaker student assistant.[9]

The definition of proficiency stated above requires that it encompass measures of morphosyntactic and lexical complexity and accuracy. Each of these aspects is dealt with in turn.[10]

5.4 How to measure morphosyntactic complexity

Following Ortega (2003:492), (morpho-)syntactic complexity is defined as 'the range of forms that surface in language production and the degree of sophistication of such forms'. By far the most common measure of morphosyntactic complexity, in research on L1 acquisition at least, is mean length of utterance (MLU).[11] In the context of child L2 ~ adult L2 comparisons, the use of MLU raises some interesting and important issues, primarily concerning whether this measure can be extended beyond its typical use with young L1 children. In this section, I present data which show that because MLU still develops in older children and is highly variable in (native-speaker) adults, it cannot be used as a basis of comparison between L2 children and L2 adults. I will argue that instead, verbal density, the average number of finite and non-finite verbs per utterance, is a more suitable measure.

5.4.1 MLU

In L1 acquisition, MLU (Brown 1973; Nice 1925) is the generally accepted means of assessing a child's stage of (morphosyntactic) development. It is also regularly used to match impaired (e.g. SLI) and typically developing L1 and/or L2 children (see e.g. Bol 2003 for relevant discussion). Its widespread use may, however, be put down to the ease with which it can be calculated and the lack of a suitable alternative as much as to its validity as a measure of linguistic proficiency. As the frequent discussions witnessed in the literature demonstrate, both its validity and reliability have been questioned. While some have claimed, for example, that MLU is unreliable beyond the two-word stage (Klee and Fitzgerald 1985; Scarborough, Rescorla, Tager-Flusberg, Fowler and Sudhalter 1991), others have shown that this is not the case (Blake, Quartaro and Onorati 1993; Rondal, Ghiotto, Bredart and Bachelet 1987; Shriner 1967). Assuming, for the moment at least, that MLU *is* a valid measure of grammatical development in L1 children, let us consider whether it would be appropriate to use MLU as a means of comparing the proficiency of L2 children with that of L2 adults. As noted in § 4, comparing L2 children (who are children at time of testing) and L2 adults means that age will vary across the groups; furthermore, L2 children are usually considerably older than the children for which MLU is generally used. Thus, in order to assess whether MLU is a suitable measure of morphosyntactic complexity for present purposes, it is necessary to determine whether its validity extends to (i) L2 acquisition, and (ii) older children. We start with the former.

As a result of their knowledge of another language, L2ers are usually capable of producing multi-word/morpheme sentences almost immediately after initial contact with the TL (Adamson 1988; Larsen-Freeman and Strom 1977:124). This means that their initial MLU is comparatively high and hence there is less room for the L2er to develop in this respect. In other words, in terms of MLU, L2ers are likely to ‘skip a stage’ (or several) in comparison with L1 children. Consequently, MLU is often measured in words (MLUw) rather than morphemes. The following discussion includes studies measuring MLU both ways.

The few available studies on MLU in L2 acquisition give mixed results with respect to the use of this measure as an indicator of grammatical complexity. On the one hand, it has been shown that MLU is a valid measure in L2 acquisition because it develops linearly with increasing proficiency level and it correlates with standardised tests (Larsen-Freeman and Strom 1977; Verhoeven and Vermeer 1989). On the other hand, however, it has been claimed that such results may stem from circular argumentation where MLU is (subconsciously) included in how the different proficiency levels are determined (Dewaele 2000).

Establishing whether MLU develops in older L1 children will allow us to determine whether – assuming for the moment that MLU is suitable for use in L2 acquisition – it can be used to compare L2 children and adults. The logic is as follows. Given that adults are older than children, it is to be expected that these two groups will probably differ in terms of their MLU. As we shall see below, this difference, in and of itself, need not be a problem. What would be problematic, however, in the sense that it would seriously complicate the child L2 ~ adult L2 comparison, is if age were found to correlate significantly with MLU *throughout* childhood, that is, beyond the age of 5, the age until which it has been established that such an age ~ MLU correlation exists (but see Conant 1987; Johnston 2001; cf. Klee and Fitzgerald 1985; Miller and Chapman 1981; Rondal *et al.* 1987). The reason is as follows: whereas for L2 adults, MLU would solely be a measure of language development, for L2 children, it could also be a function of age itself, and this would introduce an unwanted additional variable to the child L2 ~ adult L2 comparison. Implicit in this argument is the assumption that the L2 children's MLU in their L2 will not progress beyond the value of their MLU in their L1 (which, I believe, is relatively uncontroversial, at least when exposure to the L1 is maintained).

The following example serves to illustrate the argument. Imagine we have two L2 children, Tom and Ann. Both are native speakers of English. Tom was first exposed to the TL, Dutch, at age 4 and Ann at age 7. They are tested 3 years later, that is, when Tom is 7 years old and Ann is 10. Assuming that all other variables are held constant, imagine that Ann's Dutch MLU is observed to be higher than Tom's. If MLU is still developing in L1 children between the ages of 5 and 10 years, for example, it would be unfair to conclude that Ann's L2 grammar was more complex than Tom's, because this might be due to her relative cognitive maturity. In this context, cognitive maturity should be understood as maturational constraints on information processing capabilities and other cognitive processes relating to and including memory (e.g. Gavens and Barrouillet 2004; see Schneider 2002 for overview), which in turn constrain the production of lengthy and complex utterances (Berman 2007). In this sense, then, the reason for Tom's lower MLU in his L2 Dutch might not be because he is less proficient than Ann; rather, it might result from him being 3 years younger than she is and from his L1 MLU being lower, which, in turn, would mean that his L2 MLU would be lower.

In a large-scale longitudinal study on various aspects of linguistic development in English-speaking children and adolescents aged 6 through 17, Loban (1976) observes a clear increase in MLUw with age (see also Hunt 1970). Age effects beyond 5 years are also observed by Minifie, Darley and Sherman (1963) and Shriner and Sherman (1967). Shriner (1967) and Chabon, Kent-Udolf and Egolf (1982), on the other hand, fail to find such effects. Shriner (1967) relates this to an increase in variability in MLU scores after age 5. The issue of variability is discussed in more detail below.

The mixed nature of these results make it difficult to draw any firm conclusions regarding the nature of MLU in older L1 children and its potential implications for use with L2 children. In order to investigate this question further, (semi-spontaneous) data using the picture description task described at the end of the preceding section were analysed from L1 Dutch children aged 7, 9 and 11, and from L1 Dutch adults.[12] To calculate MLUw, T-unit was used instead of (the rather vague notion of) utterance. A T-unit is defined as 'one main clause plus whatever subordinate clause and nonclausal expressions are attached to or embedded within it' (Hunt 1970:14). The results are presented in Table 1.

Table 1. L1 Dutch children and adults: MLUw

Age group	**N**	**Mean**	**Range**	**SD**
7-year-olds	10	5.88	4.88-6.74	0.67
9-year-olds	10	6.41	5.49-7.39	0.60
11-year-olds	10	6.60	6.19-7.39	0.36
Adults	10	7.50	5.86-10.02	1.32

A significant difference is found between the three child groups (Kruskal-Wallis: $\chi^2 = 6.068$, df = 2, $p < .05$), which suggests that MLUw does develop significantly between the ages of 7 and 11.[13] The implication of this finding is that, on the logic laid out above, and assuming that L1 children acquiring other languages will pattern similarly to L1 Dutch-speaking children, using MLU as a measure of morphosyntactic complexity would introduce an additional variable into the child L2 ~ adult L2 comparison: for L2 adults, it would measure language development, whereas for L2 children it would measure language development and it would also be a function of age.

The data from the adults in Table 1 indicate that there is considerable variability in MLUw amongst native speakers. Most (80%) of the adults had MLUw values above 6.5, indicating a reasonable amount of consistency across speakers. Nevertheless, the large range of 5 words indicates that even when native-speaker adults perform the same task, their MLUw may vary considerably. This finding has two implications. Firstly, it suggests that MLUw as a measure of grammatical complexity for native-speaker adults may not be very reliable. Secondly, as a consequence, pinpointing a value at which an L2er's MLUw value can be considered native-like, or highly proficient, would be rather difficult. If native speakers vary from 5 to 10 in terms of their MLUw, it is not clear whether, for example, an L2er with a MLUw of 8 should be considered more targetlike/proficient than an L2ers with a MLUw of 6.

To summarise: Although there is previous research showing that MLU is a relatively reliable indicator of grammatical complexity in early child L1 acquisition and to a certain extent in L2 acquisition as well, this measure remains, on the whole, rather unreliable. The analysis of new L1 data from older children suggests that MLUw may develop well into the first decade of life. This is a potentially interesting finding in and of itself, although further research (for example with more subjects and for different languages) is needed to confirm it. In the present context, however, where L2 children are compared with L2 adults, this result, coupled with the observation that even on the same task, MLUw in adult native speakers is highly variable, means that MLUw must be regarded as an unsuitable measure of grammatical complexity.

5.4.2 Verbal density

An alternative measure of morphosyntactic complexity is verbal density, defined as the average number of finite and non-finite verbs per T-unit (see above for definition).[14] This measure was adopted by Chaudron and Parker (1990), following Pica and Long (1986), in an analysis of free and elicited production data in Japanese-English adult L2 subjects. There are several reasons to believe that it is a valid means of measuring grammatical complexity, at least in a rudimentary fashion. First of all and most importantly, it captures complexity in a central aspect of grammatical development, that is, the use of different verb forms, such as non-finite dependent clauses ((1)-a), relative clause modification ((1)-b), modals, auxiliaries and complex tense forms ((1)-c and (1)-d), and durative constructions ((1)-e).

(1) a. She decides to go for a swim

b. The girl who is wearing a green pullover fell down

c. She shouldn't have done that

d. After the girl had eaten, she went out to play

e. The boy sits reading a book

Verbal density differs from MLUw because it does not just measure length but arguably also depth, that is, the utterances illustrated in (1) demonstrate grammatical complexity at the clause level and this is different from simply stringing words together. This is particularly important for L2ers who, as noted above, will as a result of their L1 be able to produce multi-word utterances from early on.

Verbal density scores were calculated for both children and adults. All verbs were counted in this calculation, including copula *zijn* 'to be' and the auxiliaries *hebben* 'to have' and *zijn* 'to be', *gaan* 'to go' and modals.[15] The score was calculated by dividing the total number of finite and non-finite verbs by the number of T-units. Importantly, unlike MLUw, verbal density was not found to develop significantly as a function of age in the child L1 Dutch data. The results are presented in Table 2.

Importantly, unlike MLUw, verbal density was not found to develop significant as a function of age in the child L1 Dutch data.

Table 2. L1 child and adult: Verbal density

Age group	N	Mean	Range	SD
7-year-olds	10	1.29	1.07-1.48	0.13
9-year-olds	10	1.41	1.24-1.65	0.12
11-year-olds	10	1.40	1.23-1.51	0.08
Adults	10	1.58	1.32-1.97	0.22

The three child L1 groups do not significantly differ from each other (Kruskal-Wallis: $\chi^2 = 4.846$, df = 2, $p > .05$). This means that, unlike MLU, verbal density appears not to develop significantly with age in L1 children such that it would become a confounding factor when used with L2 children. The L1 children do differ from the L1 adults as a group, however (Mann-Whitney: $Z = -2.936$, $p = .003$).[16] How to deal with this difference when using verbal density as an indicator of L2 proficiency is dealt with in §6.[17]

Thus far, only measures of morphosyntactic complexity have been considered; the definition of proficiency adopted here also includes lexical complexity and morphosyntactic and lexical accuracy. We now turn to how these aspects of proficiency should be measured.

5.5 How to measure lexical complexity

5.5.1 Type/Token Ratio

Lexical complexity, which is understood as synonymous with lexical diversity or richness, is traditionally measured using the 'Type/Token Ratio' (TTR). TTR is calculated by dividing the number of types (V) in a sample by the number of tokens (N) and the higher the TTR, the more diverse a lexicon a learner is claimed to have. For example, a TTR of .5 (e.g. 10 types in a sample of 20 tokens) is assumed to reflect a more diverse lexicon than a TTR of .25 (e.g. 5 types/20 tokens). This score is assumed to provide an index of lexical development independent of sample size, and consequently, it is widely used in both L1 and L2 acquisition studies.

The TTR has, however, been shown to be inadequate in several ways. Richards (1987) demonstrates that TTR *is* affected by sample size. TTR is artificially deflated as a result of an increase in tokens in a sample. Imagine that one learner produces five different types in a sample of 20 tokens and another learner produces the same five types, but her sample consists of 30 tokens. The TTR for the second learner (.167) would actually suggest that this learner's lexicon is less diverse than that of the first learner (.25), but this is not the case: they both produce the same five types. When the increase in sample size results from linguistic development within the same learner, for example when determiners are acquired, this slight dip in TTR would incorrectly suggest a lack of development (or even backsliding) whereas in reality, the learner will have made considerable steps in his or her linguistic abilities, even though the TTR does not reflect this (van Hout and Vermeer 1988).

5.5.2 Guiraud's index

Broeder, Extra and van Hout (1993) argue that the measure known as Guiraud's index (Guiraud 1960), or the 'Indice de Richesse', is a more successful measure of lexical richness, because unlike TTR, it increases over time (see also Vermeer 1986). Guiraud's index is calculated by dividing the number of types by the square root of the number of tokens ($V\sqrt{N}$). By taking the square root of the number of tokens, the problem of a negative correlation with increasing sample size (as with TTR) is obviated.

Similar to our discussion of MLUw and verbal density, in order to assess the suitability of Guiraud's index as an indicator of lexical complexity when comparing L2 children and adults, it is necessary to determine the extent to which native-speaker children and adults differ with respect to this variable (especially given that vocabulary size is known to vary across speakers). To this end, Guiraud's index was calculated for the same L1 Dutch children and adults. The results are presented in Table 3.

Table 3. L1 children and adults: Lexical complexity

Age group	**N**	**Mean**	**Range**	**SD**
7-year-olds	10	6.04	5.01-7.44	0.72
9-year-olds	10	6.27	4.97-7.30	0.59
11-year-olds	10	5.80	5.25-6.14	0.74
Adults	10	7.07	6.07-8.22	0.74

A one-way ANOVA indicates that there is a significant difference between groups ($F = 8.161$, $p < .001$). Post-hoc Bonferroni tests show that the child groups do not differ from each other but that they each differ from the adults (7-year-olds vs. adults, $p < .01$; 9-year-olds vs. adults, $p < .05$; 11-year-olds vs. adults, $p < .001$). Thus, as with verbal density, lexical complexity appears not to vary significantly with age in L1 children such that it would become a confounding factor when used with L2 children.[18] How the difference between the children and adults is dealt with is postponed until §6. First, we consider how to measure accuracy.

5.6 How to measure morphosyntactic and lexical accuracy

An assessment of L2 proficiency made on the basis of complexity alone would be insufficient because complexity closely interacts with accuracy (Lalleman 1986). L2ers who produce complex yet inaccurate utterances should not be considered more proficient than L2ers who produce less complex but more accurate utterances. To take the interaction between these two factors into account, a measure of proficiency should incorporate a measure of accuracy as well as a measure of complexity.

Accuracy is generally measured using rate of errorfree utterances (e.g. Larsen-Freeman and Strom 1977; Larsen-Freeman 1978), that is, the number of errorfree utterances (or in this case, T-units) divided by the total number of utterances. The rationale behind this calculation is simple: the more developed/advanced L2ers are, the less likely they are to commit errors.[19] All errors are not equal, however, and in adopting this measure, the researcher is faced with certain choices regarding what should (not) be counted as an error, including how to deal with repeated errors and with pervasive errors which do not differentiate learners at different proficiency levels.

When subjects make repeated errors, for example, consistently using a particular word incorrectly or systematically using a non-targetlike verb form, we might consider only counting the first instance of such an error as an error so as not to 'overpunish' them, as this would potentially artificially deflate their rate of errorfree utterances as a result of one persistent error. However, the variability which is often pervasive in L2 grammars means that subjects will sometimes produce a certain error and at other times they will not. Consequently, if an error were initially counted as such, the subject in question would not be 'rewarded' when s/he does produce a targetlike form, but if the error were ignored, the L2er would be attributed with more knowledge than s/he has. Hence, repeated errors should be counted separately.

Certain errors may be so pervasive that they fail to differentiate learners at different proficiency levels. In this case, incorporating them into a measure of accuracy would be pointless. Consequently, such errors should be ignored. The nature of such errors may vary according to TL. In the L2 Dutch data reported on below, for example, grammatical gender errors fell into this category, and hence, they were excluded from the accuracy count.

6. Introducing the Age-Sensitive Composite Proficiency Score (ASCOPS)

The review of the available literature, coupled with the analysis of L1 child/adult data, suggest that morphosyntactic and lexical complexity and accuracy, the core elements of proficiency as defined here, are best measured by verbal density, Guiraud's index and rate of errorfree utterances. This section presents a newly developed composite proficiency measure which combines these three elements into one score such that the inherent differences between children and adults are taken into account. An overview of how this was achieved is given in Figure 1.

<table>
<tr><th>Type of measure</th><th>Linguistic domain</th><th>Measure</th><th>Resulting sub-score</th><th>Final score</th></tr>
<tr><td rowspan="2">Complexity</td><td>Morphosyntax</td><td>Verbal density:
number of finite and non-finite verbs divided by total number of T-units</td><td rowspan="2">Scores converted into standardised (z) scores for L2 child and L2 adult groups separately</td><td rowspan="4">Sub-scores are combined into a single score using principal components analysis. Each subject has a single standard normal (z) score as final proficiency score</td></tr>
<tr><td>Lexicon</td><td>Guiraud's index:
V/√N (where V=type and N=token)</td></tr>
<tr><td rowspan="2">Accuracy</td><td>Morphosyntax</td><td rowspan="2">Rate of errorfree utterances:
number of errorfree utterances divided by total number of T-units</td><td rowspan="2">%</td></tr>
<tr><td>Lexicon</td></tr>
</table>

Figure 1. Overview of proficiency measure

Recall that the L1 children and adults who completed this task differed in terms of verbal density and lexical diversity. Assuming that this difference results from differences in cognitive maturity (as outlined in §5.4.1), this has serious implications for how this measure is implemented as part of a child L2~adult L2 comparison.[20] As noted above with respect to MLU, it would be unreasonable to expect that an L2 child exceed the verbal density or lexical diversity score achieved by an L1 child of the same age. Furthermore, if we assume that the findings for L1 Dutch children and adults can be extended to other languages – English is the L1 of (all) the L2ers in this case – it would also be unreasonable to expect an L2 child's interlanguage verbal density or lexical diversity score to exceed that of her L1. Thus, it is highly likely that L2 children will intrinsically differ from L2 adults in terms of these measures.

A distinction between L2 children, on the one hand, and L2 adults, on the other, is usually made on the basis of their age at first exposure. However, in the present context, given the constraints which relative cognitive maturity may place on potential verbal density and lexical complexity scores, the difference between the two groups must be determined in terms of age at time of testing. The age at time of testing thus ranges from 7;3 to 14;11 (mean = 10;9; SD = 2;2) for the L2 children and from 22;2 to 50;0 (mean = 36;2; SD = 8;0) for the L2 adults. The length of exposure ranges from 0;2 to 10;4 (mean = 4;1; SD = 2;11) for the L2 children and from 0;4 to 25;0 (mean = 7;3; SD = 6;5) for the L2 adults. All learners were resident in The Netherlands at time of testing.

Data were collected using a picture description task (see §5.3 for details). The following types of utterance were excluded from analysis: one-word utterances consisting of a proper name or of *ja* ‘yes’ or *nee* ‘no’; incomplete utterances; utterances containing unidentifiable material; formulaic utterances such as *weet ik niet* ‘I don’t know’; utterances where more than half the words were in English. Verbal density and lexical complexity scores were calculated using the procedures outlined in §5.4.2 and §5.5.2, respectively. The rate of errorfree utterances was calculated by dividing the total number of errorfree utterances by the total number of utterances produced by a given subject. To be categorised as errorfree, an utterance had not to contain a number of morphological, syntactic and lexical errors. Morphological errors included non-target subject-verb agreement, non-target form of the verb stem or of a noun. Syntactic errors were: non-target work order such as failure to use verb second in matrix clauses and non-final finite verbs in subordinate clauses; failure to split up particle and verb in particle verbs; missing functional elements such as determiners, complementisers and copula/auxiliary verbs. Lexical errors included the use of non-target prepositions, subordinating conjunctions, collocations and a target word used with a non-target meaning. For examples of each type, the reader is referred to Unsworth (2005: 203-206). Cohen's Kappa statistic was calculated for a random sample of 10% of the transcripts and was found to be 0.829 ($p < .001$ (approx.)), indicating that there was almost perfect agreement between different coders for errorfree utterances.[21]

All three sub-scores for the L2 children and adults and for the L1 children and adults (discussed in previous sections) are presented in Table 4.

Table 4. L1/L2 children and adults: Verbal density and lexical diversity

		Verbal density			**Lexical diversity**			**Rate of errorfree utterances (%)**		
Group	**N**	**Mean**	**Range**	**SD**	**Mean**	**Range**	**SD**	**Mean**	**Range**	**SD**
L1 children	30	1.37	1.07-1.65	0.12	6.04	4.97-7.44	0.58	99.6	98.8-100	0.25
L2 children	47	1.30	0.74-1.68	0.21	5.34	3.58-7.54	1.08	54.8	12.5-96.5	22.2
L1 adults	10	1.58	1.32-1.97	0.22	7.07	6.07-8.22	0.74	100	--	0
L2 adults	18	1.38	0.91-1.88	0.28	6.89	3.96-8.94	1.55	65.6	40.0-95.3	17.0

Let us first consider the results for verbal density and lexical diversity. Just eyeballing the data, both L2 groups pattern similarly with respect to their L1 age-equivalents: the mean for the L2 children is lower than the L1 children but the highest score is comparable, and a similar pattern obtains for the L2 adults when compared with the L1 adults. This suggests that, as expected, the adults (are able to) obtain higher scores. Both the adult L2 and the child L2 data have a normal distribution (Shapiro-Wilk: $p > .05$ for both groups on both scores). Thus, it is not the case that the child L2 data are negatively skewed and the adult L2 data positively skewed such that this could explain the difference between the two groups in terms of highest scores. Note also that the maximum scores achieved by the L2 children is virtually identical to the L1 children

(1.68 and 1.50, respectively, for verbal density, and 7.54 and 7.44, respectively, for lexical diversity). As noted in the preceding sections, there is a significant difference between the L1 children and adults for both verbal density ($Z = -2.936$, $p = .003$) and lexical diversity ($Z = -3.498$, $p = .000$). It is assumed that this difference relates to the age differences between these two groups, either in terms of 'internal cognitive and social developments' (Berman 2007:359) or length of exposure to the language. If this is the case, similar differences are expected between the L2 children and adults. For the L2 children and adults, there is a significant difference between groups for lexical diversity ($Z = -3.650$, $p = .000$) but not for verbal density ($Z = -.916$, $p = .360$). The lack of a significant difference between the two L2 groups with respect to verbal density is unexpected. It appears, however, that this may be the result of the L2 adults' comparatively poorer performance on verbal density than on lexical diversity: there is no difference between the L2 and L1 adults on lexical diversity ($Z = .048$, $p = .962$), whereas the difference between the two groups on verbal density is approaching significance ($Z = -1.822$, $p = .068$). It is therefore not possible to rule out that, like the L1 groups, the L2 children's scores on verbal density are significantly lower than those of the L2 adults. For this reason, the verbal density data will be treated similarly to the lexical diversity data in calculating the final proficiency score.

As noted above, the differences between the L1/L2 children, on the one hand, and L1/L2 adults, on the other, may result from differences in their cognitive maturity (e.g. working memory or information processing capacities) or from differences in their length of exposure to the language.[22] When L2 children are tested when they are children, their possible maximum length of exposure is by definition shorter than the possible maximum length of exposure for L2 adults. In the data presented here, the average length of exposure is 49.1 months for the L2 children (SD: 35.3) and 86.9 months for the L2 adults (SD: 77.3). For the methdological point at hand, it does not matter which of these two factors is the cause of the observed differences. What is crucial is that both these factors systematically distinguish the two groups.[23]

What the intrinsic differences between children and adults in terms of potential verbal density and lexical diversity scores mean is that L2 adults who score higher on these measures should not automatically be regarded as more proficient than L2 children whose scores are slightly lower. Such a disparity might stem from the inherent differences either in cognitive maturity or length of exposure between the child and adult subjects rather than from a contrast in their interlanguage grammars. In order to ensure that this age-related factor does not adversely influence the L2 subjects' proficiency scores, the verbal density and lexical diversity scores for this variable are converted into standardised (z-)scores for the child and adult groups separately, before they are combined into the final proficiency score (see Unsworth 2005:189 for more details). Before considering how this is achieved, let us briefly turn to the scores for the rate of errorfree utterances.

The rate of errorfree utterances presented in Table 4 is at ceiling for both the L1 adults and children. The rate of errorfree utterances for the L2 adults is on average higher than for the L2 children, but the difference between the two groups is not significant ($t = -1.861$, $p = .067$). The difference is approaching significance, however. This observation could be interpreted in a similar fashion to the data for the other two scores, that is, the children's scores are lower than the adults' as a result of their relative cognitive immaturity. This, however, is unlikely, given that in contrast to the other two scores, the upper end of the range of scores for accuracy is comparable across the L2 child and L2 adult groups: there are learners in both groups who have a rate of errorfree utterances of around 96%. If the L2 children's scores were constrained by their age relative to the adults, this should not be the case.

We now turn to how the three sub-scores are amalgamated into one final proficiency score. The two sets of z-scores for lexical diversity and verbal density are combined with the accuracy scores using a principal components (or factor) analysis. Principal components analysis is a means of reducing the number of variables in a data set, and of detecting structure between these variables (StatSoft 2004). This is achieved by modelling the data on a three-dimensional scatterplot to obtain one or more new variables (or components) which account for as much variance amongst the original variables as possible. It is assumed that the original variables correlate, that is, that they measure the same construct. In our case, it is assumed that they all measure some aspect of L2 proficiency. This is the case. For the proficiency scores given in Table 4, there is a significant moderate correlation between verbal density and lexical diversity ($r = .587$, $p <.001$), between verbal density and rate of errorfree utterances ($r = .412$, $p <.001$) and between lexical diversity and rate of errorfree utterances ($r = .667$, $p <.001$).

Note that the original variables are not identical; if they were, two would be redundant. Principal components analysis is a means of extracting the commonalities between several, correlated variables in such a way that as much variance as possible amongst these variables is accounted for by the resulting components or factors. In the proficiency data used here, only one component was extracted and this accounts for approximately 70% of the variance in the data.[24] The factor loadings for the three original variables are as follows: verbal density = .779, lexical diversity = .912 and rate of errorfree utterances = .828. The higher the factor loading, the more that original variable contributes to the extracted component. Importantly, when the same proficiency score was calculated for a different group of subjects (see § 7), highly similar factor loadings were observed (.781 for verbal density, .905 for lexical diversity and .829 for rate of errorfree utterances). This suggests that the ASCOPS works consistently across different subjects. On the assumption that the sub-scores are valid measures of L2 proficiency, the resulting proficiency scores can be considered to provide an approximate yet relatively robust indication of the subjects' L2 proficiency.[25]

The values for this extracted component are saved as standardised variables and these scores (having a standard normal distribution) are used as the proficiency score for each person. On the basis of these scores, it is possible to divide subjects into different proficiency level groups and subsequently, the L2 children and adults can be compared with each other within each of these levels, that is, the 'right' children can now be compared with the 'right' adults. Before examining ASCOPS in action, the advantages and disadvantages of the score are discussed.

6.1 Disadvantages of ASCOPS

The first, practical disadvantage of ASCOPS when compared to proficiency measures such as standardised tests is that the transcription and analysis of the semi-spontaneous data which are needed to calculate the score are very time-consuming.[26] Second, there is as yet no independent evidence that verbal density measures linguistic proficiency. Note in this regard, however, that such a lack of construct validity is a problem with most measures of morphosyntactic complexity, particularly in L2 acquisition research: no-one (to my knowledge) has systematically investigated whether and how such measures of morphosyntactic complexity relate to developmental sequences. This means that any such measure faces this criticism and as such, it cannot be used as an argument against one over the other (Ortega 2000). One final disadvantage is that cross-study comparisons based on this measure should be made with caution. This is because the measure is based upon standardised scores which relate to a particular group of subjects tested at one particular time.

6.2 Advantages of ASCOPS

The most important advantage of ASCOPS is that it takes into account the intrinsic differences between children and adults (either in terms of cognitive maturity or length of exposure) and the consequences this has on linguistic development. This is essential if L2 children and adults are to be compared on the basis of their linguistic capabilities alone (to the extent that this is possible). Also, in contrast to other studies using rate of errorfree utterances (e.g. Larsen-Freeman 1983), the types of non-targetlike forms which are counted as errors, as well as how this measure of proficiency is combined with others, is stated very explicitly (see Unsworth 2005 Chapter 4). As a consequence, other researchers will be able to make informed judgements about the extent to which two or more proficiency measures are comparable and furthermore, if desired, exact replication will also be possible. In an attempt to increase the validity and reliability of this measure, independent motivation is provided (where possible) for the decisions to include or exclude particular elements from the final score. Another important advantage is that – unlike standardised tests – this measure can in principle be used for any language.

There are also numerous advantages to the type of task used to collect data. Firstly, it is based on spoken language only and is therefore suitable for use with younger subjects. Secondly, it involves an activity, namely describing pictures/telling a story, which most (if not all) subjects are familiar with. Thirdly, it requires the subject to focus on content rather than on form, which also serves to help the subject relax. Fourthly, it can be worked into the test procedure quite easily and it can also be used as a sort of distracter to split up parts of the experimental procedure. Furthermore, it is not very time-consuming to carry out (10-15 minutes at most).

7. ASCOPS in action

The ASCOPS was implemented as part of a study on the acquisition of direct object scrambling in Dutch by English-speaking children and adults (Unsworth 2005). This section briefly summarises the results of this study to illustrate how ASCOPS can be used as the basis for a comparison between L2 children and adults.

Scrambling in Dutch involves moving the direct object NP from its base position, directly adjacent to the verb as in (2)-a, to an adjoined position to the left of other sentential constituents, such as adverbials or negation, as in (2)-b.

(2) a. Jan heeft niet [de boom] geplant
John has not the tree planted

b. Jan heeft [de boom]$_i$ niet t_i geplant
John has the tree not planted
'John didn't plant the tree yesterday.'

Whether or not objects scramble depends on various semantic and discourse/pragmatic factors (de Hoop 1992; Neeleman and Reinhart 1998; Van Geenhoven 1998). To test whether the L2ers, whose L1, English, does not have scrambling, had acquired this property of Dutch, an elicited production task was designed (following Schaeffer 2000). In this task, subjects were presented with contexts in which scrambling was required (viz. specific indefinite objects and definite objects in sentences containing negation).

In addition to this experimental task, the L2 children and L2 adults completed the picture description task described above and on the basis of the proficiency scores derived from the data collected using this task, they were divided into three different proficiency levels, low, mid and high. The details are given in

Table 5.[27] Importantly, to ensure the independence of the two tasks, all utterances including (non-)scrambled objects were excluded from the data used to calculate the proficiency score.

Table 5. Overview of proficiency groups (for production experiment)

	L2 children				L2 adults				Mann-Whitney tests comparing child/adult groups
	n	**mean**	**range**	**SD**	**n**	**mean**	**range**	**SD**	
Low[28]	12	-1.06	-2.06 to -0.60	47	9	-0.92	-1.60 to -0.51	34	Z = -.444, p = .657
Mid	6	0.10	-0.34 to 0.44	34	6	0.08	-0.33 to 0.46	28	Z = -.160, p = .873
High	7	1.35	0.57 to 2.12	55	8	0.94	0.63 to 1.25	23	Z = -1.390, p = .165

As the results in the final column of

Table 5 indicate, there is no significant difference between the child and adult groups at any of the three levels. Furthermore, the relative distribution of the three proficiency levels within the L2 child group does not differ significantly from that of the L2 adult group (Chi-squared: $\chi^2 = .413$, df = 2, p = .814). We can thus with some certainty claim that the 'right' children will be compared with the 'right' adults.

When we compare the scrambling behaviour of the L2 children with that of the L2 adults for each of the different proficiency levels, the following pattern emerges: the low proficiency children and adults consistently fail to scramble (in contexts where they should), whereas the mid and high proficiency children and adults (more or less) consistently scramble. There were no significant differences between the L2 child and L2 adult groups in any of the proficiency levels in any of the conditions. These data suggest that the L2 children and L2 adults pass through the same developmental sequence. In addition to serving as a basis to compare the L2 children and L2 adults, the proficiency data are thus also used to infer developmental sequence data from cross-sectional data (Thomas 1994). Recall the rationale for the child L2 ~ adult L2 comparison put forward by Schwartz (1992): assuming that child L2 acquisition is driven by UG, comparing developmental sequences of L2 children with those of L2 adults, while holding the L1 constant, will provide evidence for or against UG involvement in adult L2 acquisition. The L2 children and adults were observed to pass through the same developmental sequence and hence, they are consistent with the claim that UG is involved in adult L2 acquisition (see Unsworth 2004; 2005 for more details).

8. Conclusion

Careful comparisons of child L2 development with adult L2 development can inform our understanding of these two types of acquisition. Carrying out such a child L2 ~ adult L2 comparison brings with it a number of methodological and conceptual problems, however. In order to evaluate the differences and similarities between these two learner groups in their acquisition of a particular TL property, it is necessary to control for as many other variables as possible. L1 transfer is easily controlled for by comparing children and adults from the same language background: if there is transfer, it should be the same for both groups. Investigating child L2 development requires testing L2 children when they are children. Thus, in terms of age at time of testing, L2 children and adults will automatically differ from each other. This has significant consequences for the type of tasks, experimental or otherwise, which can be used.

In order to ensure that the 'right' children are compared with the 'right' adults, L2 proficiency should be carefully controlled for. Low proficiency children should thus be compared with low proficiency adults, high proficiency children with high proficiency adults, and so on. Proficiency, as it is understood here, involves morphosyntactic and lexical complexity and accuracy. MLU was argued to be an inadequate measure of morphosyntactic complexity and instead, verbal density, is used for this purpose. This is combined into one single score with Guiraud's index ($V\sqrt{N}$) as a measure of lexical complexity and the rate of errorfree utterances as a measure of lexical and morphosyntactic accuracy in such a way that the intrinsic differences between children and adults are taken into account. This score, the Age-Sensitive Composite Proficiency Score (ASCOPS), constitutes a first attempt at developing a proficiency measure suitable for both L2 children and L2 adults which is not specific to any particular language. Future research using this method with other languages and systematically comparing it to other measures will determine the extent to which ASCOPS constitutes a valid and reliable indicator of proficiency in L2 children and adults.

Adamson, H.D. 1988. *Variation Theory and Second Language Acquisition*. Georgetown University Press, Washington D.C.

Appel, R. 1984. *Immigrant children learning Dutch. Sociolinguistic and psycholinguistic aspects of second-language acquisition*. Foris, Dordrecht.

Asher, J. and Price, B. 1967. The learning strategy of total physical response: Some age differences. *Child Development* 38: 1218-1227.

Belletti, A. and Hamann, C. 2000. Ça on fait pas! On the L2-Acquisition of French by two young children with different source languages. In *Proceedings of the 24th Boston University Conference on Language Development,* S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.), 116-127. Cascadilla Press: Somerville, MA.

Berman, R.A. 2007. Developing linguistic knowledge and language use across adolescence. In *Handbook of Language Development,* E. Hoff and M. Schatz (eds.), 347-367. Blackwell: Oxford.

Bialystok, E. 1993. Metalinguistic awareness: The development of children's representations of language. In *Systems of Representation in Children: Development and Use,* C. Pratt and A.F. Garton (eds.), 211-233. John Wiley and Sons: Chichester.

Bialystok, E. and Hakuta, K. 1994. *In Other Words: The Science and Psychology of Second Language Acquisition*. Basic Books, New York.

Bialystok, E. 1997. The structure of age: In search of barriers to second language acquisition. *Second Language Research* 13: 116-137.

Bialystok, E. and Miller, B. 1999. The problem of age in second-language acquisition: Influences from language, structure and task. *Bilingualism: Language and Cognition* 2: 127-145.

Birdsong, D. 1999. Introduction: Whys and why nots of the Critical Period Hypothesis for second language acquisition. In *Second Language Acquisition and the Critical Period Hypothesis,* D. Birdsong (ed.). 1-22. Lawrence Erlbaum: Mahwah, NJ.

Birdsong, D. and Molis, M. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44: 235-249.

Blake, J., Quartaro, G. and Onorati, S. 1993. Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language* 20: 139-152.

Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? In *Linguistic Perspectives on Second Language Acquisition,* S.M. Gass and J. Schachter (eds.), 42-68. Cambridge University Press: Cambridge.

Blom, E. and Polišenskà, D. 2006. Verbal inflection and verb placement in first and second language acquisition. In *Proceedings of the 29th Linguistics Colloquium,* M. Vliegen (ed.). 1-10. Peter Lang: Amsterdam.

Bol, G.W. 2003. MLU-matching and the production of morphosyntax in Dutch children with Specific Language Impairment. In *Language Competence Across Populations: Toward a Definition of Specific Language Impairment.,* Y. Levy and J.C. Schaeffer (eds.), 259-271. Lawrence Erlbaum: Mahway, NJ.

Broeder, P., Extra, G. and van Hout, R. 1993. Richness and variety in the developing lexicon. In *Adult language acquisition: Cross-linguistic perspectives.,* C. Perdue (ed.). 145-163. Cambridge University Press: Cambridge.

Brown, R. 1973. *A First Language*. Harvard University Press, Cambridge, MA.

Chabon, S.S., Kent-Udolf, L. and Egolf, D.B. 1982. The temporal reliability of Brown's mean length of utterance (MLU-m) measure with post-stage V children. *Journal of Speech and Hearing Research* 25: 117-124.

Chaudron, C. and Parker, K. 1990. Discourse markedness and structural markedness. The acquisition of English noun phrases. *Studies in Second Language Acquisition* 12: 43-64.

Clahsen, H. 1986. Verb inflections in German child language: acquisition of agreement markings and the functions they encode. *Linguistics* 24: 79-121.

Clahsen, H. 1988. Critical phases of grammar development. A study of the acquisition of negation in children and adults. In *Language Development,* P. Jordens and J. Lalleman (eds.), 123-148. Foris: Dordrecht.

Conant, S. 1987. The relationship between age and MLU in young children: A second look at Klee and Fitzgerald's data. *Journal of Child Language* 14: 169-173.

Curtiss, S. and Yamada, J. 1985. *Comprehensive Language Evaluation - Receptive Measures (CYCLE-R) and the Comprehensive Language Evaluation - Elicited Production (CYCLE-E)*.

de Hoop, H. 1992. Case Configuration and Noun Phrase Interpretation. Phd dissertation, published (1996) New York: Garland. University of Groningen.

De Jong, J.H.A.L. and Verhoeven, L. 1992. *The Construct of Language Proficiency*. John Benjamins, Amsterdam.

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

Dewaele, J.-M. 2000. Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée [Seize the unseizable? Measures of Utterance Length and Applied Linguistics]. *International Review of Applied Linguistics* 38: 17-33.

Dulay, H. and Burt, M. 1974. Natural sequences in child second language acquisition. *Language Learning* 24: 37-53.

Ellis, R. 2004. The definition and measurement of L2 explicit knowledge. *Language Learning* 54: 227-275.

Ervin-Tripp, S.M. 1978. Is second language learning like the first? In *Second Language Acquisition. A Book of Readings,* E.M. Hatch (ed.). 190-206. Newbury House: Rowley, MA.

Felix, S.W. 1985. More evidence on competing cognitive systems. *Second Language Research* 1: 47-72.

Gavens, N. and Barrouillet, P. 2004. Delays of retention, processing efficiency, and attentional resources in working memory span development. *Journal of Memory and Language* 51: 644-657.

Gilkerson, J. 2005. Testing for UG Access: An Investigation of How Native Spanish Speakers Acquire English Particle Verbs. Unpublished doctoral dissertation. UCLA.

Guiraud, P. 1960. *Problèmes en méthodes de la statistique linguistique*. D. Reidel, Dordrecht.

Haberzettl, S. 1999. *Katze Maus essen* vs. *Katze essen Maus*: Die L1 als Königs- oder Holzwegbereiter zur L2? Zum Einfluss des L1-Wissens im Erwerb der deutschen Verbstellung durch türkische und russische Kinder. In *Internationale Tendenzen der Syntaktik, Semantik und Pragmatik [International Trends in Syntax, Semantics and Pragmatics],* H.O. Spillmann and I. Warnke (eds.), 157-165. Peter Lang: Frankfurt.

Harrington, M. and Sawyer, M. 1992. L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14: 25-38.

Haznedar, B. 1997. L2 acquisition by a Turkish-speaking child: Evidence of L1 influence. In *Proceedings of the 21st Boston University Conference on Language Development,* E. Hughes, M. Hughes and A. Greenhill (eds.), 245-256. Cascadilla Press: Somerville, MA.

Hunt, K.W. 1970. *Syntactic Maturity in Schoolchildren and Adults*. University of Chicago Press, Chicago.

Hyltenstam, K. and Abrahamsson, N. 2003. Maturational constraints in SLA. In *The Handbook of Second Language Acquisition,* C. Doughty and M.H. Long (eds.), 539-588. Blackwell: Oxford.

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

Johnson, J.S. and Newport, E.L. 1991. Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition* 39: 215-258.

Johnston, J.R. 2001. An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language and Hearing Research* 44: 156-164.

Karmiloff-Smith, A. 1986. From metaprocess to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* 28: 95-147.

Kellerman, E. 1995. Age before beauty: Johnson and Newport revisited. In *The Current State of Interlanguage: Studies in Honor of William E. Rutherford,* L. Eubank, L. Selinker and M. Sharwood-Smith (eds.), 219-232. John Benjamins: Amsterdam.

Klee, T. and Fitzgerald, M.D. 1985. The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language* 12: 251-269.

Krashen, S. 1982. Accounting for child-adult differences in second language rate and attainment. In *Child-Adult Differences in Second Language Acquisition,* S. Krashen, R. Scarcella and M.H. Long (eds.), 202-206. Newbury House: Rowley, MA.

Lakshmanan, U. 1995. Child second language acquisition of syntax. *Studies in Second Language Acquisition* 17: 301-329.

Lalleman, J. 1986. *Dutch Language Proficiency of Turkish Children Born in The Netherlands*. Foris, Dordrecht.

Landis, J.R. and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174.

Larsen-Freeman, D.E. and Strom, V. 1977. The construction of a second language acquisition index of development. *Language Learning* 27: 123-134.

Larsen-Freeman, D.E. 1978. An ESL index of development. *TESOL Quarterly* 4: 439-448.

Larsen-Freeman, D.E. 1983. Assessing global second language proficiency. In *Classroom Oriented Research in Second Language Acquisition,* H.W. Seliger and M.H. Long (eds.), 287-304. Newbury House: Rowley, MA.

Lenneberg, E. 1967. *Biological Foundations of Language*. Wiley & Sons, New York.

Loban, W. 1976. *Language Development: Kindergarten Through Grade Twelve*. National Council of Teachers of English, Urbana, Illinois.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ.

Meisel, J.M., Clahsen, H. and Pienemann, M. 1981. On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition* 3: 109-135.

Miller, J.F. and Chapman, R.S. 1981. The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research* 24: 154-161.

Minifie, F.D., Darley, F.L. and Sherman, D. 1963. Temporal reliability of seven language measures. *Journal of Speech and Hearing Research* 6: 139-148.

Neeleman, A. and Reinhart, T. 1998. Scrambling and the PF Interface. In *The Projection of Arguments: Lexical and Compositional Factors,* M. Butt and W. Geuder (eds.), 309-353. CSLI Publications: Stanford.

Newport, E.L. 1990. Maturational constraints on language learning. *Cognitive Science* 14: 11-28.

Nice, M.M. 1925. Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology* 16: 370-379.

Ortega, L. 2000. Understanding Complexity: The Measurement of Change in the Syntax of Instructed L2 Spanish Learners. PhD dissertation. University of Hawai'i.

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24: 492-518.

Penfield, W. and Roberts, L. 1959. *Speech and Brain Mechanisms*. Princeton University Press, Princeton, NL.

Pica, T. and Long, M.H. 1986. The linguistic and conversational performance of experienced and inexperienced teachers. In *Talking to Learn: Conversation in Second Language Acquisition.,* R.R. Day (ed.). 85-98. Newbury House: Rowley, MA.

Pulvermüller, F. and Schumann, J.H. 1994. Neurobiological mechanisms of language acquisition. *Language Learning* 44: 681-734.

Richards, B. 1987. Type/Token Rations: What do they really tell us? *Journal of Child Language* 14: 201-209.

Rondal, J.A., Ghiotto, M., Bredart, S. and Bachelet, J.-F. 1987. Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language* 14: 433-446.

Scarborough, H.S., Rescorla, L., Tager-Flusberg, H., B., Fowler, A.E. and Sudhalter, V. 1991. The relation of utterance length to grammatical complexity in normal and language-disorderd groups. *Applied Psycholinguistics* 12: 23-45.

Schaeffer, J.C. 2000. *The Acquisition of Direct Object Scrambling and Clitic Placement: Syntax and Pragmatics.* John Benjamins, Amsterdam.

Schneider, W. 2002. Memory development in childhood. In *Handbook of Childhood Cognitive Development,* U. Goswami (ed.). 236-256. Blackwell: Oxford.

Schumann, J.H. 1975. Affective factors and the problem of age in second language acquisition. *Language Learning* 25: 209-235.

Schwartz, B.D. 1992. Testing between UG-based and problem-solving models of L2A: Developmental sequence data. *Language Acquisition* 2: 1-19.

Schwartz, B.D. 2003. Child L2 acquisition: Paving the way. In *Proceedings of the 27th Boston University Conference on Language Development,* B. Beachley, A. Brown and F. Conlin (eds.), 26-50. Cascadilla Press: Somerville, MA.

Schwartz, B.D. 2004. Why child L2 acquisition? In *Proceedings of Generative Approaches to Language Acquisition 2003,* J. Van Kampen and S. Baauw (eds.), 47-66. LOT Occasional Series: Utrecht, The Netherlands.

Schwartz, B.D. to appear. Unraveling inflection in child L2 development. *Morphology*.

Shriner, T.H. 1967. A comparison of selected measures with psychological scale values of language values. *Journal of Speech and Hearing Research* 10: 828-835.

Shriner, T.H. and Sherman, D. 1967. An equation for assessing language development. *Journal of Speech and Hearing Research* 10: 21-40.

Snow, C.E. and Hoefnagel-Höhle, M. 1982. The critical period for language acquisition: Evidence from second language learning. In *Child-Adult Differences in Second Language Acquisition,* S. Krashen, R. Scarcella and M.H. Long (eds.), 93-111. Newbury: Rowley, MA.

StatSoft, I. 2004. *Electronic Statistics Textbook*. Statsoft, Tulsa, OK. http://www.statsoft.com/textbook/stathome.html

Thomas, M. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44: 307-336.

Unsworth, S. 2004. On the syntax-semantics interface in Dutch: Adult and child L2 acquisition compared. *International Review of Applied Linguistics* 42: 173-187.

Unsworth, S. 2005. Child L2, Adult L2, Child L1: Differences and Similarities. A Study on the Acquisition of Direct Object Scrambling in Dutch. Phd thesis. Utrecht University.

Valian, V. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40: 21-81.

Van Geenhoven, V. 1998. *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic.* CSLI Publications, Stanford.

van Hout, R. and Vermeer, A. 1988. Spontane taaldata en het meten van lexicale rijkdom in tweede-taalverwerving. [Spontaneous speech data and measuring methods of lexical richness in second language acquisition.]. *Toegepaste Taalwetenschap in Artikelen* 32: 108-122.

Verhoeven, L. and Vermeer, A. 1989. *Diagnose van Kindertaal. Nederlandse Taalvaardigheid van Autochtone en Allochtone Kinderen. [Child Language Diagnosis. Dutch Language Proficiency of Native and Non-Native Children]*. Zwijsen, Tilburg.

Verhoeven, L. and Vermeer, A. 2001. *Taaltoets Alle Kinderen. Diagnostische toets voor de mondelinge vaardigheid Nederlands bij kinderen van groep 1 tot en met 4. [Language Test All Children. Diagnostic test for Dutch oral proficiency of children from grade 1 to grade 4.]*. Cito, Arnhem.

Vermeer, A. 1986. Tempo en Stuktuur van Tweede-taalvewerving bij Turkse en Marokkaanse Kinderen [Tempo and Structure of Second Language Acquisition of Turkish and Moroccan Children]. PhD dissertation. University of Tilburg.

Weerman, F., Bisschop, J. and Punt, L. 2003. L1 and L2 acquisition of Dutch adjectival inflection. Paper presented at Generative Approaches of Language Acquisition 2003, Utrecht, The Netherlands [4th September]

Whong-Barr, M. and Schwartz, B.D. 2002. Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition* 22: 579-616.

· This paper is based on Chapters 4 and 5 of my PhD thesis (Unsworth 2005), completed with financial support from Utrecht institute of Linguistics (OTS), which I gratefully acknowledge. I would like to thank my supervisors Peter Coopmans and Bonnie D. Schwartz for their comments on these chapters, as well as Hugo Quené for his help with principal components analysis. Thanks also to two anonymous reviewers.

[1] A secondary assumption is that the general problem-solving principles in question are only relevant to adults (because (L1/L2) children make use of UG).

[2] It might be objected that similar developmental sequences could also be the result of both groups using problem-solving or general learning mechanisms. If this were the case, we would expect the more cognitively mature adult L2ers to be more successful than the less cognitively mature child L2ers (Schwartz 1992:8, fn.6). As mentioned above, the L2 data indicate otherwise, however.

[3] Phonology is not included as part of proficiency here. This is not unusual. For example, cloze tests, which are regularly used to assess (adult) L2 proficiency do not assess phonology.

[4] This definition focuses on knowledge, that is, linguistic competence, rather than use/control (although when testing knowledge some aspect of use/control will inevitably be involved) and hence it excludes pragmatic skills, many of which would be classified as properties of language use.

[5] Length of exposure could also be used as a basis for equivalence between L2 children and L2 adults. Given the potential differences in rate of acquisition between children and adults (Asher and Price 1967; Snow and Hoefnagel-Höhle 1982), as well as between individual L2ers in general, this variable would be unsuitable for this purpose, however. An anonymous reviewer argues that using proficiency as the basis of L2 child ~ L2 adult comparisons masks important differences in acquisitional speed which *would* be captured were the two groups compared using length of exposure. Determining the relative rates of acquisition for L2 children and adults is not the goal of the present study, however. In fact, as Hyltenstam and Abrahamsson (2003:547) point out, the informative value of acquisitional speed with respect to age effects in L2 acquisition, and more specifically, with respect to the existence of a critical period, is unclear. Any comparison of L2 children and adults matched for length of exposure would undoubtedly

involve comparing L2ers at different proficiency levels and this could lead to the kinds of problems outlined in the main text. A comparison of L2 children and adults matched on proficiency may well include learners with varying lengths of exposure, but if rate of acquisition is not the focus of enquiry, this is essentially irrelevant. Having said that, the set-up proposed in this paper can be used to capture potential rate differences if desired: one could simply determine the length of exposure for the children compared to the adults at the same proficiency level.

[6] An extended definition of explicit knowledge is provided by Ellis (2004:244): "Explicit L2 knowledge is the declarative and often anomalous knowledge of the phonological, lexical, grammatical, pragmatic, and sociocritical features of an L2 together with the metalanguage for labelling knowledge. It is held consciously and is learnable and verbalisable. It is typically accessed through controlled processing when L2 learners experience some kind of linguistic difficulty in the use of the L2. Learners vary in the breadth and depth of their L2 explicit knowledge."

[7] This is not to say of course that metalinguistic knowledge is not used in more naturalistic tasks (for example, by 'monitoring' or applying rules online, etc.) or that this does not vary across speakers, but more naturalistic tasks at least do not *encourage* the use of such knowledge.

[8] The journals were: *Applied Linguistics*, *Language Learning*, *Second Language Research* and *Studies in Second Language Acquisition.* All volumes from 1988 until 1992 were surveyed.

[9] The score developed here is based on production data only, but the assumption is that this task taps the L2ers' underlying linguistic competence.

[10] What follows is an abbreviated version of Unsworth (2005 Chapter 4).

[11] See also Appel (1984) and Verhoeven and Vermeer (1989) for comparative discussion of a variety of measures of morphosyntactic complexity in child L2 Dutch.

[12] These children and adults functioned as controls for the study discussed in § 7.

[13] It is not possible to carry out an ANOVA with these data because they do not conform to the assumptions of equal variances (Levene's statistic = 5.136, $p = .006$).

[14] Other alternatives include rate of subordination (Hunt 1970; Verhoeven and Vermeer 1989) and rate of verbal utterances (Belletti and Hamann 2000; Valian 1991); these are unsuitable measures of proficiency, however, because they do not sufficiently differentiate proficiency levels (see Unsworth 2005:183-184 for relevant discussion).

[15] It could be objected that including auxiliaries would unfairly favour subjects who produce periphrastic tenses over subjects who use synthetic tenses only. This objection does not hold, however. Firstly, the use of a variety of tenses is an indication of grammatical sophistication, which is what verbal density is designed to measure. Secondly, excluding auxiliaries would mean that learners who used the more complex structures, such as (1)-d, which indicate grammatical complexity rather than simply choice of tense, would not be credited for this.

[16] The non-parametric Mann-Whitney test is used for data which do not conform to the prerequisites for a t-test (Levene's statistic = 7.154, p = .011).

[17] One caveat concerning the native-speaker adult data is in order. The standard deviations given in Table 2 indicate that, as was observed for MLUw, the variation in the adult data is almost twice as much as in the child groups. This is less than ideal, but it is my opinion that the advantages of this measure over the alternatives outweigh this potential disadvantage.

[18] Note that the claim regarding the lack of difference between these three child L1 groups is only intended to hold for this particular task: in this context, where learners have to describe the same sets of pictures, these three groups produce a similar range of lexical items, one which is significantly more restricted than the range of items produced by adults on the same task. I wish to make no claims regarding L1 vocabulary development in general.

[19] See e.g. Larsen-Freeman and Stom (1977) for arguments against alternative measures such as the proportion or number of errors.

[20] An anonymous reviewer notes that adults also vary in terms of their working memory capacities and that working memory has furthermore been shown to correlate with L2 proficiency (e.g. Harrington and Sawyer 1992). This observation may indeed account for the variability attested in both the MLU and verbal density data amongst the native adults (cf. fn 17). However, the point at issue here is that the L2 adults' cognitive capacities, including working memory, are not capped as a consequence of their maturational state in the same way as those of L2 children.

[21] Cohen's Kappa statistic measures the extent to which two (or more) coders' judgements agree while taking into account the likelihood that any agreement would be due to chance (see Landis and Koch 1977).

[22] I thank an anonymous reviewer for pointing this out.

[23] A regression analysis for lexical diversity indicates that both age at time of testing and length of exposure contribute roughly equally to the observed scores (r^2 = .460, ANOVA F = 26.418, p = .000; standardised coefficients = .338 for age at time of testing and .437 for length of exposure).

[24] Two more components were extracted, but these had eigenvalues of less than 1.00. This means that they explain less variance than the original variables. In other words, everything common to the original variables is contained within the first component. (See StatSoft (2004) for more details about eigenvalues.)

[25] Ideally, this would be confirmed by comparing the ASCOPS scores with those obtained using one of the standardised tests listed in §5.3.

[26] Once the data have been transcribed, however, they can provide a useful source of data for further purposes.

[27] There were 25 L2 children. Their age at first exposure to Dutch was between 4;0 and 7;1 (mean = 5;6; SD = 1;0), their age at the time of testing ranged from 5;3 to 17;4 (mean = 9;3; SD = 2;4) and their length of exposure from 0;2 to 13;0 (mean = 3;8; SD = 2;6). There were 23 L2 adults. Their age of first exposure to Dutch was between 8;0 and 32;0 (mean = 19;3; SD = 8;8), their age at the time of testing between 10 and 50 years (mean = 23;10; SD = 11;0), and their length of exposure from 3 months to 27 years (mean = 4;4; SD = 6;0).

[28] This includes two L2 children and one L2 adult whose proficiency in Dutch was so low, they were unable to carry out the picture description task.

References

Adamson, H.D. 1988. *Variation Theory and Second Language Acquisition*. Georgetown University Press, Washington D.C.

Appel, R. 1984. *Immigrant children learning Dutch. Sociolinguistic and psycholinguistic aspects of second-language acquisition*. Foris, Dordrecht.

Asher, J. and Price, B. 1967. The learning strategy of total physical response: Some age differences. *Child Development* 38: 1218-1227.

Belletti, A. and Hamann, C. 2000. Ça on fait pas! On the L2-Acquisition of French by two young children with different source languages. In *Proceedings of the 24th Boston*

University Conference on Language Development, S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.), 116-127. Cascadilla Press: Somerville, MA.

Berman, R.A. 2007. Developing linguistic knowledge and language use across adolescence. In *Handbook of Language Development,* E. Hoff and M. Schatz (eds.), 347-367. Blackwell: Oxford.

Bialystok, E. 1993. Metalinguistic awareness: The development of children's representations of language. In *Systems of Representation in Children: Development and Use,* C. Pratt and A.F. Garton (eds.), 211-233. John Wiley and Sons: Chichester.

Bialystok, E. and Hakuta, K. 1994. *In Other Words: The Science and Psychology of Second Language Acquisition*. Basic Books, New York.

Bialystok, E. 1997. The structure of age: In search of barriers to second language acquisition. *Second Language Research* 13: 116-137.

Bialystok, E. and Miller, B. 1999. The problem of age in second-language acquisition: Influences from language, structure and task. *Bilingualism: Language and Cognition* 2: 127-145.

Birdsong, D. 1999. Introduction: Whys and why nots of the Critical Period Hypothesis for second language acquisition. In *Second Language Acquisition and the Critical Period Hypothesis,* D. Birdsong (ed.). 1-22. Lawrence Erlbaum: Mahwah, NJ.

Birdsong, D. and Molis, M. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44: 235-249.

Blake, J., Quartaro, G. and Onorati, S. 1993. Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language* 20: 139-152.

Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? In *Linguistic Perspectives on Second Language Acquisition,* S.M. Gass and J. Schachter (eds.), 42-68. Cambridge University Press: Cambridge.

Blom, E. and Polišenskà, D. 2006. Verbal inflection and verb placement in first and second language acquisition. In *Proceedings of the 29th Linguistics Colloquium,* M. Vliegen (ed.). 1-10. Peter Lang: Amsterdam.

Bol, G.W. 2003. MLU-matching and the production of morphosyntax in Dutch children with Specific Language Impairment. In *Language Competence Across Populations: Toward a Definition of Specific Language Impairment.,* Y. Levy and J.C. Schaeffer (eds.), 259-271. Lawrence Erlbaum: Mahway, NJ.

Broeder, P., Extra, G. and van Hout, R. 1993. Richness and variety in the developing lexicon. In *Adult language acquisition: Cross-linguistic perspectives.,* C. Perdue (ed.). 145-163. Cambridge University Press: Cambridge.

Brown, R. 1973. *A First Language*. Harvard University Press, Cambridge, MA.

Chabon, S.S., Kent-Udolf, L. and Egolf, D.B. 1982. The temporal reliability of Brown's mean length of utterance (MLU-m) measure with post-stage V children. *Journal of Speech and Hearing Research* 25: 117-124.

Chaudron, C. and Parker, K. 1990. Discourse markedness and structural markedness. The acquisition of English noun phrases. *Studies in Second Language Acquisition* 12: 43-64.

Clahsen, H. 1986. Verb inflections in German child language: acquisition of agreement markings and the functions they encode. *Linguistics* 24: 79-121.

Clahsen, H. 1988. Critical phases of grammar development. A study of the acquisition of negation in children and adults. In *Language Development,* P. Jordens and J. Lalleman (eds.), 123-148. Foris: Dordrecht.

Conant, S. 1987. The relationship between age and MLU in young children: A second look at Klee and Fitzgerald's data. *Journal of Child Language* 14: 169-173.

Curtiss, S. and Yamada, J. 1985. *Comprehensive Language Evaluation - Receptive Measures (CYCLE-R) and the Comprehensive Language Evaluation - Elicited Production (CYCLE-E)*.

de Hoop, H. 1992. Case Configuration and Noun Phrase Interpretation. Phd dissertation, published (1996) New York: Garland. University of Groningen.

De Jong, J.H.A.L. and Verhoeven, L. 1992. *The Construct of Language Proficiency*. John Benjamins, Amsterdam.

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

Dewaele, J.-M. 2000. Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée [Seize the unseizable? Measures of Utterance Length and Applied Linguistics]. *International Review of Applied Linguistics* 38: 17-33.

Dulay, H. and Burt, M. 1974. Natural sequences in child second language acquisition. *Language Learning* 24: 37-53.

Ellis, R. 2004. The definition and measurement of L2 explicit knowledge. *Language Learning* 54: 227-275.

Ervin-Tripp, S.M. 1978. Is second language learning like the first? In *Second Language Acquisition. A Book of Readings,* E.M. Hatch (ed.). 190-206. Newbury House: Rowley, MA.

Felix, S.W. 1985. More evidence on competing cognitive systems. *Second Language Research* 1: 47-72.

Gavens, N. and Barrouillet, P. 2004. Delays of retention, processing efficiency, and attentional resources in working memory span development. *Journal of Memory and Language* 51: 644-657.

Gilkerson, J. 2005. Testing for UG Access: An Investigation of How Native Spanish Speakers Acquire English Particle Verbs. Unpublished doctoral dissertation. UCLA.

Guiraud, P. 1960. *Problèmes en méthodes de la statistique linguistique*. D. Reidel, Dordrecht.

Haberzettl, S. 1999. *Katze Maus essen* vs. *Katze essen Maus*: Die L1 als Königs- oder Holzwegbereiter zur L2? Zum Einfluss des L1-Wissens im Erwerb der deutschen Verbstellung durch türkische und russische Kinder. In *Internationale Tendenzen der Syntaktik, Semantik und Pragmatik [International Trends in Syntax, Semantics and Pragmatics],* H.O. Spillmann and I. Warnke (eds.), 157-165. Peter Lang: Frankfurt.

Harrington, M. and Sawyer, M. 1992. L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14: 25-38.

Haznedar, B. 1997. L2 acquisition by a Turkish-speaking child: Evidence of L1 influence. In *Proceedings of the 21st Boston University Conference on Language Development,* E. Hughes, M. Hughes and A. Greenhill (eds.), 245-256. Cascadilla Press: Somerville, MA.

Hunt, K.W. 1970. *Syntactic Maturity in Schoolchildren and Adults*. University of Chicago Press, Chicago.

Hyltenstam, K. and Abrahamsson, N. 2003. Maturational constraints in SLA. In *The Handbook of Second Language Acquisition,* C. Doughty and M.H. Long (eds.), 539-588. Blackwell: Oxford.

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

Johnson, J.S. and Newport, E.L. 1991. Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition* 39: 215-258.

Johnston, J.R. 2001. An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language and Hearing Research* 44: 156-164.

Karmiloff-Smith, A. 1986. From metaprocess to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* 28: 95-147.

Kellerman, E. 1995. Age before beauty: Johnson and Newport revisited. In *The Current State of Interlanguage: Studies in Honor of William E. Rutherford,* L. Eubank, L. Selinker and M. Sharwood-Smith (eds.), 219-232. John Benjamins: Amsterdam.

Klee, T. and Fitzgerald, M.D. 1985. The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language* 12: 251-269.

Krashen, S. 1982. Accounting for child-adult differences in second language rate and attainment. In *Child-Adult Differences in Second Language Acquisition,* S. Krashen, R. Scarcella and M.H. Long (eds.), 202-206. Newbury House: Rowley, MA.

Lakshmanan, U. 1995. Child second language acquisition of syntax. *Studies in Second Language Acquisition* 17: 301-329.

Lalleman, J. 1986. *Dutch Language Proficiency of Turkish Children Born in The Netherlands*. Foris, Dordrecht.

Landis, J.R. and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174.

Larsen-Freeman, D.E. and Strom, V. 1977. The construction of a second language acquisition index of development. *Language Learning* 27: 123-134.

Larsen-Freeman, D.E. 1978. An ESL index of development. *TESOL Quarterly* 4: 439-448.

Larsen-Freeman, D.E. 1983. Assessing global second language proficiency. In *Classroom Oriented Research in Second Language Acquisition,* H.W. Seliger and M.H. Long (eds.), 287-304. Newbury House: Rowley, MA.

Lenneberg, E. 1967. *Biological Foundations of Language*. Wiley & Sons, New York.

Loban, W. 1976. *Language Development: Kindergarten Through Grade Twelve*. National Council of Teachers of English, Urbana, Illinois.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ.

Meisel, J.M., Clahsen, H. and Pienemann, M. 1981. On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition* 3: 109-135.

Miller, J.F. and Chapman, R.S. 1981. The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research* 24: 154-161.

Minifie, F.D., Darley, F.L. and Sherman, D. 1963. Temporal reliability of seven language measures. *Journal of Speech and Hearing Research* 6: 139-148.

Neeleman, A. and Reinhart, T. 1998. Scrambling and the PF Interface. In *The Projection of Arguments: Lexical and Compositional Factors,* M. Butt and W. Geuder (eds.), 309-353. CSLI Publications: Stanford.

Newport, E.L. 1990. Maturational constraints on language learning. *Cognitive Science* 14: 11-28.

Nice, M.M. 1925. Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology* 16: 370-379.

Ortega, L. 2000. Understanding Complexity: The Measurement of Change in the Syntax of Instructed L2 Spanish Learners. PhD dissertation. University of Hawai'i.

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24: 492-518.

Penfield, W. and Roberts, L. 1959. *Speech and Brain Mechanisms*. Princeton University Press, Princeton, NL.

Pica, T. and Long, M.H. 1986. The linguistic and conversational performance of experienced and inexperienced teachers. In *Talking to Learn: Conversation in Second Language Acquisition.,* R.R. Day (ed.). 85-98. Newbury House: Rowley, MA.

Pulvermüller, F. and Schumann, J.H. 1994. Neurobiological mechanisms of language acquisition. *Language Learning* 44: 681-734.

Richards, B. 1987. Type/Token Rations: What do they really tell us? *Journal of Child Language* 14: 201-209.

Rondal, J.A., Ghiotto, M., Bredart, S. and Bachelet, J.-F. 1987. Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language* 14: 433-446.

Scarborough, H.S., Rescorla, L., Tager-Flusberg, H., B., Fowler, A.E. and Sudhalter, V. 1991. The relation of utterance length to grammatical complexity in normal and language-disorderd groups. *Applied Psycholinguistics* 12: 23-45.

Schaeffer, J.C. 2000. *The Acquisition of Direct Object Scrambling and Clitic Placement: Syntax and Pragmatics.* John Benjamins, Amsterdam.

Schneider, W. 2002. Memory development in childhood. In *Handbook of Childhood Cognitive Development,* U. Goswami (ed.). 236-256. Blackwell: Oxford.

Schumann, J.H. 1975. Affective factors and the problem of age in second language acquisition. *Language Learning* 25: 209-235.

Schwartz, B.D. 1992. Testing between UG-based and problem-solving models of L2A: Developmental sequence data. *Language Acquisition* 2: 1-19.

Schwartz, B.D. 2003. Child L2 acquisition: Paving the way. In *Proceedings of the 27th Boston University Conference on Language Development,* B. Beachley, A. Brown and F. Conlin (eds.), 26-50. Cascadilla Press: Somerville, MA.

Schwartz, B.D. 2004. Why child L2 acquisition? In *Proceedings of Generative Approaches to Language Acquisition 2003,* J. Van Kampen and S. Baauw (eds.), 47-66. LOT Occasional Series: Utrecht, The Netherlands.

Schwartz, B.D. to appear. Unraveling inflection in child L2 development. *Morphology*.

Shriner, T.H. 1967. A comparison of selected measures with psychological scale values of language values. *Journal of Speech and Hearing Research* 10: 828-835.

Shriner, T.H. and Sherman, D. 1967. An equation for assessing language development. *Journal of Speech and Hearing Research* 10: 21-40.

Snow, C.E. and Hoefnagel-Höhle, M. 1982. The critical period for language acquisition: Evidence from second language learning. In *Child-Adult Differences in Second Language Acquisition,* S. Krashen, R. Scarcella and M.H. Long (eds.), 93-111. Newbury: Rowley, MA.

StatSoft, I. 2004. *Electronic Statistics Textbook*. Statsoft, Tulsa, OK. http://www.statsoft.com/textbook/stathome.html

Thomas, M. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44: 307-336.

Unsworth, S. 2004. On the syntax-semantics interface in Dutch: Adult and child L2 acquisition compared. *International Review of Applied Linguistics* 42: 173-187.

Unsworth, S. 2005. Child L2, Adult L2, Child L1: Differences and Similarities. A Study on the Acquisition of Direct Object Scrambling in Dutch. Phd thesis. Utrecht University.

Valian, V. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40: 21-81.

Van Geenhoven, V. 1998. *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic.* CSLI Publications, Stanford.

van Hout, R. and Vermeer, A. 1988. Spontane taaldata en het meten van lexicale rijkdom in tweede-taalverwerving. [Spontaneous speech data and measuring methods of lexical richness in second language acquisition.]. *Toegepaste Taalwetenschap in Artikelen* 32: 108-122.

Verhoeven, L. and Vermeer, A. 1989. *Diagnose van Kindertaal. Nederlandse Taalvaardigheid van Autochtone en Allochtone Kinderen. [Child Language Diagnosis. Dutch Language Proficiency of Native and Non-Native Children]*. Zwijsen, Tilburg.

Vermeer, A. 1986. Tempo en Stuktuur van Tweede-taalvewerving bij Turkse en Marokkaanse Kinderen [Tempo and Structure of Second Language Acquisition of Turkish and Moroccan Children]. PhD dissertation. University of Tilburg.

Weerman, F., Bisschop, J. and Punt, L. 2003. L1 and L2 acquisition of Dutch adjectival inflection. Paper presented at Generative Approaches of Language Acquisition 2003, Utrecht, The Netherlands [4th September]

Whong-Barr, M. and Schwartz, B.D. 2002. Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition* 22: 579-616.