GCS: A grammatical coding system for natural language data

SUSAN CURTISS
University of California, Los Angeles, California

JEFF MacSWAN
Arizona State University, Tempe, Arizona

JEANNETTE SCHAEFFER
Ben Gurion University of the Negev, Beer-Sheva, Israel

MURAT KURAL
University of California, Irvine, California

and

TETSUYA SANO
Meiji Gakuin University, Tokyo, Japan

This article presents a rationale and description of GCS, or Grammatical Coding System. GCS is a general-use grammatical coding system designed for research on the language of normal and language-impaired children or adults and is especially useful for studies in which a relatively large number of participants are involved. It implements recent theoretical developments in linguistics to characterize development and/or language disorder in children and adults. In addition to the coding system, a computerized method for reading coded transcripts and calculating relevant descriptive statistics is presented. A full coded transcription is included in the Appendix. A detailed GCS manual may be downloaded from www.psychonomic.org/archive.

In this article, we describe GCS, an acronym for Grammatical Coding System; GCS is designed as a general-use grammatical coding system appropriate for research on the language of normal and language-impaired children or adults. GCS is intended for use in any study concerned with grammatical development and is especially useful for studies in which a relatively large number of participants are involved. It takes advantage of recent theoretical developments in the linguistic sciences to characterize development and/or language disorder in children and adults. In addition to our coding system, we present a computerized method for reading coded transcripts and calculating relevant descriptive statistics.

This article is organized as follows. The first section describes the scientific context in which GCS was developed. The second section outlines the theoretical framework that guided us in developing the particular coding conventions used in GCS and discusses specific ways in which MacWhinney's (2000) CHAT system needed to be extended in order to meet our needs. The third section presents GCS, our coding system, which includes examples of coded utterances throughout. The fourth section outlines our computerized analysis system. A full coded transcription is included in the Appendix.

GCS'S DEVELOPMENT CONTEXT

The research project for which GCS was developed investigated the neurology of language acquisition, both the lateralization and localization of language during development in the normal case, as well as in children who have severe brain damage. This research program has specifically examined language development in children with medically intractable epilepsy, in treatment of which they have undergone surgical resection of the diseased tissue. The surgeries have ranged from unilobar resections (e.g., temporal lobectomy) to multilobar resections (e.g., temporal-parietal-occipital lobectomy) or, in more extreme cases, hemidecortication (removal or disconnection of one cerebral hemisphere, often referred to as hemispherectomy). The effects on language acquisition of disease and removal of different parts of the left or right hemisphere at different ages have then been examined and compared—

left versus right, one region versus another, one etiology versus another, one age versus another, and, importantly, brain-damaged child versus normally developing child.

The research has focused on several topics, including (1) the capacity of each isolated developing hemisphere to subserve lexical and grammatical development alone (a comparison of left-hemispherectomized and right-hemispherectomized children with each other and with normally developing children), (2) the development of lateralization and localization of grammar (specifically, syntax and morphology), as opposed to lexicon, (3) the effects of brain damage on specific functional subsystems of the grammar (namely, the D[eterminer]-system, the I[nflectional]-system, and the C[omplementizer]-system) and other core aspects of syntax, (4) the effects of localized brain damage on the lexicon (the establishment of a mental dictionary of content words and their interrelations), as opposed to syntax and morphology, and (5) maturational constraints on the acquisition of grammar—again, syntax and morphology. The research has been part of a multidisciplinary investigation regarding whether there is a systematic association between specific patterns of linguistic delay or anomaly and specific neuropathologies.

An obvious major component of this research had to be a detailed grammatical analysis of the language of the children in the study, both normal and brain damaged. Such analysis required a rich, theoretically informed and motivated system for explicitly describing and defining the internal structure and content of language. Thus, although both formal test performance and language samples were used to evaluate language performance, it was for the analysis of the spontaneous speech samples that GCS was developed. Aspects of this research may be reviewed in Caplan, Curtiss, Chugani, and Vinters (1996), Curtiss and de Bode (1998, 1999a, 1999b, 2003), Curtiss, de Bode, and Mathern (2001), Curtiss and Schaeffer (1997a, 1997b, 2004), de Bode (1998), and de Bode and Curtiss (1999).

THEORETICAL BACKGROUND

The CHILDES System

In our search for a satisfactory system with which to code our data, CHILDES (Child Language Data Exchange System; MacWhinney, 1991, 1995, 2000) came closest to our needs. CHILDES is a computerized database system containing child language, aphasia and second-language corpora. It consists of three basic components: (1) the raw data, (2) the CHAT (Codes for the Human Analysis of Transcripts) system, and (3) the CLAN (Computerized Language Analysis) system.

The raw data consist of the literal transcription of recorded speech onto the so-called *main tier*, or the *speaker tier*. Each speaker tier starts with an asterisk (*), followed by a three (capital) letter speaker ID and a colon. Since we were interested mainly in the morphological and syntactic aspects of our data (and not in phonological or discourse characteristics), we adopted the basics of the main tier coding as proposed in CHILDES (including

false starts, unintelligible speech, punctuation, unfinished sentences, overlap, etc.) and postponed the coding of properly linguistic features to the dependent tiers, as will be discussed below.

CHAT coding conventions allow one to create as many dependent tiers as needed. The coding conventions of GCS differ from the ones proposed in CHAT, perhaps due primarily to a difference in focus and theoretical orientation. The GCS coding system allows researchers to mark numerous distinctions and generalizations that are not generally marked in the present CHILDES database. Since the analysis of morphological and syntactic structure is central to our study, our coding scheme adds three dependent coding tiers to the main speaker tier: (1) the morphological tier, (2) the syntactic tier, and (3) the lexical tier. In addition, GCS uses a comment tier, in which anything relevant (either phonological, discourse related, or visual) for the interpretation of the utterance may be expressed. Just as in CHAT, our dependent tiers all start with the percentage symbol, followed by a three-letter (lowercase) code, followed by a colon.

Insofar as GCS uses existing CHAT transcription format and is amenable to existing CLAN commands, we offer it as a useful extension to existing CHILDES capabilities. Indeed, MacWhinney and Snow (1990), describing plans for future modifications to CHILDES, invite extension:

> We encourage other researchers to join us in working toward these new goals, to make full use of the current CHILDES tools, and to propose new directions and possible improvements to the system. (p. 472)

Let us now turn to a discussion of the theoretical framework that guided us in creating GCS.

Linguistic Theoretical Framework

The conventions of GCS are theoretically grounded in and motivated by the principles and parameters theory outlined in Chomsky (1995) and subsequent work. Particularly relevant to the study of language acquisition, perhaps, is the promise of a syntactic theory in which parameters are restricted to morphological properties of the lexicon, as Chomsky (1991) has noted:

> If there were only one human language, the story would essentially end there. But we know that this is false, a rather surprising fact. The general principles of the initial state evidently allow a range of variation. Associated with many principles there are parameters with a few—perhaps just two—values. Possibly, as proposed by Hagit Borer, the parameters are actually restricted to the lexicon, which would mean that the rest of the I-language is fixed and invariant, a far-reaching idea that has proven quite productive. (p. 23)

Restricting parameters to the lexicon means that linguistic variation falls out of just the morphological properties (abstract and concrete) of the lexicon (Borer, 1984). In this model, there are two central components: C_{HL}, a computational system for human language, which is presumed to be invariant across languages, and a lexicon, to

which the idiosyncratic differences observed across languages are attributed. The suggestion that the I-language is fixed and invariant in this way introduces a version of the universal base hypothesis, the notion that phrase structure does not vary across languages; surface differences in word order relate only to the rearrangement of elements in the syntactic tree as the result of movement operations, triggered by lexically encoded morphological features.

Phrase structure is also derived from the lexicon in this framework. An operation, which Chomsky (1995) calls *Select*, picks items from the lexicon and introduces them into the numeration, an assembled subset of the lexicon used to construct a derivation. Another operation, *Merge*, takes items from the numeration and forms new, hierarchically arranged syntactic objects (substructures). The *Move* operation applies to syntactic objects formed by *Merge* to build new structures. In the minimalist program, then, phrase structure trees are built derivationally by the application of the *Select*, *Merge*, and *Move* operations, constrained only by the condition that lexically encoded features match in the course of a derivation. Phrase structure, along with configurationally defined intermediate and maximal projections, therefore has no independent status in the grammatical system (C_{HL}).

In Chomsky (1995), movements are driven by feature checking and may be of two types: A head may undergo *head movement* and adjoin to another head, or a maximal projection may move to the specifier position of a head. In either case, the element moves for the purpose of checking morphological features of case, number, person, and gender. In addition, its movement may be *overt* or *covert*. Overt movements are driven by *strong* features and are visible at *phonetic form* (PF, traditionally known as the *surface structure*) and *logical form* (LF, the interpretive level). Covert movements, driven by *weak* features, are visible only at LF.

Principles of economy select among convergent derivations. One such principle, full interpretation (FI), requires that no symbol lacking a sensorimotor interpretation be admitted at PF. Applied at LF, FI entails that "every element of the representation have a (language-independent) interpretation" (Chomsky, 1995, p. 27). Thus, uninterpretable features (denoted [-Interpretable]) must be checked and (in some proposals) deleted by LF. Such features include case, person, number, and gender. A derivation is said to *converge at an interface level* (PF or LF) if it satisfies FI at that level; it *converges* if FI is satisfied at both levels. A derivation that does not converge is also referred to as one that *crashes*. If features are not checked, the derivation *crashes*; if they *mismatch*, the derivation is canceled (i.e., a different convergent derivation may not be constructed).

Particularly within recent work in the principles and parameters framework, functional categories, which host elements such as complementizers (Cs), verbal inflections (Is), and determiners (Ds), are of crucial importance (Abney, 1987; Borer, 1984; Chomsky, 1995). The C, I, and D categories are lexical heads that project up to maximal projections under merge, as is illustrated in Figure 1.

As is shown in Figure 1, the maximal projections of the functional categories C, I, and D are CP, IP, and DP, respectively. Each maximal projection dominates a head (C, I, and D) and, when required, a complement (XP). Complementizers, Wh-elements, relativizing elements, and moved auxiliaries (interrogative formation) occupy C. Auxiliaries are base-generated in I, which also hosts phi-features (person, number, and gender), case, agreement, and tense features that trigger verb movement. Determiners occupy a D-position. Instantiations of D include determiners and nominal plural formation.

The morphological properties that drive movements within the system, thus accounting for cross-linguistic variation, are associated with functional categories that attract particular lexical categories in order to check morphological features. Thus, I triggers movement of V (=verb) to I, and D triggers movement of N (=noun) to D, to check feature agreement. DP moves into the specifier position of I in order to check nominative case, or into the specifier position of *v* (=light verb) to check accusative case. Similarly, in our coding system, functional categories are marked in all instances, along with other important, lexically encoded information.

Researchers interested in a more detailed discussion of the minimalist program are referred to Chomsky (1995), Hendrick (2003), Radford (1997), and Webelhuth (1995). Researchers less familiar with generative grammar may benefit by first reading relevant sections of Fromkin (2000).

Validity

The validity of our coding system is tied to an external criterion—namely, linguistic theory, developed out of a rich history of empirical inquiry (for a review, see Chomsky, 1995). An important subcomponent of validity is reliability, the degree to which repeated coding events of the same transcript will yield similar measures.
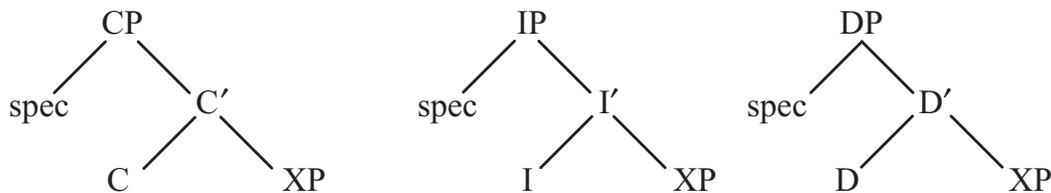
Figure 1. Projections of functional categories C, I, and D.

Two sorts of judgments are required by coders that might lead to inconsistency in coding and, hence, pose a threat to validity for any linguistic coding system. These include (1) a judgment regarding the grammaticality of a phrase or utterance and (2) a judgment regarding the correct structural or interpretative analysis of a phrase or utterance. If different coding events for the same transcript involve different grammaticality judgments on the part of coders, scores will differ with respect to the measure of error in the respective structure or category under analysis. If different coding events for the same transcript involve different structural analyses of utterances, scores will differ with respect to the measure of total occurrences of one or another particular grammatical structure or category.

To guard against the first threat to validity—involving grammaticality judgments—we invoke Labov's (1975) consensus principle and clear case principle:

The consensus principle. If there is no reason to think otherwise, assume that the judgments of any native speaker are characteristic of all speakers of the language. (Labov, p. 31)

Transcripts must be coded by native speakers, and error analysis must be proofed by at least one other native speaker who has been trained in the coding system. When judgments differ, further study of a transcript or speech community is warranted.

The clear case principle. Disputed judgments should be shown to include at least one consistent pattern in the speech community or be abandoned. If differing judgments are said to represent different dialects, enough investigation of each dialect should be carried out to show that each judgment is a clear case in that dialect. (Labov, p. 31)

At times, linguistic theory will assist in deciding such disputes, as Chomsky (1957) has suggested:

In many intermediate cases we shall be prepared to let the grammar itself decide, when the grammar is set up in the simplest way so that it includes the clear sentences and excludes the clear non-sentences. This is a familiar feature of explication. (Chomsky, 1957, p. 14)

Here we differ with Labov, who believes that the judgments of the experimenter should always be excluded when intuitions differ (1975, p. 31). (See Newmeyer, 1983, and Schütze, 1996, for further discussion.)

In short, we appeal to consensus among coders to decide cases in which grammaticality judgments differ. If consensus cannot be established through further study of the speech community of interest, the disputed datum is excluded from the corpus. In practical terms, such cases have been rare in our experience and would be of no statistical significance in studies involving sufficiently large sample sizes.

The second threat to validity, in which coders might come to different structural analyses of the same utterance, may be addressed in a way quite similar to the first:

A consensus of trained linguists is sought. If consensus cannot be obtained after discussion and further study of the linguistic community of interest, the disputed datum should be eliminated from the corpus. Again, in practical terms, we have found extremely few cases of this nature, so that no statistical significance would obtain in studies involving sufficiently large sample sizes.

In the next section, we will illustrate GCS, our coding system, and show how it differs from standard CHAT conventions.

CODING THE UTTERANCES

All linguistic coding is done on the dependent tiers, which include %mor: (morphology), %syn: (syntax), and %lex: (lexicon). In this section, we will present our coding system for the morphological tier, syntactic tier, and lexical tier and then will briefly discuss some advantages of the GCS system over CHAT.

Morphological Tier

Because of the centrality of functional structure in modern generative grammar, the analysis of functional categories was similarly of central importance in our study, as it has been for numerous other researchers in normal (Hyams, 1996; Radford, 1995; Sinka & Schelletter, 1998; Wexler, 1998) and impaired (e.g., Jakubowicz, Durand, Rigaut, & van der Velde, 2001; Leonard, 1998) language development. For this reason, our coding system was developed with a focus on the analysis of these items. Thus, on the morphological tier (%mor:), morphemes/elements related to the functional heads C, I, and D are labeled with codes that consist of information about the functional category and the grammatical function the morpheme/element fulfills. Some examples are given in (1) and (2).

(1) *she* is a **S**ubject (grammatical function) **PRO**noun with nominative case marking, which undergoes feature checking with the functional head I. It is thereby coded: **IPROS**|she

(2) *my*, the **POS**sessive **D**eterminer related to the functional head **D** is coded **DPOSD**|my

GCS also codes for phonetically overt bound morphology on this tier. Stems and affixes are divided by "-", as in (3):

(3) *comes* is a finite verb form, with verbal **A**greement inflection (expressed by a bound morpheme) and is related to **I** by movement. It is coded as: **IA**|come-s

Combining the examples in (1) and (3), an utterance such as *she comes* would be coded as in (4):

(4) *CHI: she comes
     %mor: IPROS|she IA|come-s

In addition to bound morphology, GCS codes some free grammatical morphemes on the morphological tier, such as pronouns and prepositions. Furthermore, as we will dis-

Table 1
Codes for the Morphological Tier of the I-System

Code	Definition of Code
IA|	overt agreement marked on the verb
IT|	tense marked on the verb
IAUX|	auxiliary, modal, or copula *be*
ITAGAUX|	auxiliary, modal, or copula *be* in tag question
IPROS|	subject pronoun with overt nominative case
I|to	infinitival *to*

cuss below, utterance length is calculated on the basis of the morpheme count on the morphological tier.

As for the I-system, elements that are related to tense, agreement, nominative case, auxiliaries, modals, do-support, and infinitival *to* are coded on the morphological tier. An exhaustive list is given in Table 1.

As is well known, bound morphology can be regular or irregular. For example, certain verbs are inflected regularly for past tense, others irregularly, and plural formation can also be regular or irregular. In order to distinguish between regular and irregular bound morphology, GCS uses different codes. A past tense verb form such as *bought* is coded on the morphological tier as IT|buy-**d**, whereas a regular past tense verb form such as *walked* gets an *-ed* code: IT|walk-**ed**. A regular plural form such as *cats* is coded as D|cat-pl, whereas an irregular plural form such as oxen receives the code D|ox-p.[1]

Examples in (5)–(7) illustrate some of the codes discussed so far.

(5)   *CHI: he won't eat cookies
      %mor: IPROS|he IAUX|will~NEG|not eat C|cookie-**pl**

(6)   *CHI: mom bought two cars
      %mor: mom IT|buy-**d** two D|car-**pl**

(7)   *CHI: two geese crossed the road
      %mor: two D|goose-**p** IT|cross-**ed** DART|the road

In (5), as in standard CHAT format, cliticization is indicated by a tilde (~), but in GCS cliticization is marked only on the morphological tier.

D-system elements that are coded on the morphological tier include determiners, possessives, and plurals.

An exhaustive list is provided in Table 2, where *xxx* represents an uninflected stem. For instance,

(8)   *CHI: these dolls are mine
      %mor: DDEM|this-p D|doll-pl IAUX|be-s DPOS|mine

Finally, the C-system elements we chose to code on the morphological tier include complementizers (introducing complement and adjunct clauses), *wh*-words, relativizers, and moved auxiliaries (i.e., auxiliaries moved to C). These are listed in Table 3.

Some of the codes in Table 3 are exemplified in (9):

(9)   *CHI: where is the car that I bought yesterday?
      %mor: CWH|where CAUX|is DART|the car CREL|rel that IPROS|I IT|buy-d yesterday

It is important to note that, although this tier is labeled the *morphological* tier, it encodes information relevant to the syntax, consistent with current research in syntactic theory. Thus, our morphological tier is designed to meet the principal objectives of including, but going beyond, the coding of bound and free morphemes to code (1) the functional categories C, I, D, and other relevant feature specifications and (2) errors related to these functional categories and their subtypes (i.e., omissions, misselections, and overinsertions).

Syntactic Tier

The syntactic tier (%syn:) is designed as the place for coding constituent structure (including types of embedding and internal phrasal and clausal structure), constituent order (capturing linear order and movement), and partial information regarding constituent boundaries. Each syntactic phrase is labeled with category and grammatical function labels, as is illustrated in (10)–(12):

(10) SNP = **S**ubject **N**oun **P**hrase (category = NP, grammatical function = subject)

(11) MOD = **modal**

(12) CADJ = **ADJ**unct clause, complementizer in **C**

Embedded clauses are enclosed in parentheses, and codes for embedding types are in square brackets. Verbal morphology is also included on this tier, using the

Table 2
Codes for the Morphological Tier of the D-System

Code	Definition of Code
DART|	article
DPOSD|	possessive determiner followed by noun (*my*, *his/her*, etc.)
DPOS|	independent possessive pronoun (*mine*, *yours*, etc.)
DDEM|	demonstrative determiner followed by noun
D|xxx-pl	(regular) noun plural
D|xxx-p	(irregular) noun plural
D|xxx-'s	singular possessive affix
DDEM|xxx-p	demonstrative determiner marked for plural
D|xxx-s'	plural possessive affix
DQUA|	basic quantifiers
DCAR|	cardinal quanitifiers
DPRO|	demonstrative pronoun

Table 3
Codes for the Morphological Tier of the C-System

Code	Definition of Code
CCOM|	embedding complementizer (also for embedding *wh*-words)
CWH|	*wh*-words (in main *wh*-clauses)
CADJ|	adjunct complementizer (e.g., *because*, *if*)
CREL|	relativizer
CAUX|	inverted auxiliary (e.g., in root questions)

same symbol (hyphen) as that used on the morphological tier. This allows one to capture the co-occurrence of syntactic and morphological phenomena, such as null subjects and tense marking. A list of syntactic codes is given in Table 4.

Examples of coding of well-formed utterances are given in (13)–(14). Note that <**cp** and <**ip** are used to mark CP and IP boundaries.

(13) *CHI: she can't sing if he's in the room
      %syn: <ip SNP MOD NEG V <cp ([CADJ] SNP BE PP)

(14) *CHI: I am Mannekin the good ghost
      %syn: <ip SNP BE PNP ([MDF] APNP)

Lexical Tier

The lexical tier (%lex:) should be understood as the tier for coding types, tokens, and errors of the major lexical categories—namely, N(oun), V(erb), ADV(erb), ADJ(ective) in each utterance of a speech sample. The codes of the lexical tier are listed in Table 5.

All of the morphemes that are part of the same lexical entry are placed, without spaces between them, in the same listing. A "+" is used to code morpheme bound-

Table 4
Codes for the Syntactic Tier

Code	Definition of Code
SNP	subject noun phrase
DNP	direct object noun phrase
INP	indirect object noun phrase
PNP	predicative nominal
SWH	subject *wh*-word
DWH	direct object *wh*-word
IWH	indirect object *wh*-word
AWH	adjunct *wh*
AUX	auxiliary
MOD	modal
V	main verb
Vf	finite (verb inflected for tense or agreement)
PRT	particle
BE	form of the verb *to be*
AP	adjectival phrase
PP	prepositional phrase
ADVP	adverbial phrase
NEG	negation
CCOM	complementizer introducing complement clause
CADJ	complementizer introducing adjunct clause
CREL	relativizer
MDF	modifier
APNP	appositive

aries in compounds. An "=" is again used to signal an error. The lexical coding is illustrated in (15) below.

(15) *CHI: my father poured dinner for us.
      %lex: N |father V| = pour N |dinner

This tier allows one to examine the structure, size, and productivity of the speaker's lexicon. Omitted words would be captured on one of the other tiers defining the resultant grammatical error.

GCS underwent a long period of refinement and modification. We were frequently forced to add to or modify the system to accommodate data that deviated in unexpected ways from the normal, adult grammar. By designing GCS to describe and analyze disordered as well as normal language, however, it has become a system that is useful for linguistic analysis of all kinds of speakers, including normal children, children with language disorders, adults with acquired aphasia, and of course, normal mature speakers.

Error Coding

Errors related to all structures coded are identified by the equal sign (=), whereas in the standard CHAT format an asterisk (*) is placed before the item in which the error occurs. Furthermore, because the GCS system allows much richer detail to be presented in the codes, no separate %err: (error) tier is necessary to elaborate the nature of the error under analysis, necessary in CHAT. GCS distinguishes two kinds of errors: placement errors and form errors.

Placement errors. An error is defined as a placement error when (1) an element occurs that should not have, (2) there is a missing element that should be present, (3) there is an element that is placed lower (in the structural derivation) than it should be, or (4) there is an element that is placed higher than it should be.

An added, extra element that should not be present is marked with an equal sign before it—for example, = xxx. If it is the entire category that should have been deleted, "=" goes before the category in both the %mor tier and the %syn tier, as is illustrated on (16)–(19).

(16) *XXX: I want for to go now
      %mor: IPROS|I want =**CCOM**|for I|to go now
      %syn: <ip SNP V ([=**CCOM**] [TO] V ADV)

(17) *XXX: I can to play now
      %mor: IPROS|I IAUX|can =**I**|to play now
      %syn: <ip SNP AUXM =**[TO]** V ADVP

Table 5
Codes for the Lexical Tier

Code	Definition of Code
N|	noun
V|	verb
ADV|	adverb
ADJ|	adjective

(18) *XXX: who did leave?

%mor: CWH|who =**CAUX|do-d** leave

%syn: <cp SWH =**AUXD** <ipleave

(19) *XXX: who do you think that left?

%mor: CWH|who CAUX|do PRO|you think =**CCOM|that** IT|leave-d

%syn: <cp SWHE AUXD <ip SNP V <cp (**[=CCOM]** Vf)

If it is a morpheme that should not have been added, the "=" is placed before the morpheme at %mor:, as in (20)–(22).

(20) *XXX: I sees you

%mor: IPROS|I **IA|see-=s** PRO|you

%syn: <ip SNP V DNP

(21) *XXX: she is sitting on the furnitures

%mor: IPROS|she IAUX|be-s sit-ing P|on DART|the **D|furniture-=pl**

%syn: <ip SNP AUXB V-ing PP

(22) *XXX: there are a lot of peoples

%mor: PROE|there IAUX|be-s DART|a DCAR|lot P|of **D|person-p-=pl**

%syn: <ip EXP AUXB PNP

Missing elements that should be present are marked by placing an =**0** before them. If what is missing is the entire category, the =**0** is placed before it at both %mor and %syn (at %mor here the missing word is known).

(23) *XXX: I am scared monsters

%mor: IPROS|I IAUX|be-s scared =**P|0of** D|monster-pl

%syn: <ip SNP AUXB V-en =**0P** ONP

(24) *XXX: she put on the table

%mor: IPROS|she put P|on DART|the table

%syn: <ip SNP V =**0DNP** PP

(25) *XXX: not tell mother

%mor: NEG|not tell mother

%syn: =**0SNP** =**0AUXD** NEG V INP

(26) *XXX: I sleeping

%mor: IPROS|I IAUX|=**0be** sleep-ing

%syn: <ip SNP =**0AUXB** V-ing

Since there is no evidence that the missing *be* in the preceding example has any inflection, it is coded as a bare aux (only as *be*).

If the missing material is simply a morpheme, =**0** is placed before it at %mor. If the missing morpheme is inflectional morphology on the verb, the verb is coded as a **Vn** (*n* for nonfinite, in contrast to **Vf** for finite verbs) at %syn. "Semi-auxiliaries" or "catenatives" (e.g., hafta, gonna) are coded as auxiliaries with "to" cliticized to them.

(27) *XXX: he see me

%mor: IPROS|he **IT|see-=0d** PROO|me

%syn: <ip SNP **Vn** DNP

(28) *XXX: she hafta leave now

%mor: IPROS|she **IAUX|have-=0s~to** leave now

%syn: <ip SNP **AUXH** V ADVP

(29) *XXX: I ate two sandwich

%mor: IPROS|I IT|eat-d DCAR|two **D|sandwich-=0pl**

%syn: <ip SNP Vf DNP

Material that is ungrammatically low (on the right) in the sentence is prefixed with =**L** at %syn. Movement errors are not coded at %mor.

(30) *XXX: what he can say?

%mor: CWH|what IPROS|he IAUX|can say

%syn: <cp DWH <ip SNP =**LAUXM** V

(31) *XXX: fall down tim

%mor: IT|fall=0-d PRT|down tim

%syn: Vn PRT =**LSNP**

(32) *XXX: I took down it

%mor: IPROS|I IT|take-d PRT|down PRO|it

%syn: <ip SNP Vf PRT =**LDNP**

Material that is ungrammatically high (on the left) is prefixed with an =**H** at %syn.

Form errors. By definition, form errors are morphological and marked at %mor only. Such errors are ignored at %syn. Form errors that comprise an entirely wrong morpheme are prefixed with an equal sign (=), which follows the hyphen (-).

(33) *XXX: I have listening to story

%mor: IPROS|I **IAUX|=have** V-ing P|to DART|=0the story

%syn: <ip SNP AUXH V-ing PP

(34) *XXX: she were sleeping

%mor: IPROS|she **IAUX|be-=s-d** sleep-ing

%syn: <ip SNP AUXB V-ing

(35) *XXX: he is my bestest friend

%mor: IPROS|he IAUX|be-s DPOSD|my **good-est-=est** friend

%syn: SNP AUXB PNP

All form errors that are simply misselections are coded with an exclamation sign (!) preceding the ungrammatical

form (in addition to the = sign, which marks all errors). These are instances that typically involve regular and irregular forms—that is, the choice of wrong allomorphs.

(36) *XXX: I goed to school that day
      %mor: IPROS|I **IT**|**go**=**-!ed** P|to school DDEM|that day
      %syn: <ip SNP **Vf** PP ADV

(37) *XXX: all the childs are coming
      %mor: all DART|the **D**|**child**=**-!s** IAUX|be-s come-ing
      %syn: <ip SNP AUXB V-ing

(38) *XXX: I should have went home
      %mor: IPROS|I IAUX|should AUX|have go=-!n P|0to home
      %syn: <ip SNP AUXM AUXH **Vf** 0P NP

Please refer to the Appendix for a full coded transcript illustrating this system and see Curtiss, MacSwan, Schaeffer, Kural, and Sano (2001) for a full description of the coding system.

Some Advantages of GCS Over CHAT

Our main goal in extending CHAT was to create a coding system that would capture theoretically pertinent features of our data in a consistent, systematic way. In particular, we wanted to create a system that would allow us to focus on functional categories in the language samples we collected.

Consider (39), a sample of standard CHAT coding, taken from a transcript in MacWhinney (2000).

(39) *XXX: and that's a little fisky like # I can't think of the name of that boy.
      %mor: CONJ|and PRO|that~V|be&PRES ART|a ADJ|little N |*? ADJ|like PRO|I&1S AUX|can~NEG V|think PREP|of ART|the&DEF N |name PREP|of ADJ|that&DEM N |boy

CHAT codes the pronoun "I" as **PRO|I&1S**, indicating that it is a pronoun (**PRO**) and that it fusionally inflects (**&**) for first person (1) and singular (**S**). In GCS, "I" would be coded as **IPROS|I** to represent that (1) it checks features with the functional category I, (2) is a **PRO**noun and has the grammatical function **S**ubject (i.e., that it is in the specifier position of IP). As will be seen below, this coding system allows us to associate "I" with other elements that relate to the functional category I(nflection) in calculating error rates or in searching for co-occurrence of overt nominative case-marked elements with agreement suffixes, and so on. Also note that MacWhinney (2000) codes the definite article "the" as **ART|the&DEF**, indicating that it is an article and that it fusionally carries the feature [+definite]. In our system, "the" is coded as **DART|the**, capturing the fact that the article "the" is an instance of the functional category D.

Notice that CHAT codes categorial information on the morphological tier, as can be seen in (39). Since GCS

was designed to examine the status of functional categories, we divide the labor and define a separate lexical tier for the encoding of substantive lexical categories. The morphological tier, then, is the sole tier for encoding information about the various functional projections in addition to overt morphological marking, and syntactic categories of lexical items are coded on the lexical (%lex:) tier.

This also represents an advantage over the CHAT system, given the nature of our work and our theoretical assumptions. In CHAT, a considerable amount of material on the morphological tier would have to be eliminated to evaluate lexical productivity, semantic range, and richness within the lexicon and the robustness of lexical retrieval. However, GCS has built in a tier on which only lexical category entries are represented. Here again, the theoretical framework underlying GCS makes it desirable to separate lexical category items from functional category elements. Having a separate tier, the lexical tier, on which only lexical category items are represented (i.e., nouns, verbs, adjectives, and adverbs) readily allows for an examination of a variety of parameters one may wish to scrutinize in developmental data, normal and impaired, as well as breakdown data. In DAT (dementia of the Alzheimer's type), for example, one finds impairments principally of lexical category words, the very words coded separately on the lexical tier in GCS; in addition, in DAT one finds lexical impairments of the very kind having a lexical tier makes readily analyzable (e.g., productivity or retrieval; Huff, Corkin, & Growdon, 1986; Kempler, 1988, 1995; Kempler, Curtiss, & Jackson, 1987). Such findings clearly motivate the separation of lexical category items from the mix of elements coded on the morphological line in the CHAT system.

GCS codes affixal morphology by using the traditional hyphen (-x) that linguists use for morphological segmentation. Nonetheless, in doing so, we were careful to build in a coding scheme parallel to the CHAT ampersand that distinguishes regular from irregular inflections in the nominal and verbal domains, since regular versus irregular inflectional morphology is central to potential processing distinctions that may play a role in developmental and acquired language impairments. (Hence, we use [-pl] for regular plural and [-p] for irregular plural, [-ed] for regular past, [-d] for irregular past, [-en] for regular past participle, and [-n] for irregular past participle.)

As in CHAT, we code constituent structure (including types of embedding and internal phrasal and clausal structure) and constituent order on a syntactic tier. On this tier, category and grammatical function are labeled; (40) is taken from MacWhinney (2000, p. 106):

(40) *CHI: if I don't get all the cookies you promised to give me, I'll cry.
      %ssy: <C S X V M M D < S V < R V I > [CP] > [RC] > [CC] > S V > [MC] .

In the CHAT system, a hand-coded syntactic tier is designated %ssy:, and %syn: is used for automatically coded syntactic tiers generated by a computational parser based on a phrase structure grammar (MacWhinney, 2000;

Sagae, Lavie, & MacWhinney, 2001; Sagae, MacWhinney, & Lavie, 2004).

On the syntactic tier in our system, GCS codes the fact that certain structures are base-generated in or are adjoined by movement to the C-node and are, therefore, considered C-system elements. In addition, in our system each syntactic phrase is labeled for both category and grammatical function. So, for example, what CHILDES codes as C (conjunction) or RC (relative clause), GCS codes as CCOM (**COM**plementizer, base-generated in **C**) or CREL (a **REL**ativizer, base-generated in, or moved under **C**) to capture the fact that these are elements of the C-system.

In addition to its focus on functional categories, as in contemporary syntactic theory, GCS is also easily adaptable to other languages, because the coding system uses universal, rather than language-particular, categories for specifying various morphological and syntactic properties. Consider, for instance, the following adaptation of GCS to Spanish, with translation provided in brackets (MacSwan, 2001; Valadez, MacSwan, & Martínez, 2002):

*MAR: El niño se está dormiendo, y la rana se escapó.
[The boy is going to sleep, and the frog escaped]

%mor: DART|el niño REF|se IAUX|está-3Ss dormir-DUR conj|y DART|la rana REF|se IF|escapar-pret3Ss

%syn: CP [DPi se VP] CONJ CP [DPi se VP]

%lex: N |niño N |dormir N |rana N |escapar

Functional categories and other universals are coded the same in Spanish as in English. However, where Spanish uses grammatical morphemes not present in English, new codes have been created to accommodate these language-particular features (e.g., REF for reflexive and pret3Ss for third-person singular subject preterite agreement). Similarly, the English version of GCS presented here uses PRT for verb particles and IAUXTAG for auxiliaries in English tag question constructions, structures not present in Spanish.

Finally, we believe that GCS presents one additional advantage over CHAT in that it provides a system for grammatical coding consistent with the theoretical assumptions of a large group of researchers who work broadly within the framework of generative grammar. We hope that other researchers who work within the principles and parameters tradition will find GCS to be a useful refinement of CHAT.

THE ANALYSIS SYSTEM

As was discussed earlier, the CHILDES system includes a number of programs for analyzing children's talk, called CLAN. Software for the MacIntosh and DOS were released along with MacWhinney (1991, 1995), and a Windows version of the software, which accompanies MacWhinney (2000), adds a graphical interface to the CLAN programs that makes them much easier to use.

Pediatric Epilepsy Project
Neurology of Language Acquisition
Susan Curtiss, Principal Investigator
I-System Scoresheet
I-Sys.exe (v2.00)

Please enter your first and last name for the report.
? Susan Curtiss

Would you like the screen to pause so you can monitor freq.exe during the analyses? [Yes/No] ? Yes

Please enter the three letter/digit file name extension which designates the current patient (e.g., JR1, JR2, RP3, etc.).

Wildcards are ok too (e.g., *, JR?, etc.).

? tm1

What is the child's full name?
? Tobi Myers

Figure 2. A snapshot of the opening screen of I-sys.exe.

```
Looking for OCCIPROS ... generating IPROS.TM1
FREQ.EXE +t%mor -t* +s*ipros|* acoded.TM1
Tue May 28 00:54:30 2001
FREQ.EXE (07-JUL-93) is conducting analyses on:
  ONLY dependent tiers matching: %MOR;
****************************************
From file <acoded.TM1>




~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Press [C] to end pause.
Press [Q] to quit i-sys.exe now.
Press any other key to continue monitoring freq.exe.
Action [C, Q, ...]? C
```

Figure 3. A snapshot of I-sys.exe during execution.

Freq.exe, one of the several CLAN programs associated with the CHILDES project, provided the basis for a computer analysis system that we designed to answer specific research questions regarding various subsystems of grammar. While freq.exe is a powerful tool for counting frequencies of words, morphemes, or codes, research questions posed by our project required that a variety of very different morphological, syntactic, and lexical codes be counted as evidence regarding a single subsystem of the grammar. With respect to the I-system, for instance, GCS marks a subject pronoun *she* as IPROS|she and errors in IPROS as IPROS|=she, as was explained above. Counts of IPROS in the language samples could be obtained using the standard freq.exe command shown in (42).

(42)   freq +t%mor -t* +s "*IPROS|*" acoded.ext > output1.ext

This command tells the program freq.exe to search the morphological tier (%mor), excluding all other tiers (-t*), to find the string (+ s) "IPROS|" (even where material occurs to its left and to its right, as denoted by the asterisk "wildcards"). The program searches the file acoded.ext and puts the results of its search in the file output1.ext.

In counting errors, two freq.exe commands were required, as shown in (43).

(43a)   freq +t%mor -t* +s "= IPROS|*" acoded.ext > output2.ext

(43b)   freq +t%mor -t* +s "IPROS|=*" acoded.ext > output3.ext

The code in (43a) (=IPROS|) represents an error with respect to the functional category itself, whereas the

code in (43b) (IPROS|=) represents an error with respect to an element within that functional head.

For the I-system alone, occurrences of codes for 16 different variables, plus variables for placement errors and total errors, needed to be calculated individually and sorted for statistical analysis. Because GCS codes for a variety of errors in each category, two or three freq.exe commands would be necessary for each error subtotal, depending on the range of errors coded in the data for each category. Occurrences of errors obtained in freq.exe searches, such as those in (43), would have to be added together in order to account for the total number of errors in each category under analysis. In all, 58 freq.exe searches would have been required for each data point for the analysis of the I-system alone. Although the FREQMERG command discussed in MacWhinney (1995) would be useful for obtaining subtotals for some of these variables, in the research project for which GCS was devised, we would, nonetheless, have had to keep track of 53 file names to be combined in 20 uses of the FREQMERG command.

Therefore, in order to simplify and streamline analyses for the I-system, we created a simple program, written in a variety of the BASIC programming language, called i-sys.exe. I-sys performs four operations, as outlined in (44). In our conventions, a data point is tracked with a unique computer filename extension, here denoted *.ext*.

(44) Each time i-sys.exe is run, it

(a)   calls freq.exe 58 times to conduct relevant I-system searches on coded files, creating a sep-

```
                     Pediatric Epilepsy Project
                   Neurology of Language Acquisition
                  Susan Curtiss, Principal Investigator

                          I-System Scoresheet

Data point:    TM1                    Child's name: TOBI MYERS
Date/time:     05-28-2001 01:10:26    Run by:     SUSAN CURTISS
=================================================================
%mor:          Total        Omission      Total       percent
Code           occurences   errors        errors      total errors
-----------------------------------------------------------------
IPROS          2            0             0           0.00%
IPROStag       0            0             0           0.00%

IPROS subtotal 2            0             0           0.00%
-----------------------------------------------------------------
IAUX           4            0             0           0.00%
IAUXtag        0            0             0           0.00%

IAUX subtotal  4            0             0           0.00%
-----------------------------------------------------------------
IF-ed          0            0             0           0.00%
IF-d           1            0             0           0.00%
IF-dtag        0            0             0           0.00%

TENSE subtotal 1            0             0           0.00%
-----------------------------------------------------------------
IA-s           0            0             0           0.00%

AGR subtotal   0            0             0           0.00%
-----------------------------------------------------------------
ITO            0            0             0           0.00%
-----------------------------------------------------------------
PROE           0            0             0           0.00%
-----------------------------------------------------------------
-en            0            0             0           0.00%
-n             0            0             0           0.00%
-ing           1            0             0           0.00%

ASPECT
subtotal       1            0             0           0.00%
-----------------------------------------------------------------
NEG            1            0             0           0.00%
-----------------------------------------------------------------
I|=0           n/a          0             0           n/a
-----------------------------------------------------------------
TOTAL          9            0             0           0.00%
=================================================================
Utterances (U) =  5                  Analyzed in  0.13 minutes
 Morphemes (M) =  32                    I-sys.exe version 2.00
     MLU (M/U) =  6.40
```

Figure 4. Output of I-sys.exe.

arate freq.exe file containing search results for each iteration;

(b) reads the 58 files created in (a) and extracts the "total number of words" (codes, actually) in each;

(c) computes results obtained in (b) to compile a final report called *i-report.ext*; and

(d) calculates the number of morphemes, number of utterances, and mean length of utterance (MLU).

For convenience, all of the output files created in (44a) are merged into a single file called *freq-out.ext*. The run-time of the analysis under i-sys.exe varies with the length of the transcript and computing resources available, but the average use of time in our laboratory per data point has been less than 0.5 min. (Similar analysis programs devoted to the D-system and the C-system, and to syntactic and lexical coding, are needed for each new set of research questions posed.)

Mean length of utterance (MLU) is traditionally regarded as a measure of a child's linguistic maturity (Brown, 1973). Although MacWhinney's (1995, 2000) MLU command can be manipulated to properly serve this function for our data, we created a simple algorithm in i-sys.exe to calculate the MLU for each data point. Our algorithm

counts each utterance coded for morphological information (on the %mor tier) and determines the number of morphemes on each %mor line by counting elements separated by -, ~, +, and space (our morpheme delimiters). In addition, elements coded as missing (marked =0 on the %mor tier) are excluded. Because GCS does not code false starts, sound effects, frozen phrases, and unintelligible utterances on the %mor tier, expressions of this sort are automatically excluded from the MLU calculation.

When it is run, the program asks the user for some basic information, as is shown in Figure 2, using our neurology of language acquisition study as an illustration. The user's name and the child's name are recorded for the final report. The program also asks for the file name extension in order to locate the file for the current data point. Because freq.exe reports CHAT-format violations, the user is given an opportunity to monitor freq.exe during execution.

If a wildcard (*,?) is used for the file name extension, i-sys.exe will not prompt for the child's name but will use his/her initials instead, derived from the file name extension (*tm* in Figure 2). The wildcard feature allows i-sys.exe to successively analyze multiple transcripts, which represent successive data points for the same child or for a group of children.

If the user asks to monitor freq.exe output during execution, an opportunity is presented to terminate i-sys.exe, end the monitor function, or display the next result of applying freq.exe to the coded transcript, as is shown in Figure 3.

The I-report, shown in Figure 4, displays the descriptive statistics for the data under analysis. Another small utility program reads the reports of the type shown in Figure 3 and converts them into a fixed-width text file that can then be imported into any popular statistical software for detailed analysis.

A limitation of the current computerized analysis system is that it is highly specific to our research questions. In current work, we are developing a system that will run independently of other programs and allow the user to alter the search criteria in any way desired. We also encourage other researchers to develop new tools of analysis that will permit simultaneous searches over a broad range of editable codes and produce reports such as the one shown in Figure 4, as well as a fixed-width data file that can be imported into popular statistical packages. It is hoped, too, that future development might include a computational parsing system to generate GCS codes on the syntactic tier, akin to those employed by some CHAT users (Sagae, Lavie, & MacWhinney, 2001; Sagae, MacWhinney, & Lavie, 2004).

Nonetheless, the current CLAN suite of programs will serve many researchers perfectly well in conducting searches on GCS-coded transcripts. However, some of the more specific utilities developed for CHAT may need modification. For instance, a pair of utility programs, known as POST and MOR, allow CHAT users to perform automatic coding of transcripts. This method is highly efficient and results in fewer coding errors than does manual coding (Parisse & Le Normand, 2000).

Nonetheless, with large corpora, many may find it worth their time to modify the lexicons and grammars used by POST and MOR to allow these utilities to perform automatic coding using GCS coding conventions.

We hope the coding and analysis system presented here proves useful and insightful to researchers concerned with child language acquisition, the analysis of normal and impaired language, and any other sort of linguistic research that depends on detailed and theoretically well-informed transcriptions of natural speech. A more detailed description of the GCS may be reviewed in Curtiss, MacSwan, Schaeffer, Kural, and Sano (2001), archived at http://www.psychonomic.org/archive/.

REFERENCES

ABNEY, S. P. (1987). *The English noun phrase in its sentential aspect*. Unpublished doctoral dissertation, MIT.

BORER, H. (1984). *Parametric syntax: Case studies in Semitic and Romance languages*. Dordrecht: Foris.

BROWN, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

CAPLAN, R., CURTISS, S., CHUGANI, H. T., & VINTERS, H. V. (1996). Pediatric Rasmussen encephalitis: Social communication, language, PET and pathology before and after hemispherectomy. *Brain & Cognition*, **32**, 45-66.

CHOMSKY, N. (1957). *Syntactic structures*. The Hague: Mouton.

CHOMSKY, N. (1991). Linguistics and cognitive science: Problems and mysteries. In A. Kasher (Ed.), *The Chomskyan turn*. Cambridge: Blackwell.

CHOMSKY, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

CURTISS, S., & DE BODE, S. (1998). Linguistic outcomes in hemispherectomized children. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 121-133). Somerville, MA: Cascadilla.

CURTISS, S., & DE BODE, S. (1999a). Age and etiology as predictors of language outcome following hemispherectomy. *Developmental Neuroscience*, **21**, 174-181.

CURTISS, S., & DE BODE, S. (1999b). Language outcome after hemispherectomy. *Brain & Cognition*, **41**, 111.

CURTISS, S., & DE BODE, S. (2003). How normal is grammatical development in the right hemisphere following hemispherectomy? The RI stage and beyond. *Brain & Language*, **86**, 193-206.

CURTISS, S., DE BODE, S., & MATHERN, G. (2001). Spoken language outcomes after hemispherectomy: Factoring in etiology. *Brain & Language*, **79**, 379-396.

CURTISS, S., MacSWAN, J., SCHAEFFER, J., KURAL, M., & SANO, T. (2001). *GCS: Grammatical Coding System manual*. Available at http://www.psychonomic.org/archive/.

CURTISS, S., & SCHAEFFER, J. (1997a, March). *Development of the I- and D-systems in children with hemispherectomy*. Paper presented at the annual Society for Cognitive Neuroscience meeting, Boston.

CURTISS, S., & SCHAEFFER, J. (1997b). Syntactic development in children with hemispherectomy: The INFL-system. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 103-114). Somerville, MA: Cascadilla.

CURTISS, S., & SCHAEFFER, J. (2004). *Syntactic development in children with hemispherectomy: The I-, D-, and C-systems*. Manuscript submitted for publication.

DE BODE, S. (1998). Interhemispheric language transfer and functional plasticity. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 134-140). Somerville, MA: Cascadilla.

DE BODE, S., & CURTISS, S. (1999). Neurobiological mechanisms of lan-

guage acquisition. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 150-161). Somerville, MA: Cascadilla.

FROMKIN, V. (ED.) (2000). *Linguistics: An introduction to linguistic theory*. London: Blackwell.

HENDRICK, R. (ED.) (2003). *Minimalist syntax*. Malden, MA: Blackwell.

HUFF, F. J., CORKIN, S., & GROWDON, J. H. (1986). Semantic impairment and anomia in Alzheimer's disease. *Brain & Language,* **28,** 235-249.

HYAMS, N. (1996). The underspecification of functional categories in early grammar. In H. Clahsen (Ed.), *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and crosslinguistic comparisons* (pp. 91-127). Amsterdam: Benjamins.

JAKUBOWICZ, C., DURAND, C., RIGAUT, C., & VAN DER VELDE, M. (2001). Computational complexity over time: The development of functional categories in French-speaking children with SLI. In A. H.-J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (Vol. 1, pp. 365-376). Somerville, MA: Cascadilla.

KEMPLER, D. (1988). *Syntactic and symbolic abilities in Alzheimer's disease*. Unpublished doctoral dissertation, UCLA.

KEMPLER, D. (1995). Language changes in dementia of the Alzheimer's type. In R. Lubinski (Ed.), *Dementia and communication: Research and clinical applications* (pp. 98-114). San Diego: Singular.

KEMPLER, D., CURTISS, S., & JACKSON, C. (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech & Hearing Research,* **30,** 343-350.

LABOV, W. (1975). Empirical foundations of linguistic theory. In R. Austerlitz (Ed.), *The scope of American linguistics* (pp. 77-133). Lisse: de Ridder.

LEONARD, L. B. (1998). Functional categories in the grammars of children with specific language impairment. *Auditory & Speech Sciences*, **38**, 1270-1283.

MACSWAN, J. (2001, April). *The non-non crisis: Knowledge of language and problems of construct validity in native language assessment*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Seattle.

MACWHINNEY, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

MACWHINNEY, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Erlbaum.

MACWHINNEY, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Hillsdale, NJ: Erlbaum.

MACWHINNEY, B., & SNOW, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language,* **17,** 457-472.

NEWMEYER, F. (1983). *Grammatical theory: Its limits and its possibilities*. Chicago: University of Chicago Press.

PARISSE, C., & LE NORMAND, M.-T. (2000). Automatic disambiguation of the morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments, & Computers,* **32,** 468-481.

RADFORD, A. (1995). Phrase structure and functional categories. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language acquisition* (pp. 483-507). Oxford: Blackwell.

RADFORD, A. (1997). *Syntax: A minimalist introduction*. Cambridge: Cambridge University Press.

SAGAE, K., LAVIE, A., & MACWHINNEY, B. (2001). Parsing the CHILDES database: Methodology and lessons learned. In *Proceedings of the Seventh International Workshop in Parsing Technologies*. Beijing.

SAGAE, K., MACWHINNEY, B., & LAVIE, A. (2004). Automatic parsing of parental verbal input. *Behavior Research Methods, Instruments, & Computers,* **36,** 113-126.

SCHÜTZE, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: Chicago University Press.

SINKA, I., & SCHELLETER, C. (1998). Morphosyntactic development in bilingual children. *International Journal of Bilingualism,* **2,** 301-326.

VALADEZ, C., MACSWAN, J., & MARTÍNEZ, C. (2002). Toward a new view of low achieving bilinguals: A study of linguistic competence in designated "semilinguals." *Bilingual Review*, **25**, 238-248.

WEBELHUTH, G. (ED.) (1995). *Government and binding theory and the minimalist program: Principles and parameters in syntactic theory*. Oxford: Blackwell.

WEXLER, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua,* **106,** 23-79.

NOTE

1. We prefer the use of distinct codes for regular and irregular, rather than using the ampersand (&) as a prefix denoting regular forms, as in CHAT. In CHAT, the ampersand is associated with the phonological shape of irregular forms, as well as nonce and nonsense forms (MacWhinney, 2000, pp. 16–17, 101). In GCS, we are more narrowly concerned with coding the morphological attributes of affixes.

ARCHIVED MATERIALS

The following material may be accessed through the Psychonomic Society's Norms, Stimuli, and Data Archive, http://www.psychonomic.org/archive/.

To access this material, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Curtiss), and the publication year (2004).

FILE: Curtiss-BRMIC-2004.zip.

DESCRIPTION: The compressed archive file contains GCSmanual.pdf, a monograph describing the GCS (Grammatical Coding System) in detail; the manual may be cited as Curtiss, MacSwan, Schaeffer, Kural, and Sano (2001), *GCS–Grammatical Coding System Manual* (Austin, TX: Psychonomic Society's Norms, Stimuli, and Data Archive).

AUTHOR'S E-MAIL ADDRESS: scurtiss@ucla.edu

AUTHOR'S WEB SITE: http://www.linguistics.ucla.edu/people/curtiss/index.htm

APPENDIX

```
@Begin
@Participants:          MIC Michelle Cuttler [pseudonym], INV investigator
@Age of MIC:            19;8
@Sex of MIC:            female
@Birth of MIC:          24-NOV-1972
@Date of surgery:       12-DEC-1986
@Type of surgery:       Left hemi
@Date of sample:        30-JUL-1992
@Coding:                GCS
@Filename:              acoded.mc4
@Situation:             Story retelling game.
```

```
*INV:    okay.  can you try and tell me that story?
*MIC:    um # he n the ghost came to his house.
%mor:    DART|the ghost IT|come-d P|to DPOSD|his house
%syn:    <ip SNP Vf PP
%lex:    N|ghost V|come N|house
*INV:    mm hm.
*MIC:    and he thought he was in a real life.
%mor:    ONJ|and IPROS|he IT|think-d CCOM|0that IPROS|he IAUX|be-s-d P|in DART|a real life
%syn:    [ONJ] <ip SNP Vf ( [0CCOM] <ip SNP AUXB PP )
%lex:    V|think V|be ADJ|real N|life
*MIC:    but but he kind of woke up and he th found out he was dreaming.
%mor:    ONJ|but IPROS|he kind P|of IT|wake-d PRT|up ONJ|and IPROS|he IT|find-d PRT|out
         CCOM|0that IPROS|he IAUX|be-s-d dream-ing
%syn:    [ONJ] <ip SNP ADVP Vf PRT [ONJ] <ip SNP Vf PRT ( [0CCOM] <ip SNP AUXB V-ing)
%lex:    ADV|kindof V|wakeup  V|findout V|dream
*INV:    right.  what did you think about this story?
*MIC:    it was good.
%mor:    PRO|it IAUX|be-s-d good
%syn:    <ip SNP AUXB PAP
%lex:    V|be ADJ|good
*INV:    wha, what was good about it?
*MIC:    how l # Mannekin came in and almost attacked Peter.
%mor:    CWH|how mannekin IT|come-d P|in ONJ|and almost IT|attack-ed peter
%syn:    <cp ( [AWH] <ip SNP Vf PRT [ONJ] ADVP Vf DNP )
%lex:    N|mannekin V|comein ADV|almost V|attack N|peter
%com:    Michelle laughs.
*INV:    good job.
*MIC:    that he was dreaming.
%mor:    CCOM|that IPROS|he IAUX|be-s-d dream-ing
%syn:    =[CCOM] <ip SNP AUXB V-ing
%lex:    V|dream
%com:    Michelle shrugs.
*INV:    that he was dreaming.
*MIC:    yeah.
*INV:    mm hm.  was there anything else you liked about the story?
%com:    Michelle shakes head.
*INV:    mm.  was there anything you didn't like about the story?  that he was # dreaming about bad
         things.
*INV:    mm hm.  you didn't like that?
%com:    Michelle shakes head.
*INV:    um, do you think this is a true story?
*MIC:    no.
%com:    Michelle shakes head.
*INV:    how do you know?
*MIC:    cause there's no such things as ghosts and stuff.
%mor:    CADJ|because PROE|there~IAUX|be-s DQUA|no such D|thing-pl P|as ghost-pl and stuff
%syn:    [CADJ] <ip EXP AUXB PPNP PP
%lex:    V|be ADJ| such N|thing N|ghost N|stuff
%com:    Michelle shrugs.
*INV:    how do you know?
*MIC:    <I don't know> [>].
```

APPENDIX (Continued)

%com: Michelle laughs and shrugs.
*INV: <are you sure> [<]?
*MIC: mm mm.
%com: Michelle shakes head.
*INV: mm?
*MIC: no.
*INV: you're not sure?
%com: Michelle shakes head.
*INV: wha what makes you think so?
*MIC: I've never seen one.
%mor: IPROS|I~IAUX|have never see-n PRO|one
%syn: <ip SNP AUXH ADVP V-n DNP
%lex: ADV|never V|see
*INV: you've never seen one. okay. do you know anybody else who has seen a ghost?
*MIC: mm mm.
%com: Michelle shakes head.
*INV: so what do you think? are there really such things as ghosts?
*MIC: maybe.
%mor: maybe
%syn: ADVP
%lex: ADV|maybe
%com: Michelle nods.
*INV: how come?
*MIC: cause he dreams about stuff he likes and he doesn't like.
%mor: CADJ|because IPROS|he IA|dream-s P|about stuff CREL|0that IPROS|he IA|like-s ONJ|and
               CCOM|0that IPROS|he IAUX|do-s~NEG|not like
%syn: 0CP ([CADJ] <ip SNP Vf PP ([0DREL] <ip SNP Vf [ONJ] AUXD NEG V))
%lex: V|dream N|stuff V|like V|like
*INV: mm hm. how did Peter know it was a dream?
*MIC: because he woke up and realized he was still # that there was nothing in his room.
%mor: CADJ|because IPROS|he IT|wake-d PRT|up ONJ|and IT|realize-ed CCOM|0that IPROS|he
               IAUX|be-s-d still CCOM|that PROE|there IAUX|be-s-d DQUA|no+thing P|in DPOSD|his room
%syn: 0CP ([CADJ] <ip SNP Vf PRT [ONJ] Vf ([CCOM] <ip SNP AUXB ADVP ([CCOM] <ip EXP
               AUXB PNP PP)))
%lex: V|wakeup V|realize V|be ADV|still V|be N|no+thing N|room
*INV: mm hm. so that's how he knew it was just a dream?
*MIC: uh huh.
*INV: okay. um now you say you don't believe in ghosts. but when you were a kid were you scared of
               ghosts?
*MIC: I had # on a dream that I saw a dragon or something else.
%mor: IPROS|I IT|have-d P|!on DART|a dream CCOM|that IPROS|I IT|see-d DART|a dragon or
               DQUA|some+thing else
%syn: <ip SNP Vf PP ([CCOM] <ip SNP Vf DNP)
%lex: V|have N|dream V|see N|dragon N|some+thing ADJ|else
%com: The possible false start makes it hard to tell whether "on a dream" is an error or not.
*MIC: the only thing.
%mor: DART|the only thing
%syn: =0SNP =0AUXB NP
%lex: ADJ|only N|thing
*INV: mm hm. xxx.
*MIC: mm mm.
*MIC: I've never dreamt about ghosts.
%mor: IPROS|I~IAUX|have never IT|dream-d P|about D|ghost-pl
%syn: <ip SNP AUXH ADVP Vf PP
%lex: ADV|never V|dream N|ghost
*INV: I can't hear you.
*MIC: I've never dreamt about ghosts.
%mor: IPROS|I~IAUX|have never dream-n P|about D|ghost-pl
%syn: <ip SNP AUXH ADVP V-n PP
%lex: ADV|never V|dream N|ghost
*INV: okay. you've never been scared about ghosts?
*MIC: uh uh.
*INV: how do you think the child in the story felt?

APPENDIX (Continued)

*MIC: scared.
%mor: scare-en
%syn: AP
%lex: ADJ|scared
*INV: mm hm. how would you have felt?
*MIC: scared.
%mor: scare-en
%syn: AP
%lex: ADJ|scared
*INV: you're scared.
*MIC: yeah.
*INV: um. how do you think he felt when he woke up?
*MIC: a lot better cause he found out he was just dreaming.
%mor: DART|a DCAR|lot good-r CADJ|because IPROS|he IT|find-d PRT|out CCOM|0that IPROS|he
        IAUX|be-s-d just dream-ing
%syn: 0SNP 0V ADVP AP ([CADJ] <ip SNP Vf PRT ([0CCOM] <ip SNP AUXB ADVP V-ing))
%lex: ADV|alot ADJ|good V|findout ADV|just V|dream
*INV: mm hm. okay. you know that Peter's dream wasn't very scary.
*MIC: mm hm.
%com: Michelle nods.
*INV: some children have real scary dreams. um, can you tell me one of your scary dreams. you did
        mention something about this dragon. you remember a scary dream?
*MIC: I was in a lake and um I um this dragon was chasing me.
%mor: IPROS|I IAUX|be-s-d P|in DARTa lake ONJ|and DDEM|this dragon IAUX|be-s-d chase-ing
        PROO|me
%syn: <ip SNP AUXB PP [ONJ] <ip SNP AUXB V-ing DNP
%lex: V|be N|lake N|dragon V|chase
%com: Michelle nods.
*MIC: he only had like eyes.
%mor: IPROS|he only IT|have-d like D|eye-pl
%syn: <ip SNP ADVP Vf DNP
%lex: V|have ADV|only N|eye
*MIC: really spooky eyes
%mor: really spook-y D|eye-pl
%syn: ADVP NP
%lex: ADV|really ADJ|spooky N|eye
*INV: mm hm.
*MIC: and he was chasing me and I couldn't get out, away from him.
%mor: ONJ|and IPROS|he IAUX|be-s-d chase-ing PROO|me ONJ|and IPROS|I IAUX|can-d~NEG|not
        get PRT|away P|from PROO|him
%syn: [ONJ] <ip SNP AUXB V-ing DNP [ONJ] <ip SNP AUXM NEG V PRT PP
%lex: V|chase V|get ADV|away
%com: Is "out" an error? Was the intended utterance "get away"
*INV: mm hm.
*MIC: and that's it.
%mor: ONJ|and PRO|that~IAUX|be-s PRO|it
%syn: [ONJ] <ip SNP AUXB PNP
%lex: V|be
*INV: how did your dream end?
*MIC: I kind of got away on the shore t #
%mor: I kind P|of IT|get-d PRT|away P|on DART|the shore
%syn: <ip SNP ADVP Vf PRT PP
%lex: ADV|kindof V|getaway N|shore
*MIC: that's all
%mor: PRO|that~IAUX|be-s all
%syn: <ip SNP AUXB PNP
%lex: V|be
*INV: so you got safe.
*MIC: yeah.
*INV: yeah. scary dream. now how about daydreaming. do you like to daydream?
*MIC: uh huh.
%com: Michelle nods.
*INV: yeah. what kind of things do you like to daydream about?

APPENDIX (Continued)

*MIC:	mm mm.
%com:	Michelle shakes head.
*MIC:	just # I like just dream how I would want a boy to look and stuff.
%mor:	IPROS|I like just dream CWH|how IPROS|I IAUX|would want CCOM|0for DART|a boy I|to look and stuff
%syn:	<ip SNP ADVP V <cp ([AWH] <ip SNP AUXM V ([0CCOM] <ip SNP [TO] V))
%lex:	ADV|just V|dream V|want N|boy V|look N|stuff
*INV:	how you want a boy to look?
*MIC:	yeah.
*INV:	uh huh.
*MIC:	how cute and stuff.
%mor:	how cute and stuff
%syn:	ADVP AP
%lex:	ADJ|cute N|stuff
*MIC:	yeah.
*INV:	and and that boy would be any boy or you're talking about a special boy?
*MIC:	a special one.
%mor:	DART|a special PRO|one
%syn:	AP NP
%lex:	ADJ|special
%com:	Michelle laughs.
*INV:	a special one. um, okay that was good. let's go on and hear our next story.
%com:	next story is played.
*INV:	okay. which story do you want to tell?
*MIC:	&a # a good or bad child.
%mor:	DART|a good or bad child
%syn:	AP NP
%lex:	ADJ|good ADJ|bad N|child
*MIC:	it's going to be a good child.
%mor:	PRO|it~IAUX|be-s AUX|goingto be DART|a good child
%syn:	<ip SNP AUXB AUX V PNP
%lex:	V|be ADJ|good N|child
*INV:	so which one are you doing?
*MIC:	a good child.
%mor:	DART|a good child
%syn:	AP NP
%lex:	ADJ|good N|child
*INV:	a good child. okay. xxx.
*MIC:	one time this child who lived at home and cleaned the house for her mom # didn't have to be told to do anything
%mor:	DCAR|one time DDEM|this child CREL|who IT|live-ed P|at home ONJ|and IT|clean-ed DART|the house P|for DPOSD|her mom IAUX|do-d~NEG|not AUX|haveto AUX|be tell-n I|to do DQUA|any+thing
%syn:	ADVP <ip SNP ([SWHREL] <ip Vf PP [ONJ] <ip Vf DNP PP) AUXD NEG AUX AUXB V-en ([TO] V DNP))
%lex:	ADV|one N|time N|child V|live N|home V|clean N|house N|mom V|tell V|do N|any+thing
%com:	'one time' seems out of place here.
*MIC:	she just didn't.
%mor:	IPROS|she just IAUX|do-d~NEG|not
%syn:	<ip SNP ADVP AUXD NEG 0VP
*MIC:	and did it and she got rewarded for it.
%mor:	ONJ|and IT|do-d PRO|it ONJ|and IPROS|she IAUX|get-d reward-en P|for PRO|it
%syn:	[ONJ] =0SNP Vf DNP [ONJ] <ip SNP AUXG V-en PP
%lex:	V|do ADJ|rewarded
*MIC:	and so she kept cleaning at the house.
%mor:	ONJ|andso IPROS|she IAUX|keep-d clean-ing P|at DART|the house
%syn:	[ONJ] <ip SNP Vf ([PC] V-ing PP)
%lex:	V|keep V|clean N|house
*MIC:	and &she # she um helped her mom out everywhere.
%mor:	ONJ|and IPROS|she IT|help-ed DPOSD|her mom PRT|out DQUA|every+where
%syn:	[ONJ] <ip SNP Vf DNP PRT ADVP
%lex:	V|helpout N|mom ADV|every+where
*INV:	mm hm.

APPENDIX (Continued)

*MIC:	and her mom was happy.
%mor:	ONJ|and DPOSD|her mom IAUX|be-s-d happy
%syn:	[ONJ] <ip SNP AUXB PAP
%lex:	N|mom V|be ADJ|happy
*INV:	mm hm. great. that was very good.
%com:	next story is played.
*INV:	okay. can you try and tell me that story.
*MIC:	Tim # Tim was in a school that he didn't+…
%mor:	tim IAUX|be-s-d P|in DART|a school CREL|that IPROS|he IAUX|do-d~NEG|not
%syn:	<ip SNP AUXB PP ([DREL] <ip SNP AUXD NEG =0VP)
%lex:	N|tim V|be N|school
%com:	unfinished sentence
*MIC:	all the kids didn't like him him.
%mor:	DQUA|all DART|the D|kid-pl IAUX|do-d~NEG|not like PROO|him
%syn:	<ip SNP AUXD NEG V DNP
%lex:	ADV|all N|kid V|like
*MIC:	and they kept on laughing at him and &he # he fell in the puddle
%mor:	ONJ|and IPROS|they IT|keep-d PRT|on laugh-ing PRT|at PROO|him ONJ|and IPROS|he IT|fall-d P|in DART|the puddle
%syn:	[ONJ] <ip SNP Vf PRT ([PC] V-ing PP) [ONJ] <ip SNP Vf PP
%lex:	V|keepon V|laugh V|fall N|puddle
*MIC:	and the kids still laughed.
%mor:	ONJ|and DART|the D|kid-pl still IT|laugh-ed
%syn:	[ONJ] <ip SNP ADVP Vf
%lex:	N|kid ADV|still V|laugh
*INV:	right, what did you think about this story?
*MIC:	mean, the kids were mean.
%mor:	DART|the D|kid-pl IAUX|be-s-d mean
%syn:	<ip SNP AUXB PAP
%lex:	ADJ|mean N|kid V|be
*INV:	you didn't like it.
*MIC:	mm mm.
%com:	Michelle shakes head.
*INV:	so what didn't you like about it?
*MIC:	that the kids laughed at him.
%mor:	CCOM|that DART|the D|kid-pl IT|laugh-ed P|at PROO|him
%syn:	<cp[CCOM] <ip SNP Vf PP
%lex:	N|kid V|laughat
*INV:	mm hm. and what else didn't you like about it?
*MIC:	the kids would just be mean to him.
%mor:	DART|the D|kid-pl IAUX|would just AUX|be mean P|to PROO|him
%syn:	<ip SNP AUXM ADVP AUXB PAP PP
%lex:	N|kid ADV|just V|be ADJ|mean
*INV:	how would it make you feel when you heard that? xxx
*MIC:	sad.
%mor:	sad
%syn:	AP
%lex:	ADJ|sad
*INV:	mm hm. was there anything you liked about the story?
*MIC:	mm # that he ignored him, them.
%mor:	CCOM|that IPROS|he IT|ignore-ed PROO|them
%syn:	<cp[CCOM] <ip SNP Vf DNP
%lex:	V|ignore
*MIC:	yeah.
*INV:	you liked that. what was good about that?
*MIC:	cause he didn't pay any attention to them.
%mor:	CADJ|because IPROS|he IAUX|do-d~NEG|not pay DQUA|any attention P|to PROO|them
%syn:	<cp[CADJ] <ip SNP AUXD NEG V DNP PP
%lex:	V|pay N|attention
*MIC:	and they # sometimes don't get on your nerves if they do that # if you do that.
%mor:	ONJ|and IPROS|they DQUA|some+times IAUX|do~NEG|not get P|on DPOSD|your D|nerve-pl CADJ|if PRO|you do DPRO|that
%syn:	[ONJ] <ip SNP ADVP AUXD NEG V PP ([CADJ] <ip SNP V DNP)

APPENDIX (Continued)

%lex:	ADV|some+times V|getonyournerves V|do
*INV:	so that you think that if you don't pay any attention then the kids xxx.
%com:	Michelle nods.
*MIC:	mm hm.
*INV:	um, how do you think Tim feels at school?
*MIC:	sad.
%mor:	sad
%syn:	AP
%lex:	ADJ|sad
*INV:	why's that?
*MIC:	cause the kids tease him all the time and laugh at him.
%mor:	CADJ|because DART|the D|kid-pl tease PROO|him DQUA|all DART|the time ONJ|and laugh P|at PROO|him
%syn:	0CP [CADJ] <ip SNP V DNP ADVP [ONJ] V PP
%lex:	N|kid V|tease ADV|allthetime V|laugh
*INV:	how do you think he felt when he fell?
*MIC:	um # he hurts.
%mor:	IPROS|he IT|hurt-=s
%syn:	<ip SNP Vf
%lex:	V|hurt
%err:	tense error, 3rd sing. present agreement on the verb instead of past tense
*INV:	he hurts?
*MIC:	mm hm.
*INV:	and when all the children started to laugh at him?
*MIC:	sad.
%mor:	sad
%syn:	AP
%lex:	ADJ|sad
*INV:	how would you have felt if this happened to you?
*MIC:	sad and angry and upset.
%mor:	sad and anger-y and upset
%syn:	AP
%lex:	ADJ|sad ADJ|angry ADJ|upset
*INV:	what would you have done?
*MIC:	I would have gone and tell the teacher.
%mor:	IPROS|I IAUX|would AUX|have go-n ONJ|and tell-=0n DART|the teacher
%syn:	<ip SNP AUXM AUXH V-n [ONJ] Vn INP
%lex:	V|go V|tell N|teacher
*INV:	and how would that have helped?
*MIC:	they would send him to the principal's office.
%mor:	IPROS|they IAUX|would AUX|=0have send-=0n PROO|him P|to DART|the D|principal-'s office
%syn:	<ip SNP AUXM =0AUXH Vn INP PP
%lex:	V|send N|principal N|office
*INV:	and then what would happen?
*MIC:	I don't know.
*MIC:	the principal would punish him maybe.
%mor:	DART|the principal IAUX|would punish PROO|him maybe
%syn:	<ip SNP AUXM V DNP ADVP
%lex:	N|principal V|punish ADV|maybe
*INV:	mm hm. that would make you feel better?
*MIC:	mm hm.
*INV:	why do you think the other children don't like Tim?
*MIC:	I don't know.
*MIC:	maybe he's not the kind of person they like to hang out with but+...
%mor:	maybe IPROS|he~IAUX|be-s NEG|not DART|the kind P|of person CREL|0that IPROS|they like I|to hang PRT|out P|with
%syn:	ADVP <ip SNP AUXB NEG PNP ([0OREL] <ip SNP V ([TO] V PRT P))
%lex:	ADV|maybe V|be N|kind N|person V|like V|hangout
*INV:	what would be a reason for a kid not to like another kid?
*MIC:	they look weird.
%mor:	IPROS|they look weird
%syn:	<ip SNP V PAP
%lex:	V|look ADJ|weird

APPENDIX (Continued)

*INV: why is that a reason not to like a kid?
*MIC: I don't know.
*INV: what do you think?
*MIC: I don't know.
*INV: if somebody looks weird xxx you know why do kids not like him?
*MIC: uh # cause they think they're geeks or something, but they aren't.
%mor: CADJ|because IPROS|they think CCOM|0that IPROS|they~IAUX|be-s D|geek-pl or something
          ONJ|but IPROS|they IAUX|be-s~NEG|not
%syn: 0CP ([CADJ] <ip SNP V ([0CCOM] <ip SNP AUXB PNP) [ONJ] <ip SNP AUXB NEG 0PNP)
%lex: V|think N|geeks V|be V|be
*INV: mm hm. what would you do if the children at school didn't like you?
*MIC: I would ignore them.
%mor: IPROS|I IAUX|would ignore PROO|them
%syn: <ip SNP AUXM V DNP
%lex: V|ignore
*INV: you would ignore them?
*MIC: mm hm.
*INV: and how would that help?
*MIC: they wouldn't bother me anymore.
%mor: IPROS|they IAUX|would~NEG|not bother PROO|me DQUA|any+more
%syn: <ip SNP AUXM NEG V DNP ADVP
%lex: V|bother ADV|any+more
*INV: they wouldn't bother you.
%com: Michelle shakes head.
*INV: you know when Tim saw the children talking and laughing and he thought they were talking and
          laughing about him. do you think they were?
%com: Michelle shakes head.
*INV: <no, maybe they> [>].
*MIC: <they might have and they might have not been > [<].
%mor: IPROS|they IAUX|might AUX|have ONJ|and IPROS|they IAUX|might AUX|have NEG|not be-n
%syn: <ip SNP AUXM AUXH 0VP [ONJ] <ip SNP AUXM AUXH NEG V-n 0VP
*MIC: you never know.
%mor: PRO|you NEG|never know
%syn: <ip SNP ADVP V
%lex: V|know ADV|never
*INV: right. what do you think?
*MIC: probably.
%mor: probably
%syn: ADVP
%lex: ADV|probably
*MIC: I don't know.
*INV: probably what?
*MIC: probably they were.
%mor: probably IPROS|they IAUX|be-s-d
%syn: ADVP <ip SNP AUXB 0VP
%lex: ADV|probably
*INV: they were. what kind of things do you think they were saying?
*MIC: mean things about him.
%mor: mean D|thing-pl P|about PROO|him
%syn: NP PP
%lex: ADJ|mean N|thing
*INV: like what?
*MIC: he looks ugly or something like that.
%mor: IPROS|he IA|look-s ugly or DQUA|some+thing P|like PRO|that
%syn: <ip SNP Vf PAP [ONJ] PNP
%lex: V|look ADJ|ugly
*INV: mm hm. mm hm. so you don't think it was his imagination.
*MIC: mm mm.
%com: Michelle shakes head.
*INV: do you ever know any kids who do mean and sneaky things to you?
*MIC: uh huh.
*INV: what kind of things did they do? xxx.

APPENDIX (Continued)

*MIC:	they would # um.
%mor:	IPROS|they IAUX|would
%syn:	<ip SNP AUXM
%lex:	
*MIC:	one of my friends in fact she # came to my mom's shop and said something that I didn't say, I mean.
%mor:	DCAR|one P|of DPOSD|my D|friend-pl in fact IPROS|she IT|come-d P|to DPOSD|my D|mom-'s shop ONJ|and IT|say-d DQUA|some+thing CCOM|that IPROS|I IAUX|do-d~NEG|not say
%syn:	TSNP ADVP <ip SNP Vf PP [ONJ] Vf DNP ([DREL] <ip SNP AUXD NEG V)
%lex:	N|friend V|come N|mom N|shop V|say V|say N|one
*MIC:	and I got in trouble for it.
%mor:	ONJ|and IPROS|I IAUX|get-d P|in trouble P|for PRO|it
%syn:	[ONJ] <ip SNP AUXG PPP PP
%lex:	N|trouble
*INV:	oh really. what did she say?
*MIC:	she said that I called my mom selfish stuck up which I didn't.
%mor:	IPROS|she IT|say-d CCOM|that IPROS|I IT|call-ed DPOSD|my mom self-ish ONJ|=0 stick-n PRT|up CREL|which IPROS|IIAUX|do-d~NEG|not
%syn:	<ip SNP Vf ([CCOM] <ip SNP Vf DNP PNP ([DWHREL] <ip SNP AUXD NEG 0VP))
%lex:	V|say V|call N|mom ADJ|selfish ADJ|stuckup
*MIC:	and # my mom found out and I got in big trouble.
%mor:	ONJ|and DPOS|my mom IT|find-d PRT|out ONJ|and IPROS|I IAUX|get-d P|in big trouble
%syn:	[ONJ] <ip SNP Vf PRT [ONJ] <ip SNP AUXG PPP
%lex:	N|mom V|findout V|getintrouble ADJ|big
*INV:	that is pretty mean.
*MIC:	and she was just playing a joke.
%mor:	ONJ|and IPROS|she IAUX|be-s-d just play-ing DART|a joke
%syn:	[ONJ] <ip SNP AUXB ADVP V-ing DNP
%lex:	V|play N|joke ADV|just
*INV:	oh really?
*MIC:	yeah.
%com:	Michelle nods.
*INV:	wow. what other kinds of mean things have they xxx?
*MIC:	kids have laughed at me at school when I used to have seizures and everything so+...
%mor:	D|kid-pl IAUX|have laugh-en PRT|at PROO|me P|at school CADJ|when IPROS|I IAUX|usedto have D|seizure-pl and DQUA|every+thing
%syn:	<ip SNP AUXH V-en PP PP ([CADJ] <ip SNP AUX V DNP)
%lex:	N|kid V|laughat N|school V|have N|seizure
*INV:	they did laugh?
*MIC:	uh huh.
*INV:	what did they say?
*MIC:	I can't remember.
%mor:	IPROS|I IAUX|can~NEG|not remember
%syn:	<ip SNP AUXM NEG V
%lex:	V|remember
*INV:	pretty mean.
*MIC:	yeah.
*INV:	any other sneaky things?
*MIC:	mm mm.
%com:	Michelle shakes head.
*INV:	how about you. do you ever do any sneaky things?
*MIC:	mm mm.
%com:	Michelle shakes head.
*INV:	never ever?
%com:	Michelle shakes head.
*INV:	do you think this story can be a true story?
*MIC:	mm hm.
%com:	Michelle nods.
*INV:	why's that?
*MIC:	cause kids do laugh at other kids sometimes.
%mor:	CADJ|because D|kid-pl IAUX|do laugh at other D|kid-pl DQUA|some+times
%syn:	<cp[CADJ] <ip SNP AUXD V PP ADV

APPENDIX (Continued)

%lex: N|kid V|laughat ADJ|other N|kid ADV|sometimes
%com: emphatic 'do'.
*INV: mm hm, mm hm. anything else that makes it a true story?
*MIC: not that I know.
%mor: NEG|not CREL|that IPROS|I know
%syn: 0SNP 0AUXB NEG 0PNP ([DREL] <ip SNP V)
%lex: V|know
%com: Michelle shakes head.
*INV: great. let's switch this off.
@End