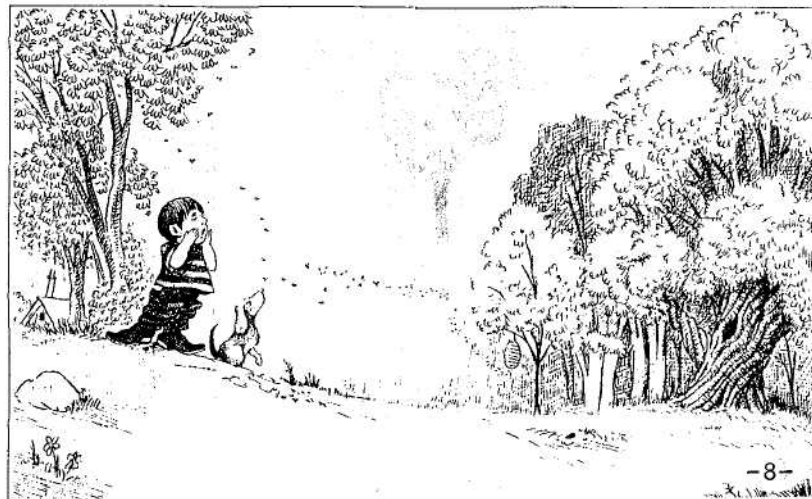# Appendix I

Frog, where are you?

-9-

-10-

-11-

-12-

-13-

-14-

-15-

-16-

-17-

-18-

-19-

-20-

-21-

-22-

-23-

-24-
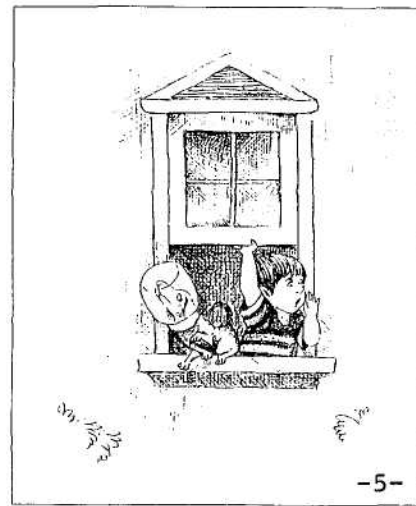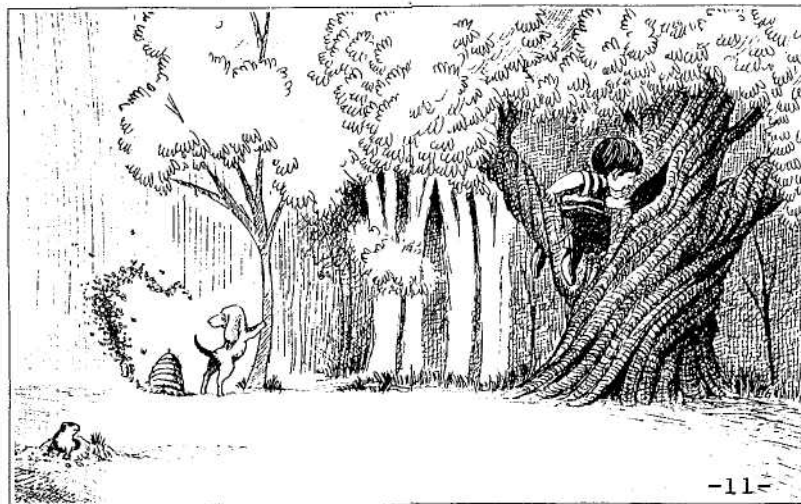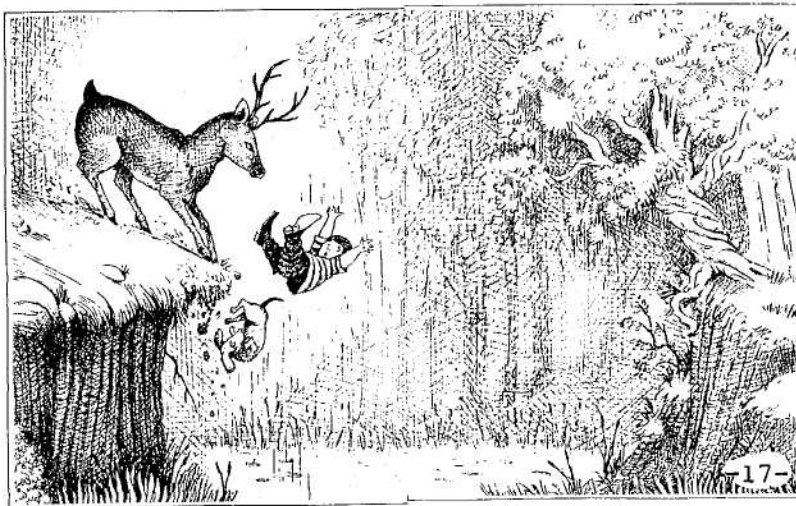
Appendix II GLOSSING AND

TRANSCRIPTION CONVENTIONS

1. GLOSSING

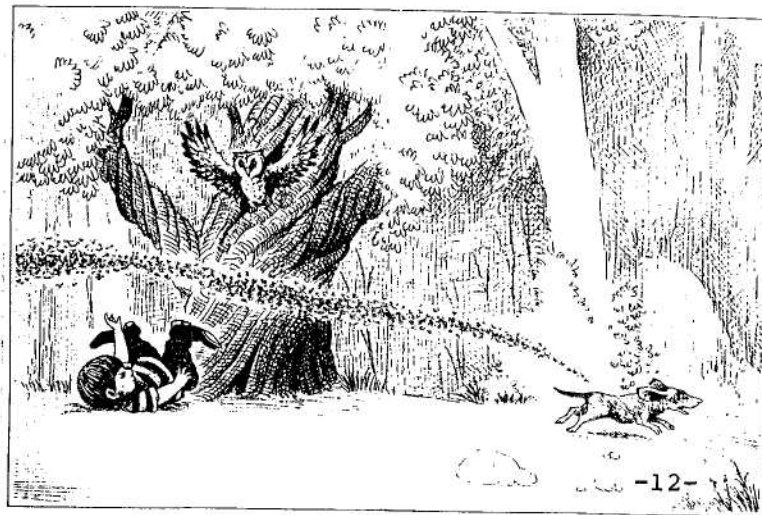All examples from the transcripts, in any language, are given in italics. Boldface (both roman and italic) is used for authors' emphasis. Words in capitals indicate intonational emphasis, based on the tape recordings. Glosses are given in single quotes, e.g., *kurbaga* 'frog'. Where an item would normally be obligatory in English but is omitted grammatically in the source-language, the "missing" element is given in parentheses, such as pronouns in "pro-drop" languages, e.g., *cayeron* '(they) fell'. In some contexts, the symbol $\varnothing$ is used to represent a zero-morpheme. Where necessary for interpretation, references of pronouns are indicated by an equal sign in square brackets, e.g., *Er fällt runter* 'He [=boy] fall down'.

The following conventions are used for glossing grammatical morphemes:

- Elements in the gloss which are expressed by a single lexical item in the original are separated in the gloss by a period, e.g., *hipil* 'made.fall', as are names for grammatical elements, e.g., PRES.PERF = present perfect.

- Grammatical morphemes are appended after a colon, e.g., *cayó* 'fell:PFV'. The same convention is used for combinations of grammatical morphemes, e.g., *düşüirüyor* 'fall:CAUS:PRES'.

- In examples for which morphological segmentation is relevant to the discussion, morphemes are separated by hyphens, corresponding to hyphens in the gloss, e.g., *geç-ir-iyor* 'get.on-CAUS-PRES.

- If an example has both a morpheme-by-morpheme and a more colloquial gloss, the latter is given in single quotes, using the following format:

  Bayku§ düşüirüyor onu, köpek de kaçvyor.

  owl fall:CAUS:PRES him dog TOPIC run.away:PRES

  "The owl knocks him [=boy] down; and as for the dog, he runs away.'

Grammatical categories a# abbreviated and given in small caps. The following abbreviations are used:

ABL = Ablative
ACC = Accusative
ART = Article
ASP = Aspect
AUX = Auxiliary
CAUS = Causative
CL = Clitic
COMPL = Completive
CONTIN = Continuous, Continuative
COP = Copula
DAT = Dative
DEF = Definite
DO = Direct Object
D.PAST = Direct Experience Past
EVID = Evidential
FEM = Feminine
FIN = Finite
FUT - Future
GEN = Genitive   :
HAB = Habitual
INCH = Inchoative
INF = Infinitive
INFER = Inferential
INSTR = Instrumental
INTRANS = Intransitive
10 = Indirect Object
IPFV = Imperfective
IRR = Irrealis
ITER = Iterative
LOC = Locative
MASC = Masculine
MOD = Modal
M.PAST = Modal/Indirect Experience Past
N = Noun
NOML = Nominalization

NEG - Negative
NOM = Nominative
NONPAST = Nonpast
NUM = Numeral, Numeric
OBJ = Object
OPT = Optative
PART = Participle
PASS = Passive
PAST = Past
PAT = Patient
PERF = Perfect
PFV = Perfective
PL = Plural
POSS = Possessive
PP = Past Participle
PREP = Preposition
PRES = Present
PRO = Pronoun
PROG = Progressive
PTL = Particle
PURP = Purposive
RECIP = Reciprocal
REFL = Reflexive
RC = Relative Clause
REPET = Repetition
RES = Resultative
SG = Singular
SIM = Simultaneity
STAT = Stative
SUBJ = Subject
TNS = Tense
TOP = Topic
TRANS = Transitive
V = Verb
VN = Verbal Noun

Narrator identification. Every example from a narrative is followed by a narrator identifier in square brackets, indicating the language, age group, specific narrator, and specific age. For example, *[E4c-4;6]* denotes an

English-speaking 4-year-old who was the third child *(4c)* in the sample, aged 4;6. Any linguistic example without such an identifier is an invented example.

2. TRANSCRIPTION

A uniform format was applied across the sample for transcribing the texts. The basic unit of analysis is the **clause,** defined for this study as "any unit that contains a **unified** predicate" by which we mean "a predicate that expresses a **single** situation (activity, event, or state), including finite and nonfinite verbs as well as predicate adjectives" (Berman, et al., 1986, p. 37; and see Chapter IB, Section 2.3). Each clause was entered on a new textline, as in the example reproduced below of the text of an English-speaking 5-year-old *(E5d-5;6).* (To aid researchers who might want to use our definition of the clause in their own transcriptions, this Appendix ends with the extended instructions to coders from our coding manual [Berman, et al., 1987.)

Transcripts, in this form, are available for public use in the CHILDES archive (MacWhinney, 1991; e-mail *info-childes@andrew.bitnet* or *info-childes@andrew.cmu.edu).* Each clause in a transcript is preceded by an ID code which identifies the subject, the picture to which the utterance refers, and a clause number. The prefix of the ID consists of six digits for children and four digits for adults, and specifies the subject's age and position in that age group (for each particular language). The data portion of the ID specifies the picture that the subject was looking at when s/he produced the utterance entered on that textline and the number of the clause in that text. Thus, for example, in the text reproduced below, the string **05;06D 03b005** refers to a child aged 5 years and 6 months who is the fourth child in that age group (in the English sample), talking about picture 3b, and the textline is the fifth clause out of the total 24 in that text. An example of an adult ID would be **20j llaO95,** standing for the tenth adult talking about picture lla, in the 95th clause in that text, and assigning an arbitrary age of 20 to all adult subjects.

The first three digits in the data portion refer, as noted, to the **page** in the picturebook which the subject was referring to when producing that line of the text. Unlike the consecutive numbering we have used for the frog-book pictures throughout this study, from 1 to 24, the original transcriptions list the pages as from 01- to 15-, with a hyphen (-) standing for a picture that fills two facing pages, the letter **a** specifying a lefthand picture and **b** a righthand picture on pages which have two separate pictures. In order to identify the picture the subject was referring to for every piece of the recording, the investigator would lightly tap on the microphone when the subject moved his or her eyes to the next part of the page, or turned the page. When the subject did not refer to each page consecutively, but either skipped or turned back one or

more pages, this was noted down by the investigator and so indicated on the transcript.

A new textline was used for each clause, even if this was not a complete sentence or a well-formed subordinate or nonfinite clause — as in clause #1 in this sample transcript. However, a single clause might contain more than one verb, when one of the verbs expresses an aspectual or modal specification of the main verb, rather than an independent event or state — as in clauses 2, 3, and 11, *and then they* **were going to sleep,** *and then he's* **gonna go out** *the window, and they* **start calling.**

All text was entered in lowercase letters, using conventional orthography and spelling, except where potentially relevant for morphosyntactic or semantic analysis. (Following German orthographic convention, German nouns were capitalized.) For example, the *-ing* form of the verbs *looking* and *calling* in the sample text would be transcribed as such even if the child pronounced them as *lookin, callin.* Another example is of initial *h* in Hebrew, which is quite typically elided in everyday adult as well as child speech, but which was always indicated so as to ensure consistency of lexical forms.

In the case of Hebrew, the only language in our sample with a non-Roman script, texts were entered in broad phonetic transcription (see Chapter HID).[1]

Material in square brackets indicates interviewer comments to the child and remarks on the situation (e.g., what page is being described in unclear or nonconsecutive references, as in clauses 7 and 16 in this text). Curly brackets indicate false starts or repairs (e.g., clause 24), uninterpretable strings (e.g., clause 8), and other material which was not included in our coding system (Chapter IB, Section 2.3). Thus, all material, except what occurs between square brackets, represents the subject's output as recorded by the investigator. Text length in clauses *{IB-Table 3)* is based only on clauses **not** totally enclosed in curly brackets. (Thus the sample in 2.1, below, has 23 codable clauses, excluding 008.)

---

[1] The main departures from fairly standard English distinctions in the Hebrew transcriptions are: *c,* which stands for the Hebrew letter *cade,* pronounced as *ts,* as in the loan English word *tsetse,* and *x,* which stands for the voiceless velar fricative (and also the historical pharyngeal *chef),* pronounced like the final consonant in words like *Bach, loch.* These occur in the Hebrew frog story, in words like *cfardea* 'frog', *cincenet* 'jar', and *rac* 'run' or *xipes* 'searched', *baxa* 'cried'. The glottal stop *alef,* where needed to represent a historical root (it is typically not pronounced in normal speech in word-initial or word-final position) is represented by a question-mark, and the apostrophe represents the historical pharyngeal *ayin* for roots, and otherwise is used to indicate a modified glottal stop between two vowels, e.g., *ne'elam* 'disappeared', *le-he'alem* 'to-disappear'.

The following standardized conventions were used to indicate (minimal) prosodic features of the text:

> Default (no indication) = steady or sustained intonation
> A dash (-) = a short pause
> Three dots (or more if needed) = a longer pause
> Comma (,) = partially falling intonation
> Period (.) = fully falling intonation, end of utterance
> Question-mark (?) = end of question-type utterance
> Single slash (/) = rising intonation or other indication of marked emphasis
> Double slash (//) = exaggeratedly marked emphasis
> Exclamation-mark (!) = on both sides of a word or series of words,
>   indicates excited delivery (e.g., ! Look !)
> Colon (:) = lengthening of a vowel (e.g., *rum* stands for a long drawn-out
>   pronunciation of the English verb 'run' and *ra:c* stands for
>   something like *raaaac* in its Hebrew equivalent
> Three exes (xxx) = an unintelligible string
> Parentheses ( ) = uncertain, but possible interpretations of a string.

Other comments on paralinguistic features such as speech quality are indicated in square brackets, e.g., [whispering], [child laughs]. As noted in various connections throughout the book (e.g., in Chapter IB-Section 2.4, Chapter IlA, Chapter IVC, Section 2), this level of transcription does not allow for in-depth analysis of the role of prosodic (or other paralinguistic features) elements in telling the frog story.

2.1. Sample Transcript

05;06D	01-001	looking in [uhhuh]
05;06D	02a002	and then {he - } they were going to sleep,
05;06D	02a003	and then he's - gonna - go out the window [singsong]
05;06D	03a004	[he's gonna go out the window / oh.]
		and then he - he had - a jar stuck on his head,...

[uhhuh/]

05;06D	03b005	and then they were - calling for the frog, [umhm/]
05;06D	04-006	and then he - breaks - the jar. [oh.]
05;06D	05-007	there's a - beehive, [yep.] and a gopher, [from page 6a]
05;06D	06-008	{ xxx }
05;06D	07-009	and ... the beehive fell down, [umhm/]
05;06D	07-010	and they look in the tree,
05;06D	09-011	and they ... and they start calling, [umhm/]
05;06D	09b012	and they think
05;06D	09b013	that's sticks,

05;06D	10a014	but it's a d - um - an antler, [antlers - ah.]
05;06D	10b015	and then they run,
05;06D	11-016	and then they - go in the water, [oh.] [also page 12]
05;06D	13-017	and behind that log is frogs, [oh, really / aha.]
05;06D	14bO18	there's some more frogs, [yeah /]
05;06D	15-019	and there's another frog.
05;06D	15-020	[What about the boy?] he goes in the water ... [uhhuh/]
05;06D	15-021	and they look...
05;06D	15-022	and watch.
05;06D	15-023	oh there's one//...
05;06D	15-024	{they can't get - } he can't get up .

2.2. Definition of " Clause" as Unit of Transcription[2]

Each clause should be transcribed on a new textline. We define a clause as any unit that contains a unified predicate. By unified, we mean a predicate that expresses a single situation (activity, event, state). Predicates include finite and nonfinite verbs, as well as predicate adjectives. In general, clauses will be comprised of a single verbal element; however, infinitives and participles which function as complements of modal or aspectual verbs are included with the matrix verb as single clauses, e.g., *want to go, started walking.* These matrix verbs plus modifiers should not be confused with utterances that clearly express two "situations," as in subordinate complement clauses, e.g., *I thought that I would go.* As illustration, each of the following phrases would be analyzed as a single clause with a unified predicate: *running through the woods; taken by surprise; (in order) to help his friends; was angry.*

In general, then, treat as a single clause those utterances that have two verbs but one subject, and treat as two separate clauses cases when each verb has a different subject, e.g., */ want to go* vs. */ want you to go.* Predicates that are clearly narrator-comments are kept with the matrix verb in a single predicate, e.g., */ assume that the boy is happy; it appears that the dog is going to fall.* These phrases constitute one clause.

2.2.1. Single Clause Examples

Single clause with two verbs and one subject: *he stopped running; they had begun to search all over.*

Single clause with different subjects (Narrator Comments): */ think the boy misses the frog; it appears that the frog is happy.*

---

[2] This section is taken directly from our coding manual (Berman, et al., 1986); therefore, "clause" is defined in terms of instructions to coders.

2.2.2. Two Clause Examples

Subordinate Complements — Same Subject NP:

> he thought
> he could get the bees .

> he said
> he would find the frog .

Subordinate Complements - Different Subject NP:

> he decided
> that it was an owl.

> he told the dog
> to be quiet.

2.2.3. Special Cases

Verbless Clauses: While a clause is defined by the presence of a "unified" predicate, a clause need not contain an overt verb. For example, copular sentences in Hebrew and Turkish which have zero-expression of the copula in the present tense are coded as a separate clause.

Ellipsis and Gapping: Treat as separate clauses strings in which the verb is lacking due to grammatical reductions such as gapping and where the verb semantics is fully recoverable from the text, or structures which can be analyzed as clauses where the copula has been deleted.

> the boy looked in his boots
> and the dog in the jar .

> with the frog not there the
> boy felt very upset.

> he began searching for the frog
> and the dog too .

Crucial Plot Information: Verbless clauses that are critical to the story-line should be coded on a separate line.

> and then he climbs over ,
> and ! little ! baby frogs .

Do not code separately as a clause speech that would otherwise be considered part of another utterance if not for considerations of intonation or the

subject's orientation to the picture book (i.e., the page at which the subject is currently looking). ($ indicates page turn in middle of utterance.)

> [pg. 5] there were bees . $ [pg. 6] and a gopher .

Do not code speech strings that are not elaborated in adjacent utterances and that do not appear to be part of the storyline. Enclose these portions of text in curly brackets and omit from coding.

> { boy } [mhm /] { dog. }[mhm /] { frog.} [
>
> mhm /] [looking at pg. 2]
>
> { ! bees ! } [looking at pg. 5]

2.2.4. Center-Embedded Clauses/Discontinuous Constituents

In order to have a constant measure of story length in number of clauses, and to allow for potential coding of each clause separately, center-embedded relative clauses are enclosed in curly brackets { } on the original textline (i.e., not coded), displaced from their original location, and coded on a new line after the associated main clause. Indicate that the clause has been displaced from its original location by enclosing it in angle brackets <> on the textline where it is coded.

> the owl { who pushed the boy } flew away .
> < who pushed the boy >
>
> little Moritz { who I just call little Moritz } sits **in front of his bed** .
> < who **I** just call little Moritz >

Some adverbial clauses may be separated from their associated main clause by an introductory adverbial phrase which sets the temporal anchor for the main clause, e.g., *in the middle of the night, one day, at night.* Because these adverbial phrases are syntactically associated with the main clause and not the subordinate adverbial clause, they should be enclosed in curly brackets { } on the original textline — i.e., not coded — displaced to the clause line containing the main clause, placed in angle brackets < > , and coded with that main clause.

> {in the middle of the night } , after they had gone to sleep
> < in the middle of the night > the frog escaped from the jar .
>
> { and after that } when they had gotten to the other side
> :   < and after that > they saw a whole lot of little frogs .

:VA Subordinate clauses of other embedded phrases that do not have adverbial anchors are treated normally and coded in order of the surface predicates.

> after they woke up
> they realized
> that the frog was gone .

> while they were sleeping and the
> moon was shining brightly the little
> frog jumped out.

2.2.5. *Additional* Coding Conventions

Task Comments: Asides which are not part of the type of text under study in the frog stories, such as task-oriented questions to the interviewer or expressions that are used to engage the listener, are not included in the clause count and are not coded.

> {I think we could be finished now . } {
> what's that called / }

> { is that a gopher / }
> { should I turn the page now / }
> { ! look ! }
> { lookit what happened . }
> { see / }
> {I don't know.}   .
> { do you know what that is / }
> {isn't it/}
> {right/}

or parts of utterances which are formulaic with respect to the story-telling situation are not counted or coded.

> { once upon a time }
> { and they lived happily ever after }
> {the end }
> { and that's all. }

Unintelligible Speech: Clauses or parts of clauses that contain a substantial portion of unintelligible speech should not be coded, especially if an accurate interpretation of the temporal quality of the clause or the predicate in general is jeopardized.

Reformulations/False Starts: The fullest version is coded in clauses containing reformulations and false starts. The repaired portion is enclosed in curly brackets { }.

{ and then he starts } ... then a gopher pops out of the hole,

{he - the dog } - the dog was barking at the deer.