

TRANSCRIPTION MANUAL

Based on modified SALT Transcription Format

Thomas Klee
Dept. of Communication Disorders, University of Canterbury
Christchurch, New Zealand

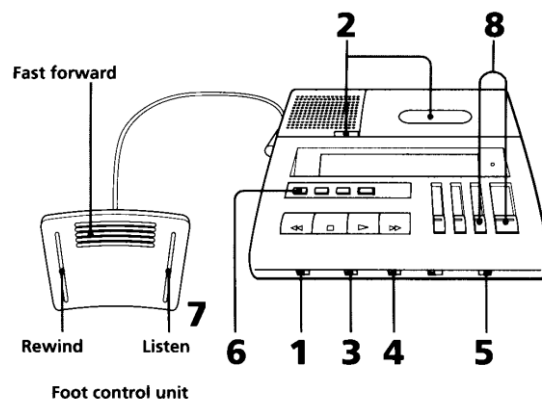
This document outlines modifications we have made to the SALT transcription format originally developed by Jon Miller and Robin Chapman at the University of Wisconsin-Madison in the 1980s. These modifications reflect coding decisions that were made for the purpose of our own research using language sample analysis (e.g., Gavin et al., 1993; Klee et al., 1989, 2007; Klee, 1992). The original SALT transcription conventions are described in the SALT program documentation and may be found by clicking on Help in SALT.

The document begins with a description of how audiotape can be transcribed using Sony transcription machines. You will want to read that section if you used a standard cassette tape when you recorded your language sample. Digital media (e.g., personal voice recorders, MP3 recorders, minidisk recorders) have for the most part replaced audiotape recordings so you will have to decide for yourself the most efficient means of transcribing your recording. You might want to try out one of the transcription software programs that can be download for free on the Web, such as Express Scribe (<http://www.nch.com.au/scribe/>). Other software programs can be found at <http://www.saltsoftware.com/salt/digital/index.cfm>.

You might also find that using your digital recorder to manually play back the recording is the easiest way to do this.

The transcription machine (for audiotape recordings)

The Sony BM-77 transcription machine has a playback frequency response of 200–8000 Hz (at 4.8 cm/s playback speed). The use of Type I tape is recommended in this machine. The use of chromium dioxide tape (Type II) or metal tape (Type IV) is not advisable. Cassettes of either 60 or 90 minutes duration are recommended (i.e., 30 or 45 minutes per side). Cassettes with a running time of more than 90 minutes are not recommended, as the digital counters on the machine may not function properly. Headphones (Beyerdynamic DT 311) are provided with each machine.



The mains power for each workstation (i.e. transcription machine and computer) is located on the wall next to each pair of workstations. Insert the tape by pressing the eject button to the left of the cassette door [2]. Slide the tape into the bracket under the door and then close the door. The transcription machine is controlled by a foot pedal, which frees your hands for

typing text into the computer. The foot pedal contains a 3-way switch. You can listen to the tape by pressing on the right side of the pedal. The tape can be rewound by pressing on the left side and fast forwarded by pressing on the back edge. (The transcriber also contains buttons on its face for these functions, but these are not normally used during transcription.) After inserting the audiotape into the transcriber, rewind it if necessary, put on the headphones, and press your foot down on the right half of the foot pedal. You should hear what you have recorded.¹ When you lift your foot, the tape will stop and automatically rewind a bit. When you press the pedal again, you will hear the same last few seconds of the recording. This allows you to listen to a segment of speech repeatedly with a single action of your foot. The amount of tape that automatically rewinds may be altered by adjusting the REVERSE TIME control. At position '9', the tape is rewound so that about a 5-second stretch can be reviewed every time you lift your foot. Less time may be selected by reducing the number ('4' or '5' is typical). At position '0', the tape stops without being rewound at all.²

The VOLUME and TONE controls [8] should be adjusted to comfortable levels. Playback speed can be changed by adjusting the speed control, although this is normally not necessary. The speed of playback can be slowed down by as much as 20% or increased by as much as 80%. To use this, set the SPEED CONTROL to ON [4] and adjust the SPEED switch on the top of the machine. Normally the SPEED CONTROL should be switched to the OFF position.

The digital counters on the face of the transcriber keep track of approximately how much tape has been transcribed. The counter labelled TOTAL is a tape counter; the one labelled UNIT is a time counter (in minutes and tenths of a minute). The counters should be set to zero at the beginning of the recording (press RESET). *A more accurate timing device should be used if precise pause or timing features are to be coded.*

Usual settings:	POWER [1]:	on
	AUTO STOP [3]:	off
	SPEED CONTROL [4]:	off
	TAPE SPEED [5]:	4.8 cm/s

Transcription notation

The transcript serves as the raw data for any linguistic analyses which will be carried out on the language sample. As such, it is a critical component in the process of language sample analysis. One must begin with an accurate account of what an individual has said (i.e. a valid transcription) for the linguistic analyses to be meaningful. There is no one 'correct' or commonly accepted system of transcription notation used in the field. Rather, the format of transcription tends to vary with the research or clinical questions being asked, with the kinds of analyses to be done or perhaps with the medium to be used (e.g., paper and pencil, text editor, word processor).

At the moment, two main analysis programs are available in the Language Analysis Lab but, regrettably, each one requires its own transcription format³. These are the *Child Language*

¹As the transcription machine is 4-track, 2-channel and monaural, recordings must be made on the right channel of the tape. If you don't hear anything on a tape you've made, it may be that you have recorded onto the left channel. Don't panic: the recording can be dubbed onto the right channel of another cassette.

² If the REVERSE TIME control doesn't seem to work, the electronic index mode switch may be incorrectly set. To disable the electronic index, stop the tape; then, press and hold RESET and STOP for more than 2 seconds.

³ CLAN contains a program for converting files to and from SALT transcription format.

Analysis (CLAN) and the *Systematic Analysis of Language Transcripts (SALT)* programs. Documentation for the CLAN program may be found in MacWhinney (1994), a copy of which is available in the lab. Many practical examples of how researchers have used CLAN are illustrated in Sokolov and Snow (1994). The documentation for the SALT program is also available in the lab, and on-line, as part of the software (Miller & Chapman, 1996). SALT is also documented in Leadholm and Miller (1992).

The remainder of this document focuses on one transcription format only—SALT. The basic notation and transcription conventions may be found in the program's own documentation and you are encouraged to read through this before using the program. Many alternative conventions may also be developed to suit the need of the investigator, offering a great degree of analytical flexibility. Over the years, we have changed and supplemented the SALT documentation to suit our own specific needs⁴. The following transcription conventions represent that effort and were compiled in collaboration with Dr William Gavin and Michael Caygill⁵.

Transcript information lines

For consistency in identifying transcript files for either clinical purposes or research projects, it is helpful to include a set of information lines at the beginning of each transcript. Although the only *requirement* of SALT is that the first column of the first line begin with the '\$' character, followed by symbols to identify the speaker(s), additional information lines (beginning with '+') are strongly recommended so that the transcript can be easily identified in the future. We insert the following lines at the beginning of each file used in our research:

```
$ P T
+ P
+ T
+ ID
+ DOE
+ Environment
+ Context
+ Transcriber
+ Audiotape review
+ A-unit check
+ Spelling check
+ Convention check
- 00:00
```

Once these lines are inserted, relevant information may be entered. For example:

```
$ P T
+ P child
+ T mother
+ ID m08
+ DOE 11-31-97 (Format: mm-dd-yy)
+ Environment Newcastle University clinic, room 1.26
```

⁴ Transcription done in SALT format is also compatible with the *Transcript Analysis System (TAS)*, Gavin & Klee, 1998).

⁵ Thanks also to Sally Hewison and many students who over the years who have raised various points of detail and helped us to clarify things. Much obliged!

```

+ Context con
+ Transcriber tk
+ Recording review tk
+ A-unit check tk
+ Spelling check tk
+ Convention check tk
- 00:00

```

In this example, ID indicates the code given to identify the individual recorded (m08). An ID number rather than the individual's name is used to protect the identity of the individual. DOE refers to the date of the evaluation - or the date the recording was made. Environment refers to the kind of interaction that took place during the recording ('con' for conversation, 'nar' for narration, etc.). The transcriber's initials are also indicated, along with the initials of the individual carrying out various post-transcription checks. Other information lines may be added (e.g. sex, date of birth, CA) if necessary.

Speaker identification

Each speaker's utterances must be identified with a unique identifier appearing at the beginning of the line. For instance, the letters C and M might be used to indicate a child and mother engaged in a conversation. Our practice is to use the letters P and T as follows:

```

$ P T
+ P child (or patient, pupil, child's name)
+ T mother (or father, therapist, teacher, etc.)
- 0:00
T what can you see here?
P it 's a farm.

```

Although any alphabetic characters may be used to identify speakers, P and T are used for consistency across transcripts and because they conform to the notation used for the LARSP analysis.

Transcript length

1. Transcribe all utterances occurring in the language sample.
2. Place timing marks at the beginning and the end of the sample, using a stopwatch to time the sample after it has been transcribed. The timing marks for a 20-minute interaction would be indicated as follows:

```

- 00:00
  (utterances here)
- 20:00

```

Terminal punctuation

.	statements and exclamations
?	questions (wh-question, yes-no question, intonational question)
!	commands
>	abandoned utterance
^	interrupted utterance
~	intonation prompt

Transcription protocol

The following steps should be followed for transcribing a language sample. In most cases, transcription from the audiotape alone should be sufficient. In the case of speakers whose speech is partially unintelligible, it may be necessary to use videotape so that the extra-linguistic context can be used as an aid to transcription.

Recording review

When finished transcribing, review the recording in its entirety while scrolling through the transcript file. Make any necessary changes to the transcription. When this verification step has been completed, initial and date the appropriate identification line in the transcript:

+ Recording review tk, 12-01-97

Videotape verification pass

If a review of the videotape is necessary (and it will be for certain clinical purposes or research projects), scroll through the transcript file and make any necessary edits to the file. When this verification step has been completed, initial and date the appropriate identification line in the transcript:

+ Videotape review tk, 12-01-97

Convention check

After the transcription is finished, a check for transcription conventions should be carried out. When the convention check has been completed, initial and date the appropriate identification line in the transcript:

+ Convention check tk, 12-01-97

1. Begin each utterance⁶ on a new line.
2. Begin each utterance with P or T. P is used to refer to the child and may also be used to refer to the patient or any other speaker in the sample. T is used to refer to the child's mother or father and may also be used to refer to the therapist or teacher or any other speaker in the sample.
3. Some utterances may require more than a single line of text. For such utterances, continue typing on additional lines, indenting 3 spaces. A speaker identification symbol should not be put on any continuation line (The SED editor that comes with SALT does this automatically) .
4. Follow transcriptional conventions set forth in Miller and Chapman (1996) (also Leadholm & Miller 1992).
5. The documentation for the SALT computer program (Miller & Chapman, 1996) provides the basic framework for transcribing language samples. However, the following conventions have been adopted for use in our research program. All of these conventions are fully compatible with SALT.

⁶ The definition of *utterance* may vary depending on the project. An operational definition should always be used if possible. One such definition has been proposed by Fletcher and Garman (1988) and Garman (1988a,b) and is further developed below (see *Utterance segmentation* section).

Utterance segmentation

The first step in transcription is to segment a speaker's conversation into units which can then be linguistically analysed. The general approach used in determining the boundaries of these 'analysable units' (A-units) is to be found in Fletcher and Garman (1988), Garman (1989a,b) and Johnson (1986). The following additions represent segmentation decisions based on, but not specifically addressed, in these reference sources.

1. When an exact repetition of a text-unit occurs, the repetition should be treated as a separate A-unit. The following example of such an utterance,

he has a hand right here and a hand right here and a hand
right here

would be segmented in the following way:

P he has a hand right here and a hand right here.
P and a hand right here.

The first use of 'and a hand right here' is considered as an ellipsis of 'he has a hand right here' and therefore is counted as being part of the first utterance.

2. In general, when either *see*, *look* or *watch* is followed by an independent clausal text-unit (i.e. an independent clause which can stand on its own) in the absence of an intervening pause, it should be placed on a separate line of the transcription and regarded as a separate A-unit. In this way, P will not be credited with using a complex sentence, as in:

P see?
P it 's a tongue.

T look!
T there 's a baby.

P watch!
P he 'sB push/ing the pram.

However, when they are followed by a dependent clausal text-unit (i.e. a dependent clause which cannot stand on its own), the whole utterance can be regarded as a single A-unit, as in:

P see what 's in here! [elliptical for *let's see what's in here!*]
P look what 's in here!
P watch the boy!

Prosody also has to be used in deciding whether such utterances count as one or two A-units. Count as one, as in the case above, when it is spoken within a single intonation contour (as a command or statement). Count as two when it is spoken with two distinct intonation contours, when the first utterance functions as a command to draw P's attention to an object and the second functions as a question:

P look!
P what 's in here?

Similarly it is necessary to listen to the prosody to decide whether to separate these words from the end of an utterance. If the utterance is spoken within the single questioning intonation contour, it is transcribed on one line.

P what 's in here look?

Alternatively, the same words could be spoken within two intonation contours, expressing a question followed by a command.

P what 's in here?

P look!

When *see* is followed by a phrasal text-unit, it should be placed on the same line as the phrase, if it appears that the two text-units represent an integrated syntactic structure, as in:

P see the boy? [elliptical for *can you / do you see the boy?*]

3. Tag questions comprised of a clausal text-unit should normally be placed on the same line of transcription as the declarative portion of the utterance:

P you 're go/ing home now are n't you?

4. 'Complex' sentences that are produced without any overt connection should be transcribed as two utterances.

T she 'sH clos/ed2 her eye/s.

T she must be very tired.

Contractions

The two words forming a contraction should be separated by a space in the transcription.

1. Verb phrase contractions⁷

's	contracted copula <i>is</i>	he 's the one over there. he 's an astronaut.
'sB	contracted auxiliary <i>is</i>	he 'sB come/ing home. he 'sB go/ing into orbit.
'm	contracted copula <i>am</i>	I 'm over here. I 'm an elephant
'mB	contracted auxiliary <i>am</i>	I 'mB go/ing to read this. I 'mB draw/ing a picture.
're	contracted copula <i>are</i>	they 're very thirsty. you 're late.
'reB	contracted auxiliary <i>are</i>	they 'reB go/ing tomorrow.

⁷ In utterances lacking a copula or auxiliary verb when it is obligated by the context (e.g. *she here, she going home*), the omitted uncontracted verbal form should be transcribed by default (e.g. *she *is here, she *is going home*).

		you 'reB eat/ing alot.
'sD	contracted auxiliary <i>does</i>	what 'sD she want? what 'sD the boy say?
'sH	contracted auxiliary <i>has</i>	he 'sH gone home. she 'sH be/en to the moon.
've	contracted auxiliary <i>have</i>	I 've left them behind we 've be/en to the park
'dD	contracted auxiliary <i>did</i>	where 'dD he go? what 'dD she say her name was?
'dH	contracted auxiliary <i>had</i>	I 'dH gone for the day. he 'dH search/ed2 everywhere.
'dW	contracted auxiliary <i>would</i>	he 'dW go if he could. he said you 'dW look for it.
'll	contracted auxiliary <i>will</i>	she 'll be here in a minute. we 'll have to hurry.
n't	contracted negative <i>not</i>	he is n't come/ing. they were n't at home.

The contraction *n't* is usually interchangeable with the full form *not*.

is not = is n't could not = could n't were not = were n't did not = did n't

Irregular contractions that do not follow this pattern are treated as one word and the *n't* is not separated from the verb.

will not = won't do not = don't
shall not = shan't can not⁸ = can't⁹

Similarly colloquial forms such as *ain't*, *in't* are also never separated. These forms are phonologically irregular in that they are not produced by adding the contraction /nt/ or /<nt/ to the affirmative form of the verb.

2. Noun phrase contractions

'sU	contraction of <i>us</i>	let 'sU eat!
'em	contraction of <i>them</i>	watch 'em get off the bus!

⁸ The uncontracted negative form *can not* is transcribed as two words rather than one, following the general pattern of negation; *will not*, *does not*, *was not*.

⁹ In British English, we treat *can't* as a being single morpheme, since the vowels in the affirmative /kɑ n/ and negative forms /kAnt/ are phonologically distinct; hence, it might be argued that these are two distinct entries in the child's mental lexicon. However, in most dialects of American English, *can't* is treated as two morphemes (and transcribed as can n't) since /kQnt/ is formed by adding the negative morpheme /t/ to the auxiliary verb.

Inflections

Inflections should be separated from the verb stem with a slash.

1. Verb inflections

/ed	regular past tense	he wish/ed he 'dH gone. she stop/ed.
/ed2	-ed past participle	I had climb/ed2 to the top. he has love/ed2 every minute.
/en	-en past participle	I had give/en him a lift. where have they be/en?
/ing	present participle	he 'sB go/ing home where 'sB the lady go/ing?
/3s	regular third-person singular ¹⁰	she go/3s every Tuesday. it look/3s like a pram.

Care must be taken when transcribing passive constructions. In the statal passive the *ed2* or *en* form refers to a state resulting from an action and functions as an adjective. In these cases the inflection is not marked.

<i>regular passives</i>	he was splash/ed2 by the duck. the girl was kiss/ed2 by the boy. the door was slam/ed2 shut.
<i>statal passives</i>	the door is closed. he is called John. her arm was broken when I saw her.

It is necessary to take context into account when deciding whether a passive is statal or dynamic. A sentence like *the door was closed* could mean either "somebody closed the door" or "the door was in a state of closure".

When the *ed2* and *en* forms are used with *get* they are treated as adjectives and the inflections are not coded:

is teddy go/ing to get dressed?

2. Noun inflections

/s	plural noun	two doggie/s. three baby/s.
----	-------------	--------------------------------

In some words ending in /f/ and /T/ the final fricative is replaced by its voiced counterpart when the word is plural. This change may be reflected in the spelling.

¹⁰ Irregular third-person singular verb forms should retain their orthographic spelling (e.g. does, has). For transcribing spoken language, whether a form is regular or irregular depends upon its phonological, rather than orthographic, form.

knife → knives calf → calves mouth → mouths - /mAYΔz/

These plurals are transcribed without changing the spelling of the noun, as the production using a voiced fricative is effectively an alternative pronunciation of the noun which only occurs in this context. Also, if the conventional orthographic spelling is retained the word count is artificially inflated when the transcript is analysed.

A common error when learning these forms is to produce them without changing the voicing, and overregularising them. This can be transcribed by attaching a code in square brackets to the noun, noting that the speaker has made a voicing error. The alternative transcriptions are outlined below.

/nAlvz/ - knife/s /nAlfs/ - knife[EO:voice]/s
/kAvz/ - calf/s /kAfs/ - calf[EO:voice]/s

Some of these words can be produced either with or without a change in voicing, such as *hoofs* - *hooves* and *scarfs* - *scarves*. This is especially true of the words ending in /T/, such as *oaths*, *youths* and *wreaths* (although not the more common *baths* and *paths*). In these cases either production should be transcribed as the singular form of the noun plus the plural inflection.

/skAfs/ or /skAvz/ - scarf/s

Another common error of overgeneralisation is to use the addition of a plural inflection with a noun that either does not have a distinct plural form, such as *sheep*, or has a plural formed by changing a medial vowel, such as *man* and *men*. The inflection is still coded, but the error is noted as well. The following examples show how to code these productions.

P all the man[EO:men]/s are there
P here are the men/s[EO:0s]
P there are alot of sheep/s[EO:0s]

/z possessive Michael/z doggie.
the lady/z bag.

/p plural-possessive noun the two girl/p bicycle/s.
the animal/p house/s

3. Adjectival inflections

/er comparative a dog is fast/er than a pig.
there 's a big/er cup.

/est superlative this is the big/est tractor.
you 're the fast/est.

4. Indeterminable segments

In some instances, it is impossible to infer the morphemic status of what appears to be a distinct morphological segment. For example, in the utterance *whats else?* the morphological status, if any, marked by the voiceless fricative in-between *what* and *else* is indeterminable. The usual inflections (plural, possessive, third-person singular) and contractions (contracted copula and auxiliary) marked with /s/ do not apply in this grammatical context.

/ss grammatically-indeterminable what/ss else?

Derivational morphemes, where a word of a different word class has been created by the addition of a suffix to an existing word, are not usually coded. Examples include *colourful*, *quickly*, *penniless*. Similarly diminutive forms of nouns that are commonly used by children are not coded, such as *doggie* and *horsie*. However, original and productive uses of suffixes should be marked, if a speaker uses a form that is not an accepted standard form. Examples of this include:

T when are you come/ing?
P soon/ish.

P what colour is it?
T it 's kind of blue/y.

Intelligibility

Unintelligible segments of any duration should be represented in the transcription as a single x. This applies to entirely unintelligible utterances, segments occurring within utterances and single unintelligible words. An x cannot be attached to any other character as it will be counted as a word in analysis.

P that 's a x.
P x.
P they 'reB go/ing x bus.

If a segment is phonologically unintelligible but its syntactic status is unambiguous, then it can be coded xz, with syntactic information included in square brackets. Utterances containing elements like this can then be counted in analysis.

P they 'reB xz[verb]/ing the bus.
P he 'sH got a xz[noun] in there.
P xz[subj] is go/ing now.

Audibility

Inaudible segments of any duration should be represented as an upper case 'IA' enclosed in braces: {IA}. These may be due to overlapping speech, very quiet speech or other noise on the recording.

Incomplete Words

Words that are abandoned during production or which become inaudible are transcribed with an asterisk following all that can be discerned. These often occur in mazes, or in abandoned or interrupted utterances.

P (I thi*) I think that go/3s there.

P it 's a ele*>
T what is it?

Note that SALT counts incomplete words (e.g. *ele**) as separate word types if the utterance occurs as part of the analysis set. For example, if *ele** and *elephant* co-occur in a transcript, they will be counted as separate word types. If you wish to avoid this, incomplete words should be enclosed in mazes.

Omissions of Lexical Items and Inflections

Code omissions of bound morphemes, copula and auxiliary verbs, prepositions and determiners using the asterisk:

he go home - he *is go/*ing home.

put it box - put it *in *the box!

he know my name - he know/*3s my name.

If a syntactic element is omitted, but it is not obvious that a specific word should have been used, then the name of the element can be used along with the asterisk.

put/3s it in there - *subj put/3s it in there.

they the bus - they *verb the bus.

Hesitations, Repetitions, Interjections

Hesitations, repetitions and reformulations are separated from the utterance by placing them within parentheses (). These include filled pauses (such as *um*, *er*) and part words.

P (I want tha*) I want that one.

P he (um) gave me the book.

P he wants the (erm the b* b*) blue one.

P (er) what do you think?

Interjections and exclamations of surprise, sympathy, anger etc. are enclosed in braces {} as they are routinized forms not counted in MLU. These include things like *oh*, *ohoh*, *ouch*, *ow*, *oh dear*, *oops*, *gosh*, *ah*.

T {ah} look at her!

P {oh} I don't know.

P {oh}.

P {ohboy}.

Some lexical items can feasibly occur as fillers, or as interjections, or as words to be included in the main body of the transcription. One example is *well*. It can occur as a filler at the start of an utterance, as a speaker prepares to start, much in the same way as *er*.

P (er) what do you think?

P (well) what do you think?

It can also be used as an exclamation of surprise or displeasure.

P {well} would you look at that!

P {well}.

P what do you make of that?

It can also be used to mean "in that case" .

P this one 's broken.

T well we 'll have to play with this one.

In these cases and with other words such as *so* and *then* it is the function the word plays in the utterance that is significant in deciding whether to enclose it in either parentheses or braces. Anything enclosed in either of these is discounted from the sample for calculation of length of utterance. Anything enclosed within parentheses can be analyzed independently to give information about problems with utterance formulation, while the braces are used to separate minor elements with no linguistic function.

Vocatives

Vocatives are treated in the same way as minor text-units. Transcribe vocatives on the same line as the rest of the utterance when the vocative and the next text-unit occur in the same intonation contour or without an intervening pause.

P Mommy come here!
P over here Mommy.

Idiosyncratic Forms, Animal Sounds and Symbolic Noises

The % character can be used to tag idiosyncratic forms, such as a child's neologistic label that has been adopted by their family, or an unusual use of an symbolic noise as a lexical form. It serves to collect these words together at the analysis stage. As it has no other effect the use of this symbol is optional, although transcribers must remain consistent.

P it 's a %pushacar.
P there 's a %vwoosh.

Animal sounds and routinized symbolic noises are transcribed as lexical items when they are used to label an animal or when they function grammatically:

P the brumbrum is go/ing down the road.
P that is a moo.

P the cow say/3s moo.
T what does the cow say?
P moo.

Symbolic noises used as sound effects in play are not counted as lexical items and are transcribed in braces, such as the following, which was said while playing with a toy cow:

P {moo}.

Routinized language

Routinized language, such as reciting the alphabet, counting and singing, should be excluded from the calculation of utterance length (Klee et al., 1989: 233). Letters of the alphabet should be capitalised and separated by a space from one another. Cardinal numbers should be represented in the transcript as numerals and separated from one another by a space. Such sequences should be placed within braces:

P {A B C D E F G}.
P {1 2 3 4 5 6 7 8 9 10}.
P {if you 're happy and you know it clap your hand/s}.

Labels for such sequences (e.g., 'the ABCs'; 'Mary Had A Little Lamb') should be treated as labels and transcribed in the following manner:

P I know my ABC/s.
P I can sing MaryHadALittleLamb.

Some forms may appear to be routinized but in fact can be glossed with a lexical item, as in the case of *huh?*. This form appears to be synonymous with *what?* when it is used as a query and therefore should not be enclosed with braces. Similarly, in the utterance

P that would be a good thing to have for breakfast, huh?

the word *huh* should appear without braces, as it appears to function as a request for confirmation.

Melodic strings

Like routinized language, sequences of syllables which are sung or hummed can also be placed in braces, thus excluding the sequence from MLU:

P {la la la la la}.

Noun-noun sequences vs. compound nouns

'A compound is a lexical unit consisting of more than one base and functioning both grammatically and semantically as a single word' (CGEL 1567). Brown's treatment of compound words (54: rule 5), which he treated as single morphemes, is consistent with Quirk's definition.

Proper nouns also count as single morphemes (Brown rule 5) and should therefore not be separated by a space:

SantaClaus	PostmanPat
CookieMonster	FiremanSam
AuntMary	KermittheFrog

Reduplicated nouns also count as one morpheme (Brown rule 5):

bunnybunny

Adnominals, a type a noun premodifier, 'are often so closely associated with the head [noun] as to be regarded as compounded with it, as indicated by the stress on the premodifying noun instead of the head' (CGEL 1330). Examples of adnominal constructions are:

fire truck	ice cream
peanut butter	Christmas tree
bunny rabbit	trash can

The point at which these noun-noun sequences come to be regarded as compound words distinct from their original bases is debatable; how they are treated changes across time and across dialects, making it difficult to come to a definitive approach to deciding whether they should be regarded as one or two words. Comparing entries in two British and one American dictionaries suggested that while *teapot* is one word on both sides of the Atlantic, there are *coffeepots* and *fish cakes* in America and *fishcakes* and *coffee pots* in Britain, and that we can't decide whether we eat *fishfingers* or *fish fingers*.

As can be seen, there is debate as to the nature of these words in adult language. This is then further complicated when looking at child language, as we have to come to some conclusion as to whether these forms have been learned and stored as single entities rather than representing creative combinations of language. There are no hard and fast rules, but there are guidelines to assist in making these decisions.

1. the adult form is accepted as a compound, even if not universally.

fishfinger, fishcake, coffeepot, bathtub

2. the combined form is more common than either of its constituent parts, (taking into account their meaning as well as their written form).

carpark, chickencoop, jacketpotato, bakingsoda, climbingframe

3. rather than one base modifying the other, they are used together to refer to something that is semantically related to both bases.

peanutbutter, macaronicheese, (but not orange juice or raspberry jam)

These are guidelines we have used in trying to maintain consistency across a series of transcripts. The compounds we decided to retain as single words are included in the spelling conventions list. However, these decisions remain quite intuitive, and individual decisions may depend on the child's usage as much as the form in question. The transcriber can only attempt to remain consistent, and to bear in mind the question 'is this one word for this child?'

It is worth noting that our approach is different to that proposed by Crystal (WL 1979: 66) who recommends analysing adnominal constructions as two words, using the 'NN' category on LARSP:

railway station
potato chips
washing machine

We have transcribed *potatochip* and *washingmachine* as single words, but *railway station* as two.

Spelling check and the transcription dictionary

The spellings of lexical items should, as far as possible, be consistent from one transcript to the next. To insure spelling consistency, a spell-checker program will be available in the lab in the near future. Once your transcript has been spell-checked, you should initial and date the following identifier line:

+ Spelling check tk 12-15-97

A list of suggested spelling conventions may be found in Appendix 1.

References

BBC English Dictionary, HarperCollins Publishers, London, 1993

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press

- Collins Gem English Dictionary, Collins, London, 1990
- Crystal, D., Fletcher, P. & Garman, M. (1989). *The grammatical analysis of language disability*, 2nd ed. London: Whurr
- Fletcher, P. & Garman, M. (1988). LARSPing by numbers. *British J. of Disorders of Communication*, 23: 309-321
- Garman, M. (1989a). The role of linguistics in speech therapy: assessment and interpretation. In P. Grunwell and A. James (Eds.), *The functional evaluation of language disorders*. London: Croom Helm
- Garman, M. (1989b). Syntactic assessment of expressive language. In K. Grundy (Ed.), *Linguistics in clinical practice*. London: Taylor & Francis
- Gavin, W. & Klee, T. (1998). *Transcript analysis system (TAS)* [Computer software, Windows 95]. Buffalo, NY: Author
- Gavin, W. J., Klee, T., & Membrino, I. (1993). Differentiating specific language impairment from normal language development using grammatical analysis. *Clinical Linguistics & Phonetics*, 7, 191-206.
- Johnson, M. (1986). A computer-based approach to the analysis of child language data. Unpublished PhD thesis, University of Reading
- Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, 12, 28-41.
- Klee, T., Gavin, W.J. & Stokes, S.F. (2007). Utterance length and lexical diversity in American- and British-English speaking children: what is the evidence for a clinical marker of SLI? In R. Paul (ed.), *Language disorders from a developmental perspective: essays in honor of Robin S. Chapman*, pp. 103-140. Mahwah, NJ: Lawrence Erlbaum.
- Klee, T., Schaffer, M., May, S., Membrino, I. & Mougey, K. (1989). A comparison of the age-MLU relation in normal and specifically language-impaired preschool children. *J. of Speech and Hearing Disorders*, 54: 226-233
- Leadholm, B.J. & Miller, J.F. (1992). *Language sample analysis: the Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.
- MacWhinney, B. (1994). *The CHILDES project: computational tools for analyzing talk*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum (CLAN computer software, Windows 95 version 2, 1997)
- Miller, J. & Chapman, R. (1981). The relation between age and mean length of utterance in morphemes. *J. of Speech and Hearing Research*, 24: 154-161
- Miller, J. & Chapman, R. (1984-2004). *Systematic analysis of language transcripts (SALT)* [Computer software, version 8]. Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison

Miller, J. & Klee, T. (1995). Computational approaches to the analysis of language impairment. In P. Fletcher and B. MacWhinney (eds.), *The handbook of child language*. Oxford: Blackwell, pp. 545-572

Sokolov, J.L. & Snow, C.E. (eds.) (1994). *Handbook of research in language development using CHILDES*. Hillsdale, NJ: Lawrence Erlbaum

Webster Dictionary, Merriam Webster, <http://www.m-w/cgi-bin/dictionary>

APPENDIX 1 Spelling Conventions

aha	dinnertime	lookie	oreo
ain't	dishwasher	lookit	oughta
allgone	dollhouse	lullaby	ow
alright	dollshouse	lunchbox	owie
alrighty	doggie	lunchtime	
ambulanceman	DonaldDuck		pandabear
Ariel	drats	ma	partykitchen
audiotape	duckie	macaronicheese	peanutbutter
auntie	dumptruck	macaroniandche	pee
		McDonalds	peekaboo
baa	ferriswheel	McMuffin	peepee
babybath	fireengine	merrygoround	pickuptruck
babyfood	fireplace	mhm (= yes)	playdough
babysit	firetruck	mic	playsuit
bakingsoda	fisherman	MickeyMouse	pocketbook
baloney	footsy	milkshake	pointy
Bambi	fryingpan	MissPotato	polarbear
Bart		MissPotatohead	polkadot
bathtub	gee	mmm (= yes)	poey
bedroom	geez	momma	Pooh
bedtime	giddyup	mommy	poopoo
betcha	gimme	monstertruck	poopy
BigBird	gonna	moo	potatochip
blah	goodbye	moocow	Potatohead
blankie	goodnight	moomoo	psst
blastoff	gootch	moosh	
boing	gosh	morningtime	racecar
booger	gotcha	MotherGoose	restroom
brr	gotta	Mr	roastbeef
bunnyrabbit		MrPotato	RobinHood
busdriver	hafta	MrPotatohead	rollingpin
busstation	hah	Mrs	Roo
busstop	halfcircle	MrsPotato	runningback
bubba	hey	MrsPotatohead	
bubblebath	hisself	Ms	SantaClaus
byebye	hissy	MsPotato	schoolbus
	hm	MsPotatohead	schoolchildren
candycane	hon		schoolguy
carpark	hometime	nah	schoolhouse
chairback	honk	neato	schoolkid
cheeseburger	horsie	nightnight	scootch
Cheyenne	hotdog	nighttime	seatbelt
chickencoop	hottub	nighty	SesameStreet
chickenkorma	huh	nightynight	sheepdog
chickenleg	HumptyDumpty	NinjaTurtle	shh
chickennugget	huuh (= no)	nono	shutup
childproof		nope	Simpson
chomp	icecream	nowayjose	sis
choochoo	icky	n't	Sizzler
choochootrain	ittybitty		sodacracker
christmastree		oclock	spilt
climbingframe	jacketpotato	oh	splat
CocaCola	jackinthebox	ohboy	sposta
cockadoodledoo	jammys	ohdear	StackAPeg
cookiecutter		oink	starlight
CookieMonster	kaboom	ok	sunvisor
cooldude	Kanga	okaydok	sweetie
	kinda	okaydoky	sweetieshop
daddy		oo	swimmingpool
DapperDan	Laramie	oop	
daycare	lightbulb	oops	teapot
daynursery	LittleMermaid	oopsy	teaspoon
decaff	LittleTykes	oopsydaisy	teatowel
DietPepsi	livingroom	oopsys	teddybear

thingy	uhuh (= no)	washingpowder	
Thumper	unsticky	washrag	
tidyuptime	upside-down	watertrough	yeah
Tigger	upstairs	WeebleWobble	yep (= yes)
Tippy		whatcha	yoo-hoo
transcriber	valentinesday	whew	yoo-hoo-hoo
tshirt	Vedauwo	whoa	yuck
	velcro	whoopsy	yucky
ucky		whoopsy-daisy	yum
uh	Walkman	wiggly	yummy
uhuh (= yes)	wanna	WinnieThePooh	
uhoh	washingmachine	woof	4X4

APPENDIX 2

Excerpt of Orthographic Transcript

\$ P T
+ P = child; T = mother
+ Univ Child Language Lab
+ ID: A23
+ Date of recording: 7/23/91
+ Sampling context: script
+ Primary transcriber: BL 9/10/91
+ Audiotape review: BL 9/10/91
+ Videotape review:
+ Secondary transcriber: TK 4/6/92
+ A-unit check: TK 4/6/92
+ Convention check: BL 9/12/91
+ Spelling check: BL 9/13/91
- 00:00
P yeah and this go/3s on the <(s*) stove>.
T <ok you can walk around there if you want>.
P what?
T where does this one go?
T do you know what that is?
P {oh} where 'sD it go?
T ok this is where you put>
T when you 're wash/ing dish/s and you dry them you can set them in there.
T ok.
T what else do we need to put out?
P this.
P pretend milk [Adj_N].
T pretend milk yeah.
T that 's pretty good is n't it?
P but it won't come out.
T it won't come out.
T I know.
P it 's stuck.
T that 's ok.
T (that 's) it 's so you can pretend you can really drink it.
T ok <you> wanna put this someplace special?
P <tea>?
P here 's our table.
T the table.
P dinner table <for pretends>.
T <yeah>.
T you wanna put out some plate/s or cup/s?
P glass.
P glass/*s for me and you.
T ok.
P we don't need milk.
T you don't want the milk now?
P pretend!
P I 'm pretend/ing now.
P I 'm a daddy.
P and you 're a mommy.