The Pye Analysis of Language

The Pye Analysis of Language is a set of routines that provide researchers or clinicians with preliminary analyses of a language sample. They provide analyses of word frequencies, phonetic frequencies and substitutions, MLU, and a lexical concordance. The analyses may then be used as a guide in eliciting further information on those aspects of language of particular interest to the researcher or clinician. They should not be used as the sole measure of the language sample. In particular, errors in transcript entry may lead to significant misanalyses of the sample.

The routines implement the procedures in Ingram (1981) ***Procedures for phonological analysis of children's language***, and Ingram (in press) ***First language acquisition: Method, description, and explanation.*** The routines were written in Spitbol, a compiled version of the Snobol4 programming language (cf. Griswold, Poage & Polonsky (1968) ***The Snobol4 programming language***). Spitbol runs on microcomputers using the DOS operating system with at least 256 K of RAM. The disk contains two programs: ***phonix.exe*** and ***syntax.exe***. The programs do not include a word processor for entering the transcript. The user will have to supply their own. It should produce a transcript in the ASCII format (ie. one that does not contain any special printer commands). Wordstar, for example, should be used in the non-document mode. I recommend PC-WRITE, a Shareware word processor distributed by Quicksoft, 219 First N. #224, Seattle, WA 98109, (206) 282-0452.

The enclosed documents include a sample transcript entry (from Ingram, in press). The transcript must begin with a name line that is marked by the dollar sign '$' at the beginning. The child's name should appear first, followed by any other participants. Comments or contextual information can be entered in two ways, either as a separate line beginning with a plus '+', or in parentheses following any line in the transcript. The format assumes that other conversational partners will only have one line per utterance. On the other hand, the child's utterances may have any number of lines (or tiers), as long as all of the child's utterances in the transcript have the same number of tiers. The sample transcript contains three tiers per utterance. The first tier must contain the phonetic transcription (or an orthographic transcription if the transcript is not being transcribed phonetically). The second tier must contain the orthographic transcription, while the third tier may contain anything. The transcript shows syntactic information, but it could just as easily contain semantic information.

Each line should begin with the speaker's initial. This initial should correspond to the first character of the names provided in the name line. Successive tiers for the child should not begin with the child's initial. Every line in the transcript should end with a punctuation mark ('.!? or ...'). The transcript

should follow Ingram's Principle of Phonetic Correspondence in that every word on the orthographic tier should have a partner on the phonetic tier. Among the enclosed documents is a suggestion for an ASCII translation for some of the IPA symbols. The phonological analysis program assumes that the phonetic tier has been typed using this alphabet. Users can add any other characters they like, but they won't be included in the phonological analysis. A version of the program is available which uses the upper ASCII characters (>127) to define actual IPA characters. It assigns each special character an ASCII value equivalent to the values shown for the ASCII IPA + 128.

The orthographic tier follows most of the conventions suggested in Ingram. Unintelligible parts of utterances have empty angle brackets ('< >') on the orthographic tier. Inaudible or indistinguishable parts of an utterance should be marked with an 'X' on the phonetic tier. Questionable parts of utterances have the questionable part of the interpretation marked with angle brackets (e.g. <truck>) on the orthographic tier. A slash ('/') in the orthographic tier separates bound morphemes. The transcript does not distinguish between stuttered words and repeated words. Both are marked with an ampersand ('&') at the beginning of each repeated attempt. Repeated words are not included in any of the analyses. Most of the routines accept words up to 20 characters long. The phonological analysis routine will only work with words up to 10 characters long. Periods ('...') mark the end of interrupted or unfinished utterances.

The phonix program contains procedures for checking the format of the transcript, analyzing lexical productivity, and doing a phonological analysis. The checking routine checks to see if the transcript begins with a name line, has speaker initials before every line, and whether every line ends in a punctuation mark. It also produces a list of the utterances with numbers so that any errors can be corrected. The lexical productivity routine displays the frequency and utterance number(s) for every word in the child's orthographic tier. It also produces what Ingram terms a 'phonological lexicon'. This lists all phonetic tokens for each word in the orthographic tier. Both of these routines make use of the original transcript. A copy of the output of the lexical productivity routine is included in the documents and on the disk in the file **daniel.phn.**

Users of the phonological analysis routine must create an abbreviated phonological lexicon. Following the procedures outlined in Ingram (1981), they should select the most representative phonetic token for each word. This can be done by making a copy of the phonological lexicon produced by the lexical productivity routine and deleting the unwanted phonetic tokens. Users will also have to type in a phonetic rendering of the presumed target. The abbreviated phonological lexicon should

therefore have one line for each word consisting of the word in standard orthography, the phonetic target, and the child's phonetic token. A sample abbreviated phonological lexicon is included in the documents and on the disk in the file **phone.lex.**

The syntax program contains a routine that produces several summary statistics, a routine that lists the child's sentence types and a concordance based on the orthographic tier, and a routine which produces a concordance based on the symbols listed on the third tier. The statistics include MLU, a classification of the child's utterances, a classification of the adult's utterances, and the proportion of the child's imitations, total and partial intelligibility, and the proportion of the adult's questions. Imitations follow the definition provided in Ingram (in press): an utterance contained in the immediately preceding utterance of the conversational partner. In particular any morpheme indications in the child's tier (using '/') should also appear in the adult's utterance if it is to be counted as an imitation. The MLU count does not analyze repeated words in the child's utterance or such fillers as 'oh, o, mm, mhm, and huh'. If the proportion of the child's imitated utterances is greater than .20, the imitations will not be included in the MLU calculation. A sample statistical analysis is provided in the documents and on the disk file **daniel.mlu.**

The lexical productivity routine in the syntax program lists the child's sentence types and their utterance numbers. It also provides a concordance for the orthographic tier. This makes it easy to see the basic syntactic patterns in the child's sample and check to see if the words are lexically free. A sample analysis is provided in the documents and on the disk file **daniel.lex.** The final syntactic analysis routine will produce a concordance for the symbols listed on the third tier. In the sample transcript I have provided a preliminary analysis of Daniel's syntactic structure. The output consists of all of Daniel's utterances which contain each syntactic category, e.g. all utterances which contain verbs or adjectives. This routine makes it easy to check on the syntactic freedom of each word. A sample analysis is provided in the documents and in the disk file **daniel.syn.**

References

Griswold, R. E., Poage, J. F. & Polonsky, I. P. 1968. ***The Snobol4 programming language.*** Englewood Cliffs, NJ: Prentice-Hall.

Ingram, D. 1981. ***Procedures for phonological analysis of children's language.*** Baltimore, MD: University Park Press.

------. in press. ***First language acquisition: Method, description, and explanation.*** Cambridge: Cambridge University Press.

An Abbreviated Phonological Lexicon

I	ay	ow
animal	An6ml	Amow
another	an^D&	An^
back	bAk	bA
ball	bal	bOw
bigbird	bi6gb&d	bab&
book	bUk	bow
boom	bu:m	bu
burt	b&t	b&
byebye	baybay	baybay
clown	klawn	haw
color	k6l&	hAd&
down	dawn	daw
draw	dra	wa
ernie	Erni	w^ni
fell	fEl	Eu
help	hElp	hEup
here	hir	hir
hewent	hiwEnt	hEup
hithere	hayDeyr	haydeyr
kiss	kIs	tI
kitty	kIFi	tIti
momma	mama	mama
mommy	mami	mani
monster	manst&	madi
more	mor	mow
nightnight	naytnayt	nayna
oscar	ask&	ad&
pretty	prIFi	pIdi
red	rEd	Ed
shoes	Suwz	Suw
that	DAt	^nd^
toe	tow	dow
under	^nd&	ay&
up	^p	^p
wait	weyt	wi
what	w^t	w^
what's	w^ts	^nd^
whatthat	w^tDAt	wadi

****
****

LEXICAL PRODUCTIVITY

Written by

Clifton Pye
The University of Kansas

****

Daniel's utterances:

1. hI:b^bgagi.
   < >.
2. 767 kHIpH.
   a tape.
3. kHipHukHOr6.
   taperecorder.
4. kHey.
   okay.
5. nu nu.
   &no, &no, &no, no.
6. sebawu:kH.
   <book>.
7. m::fb:.
   < >.
8. 7a7kHeytH.
   okay.
9. pHiFu.
   Peter.
10. eyy^vkix::.
    < >.
11. kHuA.
    car.
12. In 6 bowtH.
    it a boat.
13. E7 si:.
    let/'s see.
14. n: nAnA.
    a banana.
15. sey nAnA?
    see banana?
16. oygowpm:.
    < >.
17. dAnsin.
    dance/ing.
18. eytHeyn.
    airplane.
19. tHE:n.
    plane.

20. hEn:.
  plane.
21. ey7tE7nts.
  airplane/s.
22. eynd 6 kHats.
  and <a> truck/s.
23. 6d 6 bowt.
  and a boat.
24. 6vmaynE7.
  < >.
25. 6s6mb^nEt.
  < >.
26. kHey.
  okay.
27. kHey.
  okay.
28. sonfanEtH.
  < >.
29. Iy6 bowt.
  little boat.
30. BeykH.
  bike.
31. kHey.
  okay.
32. n6 n6nE bEkH.
  want another book.
33. now may b6k.
  <not> my book.
34. su.
  shoe.
35. su.
  shoe.
36. gA:mow.
  < >.
37. maysEn.
  lace/s.
38. ha:yi.
  < >.
39. s6veysEn.
  < >.
40. ow bEkH.
  oh, book.
41. gowbu.
  < >.
42. peytH.
  pig.
43. peyg.
  pig.
44. kHeypHukHoyd6.
  taperecorder.

45. h676 z6 kHow.
  <what's it call>?
46. kHikHowd6.
  taperecorder.
47. kH^d6.
  recorder.
48. kown6.
  corn.
49. hayhI kHow.
  < > <corn>.
50. d6 bAsAtH.
  the basket.
* Number of Daniel's lines = 50
* Total number of lines = 106

Daniel's WORD INDEX:

WORD	FREQ	SENTENCES
a:	1	<22>,
airplane:	1	18,
airplane/s:	1	21,
and:	2	22, 23,
another:	1	32,
banana:	2	14, 15,
basket:	1	50,
bike:	1	30,
boat:	3	12, 23, 29,
book:	4	<6>, 32, 33, 40,
call:	1	<45>,
car:	1	11,
corn:	2	48, <49>,
dance/ing:	1	17,
it:	2	12, 45,
lace/s:	1	37,
let/'s:	1	13,
little:	1	29,
my:	1	33,
no:	4	5, 5, 5, 5,
not:	1	<33>,
okay:	5	4, 8, 26, 27, 31,
peter:	1	9,
pig:	2	42, 43,
plane:	2	19, 20,
recorder:	1	47,
see:	2	13, 15,
shoe:	2	34, 35,
tape:	1	2,
taperecorder:	3	3, 44, 46,
the:	1	50,

truck/s: 1 22,
want: 1 32,
what's: 1 45,

Daniel's PHONOLOGICAL LEXICON

a <6>

airplane eytHeyn

airplane/s ey7tE7nts

and eynd
6d

another n6nE

banana nAnA
nAnA

basket bAsAtH

bike BeykH

boat bowtH
bowt
bowt

book <sebawu:kH>
bEkH
b6k
bEkH

call <kHow>

car kHuA

corn kown6
<kHow>

dance/ing dAnsin

it In
z6

lace/s maysEn

let/'s E7

little Iy6

my may

no nu
nu
nu
nu

not <now>

okay kHey
7a7kHeytH
kHey
kHey
kHey

peter pHiFu

pig peytH
peyg

plane tHE:n
hEn:

recorder kH^d6

see si:
sey

shoe su
su

tape kHIpH

taperecorder kHipHukHOr6
kHeypHukHoyd6
kHikHowd6

the d6

truck/s kHats

want n6

what's h676

****
SUMMARY STATISTICS

Written by

Clifton Pye
The University of Kansas
****

* Number of Daniel's utterances = 50
* Number of Daniel's imitations = 7
* Number of Daniel's intelligible utterances = 27
* Number of Daniel's partially intelligible utterances =
* Number of Daniel's partially questionable utterances = 2
* Number of Daniel's totally questionable utterances = 3
* Number of Daniel's totally unintelligible utterances = 11
* Proportion of total intelligibility = .627
* Proportion of partial intelligibility = .674
* Proportion of imitation = .14
* Proportion of adult questions = .303

Sentence classification for Daniel

Sentence Type	Sentence Numbers
Imitation	3,9,18,31,37,43,48
Totally Intell.	2,4,5,8,11,12,13,14,15,17,19,20,21,23,26,27,29, 30,32,34,35,40,42,44,46,47,50
Partially Int.	
Partially Quest.	22,33
Questionable	6,45,49
Unintelligible	1,7,10,16,24,25,28,36,38,39,41

Adult sentence classification

Sentence Type	Sentence Numbers
Imitations	9,16,29,34,37,52
Declaratives	3,6,10,11,15,17,18,19,21,23,25,26,27,28,30,31, 33,40,45,47,49,50,51,53,55,56
Questions	1,2,4,7,13,14,20,22,32,35,38,41,42,43,44,46,54
Imperatives	5,8,12,24,36,39,48

MLU count for Daniel

Number of Morphemes	Number of Sentences	Sentence Numbers
1	22	4,5,8,11,19,20,26,27,30,34,35,40,42,44,46, 47,3,9,18,31,43,48
2	8	2,14,15,17,21,29,50,37
3	5	12,13,23,32,33
4	1	22

Total Morphemes = 57; MLU = 1.583

****

LEXICAL PRODUCTIVITY

Written by

Clifton Fye
The University of Kansas

****

Daniel's SENTENCE INDEX

Number of Daniel's syntactic types = 13

Daniel's SYNTACTIC LEXICON

basket
  the basket.

boat
  and a boat.
  it a boat.
  little boat.

book
  <not> my book.
  oh, book.
  want another book.

call
  <what's it call>?

it
  <what's it call>?
  it a boat.

let/'s
  let/'s see.

little
  little boat.

my
  <not> my book.

not
  <not> my book.

oh
  oh, book.

see
  let/'s see.
  see banana?

tape
  a tape.

the
  the basket.

truck/s
  and <a> truck/s.

want
  want another book.

what's
  <what's it call>?

****
SYNTACTIC PRODUCTIVITY

Written by

Clifton Pye
The University of Kansas
****

Daniel's SYNTACTIC CATEGORIES

ADJ
29. little boat.

CONJ
22. and <a> truck/s.
23. and a boat.

DET
2. a tape.
12. it a boat.
14. a banana.
22. and <a> truck/s.
23. and a boat.
50. the basket.

EMBEDDED CLAUSE(S)
13. let/'s see.

EXC
4. okay.
5. &no, &no, &no, no.
8. okay.
26. okay.
27. okay.
31. okay.

FIL
40. oh, book.

N
2. a tape.
3. taperecorder.
6. <book>.
9. Peter.
11. car.
12. it a boat.
14. a banana.
15. see banana?
18. airplane.
19. plane.
20. plane.
23. and a boat.
29. little boat.
30. bike.
32. want another book.
33. <not> my book.
34. shoe.
35. shoe.
40. oh, book.
42. pig.
43. pig.
44. taperecorder.
46. taperecorder.
47. recorder.
48. corn.
49. < > <corn>.
50. the basket.

N/PL
21. airplane/s.
22. and <a> truck/s.
37. lace/s.

NEG
33. <not> my book.

OBJ
15. see banana?
32. want another book.

POSS-1
33. <not> my book.

PRO-NOM
12. it a boat.
45. <what's it call>?

QUANT
32. want another book.

V
13. let/'s see.
15. see banana?
32. want another book.
45. <what's it call>?

V/PRO-ACC
13. let/'s see.

V/PROG
17. dance/ing.

WH/AUX
45. <what's it call>?

(COP)
12. it a boat.