FUNCTOR ANALYSIS: A METHOD OF QUANTIFYING FUNCTION WORDS FOR COMPARING AND CLASSIFYING LANGUAGES

R. David Paul Zorc

School of Australian Linguistics
Darwin Community College

It would seem quite reasonable that function words, or functors, should play an important role in subgrouping or classifying languages, because of:
(1) their obvious importance within any given speech variety, (2) their high text frequency, and (3) their tendency towards stability and a low rate of replacement. While functors can be shown to yield qualitative results in both synchronic and diachronic studies, it would also seem advantageous to have a quantitative method for dealing with them. I would like to expound briefly on these three reasons why functors are so important to subgrouping, and suggest a method that can deal with them quantitatively.

Some might wonder why it is necessary to belabor the obvious. All too often comparative studies have centered on phonological or lexical evidence, taking a "shortcut" when there isn't one in comparative linguistics [Teeter (1963:648)], and some have even ignored or dismissed counterevidence presented by grammatical structures.

1. IMPORTANCE OF FUNCTORS.

A language is more readily defined by its grammar than by its lexicon. Let us take the following two examples:

Naka-*pay* na 'akú sa *cashier*, *ásk*-a 'imáw.

I've already *bayád*-ed the *manòg-balígya'*, *kutána* her.

The grammar of the first is Aklanon (Bisayan, Philippine), while the lexicon is English; the grammar of the second is English, while the lexicon is Aklanon.

Even if 50% of the vocabulary of any given language were to go, that language would still be that language (how much of English is still *English*?!); if even 10% of the grammar (or functors) went, one would suspect he was dealing with a pidgin.

There has been much outrage expressed in the Philippines over a speech variety called *"Taglish"* (Tagalog with a heavy English overlay). While the grammar is still very much Tagalog, it is the lexicon that some find objectionable. With all the tinkering that goes on over national languages, it is rare if so much as one functor is involved (conjunctions excepted), while a massive part of the lexicon becomes "purer" or "more international" as the case may be. All this is so because the grammar and functors are "assumed"; the language is still the language--which is why the label *Taglish* was chosen instead of *Englog*. One never

doubts that it is still Tagalog being spoken--just the *kind* of Tagalog.

2. HIGH TEXT FREQUENCY OF FUNCTORS.

Hockett (1958:264-5) lists four types of grammatical forms that may be classed as functors: *substitutes, markers, inflectional affixes,* and *abstract governing derivational affixes.* While distinguishing between contentives and functors may be difficult from language to language, forms that have any of the above four functions or attributes may be usefully classed as functors, especially if they form a *paradigm.* The important thing about functors is that a few hundred of them form the grammatical core of a language, while thousands of contentives make up the lexicon. Perhaps an added advantage of working with functors is the limitation on choice. Selecting basic functors is far less tedious--and argumentative--than selecting basic vocabulary.

If one records any text and collates the transcribed results, chances are that even the most basic lexical items (e.g., *eat, sleep, eye, tongue, full,* etc.) may not occur more than a few times, but the text would be replete with pronouns, deictics, articles or markers, negatives, interrogatives, and the like. It is thus functors, rather than contentives, that form the core or basic vocabulary of any given language.

McFarland (1974:313-9) ranked 150 morphemes that occurred with the highest frequency out of six texts in each of twelve Bikol area dialects. Of these only twelve were strictly lexical: *say, arrive, person, tell, finish, name, happen, house, time, see, good, man/male*; the remaining 138 were functors.

3. STABILITY AND LOW PROBABILITY OF REPLACEMENT OF FUNCTORS.

McFarland has stated that the above two factors (high frequency of occurrence and syntactic importance) "would seem to predict high stability, that is, low probability of replacement, for the functors and other restricted-class morphemes." (1974: 121-2). Teeter has also stressed an important fact: "...[D]irect outside influence on grammars is literally impossible, since... each child constructs his own grammar by extrapolating from the utterances he hears. Words are learned, but grammatical rules are invented. (1963:646)

One of the best examples of the stability of functors is the case of Ilongot in the Philippines. Dyen (1965:32) classified it outside of all Philippine groups as an independent member of the Northwest Hesion. Walton (1977:18) classified it as the first to split from Northern Philippine languages; he discounted its higher percentages shared with Southern Cordilleran (Pangasinan and Karaw), with which it does belong, as attributable to mutual borrowing from Pangasinan.

The Ilongot lexicon shows much independent innovation, and hence the interference with its lexicostatistical scores. But if one looks at Ilongot functors, particularly the pronouns and deictics, the similarity to other Southern Cordilleran languages becomes apparent. [See Table 1.]

TABLE 1. ILONGOT PRONOUNS AND DEICTICS COMPARED WITH SOME OTHER MEMBERS OF SOUTHERN CORDILLERAN.

PRON	Kayapa	Inibaloi	Pangasinan	Ilongot	COG?	PROTO-SC
1	hi'gak	si'kaḳ	siák	si'ak	+	*si'ka-ak
1+2	hi'gata	si'kata	sikatá	sikisi	+	*si'ka-ta
2	hi'gam	si'kam	sikáˊ	sika	+ -	*si'ka-ka *si'ka-m(u)
3	hi'gatu	si'kato	sikatúˊ	siya	-	*si'ka-tu
1+1	hi'gami	si'kami	sikamíˊ	sikami	+	*si'ka-mi
1+2+	hi'gatayu	si'kito	sikatayúˊ	sikisi	+	*si'ka-tayu
2+2	hi'gayu	si'kayo	sikayúˊ	siki	+	*si'ka-yu
3+3	hi'gada	si'kara	sikaráˊ	siyay-də	-	*si'ka-da
DEC-Topic forms						
1	hi'aday	sajay	iyáˊ/sáyay	tu	-	*s()-yay
2	hi'atan	satan	itán/sátay	ta	+	*s()-tan
3	hi'amman	saman	imán/sámay	ma	+	*s()-man
DEC-Locative forms						
1	diyay	ciyay	diyáˊ	'itut	-	*di-yay
2	ditan	citan	ditán	'itat	+	*di-tan
3	diman	ciman	dimán	'imat	+	*di-man

TABLE 2. SOME DIFFERENCES IN FUNCTORS BETWEEN NORTHERN-SAMAR AND WARAY-WARAY(BISAYAN).

Northern-Samar	Waray-Waray	GLOSS
siyáˊ	hiyáˊ	*he/she*
siráˊ	hiráˊ	*they*
si	hi	nominative person marker
si(n)	hin	indefinite genitive marker
sa(n)	han	definite genitive marker
sa	ha	oblique marker
'a:k(u')	'a:kun	*mine*
'a:m(u')	'a:mun	*ours (exclusive)*
'a:t(u')	'a:tun	*ours (inclusive)*

This subgrouping of Ilongot with other Southern Cordilleran languages is further substantiated on the basis of exclusively shared innovations: the replacement of PPH *ña *his/her* by *tu, the *si'ka- nominative pronoun formative, the deictic *tan denoting position near addressee, the deictic *man denoting remote position; the assimilation of PPH *a in the penult to the vowel in the ultima (PPH *ta:[']uh *person* > SC, Ilongot tu'u) or to a final diphthong (PPH *ka:yuh *tree, wood* > PSC *kiyəw > Ilongot kiyu); lexical innovations including Ilongot də:gin, Inibaloi cadin, Pangasinan dálin *earth* (replacing PPH *ta:naq, *lu:paq), Ilongot tawən, Inibaloi tabən, Pangasinan táwən *sky* (replacing PPH *la:ŋit; note PPH *taqwən *year*).

While any innovation can be borrowed or can spread across language boundaries, functors tend to be less open to large scale borrowing or systematic replacement since functors consist of closed paradigms or restricted-class morphemes. Thus, a single pronoun or verb affix might be borrowed, but not an entire paradigm. The Ilongot functors that do not agree with those of the other SC languages turn out to be either retentions (e.g., siya *he/she* < PPH *siyá, or tu *this* < PPH *'i-tú), or independent innovations (e.g., Ilongot siyay-də < PSC *siyay *this* + *-da *they*, or the final -t on the locatives replacing the final consonant of the stem) [note a similar paradigmatic replacement by -y in the Pangasinan topic deictic alternates].

4. THE QUANTITATIVE USE OF FUNCTORS.

A number of scholars, past and present, have used functors qualitatively. Some admirable studies include those of Greenberg (1963) on African languages, and Schebeck (n.d.) on Yuulngu (Australian) languages. However, only two studies to my knowledge have developed a method for dealing with functors quantitatively [McFarland (1974) and Zorc (1977)]. Those interested in the independent evolution and rationale of these methods are referred to those studies.

Basically, McFarland's method, called *morphemic differentiae analysis*, systematically compares all paradigms of all functors between speech varieties. Each difference is scored negatively [see below], and the score reflects the total number of differences observed. Thus the *lower* the numerical score, the closer the genetic relationship posited. Zorc's method, originally called *functor classification*, selects one-hundred basic functors specific to a language family which are observed (or likely) to differ from one speech variety to another. Pairs are scored according to a strict principle of morphological identity (i.e., any difference not directly attributable to a sound change is scored negatively). Thus, the resultant scores reflect the total number of exact cognates, so that the *higher* the numerical score, the closer the genetic relationship posited.

Both methods agree in one principle: once counted, a difference is not counted again. For example [Table 2], the common-noun case-marking particles and the third person pronouns of Northern Samar and Waray-Waray (Bisayan) differ in the replace-

ment of s- in the former by h- in the latter. Further, the first person possessive pronouns in the former end in -' (glottal stop), while those of the latter end in -n. None of these are regular sound changes or correspondences; they don't occur elsewhere in the lexicon or grammar. Rather than subtract a point for *each pair* with the discrepancy, only one point is subtracted for *each discrepancy*. Thus the total negative score by both Zorc's and McFarland's methods would be -1 for the s-:h- difference, -1 for the -':-n difference, -1 for the alternate (short) forms in the N-S case markers, and -1 for the alternate forms of N-S genitive pronouns. A stricter (and less defensible) system of scoring would yield up to -9, instead of -4 for the paradigmatic differences.

The method advocated here may be devised and applied in either of two ways, which may be termed *fine tuning* and *broad-band tuning*. If one is working with closely related speech varieties, one may wish the overall scores among dialect pairs to reflect the greatest amount of difference. Hence, Zorc (1977: 186-91) selected 50 out of the 100 functors which were found to differ. Forms observed to be the same (cognate in every regard) were excluded, e.g., Pan-Bisayan [limá] *five* and [pitú] *seven*; while *one* (reflecting *'əsá, *'isá, *'isa-rá, *sayú', *'isád, *'usád), *two* (reflecting *duhá, *duwá [with unexplained loss of *-h-], or *da-rwá), *three* (reflecting *təlú or *ta-tlú) were included. This, particularly if taken with lexicostatistical comparison [see 5.2 below], resulted in a fine tuning effect.

With widely divergent languages one might like broad-band tuning, i.e., selecting functors that are basically cognate, possibly relaxing the requirement of strict morphological identity. For example, in doing fine tuning, the comparison of Aklanon sánda : Masbate sinda *they* would be scored negatively; in broad-band tuning they could be scored positively [the differences are not the product of regular sound change, but are based on an analogy: si- (singular name marker) : sa- (plural name marker) + -n- ligature + -da *they* (enclitic); they are cognate in part]. However, the principle regarding paradigmatic differences never being counted more than once is applied in both comparisons.

Table 3 is a list of 100 functors devised on the fine tuning model for the Bisayan subgroup of closely related dialects; Table 4 is devised on the broad-band tuning model for the Yuulngu group of distantly related Australian languages. The following classes are useful in drawing up similar lists:

SUBSTITUTES

pronouns (including various case forms, enclitics, alternates),
deictics (including various case forms and verbal derivatives, e.g., *go there, come here*),
locationals (*right, left, this side, otherside, above, below, downriver, upriver, downhill, uphill*, etc.),
temporals (*today/now, yesterday, tomorrow, temporarily, later on, earlier, late, early, last (night)*, etc.),
low numbers (including indicators of number, e.g., dual or plural affixes) and quantifiers (*all, many, some, few*, etc.),

TABLE 3. LIST OF 100 FUNCTORS CHOSEN FOR BISAYAN (PHILIPPINE) CLASSIFICATION.

1. Top/pro-1	26. Verb/dec-1	51. because	76. day(time)
2. 2	27. Verb/dec-3	52. what?	77. year
3. 3	28. Neg-nominal	53. who?	78. today/now
4. 1+1	29. Neg-existential	54. whose?	79. tomorrow
5. 1+2	30. Neg-past	55. when (fut)?	80. yesterday
6. 1+2+	31. Neg-future	56. when (past)?	81. later on
7. 2+2	32. Neg-prohibitive	57. where (past)?	82. earlier
8. 3+3	33. CN/topic	58. where (fut)?	83. morning
9. Gen/pro-1	34. CN/indef.gen.	59. why?	84. afternoon
10. 2	35. CN/defin.gen.	60. how many?	85. act.intr.prog.
11. 3	36. CN/locative	61. how much?	86. act.intr.fut.
12. 1+1	37. Existential	62. how (degree)?	87. act.trans.prog.
13. 1+2	38. Name/topic.sg.	63. one	88. act.trans.past
14. 1+2+	39. Name/gen.sg.	64. two	89. act.trans.fut.
15. 2+2	40. Name/obl.sg.	65. three	90. act.trans.perf.
16. 3+3	41. Name/topic.pl.	66. four	91. passive progressive
17. Obl/formative	42. Name/gen.pl.	67. six	92. passive past
18. Top/dec-1	43. Name/obl.pl.	68. ten	93. passive imperative
19. 1+2	44. now, already	69. on top of	94. passive neg. imper.
20. 2	45. still, yet	70. under	95. instrumental future
21. 3	46. first, please	71. across	96. instrumental imperative
22. Loc/dec-1	47. because (excuse)	72. left	97. instrumental potential
23. 1+2	48. don't know	73. right	98. instrumental perfect
24. 2	49. and	74. within	99. local imperative
25. 3	50. if, when(ever)	75. night	100. local neg. imperative

Composition: pronouns (1-17), deictics (18-27), negatives (28-32), common-noun case-marking particles (33-37), personal-noun case-marking particles (38-43), discourse particles (44-48), conjunctions (49-51), interrogatives (52-62), numerals (63-68), locationals (69-74), temporals (75-84), verb suffixes (85-100). Note: a number of forms that are cognate in every regard have been omitted from this list, e.g., oblique pronouns (cognate with genitive), numbers *limá five, *pitú seven, *walú eight, etc.

TABLE 4. LIST OF 100 FUNCTORS CHOSEN FOR YUULNGU (AUSTRALIAN) CLASSIFICATION.

1. Nom/pro-1
2. 2
3. 3
4. 1+1
5. 1+2
6. 2+2
7. 3+3
8. 1+1+
9. 1+2+
10. 2+2+
11. 3+3+
12. Acc/pro-2
13. 3
14. 1+2
15. Gen/pro-1
16. 2
17. 3
18. 1+1
19. 1+2
20. 2+2
21. 3+3
22. 1+1+
23. 1+2+
24. Obl/pro-2
25. -3
26. Nom/dec-1
27. 1+2
28. 2
29. 3
30. Loc/dec-1
31. 1+2
32. 2
33. 3
34. Topic suffix
35. Nominative
36. Ergative/Instr.
37. Accusative
38. Genitive/Dative
39. Originative (done by)
40. Locative (in/at)
41. Ablative (from)
42. Allative (to/towards)
43. Pergressive (through)
44. Associative (with/by)
45. Locative Increment
46. all
47. one
48. two
49. three
50. Dual suffix
51. Plural
52. that way
53. this way
54. on this side
55. up/above
56. other side
57. downward/bottom
58. down-hill/river
59. up-hill/river
60. Question particle
61. what?
62. why? for what?
63. when?
64. how? by what?
65. who?
66. who? [ergative]
67. what-you-call-it
68. which?
69. from where? [assoc]
70. which way?
71. do what? [verb]
72. none [existential]
73. not [preverbal]
74. not having [suffix]
75. having [suffix]
76. now
77. by the way
78. temporarily
79. only, merely
80. vainly
81. always
82. other [same kind]
83. other [diff.kind]
84. probably, indeed
85. there! look!
86. then, after(wards)
87. and
88. indeed, to be sure
89. because
90. later on [same day]
91. tomorrow
92. today
93. earlier [same day]
94. reflexive/reciprocal
95. causative
96. nominalisation [verb]
97. greedy-for
98. comitative [prefix]
99. past potential
100. having-many [suffix]

Composition: pronouns (1-25), deictics (26-33), case-marking suffixes (34-45), numerals and quantifiers (46-51), locationals (52-59), interrogatives (60-71), negatives (72-74), existentials (74-75), discourse particles (76-89), temporals (90-93), verb suffixes (94-100).