NATIVE TEXTS AND FREQUENCY COUNTS AS AIDS TO THE TRANSLATOR

DOW F. ROBINSON

In this article the procedures outlined in the foregoing article are applied to a particular draft version. It will be noted that here a sample size of 1000 instead of 100 is taken, so that the confidence intervals given here are more precise than those in Table 2 on page 52. Ed.

Introduction

Rigid literalism is to be studiously avoided by the translator because of its undesirable results in that a word-for-word translation not merely fails to convey the intended message but actually introduces false information. Such literalism may be obviated by several methods suggested by Dr. E. A. Nida of the ABS: a thorough speaking knowledge of the receptor language, extensive acquaintance with native folklore and legends, the aim of translating into the "nearest cultural equivalent", discussion of all problems with the informant in *his* language, proper use of commentaries, etc. Especially in the analysis of native modes of thought and expression as revealed in folklore will the translator find the idioms and comfortable ways of vernacular speech. The purpose of this article is to demonstrate a procedure whereby one can analyze salient features of native folklore in terms of the frequency of occurrence of certain word classes.[1] The results can be compared with the frequency of occurrence of the same word classes for a document translated into the same language.[2] Granted the same general type of content, one assumes that a document translated into the receptor language and a document originating in the receptor language should compare favorably as to structure.

1. Procedure

Reasoning from the fact that a translation of Mark into Aztec should sound something like narrative materials of Aztec folklore, a

[1] This article is the culmination of observations and analyses made intermittently during the author's linguistic and translation work among the Aztecs from 1958 to 1962. I am indebted to Dr. William Wonderly of the American Bible Society and Dr. Joseph Grimes of Wycliffe Bible Translators for their suggestions in this and earlier drafts of the paper.

[2] For other linguistic studies of frequency counts, cf.:

Voegelin, C. F., Ramanujan, A. K., and Voegelin, F. M. "Typology of Density Ranges, I: Introduction," *International Journal of American Linguistics,* XXVI (July, 1960) 198-205.

—. "Typology of Density Ranges, II: Contrastive and Non-Contrastive Syntax," *International Journal of American Linguistics.* XXVII (October, 1961) 287-297.

Moore, Bruce R. "A Statistical Morpho-Syntactic Typology Study," *International Journal of American Linguistics,* XXVII (October, 1961) 298-307.

Herdan, Gustav. "Vocabulary Statistics and Phonology: A Parallel," *Language,* XLVII (April, 1961) 247-255.

detailed study of native texts was undertaken to ascertain at what points the translated Mark may have failed to conform to the Aztec mode of expression. Certain trial portions of the N.T. translated into Aztec had been poorly received by the various congregations. A common complaint was that its structure was too much like that of Spanish and, therefore, too difficult to understand.

For this study, 1,000 clauses were taken from Aztec folklore (hereafter referred to as Folklore) and compared with 1,000 clauses from the translated Mark (hereafter referred to as Mark). The clause here is considered in its traditional sense as that part of a sentence which contains a complete thought in itself. [3] More precisely, the subject of the study was the components of the clause. The methodology employed was to enquire:

1. What different word classes occurred within the clause, e.g., subject, object, verb as predicate, etc.
2. How many times, by actual count, did these different word classes occur within 1,000 clauses in Folklore and how many times within 1,000 clauses in Mark.
3. In what order did the various word classes occur.

With this methodology it proved possible to demonstrate the clause structure of Folklore and show how Mark compared with this. The following items are dealt with: the verb, introductory particles, nominalized forms, compounded verbs, the Spanish loan word *de*, and word order within the clause. Both procedure and results of the study are set forth in each section. Based on the results of this study, positive guidelines have now been set up which will serve for the translation of at least the narrative portions of the N.T.

2. The verb in relation to the clause

The verb is the keystone for the structuring of the clause in Aztec. A clause cannot occur unless there be at least one verb. Clauses often have two or three verbs. It is not uncommon to hear in daily speech four verbs in succession comprising but one clause. Thus, the Aztec language tends to express itself more in terms of action words than substantives which are so common to English or Spanish. The verbal components of the clause are set forth in the following chart (all figures in terms of 1,000 clauses).

[3] A more technical and ultimately more fruitful study of the clause can be made by using the tagmemic approach developed by Dr. Kenneth L. Pike and delineated by his associates of the Summer Institute of Linguistics.

Pike, K. L. *Language in Relation to a Unified Theory of the Structure of Human Behavior.* Glendale, Calif. Summer Institute of Linguistics, 1954. (cf. especially Chapter VII).

Elson, Benjamin and Pickett, Velma. *Beginning Morphology-Syntax,* Santa Ana, California. Summer Institute of Linguistics, 1960.

Longacre, R. E. *Syntax Procedures* (mimeographed). University of Oklahoma, Summer Institute of Linguistics, 1958.

CHART A

	1-verb clause	2-verb clause	3-verb clause	total no. of verbs	total no. of words
Folklore	722	241	37	1315	3926
Mark	886	103	11	1125	4025

The chart shows that Folklore has 16.5 % less 1-verb clauses than Mark. Folklore has more than twice the number of 2-verb clauses and three times the number of 3-verb clauses. One of every four clauses in Folklore has more than one verb but only one of every nine clauses in Mark. In folklore the ratio of verb to number of words in the clause is 1 to 3 while in Mark it is almost 1 to 4.

Another very fruitful means of evaluating these comparative figures is in terms of tables of confidence intervals. [4] Referring to the table of 95 % confidence intervals for sample size of 1,000, the occurrences for 1-, 2-, and 3-verb clauses (chart A) should fall within the following limits: 1-verb, 722: limits 690-750; 2-verb, 241: limits 210-270; 3-verb, 37: limits 34-40. In no case does the verbal structuring of the clauses in Mark fall within the confidence interval limits. Thus, the propensity for the multiple use of verbs in Folklore is not adequately matched by the verbal structuring in Mark.

3. Introductory particles

3.1 The particle in relation to the clause

Another integral component of the clause in Aztec is the occurrence of the introductory particles. This type of particle functions in such a way that it orients the clause as a whole to the sentence or paragraph. Chart B shows the relative occurrence of particles to introduce clauses in Folklore and Mark.

CHART B

	particle	no particle
Folklore	784	216
Mark	907	93

The confidence intervals for 784 range from 750-810 and for 216: 190-250. Here Mark has an overabundance of introductory particles. In Folklore, one of every four clauses has no introductory particles. In Mark, the ratio is one to ten.

[4] Tables of confidence intervals are based on the probability that random samplings from a group of persons or things reflect with 95 % accuracy the entire group. A count of persons or things is applied to the tables in order to determine its reliability. If the count falls within the respective limits provided by the tables, one may have "confidence" that the count represents with at least 95 % accuracy the entire group of things counted. Thus, by counting certain words and word classes in 1000 clauses of Folklore, we can calculate with 95 % accuracy how many times the same items should occur in 1000 clauses in Mark. For further reference, see e.g. *An Introduction to Scientific Research,* by E. Bright Wilson, New York, McGraw Hill, 1952.

3.2 The variety of particles

The Aztec language has a wide range of usage for a great variety of particles. At least five of these have a different form but with the same meaning. Their respective usage depends upon the amount of doubt which the speaker feels concerning the validity of what he is to say. Many particles occur in combination with each other, a type of phrase made up of particles. In Folklore 56 different particles occur; in Mark only 36 of these occur. Hence, Mark uses the introductory particle too many times (cf. Chart B) but with not enough variety.

3.3 Independent and dependent clauses

Here it will be of great help in understanding the Aztec mode of expression to divide the introductory particles into two functional groups: those which introduce independent clauses and those which introduce dependent clauses.

3.31 Particles which introduce independent clauses

Folklore has 22 different particles which introduce independent clauses whereas Mark uses only 10 of these. Chart C shows how many times each of the four more commonly used particles occur.

CHART C

	'and'	'and then'	'later'	'or'
Folklore	181	35	19	28
Mark	445	5	2	4

Confidence intervals are as follows: 'and', 181: from 160-210; 'and then', 35: from 32-38; 'later', 19: from 17-22; 'or', 28: from 25-31. Mark has an exaggerated use of 'and' coupled with a paucity of occurrence of these other particles.

3.32 Particles which introduce dependent clauses

Folklore has 29 particles which introduce dependent clauses and Mark uses 26 of these. Chart D shows how many times each of the eight more commonly used particles occurs.

CHART D

	'perhaps'	'until'	'if'	'when'	'when'	'where'	'who'	'which'
Folklore	14	15	45	61	0	16	8	31
Mark	2	3	0	2	41	36	54	74

First, notice the occurrence of the words meaning 'when'. These are two different forms of the word meaning 'when'. However, their usage depends upon the amount of doubt in the speaker's mind concerning when the following action will occur. Mark has consistently used the word 'when' which implies complete doubt that the action will ever occur. Here is a fine shade of meaning ("fine", that is, to the *foreigner*) which the Aztec reacts to as being vital.

Secondly, Mark employs the last four particles excessively and the first four particles too rarely according to normal Aztec usage. The

Aztec language has a wide variety of particles and most of them have a corresponding translation in Spanish. However, the manner in which these particles function in Aztec does not correspond in a one-to-one fashion with their function in Spanish.

It is our conclusion that the translation of Mark suffers syntactically from being forced into the Spanish practice of using particles. A sensitivity to the function of these particles in Aztec—particles which introduce both dependent and independent clauses—is of prime importance in order to have the translation sound like Aztec.

3.4 The particle huan, *'and'*

That the particle *huan* has a different function in Aztec from in Spanish is amply demonstrated by its range of occurrence in Folklore. Chart E shows the use of *huan* alone to introduce clauses and *huan* in combination with other particles to introduce clauses.

CHART E

	huan alone	*huan* in combination
Folklore	181	68
Mark	445	27

The limits of the confidence intervals are as follows: for 181: limits 160-210; for 68: limits 65-71. Mark uses *huan* to introduce clauses two and one half times more than Folklore. Whereas, Mark uses *huan* in combination with other particles (e.g., 'and then', 'and now', 'and also') less than half as many times as Folklore. This over-use of *huan* in Mark serves to impede the flow of verbal thought.

Ostensibly the use of *y*, 'and', in Spanish serves to translate the *kai* of Koine Greek. Yet the Aztec language needs no such particle to insure the proper flow of thought. In effect, this excessive use of *huan* as clause introducer chops up the flow of Aztec thought into small distinct units, the result being that it sounds like "baby talk". On the other hand, the use of *huan* in combination with other particles gives an Aztec flavor of speaking which Mark does not have. In Folklore one of every four occurrences of *huan* is in combination with other particles whereas in Mark it is one of every 17 occurrences.

4. Nominalized forms

4.1 Meaning of nominalized forms

In order to translate the many substantives of Spanish, e.g., love, joy, belief, faith, etc., the verb form in Aztec was used along with a suffix which causes the verb to function like a noun. However, it was discovered through the study of Folklore that a verb in its nominalized form has the meaning of action-in-process, much like that of a participle or gerund. Thus, in Mark 1:4 the nominalized forms are used of baptize, repent, forgive, and sin. Perhaps by using the participle form of these verbs in Spanish or English it would be understandable. But in Aztec

the nominalized verb forms, four in a row, yield simple confusion. Retranslated back into English the meaning is something like this: "John was preaching the action of baptizing which belongs to the action of repenting which results in the action of forgiving which belongs to the action of sinning." To be understood in Aztec, the active verb forms should be used with the implied subject added: "John was preaching so that one must repent and be baptized so that God will forgive the sinful things that he has done."

4.2 Frequency of occurrence of nominalized forms

In Folklore the nominalized forms occur 20 times within the 1,000 clauses. In Mark they occur 54 times. Confidence intervals give the limits of 20 as 18-23. Nominalized forms in Folklore are a scarce item. However, the main point of disjunction between Folklore and Mark is the type of verb put into this nominalized form. Folklore uses verbs which are related to the common things of life: farming and work. Mark uses the nominalized forms of such ideas, abstract concepts as "thinking", "forgiving", "baptism", "repentance", "reign", "life", and many others. True, these concepts are not the most common topic of conversation between two Aztecs. But if they were to be discussed they would most certainly appear in verb form in order that they be readily understood. This over-use of abstract ideas in nominalized form contributes to the so-called Spanish flavor of Mark.

5. Compounded verbs

Another feature of the Aztec language is the occurrence of root-compounding, especially in the verbs. Of the 1315 verbs occurring in the 1,000 clauses of Folklore, 106 of these occur as compounded verbs. This is a ratio of 1 to 11.6. Of the 1125 verbs occuring in the 1000 clauses of Mark, 92 of these are compounded. This is a ratio of 1 to 12.2. In this aspect, Mark squares well with Folklore. To go a bit further, compounded verbs may be broken into three types: verb-verb, noun-verb, and adjective-verb. Chart F shows the frequency of occurrence of these three types.

CHART F

	verb-verb	noun-verb	adjective-verb
Folklore	44	39	23
Mark	39	26	27

Although they appear to correlate closely numerically, none falls within the limits of confidence intervals. Nevertheless, in this aspect, Mark conforms quite closely to the mode of expression of Folklore.

6. Use of Spanish d e, 'of'

Another contributing factor to the Spanish flavor of Mark is the uncritical use of the Spanish *de*, 'of'. It occurs only 11 times in the 1000

clauses of Folklore and 66 times in Mark. However, in Folklore three of the occurrences are in phrases borrowed as a whole from Spanish. No such borrowed phrases occur in Mark. Four more occurrences of *de* in Folklore are within a phrase of introductory particles, *huan de ompa,* "and from there". This does not occur in Mark. Hence, it is more accurate to say that *de* occurs four times in Folklore and 66 times in Mark.

At least three syntactical relationships are expressed by the use of *de* in Mark: kinship, location, and ownership. Now in these three areas the *de* is simply not needed in Aztec; it has its own syntax for such items. The use of *de* in Mark implies a vague type of relationship, not always made clear by the context. Because the N.T. introduces so many new concepts into a language, to use *de* as a loan word serves to spread confusion. A very concrete example of this confusion occurred in the phrase in Aztec meaning "reign of God". In Mark its usage with *de* is such that it is not clear if the reign belongs to God, emanates from God, or if God is subject to the reign. Another area of confusion occurs when *de* is used to mean "characteristic of", as in "Father of glory". However, one of the legitimate uses of *de* as a loan word in Aztec is to indicate the direct object.

7. Ordering of words in the clause

This study of word order within the clause is limited to three main items: the verb, the subject, and the object. The words which express time, location, manner, etc., are not in focus. The structure of the Aztec language is such that whenever a verb occurs the subject is included within the verb. With transitive verbs, both the subject and object are included within the verb. On other occasions when a verb occurs, the subject occurs as a free word, i.e., the free subject specifies the included subject. Also, the free object may occur to make more specific the included object. Chart G shows the frequency of occurrence of the verb with free subject and free object.

CHART G

	verb alone	verb with free subj.	verb with free obj.	verb with both
Folklore	524	354	220	51
Mark	477	333	229	48

The figures for all columns fall within the confidence intervals except that of the verb-alone. Mark has too few clauses in which the verb-alone occurs. Although the frequency of occurrence of these items parallels each other, it is in the ordering of these items within the clause where the disjunction between Folklore and Mark appears.

7.1 Verb with free subject

Chart H shows the frequency of occurrence of the free subject preceding and following the verb.

CHART H

	subject precedes	subject follows
Folklore	204	143
Mark	246	87

In Folklore, 59 % of free subjects precede the verb; 41 % follow. In Mark, 74 % of free subjects precede the verb and 26 % follow. That is to say, in Folklore 2 out of 3 occurrences, the free subject precedes the verb. In Mark, in 3 out of 4 occurrences, the free subject precedes the verb.

7.2 Verb with free object

Chart I shows the frequency of occurrence of the free object preceding and following the verb.

CHART I

	object precedes	object follows
Folklore	60	160
Mark	24	205

In Folklore, 27 % of the free objects precede the verb; 73 % follow. In Mark, 10 % of the free objects precede the verb and 90 % follow. Thus, in Folklore, in 1 of 3 occurrences, the free object preceded the verb. And in Mark, in 1 of 10 occurrences, the free object preceded the verb.

7.3 Interpretation of 7.1 and 7.2

In general, as to the number of free subjects and free objects occurring in this study, both Folklore and Mark are in close agreement. Also, in both Folklore and Mark the preference is that the free subject precede the verb and the free object follow the verb. The area of non-agreement between Folklore and Mark is in the amount of freedom shown by the usage in Folklore to move the free subject and object so that they may occur before and after the verb. In Folklore approximately twice as many free subjects follow the verb as do in Mark. And three times as many free objects precede the verb as do in Mark. This freedom to move the free subject and object around is more than a matter of emphasis; it is one of the distinctive features of the Aztec language.

8. Guidelines for future translations

The following list contains recommendations, as it were, a check list to guide the translators as they work on the N.T. in Aztec.

1. One of every four clauses should have two verbs.
2. One of every 25 clauses should have three verbs.
3. The ratio of verb to number of words per clause should be one to three.

4. One of every five clauses should have no introductory particle.
5. Use more the various 56 introductory particles.
6. Use *huan* sparingly to introduce slauses.
7. Use *huan* more in combination with other particles.
8. Use active verb forms for abstract concepts like love, hope, faith, etc., not nominalized verb forms.
9. Make wider use of compounded verbs.
10. The use of the Spanish *de* as a loan word is equivocal at almost any point. Better to avoid it.
11. The freedom of occurrence of the free subject and free object should be maintained.

With these as guidelines, the future rendering of the N.T. into Aztec holds the promise that its mode of expression will convince the hearers that Paul and the other authors of the N.T. were Aztec speakers.

LIST OF EXCUSES

We are indebted to the Rev. John Beekman of the Wycliffe Bible Translators for the following list of "excuses" which a translator might be tempted to make when his draft is being checked by a translations consultant! Ed.

1. But that's what the KJV says.
2. But they will learn what it means.
3. But Joe told me to do it this way.
4. I know I left it out, but I reached my quota of 10 word sentences.
5. The Amplified N.T. does it this way.
6. Someone else had that commentary out of the library.
7. I left my manuscript in the village.
8. The kids tore up that page.
9. I can't translate that until I've translated Leviticus and Deuteronomy.
10. But I don't know any Greek or Hebrew.
11. But "dunk" is our word for baptism.
12. My informant says that it should be "John the Methodist".
13. It says "and" in the Greek.
14. Oh yeah?
15. I'm *trying* to get my informant to say that.