A Phono-lexicostatistical Analysis of Bikol-Sorsogon Varieties

Ryan G. de la Torre
rgdelatorre@bicol-u.edu.ph
Bicol University College of Education
Philippines

Gina O. Gonong
gonong.go@pnu.edu.ph
Philippine Normal University,
Manila**, Philippines**

Abstract This descriptive-comparative research analyzed 11 Bikol-Sorsogon varieties. A word list of 200 basic vocabulary words served as instrument in gathering the data through semi-structured interviews. This study used purposive sampling in selecting 22 participants based on the criteria that screened for Bikol-Sorsogon native speakers. Phono-lexicostatistics finalized the data analysis which comprised of phonetic and lexical comparison, phonetic and lexical similarity calculation, and data interpretation with regard to the relationship between the speech varieties. The percentages of apparent cognates revealed four main groups among the 11 Bikol-Sorsogon varieties. This relationship also validated the findings of preceding researches which classified Bikol-Sorsogon varieties into four subgroups. Thus, phono-lexicostatistics proved to be a systematic and effective method in analyzing several language varieties with only limited data available. With the dearth of literature and studies in Bikol-Sorsogon varieties, this research should serve as starting point for a more exhaustive linguistic investigation.

Keywords: Bikol-Sorsogon varieties, inherent intelligibility, language variation, linguistic similarity, MTB-MLE, phono-lexicostatistics

Introduction

Linguistic diversity is an essential aspect of a culturally diverse world. Languages shape the cultural identity of people and allow them to construct, understand and express

their thoughts, emotions, and perceptions of the world (Stark, 2009). In a contemporary global system, however, homogenization diminishes cultural and linguistic diversity which is essential for sustainable development. According to Ethnologue (Lewis, Simons & Fennig, 2018), there are over 7,106 known living languages in the world and every 14 days a language dies out, and so do the speakers' identity, culture, history, and knowledge of the world.

The Philippines has 187 listed individual languages based on Ethnologue (Lewis et al., 2018). Of these languages, 41 are institutional, 72 are developing, 45 are vigorous, 14 are in trouble, 11 are dying, and 4 are extinct. The Philippines also has more indigenous languages—175 are listed—than non-indigenous ones. There are still many linguistic areas in the country that remain unexplored or inadequately investigated.

This study examined the variations of the Bikol language spoken in Sorsogon Province which is located at the southernmost tip of the Bicol Region—one of Central Philippines' most dialectally diverse areas (Lobel & Tria, 2000). Bikol is an Austronesian language and a coordinate with the Tagalog and Bisayan branches of Central Philippine languages. According to Lobel and Tria, in the Central Philippine language family, Sorsogon varieties are generally a transition between the Bisayan and Bikol languages, hence the alternate name Bisakol.

Language variation is the study of linguistic features that differ systematically among different groups of speakers or the same speaker in different contexts. According to Holmes (2013), speakers may vary pronunciation, vocabulary or word choice, word-structure or morphology, and grammar or syntax. Moreover, language variation studies the regional varieties of the same language as well as social, ethnic, gender-related and stylistic varieties.

Mesthrie (2009) postulated that there is no objective and scientific way of determining when to use the terms 'language' and 'dialect.' On the other hand, Vajda (2013) argued that distinguishing languages from dialects depends on at least three factors, namely, mutual intelligibility, culture or opinion of the speakers, and political status. Culturally and politically, the varieties spoken in Sorsogon are considered dialects of Bikol. Based on a purely linguistic criterion, however, there is still no sufficient empirical evidence to prove that all Sorsogon varieties are mutually intelligible to the Standard Bikol.

According to Lobel and Tria (2000), one important factor for the existence of several speech varieties in Bicol is the influence of geographical barriers. Sorsogon Province has mountains that sprawl over the northeast, southwest and west portions which separate the towns keeping them isolated over a long period of time. Sorsogon is also surrounded by water and has ferry terminals that facilitate the migration of locals from Visayas to Bicol, hence the influence of Bisayan languages.

There are few extensive studies that aimed to classify Bikol language varieties into subgroups as shown by Cunanan (2015) in her synthesis of reviewed literature. Only the studies of Mintz (1973), McFarland (1974, 1983), and Lobel and Tria (2000) included Sorsogon varieties in their subgrouping of Bikol varieties. Cunanan examined the internal relationship of Bikol-Sorsogon varieties which generally corresponds to the subgroupings of Mintz and Lobel and Tria. Cunanan's study provides useful information on the present status of Bikol-Sorsogon varieties since developments in transportation and communication may affect the speech of native speakers.

This study provides a present-day picture of the language situation in Sorsogon Province through the analysis of its varieties using phono-lexicostatistics method. Phono-lexicostatistics is a term used by Mann (2005) to describe

Blair's (1990) method. Phono-lexicostatistics is technically a phone-based lexicostatistical method which is used to measure the relative degree of similarity between two or more languages through comparison of their common vocabularies. Phono-lexicostatistics is the result of years of refinements applied to the traditional lexicostatistics method developed by Morris Swadesh in the 1950s. To avoid making cognancy decision mainly by simple inspection, phono-lexicostatistics includes explicitly defined criteria to follow consistently in judging lexical pairs.

Before the application of phono-lexicostatistics, phonetically similar segments of each speech variety are individually identified, examined and chosen to be compared. Burquest (2006) described the process of determining phonetically similar segments based on the principles of phonological analysis. He claimed that normally phones differing by one feature are considered phonetically similar. He also stated that features do not all have the same weight. Therefore, it will not suffice to count the common features between segments to be considered similar.

The theoretical basis for the comparison of Bikol-Sorsogon varieties was grounded on the basic assumption that genetically related languages exhibit similarities in sound, form and meaning . The similarities occurred due to descent from a common ancestor or proto-language (Crowley, 1998). In this study, the cognates were determined based on the form and meaning of the word pairs rather than on the historical development of the speech varieties. Thus, the term 'apparent cognates' distinguishes it from 'true cognates' which are established by comparative method.

The present study, while in no way extensive, examined the present status of Bikol-Sorsogon varieties and attempted to elucidate the occurrence of discrepancies in the classification of certain varieties. Further, this study demonstrated the application of phono-lexicostatistics

method which has not been applied yet in the analysis of Bikol language varieties.

The general aim of this study was to analyze the varieties of Bikol language spoken in Sorsogon Province using phono-lexicostatistics. Specifically, the study aimed to attain the following objectives: 1) Identify the similarities and differences between the phonetic features of Bikol-Sorsogon varieties; 2) Determine the degree of phonetic and lexical similarity, regarded as apparent cognates, between Bikol-Sorsogon varieties; and 3) Explain how the phono-lexicostatistical measures may indicate the relationship between Bikol-Sorsogon varieties, the potential level of inherent intelligibility and the possible need for separate language programs.

Methodology

Research Design

This study adopted a descriptive-comparative research design. The descriptive aspect required the illustration and classification of the phonetic features and lexical forms of Bikol-Sorsogon varieties. Meanwhile, the comparative aspect had the purpose of determining the similarities and differences between the phonetic features and lexical forms. In similar studies, Cunanan (2015) classified Bikol-Sorsogon varieties into four groups by describing and comparing lexical items and determining the patterns of isogloss formed. Dio and Jamora (2014) utilized a descriptive–comparative research design to describe, illustrate and compare dialects in Sorsogon Province as medium of instruction in grade school Mathematics. Further, the qualitative method involved the description and comparison of the speech varieties, while the quantitative method warranted the calculation of the degree of phonetically similar lexical items or apparent cognates.

Instrument

The researchers gathered the data through a semi-structured interview with each language resource person (LRP). The interview questionnaire consisted of three parts, namely, 24 questions for the background information of the LRP, 20 questions about language attitude and perception, and 200-word list elicitation (see appendix). Local government unit officials validated the background information of the LRPs to ensure that they passed the eligibility criteria.

The researchers finalized the word list of 200 basic vocabulary words, adapted from Swadesh (1955) and Blair (1990), through pilot tests in select localities. They used the Swadesh list during the initial pilot test in the districts of Sorsogon and Bacon where the LRPs validated the presence of 22 suspicious transcriptions in the form of Spanish loanwords, semantic overlap, absence of direct equivalence, and supplemental semantic morphemes. Twenty two words from the SIL (Summer Institute of Linguistics) South Asia survey word list substituted the omitted items. Two pilot tests followed in Sorsogon, Bacon, and Gubat until the LRPs did not anymore recognize any suspicious transcription from the data. The contextualized 200-item word list for the survey of Bikol-Sorsogon varieties included 76 items from the Swadesh list, 22 items from the SIL South Asia survey word list, and 102 items from both word lists.

Participants

The target population was native Sorsoganon people who were representative speakers of the variety spoken in each of the eleven municipalities, namely, Donsol, Pilar, Castilla, Sorsogon, Bacon, Casiguran, Prieto Diaz, Gubat, Barcelona, Bulusan, and Bulan. This study used purposive sampling in selecting two LRPs from each municipality. The LRPs were 9 males, ages 28-73, and 13 females, ages 18-71. The researchers adapted the eligibility criteria, that screened for

Bikol-Sorsogon native speakers, from Nahhas (2007) as follows: (a) The LRP grew up in the municipality, is living in the municipality at present, and has not lived elsewhere or has lived elsewhere for a short amount of recent time; (b) The LRP spoke the speech variety first and currently speaks the speech variety as his or her best language; and (c) The LRP has at least one parent from the municipality and that parent spoke the speech variety with him or her since childhood.

Data Collection

In the data elicitation, the researchers asked the LRPs to utter twice the equivalence of each word from the 200-item word list to their own native speech varieties. Computer software, such as GoldWave Digital Audio Editor and Speech Analyzer 3.1, aided in the recording and phonetic transcription of the data. Speech Analyzer 3.1 linked the phonetic transcriptions to Phonology Assistant 3.5.2 which helped in detecting suspicious transcriptions, organizing phones into charts, and analyzing the distribution of each phone. SIL International produced mainly these software along with user guides that enabled the researchers to operate without formal training.

Data Analysis

The researchers used phono-lexicostatistics method (Blair, 1990; Mann, 2005) in the data analysis comprised of phonetic and lexical comparison, phonetic and lexical similarity calculation, and data interpretation. Unlike the methods used in similar studies, phono-lexicostatistics includes explicitly defined criteria adapted from Blair to follow consistently in determining apparent cognates as follows:

Category 1:

> (a) Identical consonants which occur in the same position in each word

(b) Identical vowels or phonetically similar vowels which occur in the same position in each word

Category 2:

(a) Phonetically similar consonants which occur in the same position
(b) Vowels which are not phonetically similar and occur in the same position in each word

Category 3:

(a) Consonants which are not phonetically similar and occur in the same position in each word
(b) A phone which corresponds to nothing in the second word of the pair

Ignore:

(a) suprasegmentals such as stress, vowel length, tones
(b) supplemental semantic morphemes
(c) reduplicated syllables
(d) phonetic processes such as metathesis, phone substitution

The matrix below functioned as reference for the acceptable category combinations for lexical similarity:

Table 1.
Acceptable Category Combinations for Lexical Similarity (Blair, 1990).

Number of phones	Category		
	1	2	3
1	1	0	0
2	2	0	0
2	1	1	0
3	3	0	0
3	2	1	0

4	4	0	0
4	3	1	0
4	3	0	1
4	2	2	0
4	2	1	1
5	5	0	0
5	4	1	0
5	4	0	1
5	3	2	0
5	3	1	1

The criteria for comparing phone pairs were not applicable to all lexical pairs since some were apparently similar or different by simple inspection alone. Instead, the researchers initially categorized the lexical pairs into three: a) apparent cognates, b) apparent noncognates, and c) indeterminate. The analysis of indeterminate lexical pairs used the criteria for comparing phone pairs and the matrix for the acceptable category combinations for lexical similarity as exemplified in the following table:

Table 2.
Analysis of Indeterminate Lexical Pairs.

English Gloss	Lexical Pairs		Indeterminate			Apparent Cognates
			Number of Phones	Category by Phone	Category Combination	
animal	ʔa.ˈjəp	ˈha.jʊp	5	2a 1b 1a 1b 1a	4-1-0	YES
head	pa.ˈjʊ	ˈʔʊ.lʊ	4	2a 2b 3a 1b	1-2-1	NO

This study generated 55 speech variety pairs, collected 2200 lexical items, and analyzed the cognancy of 11000 lexical pairs. Thus, the researchers employed WordSurv 7 by the SIL to enhance the accuracy and speed of phono-lexicostatistical analysis. Wimbish (1986) developed WordSurv 2.5 which he first tested in a survey of the languages spoken in the Zambales Mountains of the Philippines. His study required nearly 800 comparisons that led to the definition of six different Negrito languages.

Ethical Consideration

The researchers followed ethical guidelines, as specified by the Research Ethics Committee, Philippine Normal University. This included undergoing an ethics review process before engaging interview participants to ensure that the procedures were fair and unbiased to all involved. The communication letters sent to the participants provided complete information regarding the nature of the study. The data collection tools in the form of questionnaires and word list, as well as the criteria for the selection of participants, were free from gender, class, ethnic, and cultural biases. The researchers kept the participants anonymous and their forms secure and accessible by authorized persons only. Also, the researchers observed the informed consent process by ensuring that the participants were aware that they were participating in a research and by asking their consent to participate with the option of withdrawing anytime. Finally, the participants and community officials had the choice of being sent the transcription of their interviews and a summary of the results of this research.

Results and Discussion

The discussion begins with description and comparison of the phonetic features which serve as basis for the quantification of apparent cognates between Bikol-Sorsogon varieties. An interpretation of the phono-lexicostatistical measures with regard to the closeness of the varieties, potential level of inherent intelligibility, and local language development program completes the analysis.

Similarities and Differences between the Phonetic Features of Bikol-Sorsogon Varieties

All of the eleven Bikol-Sorsogon varieties analyzed contain sixteen native consonant phones. This shows that Bikol-

Sorsogon varieties are phonetically similar. On the other hand, Cunanan's (2015) identification of the consonant [£] in the Pilar variety is not supported by the present study. Lobel and Tria (2000) classified this sound as an interdental-alveolar lateral which is phonemic and found only in South Catanduanes. The present study does not provide data that show the occurrence of [£] in Pilar or in any of the eleven Bikol-Sorsogon varieties analyzed.

Table 3.
Consonant Chart of Bikol-Sorsogon Varieties

	Labial	Dental	Palatal	Velar	Glottal
Plosive					
Voiceless	p	t		k	ʔ
Voiced	b	d		g	
Nasal					
Voiceless					
Voiced	m	n		ŋ	
Flap					
Voiceless					
Voiced		r			
Fricative					
Voiceless		s			h
Voiced					
Liquid					
Voiceless					
Voiced		l			
Glide					
Voiceless					
Voiced	w		j		

With regard to the distribution of consonants, fourteen phones occur in all positions. Consonants [r] and [h] do not appear in word final and syllable final positions of native words. Also, [h] and [ʔ] have different occurences in Donsol and Pilar compared to other varieties. As shown by the

present data, [h] does not usually occur in Donsol and Pilar especially in the initial position of words where it appears in other varieties. The glottal fricative is instead articulated as a glottal stop [ʔ], such as in [ʔa.ˈlas] 'snake,' [ʔɪg.ˈdaʔ] 'lie down' and [ʔʊ.ˈbag] 'swell' that correspond to [ˈha.las], [ˈhɪg.daʔ] and [ha.ra.ˈnɪ] in other varieties. Lobel and Tria (2000) and Cunanan (2015) indicated that the absence of [h] is one of the features of Southern Bikol, which includes Donsol and Pilar.

The position of the glottal stop [ʔ] in Donsol and Pilar varies from that of the other speech varieties. Cunanan (2015) pointed out that in Pilar, the glottal stop switches position with other consonants. In the present data, this phonetic process frequently takes place between the syllable initial [ʔ] and the preceding syllable final consonant as in [ˈtʊl.ʔaŋ] 'bone' (other varieties) and [ˈtʊʔ.laŋ] 'bone' (Donsol and Pilar), although [pa.ˈhaʔ] 'thirsty' (Pilar and other varieties) and [ʔa.ˈpaʔ] 'thirsty' (Donsol) show a process where the glottal fricative [h] interchanges position with the word initial consonant then becomes a glottal stop [ʔ].

Table 4.
Vowel Chart of Bikol-Sorsogon Varieties

	Front		Central		Back	
	Unrounded	Rounded	Unrounded	Rounded	Unrounded	Rounded
High						
Tense	i					u
Lax	ɪ					ʊ
Mid						
Tense						
Lax			ə			ɔ
Low						
Tense						
Lax			a			

Bikol-Sorsogon varieties have seven core vowel phones. Six vowels [i], [ɪ], [a], [u], [ʊ], and [ɔ] appear in all eleven varieties but the central-mid [ə] occurs only in Donsol variety. This finding does not agree with Cunanan's (2015) identification of the vowel [ə] in Pilar variety. Lobel and Tria (2000) stated that [ə] occurs as a true phoneme only in Buhi-non, Libon, and Miraya in which Donsol and Pilar varieties are classified. They also indicated that in all Bikol languages, except some dialects of Southern Bikol, the Proto-Central Philippine [ə] merged with [u]. In Libon, [ə] became [o], while in the remaining varieties or dialects of Buhi-non, Iriga, and Miraya, [ə] was retained.

The present analysis shows correspondence between [ə] in Donsol variety and [ʊ], [a], and [ɪ] in other Sorsogon varieties similar to the study of Cunanan (2015.) On the other hand, data reveal that the number of correspondences between [ə] and [ʊ] is greater than between [ə] and [a] or [ɪ]. None of these correspondences were established by the present study due to limited data.

Although phonology is not covered by this research, Lobel and Tria (2000) reported that the vowels /a/, /i/ and /u/ are contrastive in Bikol. The vowels [ʊ], [o], [ɔ] and [ɪ], [ɛ] are allophones of the phonemes /u/ and /i/ respectively. Similary, Sorsogon data reveal contrasts among the vowels [i], [a], and [u] and free variations among [ʊ], [ɔ], [u] and [ɪ], [i]. The contrastive vowels /a/, /i/, and /u/ occur only in medial and final positions. The word initial vowel that appears orthographically is phonetically transcribed with a glottal stop [ʔ].

Figure 1 presents the phonetically similar segments which are connected by a single line in Bikol-Sorsogon varieties. Each of the sixteen consonants and seven vowels is composed of features that served as bases for the identification of phonetically similar segments. The

phonetically similar segments were identified according to the principles of phonemic analysis that often render such segments as allophones (Burquest, 2006). Thus, the primary consideration was the number of features that the segments have in common.
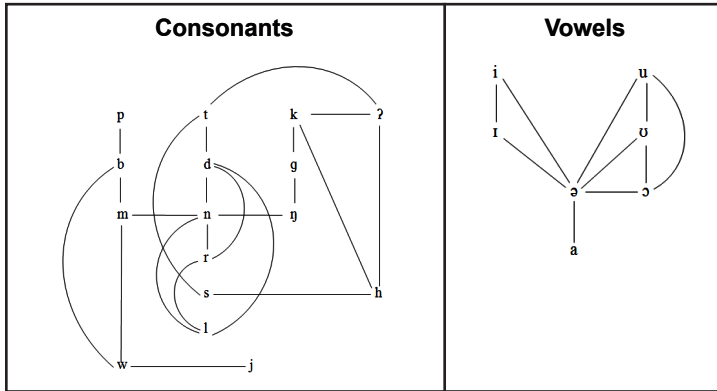

Figure 1. Phonetically Similar Segments

Table 5.
Phonetic Processes in Bikol-Sorsogon

Phonetic Processes	Examples		
	Gloss	Form 1	Form 2
Epenthesis and Elision of Consonants	'here'	[dɪ.ˈdɪ] (Don, Cast, Sor, Casi, Bula)	[ˈdɪg.dɪ] (SB, Pil, Bac, Pri)
Elision and Glottal Stop Substitution	'animal'	[ˈha.jʊp] (other varieties)	[ˈʔa.jʊp] (SB, Don, Pil)
Consonant Substitution	'burn'	[ˈsʊ.nʊg] (Sor, Gub, Bar, Bulu, Bula)	[sʊ.ˈlʊʔ] (SB, Don, Pil, Bac, Casi, Pri)
Nasal Assimilation	'dull'	[ma.ˈtaŋ.pʊl] (Casi)	[ma.tam.ˈpʊl] (Gub, Bar, Bulu, Bula)
Epenthesis and Elision of Vowels	'woman'	[ba.ˈbaj] (Don)	[ba.ˈba.jɪ] (other varieties)

Vowel Harmony	'nose'	[ʔɪ.ˈrʊŋ]	[ʔʊ.ˈrʊŋ]
		(Gub, Bar, Bulu, Bula)	(Don)
Syllable Reduction	'who'	[sɪ.ˈʔɪ.saj]	[ˈsɪ.saj]
		(Casi)	(Don, Pil, Bac, Pri)
Syllable Substitution	'where'	[dɪ.ˈʔɪn]	[sa.ˈʔɪn]
		(other varieties)	(Pil, Bac, Pri)
Metathesis	'bone'	[ˈtʊl.ʔaŋ]	[ˈtʊʔ.laŋ]
		(other varieties)	(SB, Don, Pil)
Stress Shift	'feather'	[ba.lʊ.ˈkag]	[ba.ˈlʊ.kag]
		(Don)	(other varieties)

Note. Don = Donsol variety, Pil = Pilar variety, Cast = Castilla variety, Sor = Sorsogon variety, Bac = Bacon variety, Casi = Casiguran variety, Pri = Prieto Diaz variety, Gub = Gubat variety, Bar = Barcelona variety, Bulu = Bulusan variety, and Bula = Bulan variety.

Burquest (2006) indicated that comparing words from different varieties is more complicated than a one-to-one comparison of individual phones, because sometimes language change involves linguistic processes. The phonetic processes that occur in Bikol-Sorsogon varieties formed words that are considered lexically similar. On the other hand, Bikol-Sorsogon varieties also have several words that are lexically different. The phonetic and lexical differences between the varieties may affect the speakers' writing ability and comprehension of other varieties.

Degree of Phonetic and Lexical Similarity between Bikol-Sorsogon Varieties

Table 6.

Percentages of Apparent Cognates between Bikol-Sorsogon Varieties

	Donsol	Pilar	Castilla	Sorsogon	Bacon	Casiguran	Prieto Diaz	Gubat	Barcelona	Bulusan	Bulan
Donsol	100	86	80	80	84	82	86	75	74	74	78
Pilar	86	100	78	80	95	88	96	76	75	75	77

Castilla	80	78	100	98	81	87	82	88	87	87	92
Sorsogon	80	80	98	100	82	88	83	88	88	88	92
Bacon	84	95	81	82	100	92	99	78	79	80	80
Casiguran	82	88	87	88	92	100	92	81	80	81	83
Prieto Diaz	86	96	82	83	99	92	100	78	78	78	80
Gubat	75	76	88	88	78	81	78	100	98	95	96
Barcelona	74	75	87	88	79	80	78	98	100	98	95
Bulusan	74	75	87	88	80	81	78	95	98	100	96
Bulan	78	77	92	92	80	83	80	96	95	96	100

The percentages of apparent cognates between Bikol-Sorsogon varieties were calculated using phono-lexicostatistics with the aid of WordSurv 7. The ranges for high, medium, and low degrees of relationship between Bikol-Sorsogon varieties are 87-99%, 81-86%, and 74-80% respectively. The varieties of Bacon, Prieto Diaz, and Pilar have mutually high percentages (95-99%) as well as Gubat, Barcelona, Bulusan, and Bulan (95-98%). Sorsogon and Castilla share a considerable amount of apparent cognates (98%) and relatively less high values (87-92%) to Gubat, Barcelona, Bulusan, and Bulan. Casiguran manifests similarity to a number of varieties including Bacon (92%), Prieto Diaz (92%), Pilar (88%), Sorsogon (88%), and Castilla (87%) but the percentages are relatively less high compared to the aforesaid varieties.

At the mid part of the scale, Donsol somehow manifests relationship to Pilar (86%), Prieto Diaz (86%), Bacon (84%) and Casiguran (82%) although the percentages may not be relatively significant enough. Sorsogon and Castilla exhibit a middle mutual similarity to Bacon (82, 81%) and Prieto Diaz (83, 82%), as well as Casiguran and the varieties of Bulan (83%), Gubat (81%), and Bulusan (81%).

Significantly low percentages can be observed between Donsol and the varieties of Bulan (78%), Gubat (75%), Barcelona (74%), Bulusan (74%), Sorsogon (80%), and Castilla (80%). The same can be drawn from the relationship between Bacon, Prieto Diaz, and Pilar and the varieties of Gubat, Barcelona, Bulusan, and Bulan (75-80%).

Relationship between Bikol-Sorsogon Varieties and Its Implications to Inherent Intelligibility and Language Development Programs

Among the eleven Bikol-Sorsogon varieties, four major groups are formed based on phonetically similar features and apparent cognate percentages. Group A includes Bacon, Prieto Diaz, and Pilar; group B comprises Sorsogon, Castilla, and Casiguran; group C consists of Gubat, Barcelona, Bulusan, and Bulan; and group D solely has Donsol. Variety groups A and D (85%) are closely similar as well as groups B and C (86%). Variety group B also exhibits some affinity to groups A (84%) and D (81%). Variety groups A and C (78%) as well as C and D (75%) are similar to a relatively low degree. The closeness of these varieties was also validated by the LRPs.

Table 7.

Average Cognate Percentages between Bikol-Sorsogon Variety Groups

	A	B	C	D
A	100	84	78	85
B	84	100	86	81
C	78	86	100	75
D	85	81	75	100

Blair (1990) defined inherent intelligibility as "the degree of understanding which speakers of one variety have of a similar variety because two varieties spring from

the same linguistic stock, not acquired by exposure to it" (p. 24). Simons (1979), Grimes (1988), and Blair have set a standard of 60% lexical similarity to screen out varieties that lack inherent intelligibility. A lexical similarity higher than 60% would render the intelligibility of two speech varieties inconclusive.

The range of lexical similarity percentages between Bikol-Sorsogon variety groups is 75% to 86%. Since no two variety groups are lexically similar by 60% and below, nothing can be concluded about inherent intelligibility. It is anticipated, however, that speakers of Bikol-Sorsogon varieties within the same subgroup can fairly comprehend one another, whereas speakers of the varieties separated by language boundaries should prove comprehension unlikely.

Consequently, since no two Bikol-Sorsogon varieties have a percentage below 60%, nothing can be assumed regarding the need for the varieties to have separate language development programs. It is anticipated, then, that Bikol-Sorsogon varieties within the same subgroup can use a single language program, while varieties of different subgroups should need separate language programs such as the Mother Tongue-Based Multilingual Education (MTB-MLE). In a recent study, Estremera (2017) identified the need to translate and localize MTB-MLE materials in Sorsogon City to aid the pupils' comprehension. Likewise, Dio and Jamora (2014) suggested that teachers and instructional developers collaborate in the development and validation of the localized MTB-MLE instructional materials in Sorsogon Province. These recommendations necessitate intelligibility testing to establish the mutual intelligibility between Bikol-Sorsogon varieties and to ascertain the need for separate MTB-MLE programs in Sorsogon Province.
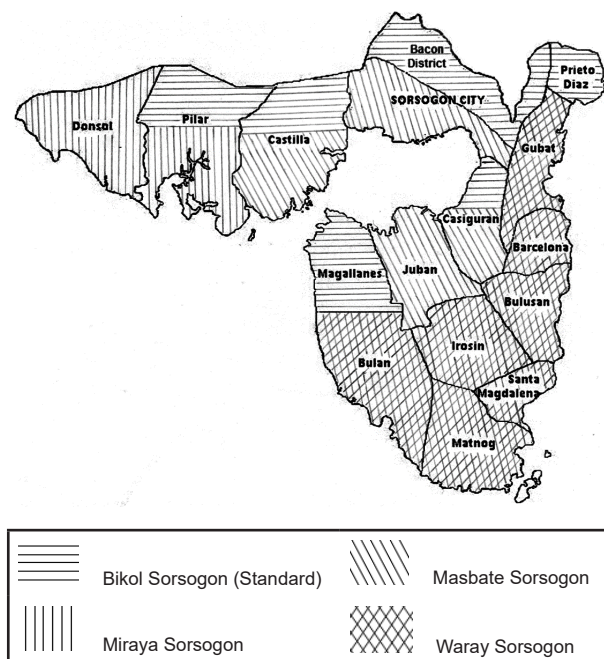
Figure 2. Geographical Distribution of Bikol-Sorsogon Varieties Based on Subgrouping Studies.

Figure 2 presents the geographical distribution of Bikol-Sorsogon varieties based on the preceding and present subgroupings. The subgrouping of Bikol-Sorsogon varieties in this study coincides with some of the works of previous researchers. McFarland (1983) classified Bacon, Prieto Diaz, and Pilar mainly spoken in Northern Sorsogon as similar to the Standard Bikol of Naga or Legazpi, and Sorsogon, Castilla, and Casiguran spoken in Central Sorsogon as related to Masbatenyo. Both McFarland and Lobel and Tria (2000) found that Gubat, Barcelona, Bulusan, and Bulan, which cover Southern Sorsogon, show relationship to Waray. Finally, Mintz (1973), McFarland, and Lobel and Tria have classified Donsol variety as similar to East Miraya of Daraga, Albay.

The present subgrouping does not exactly correspond to any of the preceding studies since methods and choice of localities vary. On the other hand, the present subgrouping apparently aligns with the works of Lobel and Tria (2000) and Cunanan (2015) except the classification of Pilar. Lobel and Tria's slight distinction of Casiguran from Sorsogon and Castilla parellelizes only with the present study. Both Cunanan and the present study cannot present a subgrouping that includes all Bikol-Sorsogon varieties. Sufficient data validate the existence and classification of the four subgroups of Bikol-Sorogon varieties nonetheless.

In relation to language development program, several native speakers in Sorsogon Province reported that the MTB-MLE materials used by the primary learners are written in the Standard Bikol of Naga which contain some unintelligible lexical items. Lorenzana (2018) mentioned in her study that Bikol Naga was assigned as L1 for the MTB-MLE instruction in Bicol Region which forces pupils, whose mother tongues are not adopted for instruction, to learn another language. Based on the PSA 2015 Census of Population, approximately 70 percent of the people in Sorsogon Province speak Bisakol (Masbate Sorsogon and Waray Sorsogon) as mother tongue, thus, Standard Bikol is not the language of wider communication.

Conclusion

This study aimed to analyze the varieties of Bikol language spoken in Sorsogon Province. It demonstrated the application of a relatively modern and hybrid method called phono-lexicostatistics. Different from the preceding researches, the method used in this study takes the middle ground between simple lexicostatistical inspection and the rigorous comparative method by adapting explicitly defined criteria in the analysis of apparent cognates. This study serves as basis

for an in-depth intelligibility assessment and development of language programs in Sorsogon Province.

Findings show that among the eleven Bikol-Sorsogon varieties, four major groups are formed based on phonetically similar features and apparent cognate percentages: (A) Bacon, Prieto Diaz, Pilar; (B) Sorsogon, Castilla, Casiguran; (C) Gubat, Barcelona, Bulusan, Bulan; and (D) Donsol. The closeness of the varieties was also validated by the LRPs. Generally, the greater the distance between two speech communities, the greater the variation. Regions that have relatively small populations and less restricted means of communication between locals may inhibit the development of marked differences in speech (Holmes, 2013). In Sorsogon Province, however, some geographically distant speech communities, such as Pilar, Bacon, Prieto Diaz and Magallanes, have closely related speech varieties.

High percentages of apparent cognates that occur across the varieties indicate that the Bikol-Sorsogon varieties belong to the same language family, while low percentages of apparent cognates reveal the influences of other languages. The relationship between Bikol-Sorsogon variety groups also validates some of the preceding studies that classified them as similar to Standard Bikol, Masbatenyo, Waray, and East Miraya. In general, Standard Bikol and East Miraya are Bikol languages, whereas Masbatenyo and Waray are Bisayan languages. Therefore, Bikol Sorsogon and Miraya Sorsogon, as well as Masbate Sorsogon and Waray Sorsogon, also possess close relationship. Despite the speakers' awareness of these variations, they generally call their language 'Bikol.' This may imply that the speakers prefer to be associated with the prestige language rather than the minority which may lead them to shift to Standard Bikol.

The persistent discrepancy in the description and subgrouping of Bikol-Sorsogon varieties, particularly Pilar and Castilla, is attributable to the differences in the bases of subgrouping, problems with methodology, and choice of localities since separate speech communities may exist within a municipal division. Moreover, closely related Bikol-Sorsogon varieties are not necessarily uniform in terms of phonetic features and lexical forms. Casiguran, for instance, shows similarities to several varieties. The level of inherent intelligibility between the varieties, as well as the status of the local language program, remains inconclusive since none of the percentages of apparent cognates goes below the standard of 60%. Comparing the results of the present study to the preceding ones validates the effectiveness and efficiency of phono-lexicostatistics method in analyzing several language varieties with limited data available.

Recommendations

This study was limited to the phonetic features and basic lexicon of the varieties for analysis and comparison. It was mainly based on actual spoken data which were analyzed in articulatory level. Other linguistic features, such as semantic morphemes and syntactic structures, may be described and compared to support or improve the subgrouping and classification of Bikol-Sorsogon varieties. Moreover, only eleven Bikol-Sorsogon varieties were analyzed in this study. Other Bikol-Sorsogon varieties, namely, Magallanes, Juban, Sta. Magdalena, Matnog, and Irosin, may be investigated using phono-lexicostatistics to come up with a general classification or subgrouping. Further, the phono-lexicostatistical percentages for relative comparison may be revalidated by conducting a replication study. This will help in the refinement of the methodology, especially in judging apparent cognates based on the adapted criteria, so as to widen its applicability to several Philippine varieties.

Linguistic similarity was measured as the relative degree of phonetically similar lexical items between Bikol-Sorsogon varieties. A full sociolinguistic profile of the province of Sorsogon can be established by the analysis of four other linguistic and sociolinguistic phenomena—dialect intelligibility, multilingualism, language use, and attitudes. Also, the phonetic data may be expanded for a phonological analysis of the varieties spoken in Sorsogon Province. Establishing the phonemes of the varieties may provide bases for orthography development that unifies the features common to all Bikol-Sorsogon varieties. In the field of instruction, the word lists in different varieties of Bikol-Sorsogon produced by this study may be developed into a multilingual glossary of basic literacy concepts which may be used as supplement material for MTB-MLE.

■ ■ ■

References

Blair, F. (1990). *Survey on a shoestring*. Dallas: Summer Institute of Linguistics and The University of Texas at Arlington.

Burquest, D. A. (2006). *Phonological analysis: A functional approach* (3rd ed.)*.* Dallas: Summer Institute of Linguistics International.

Crowley, T. (1998). *An introduction to historical linguistics* (3rd ed.). New Zealand: Oxford University Press.

Cunanan, F. (2015). *Ang dialect area ng Bikol-Sorsogon: Isang paunang suri* [The dialect area of Bikol-Sorsogon: A Preliminary Investigation]. University of The Philippines Diliman.

Dio, R. & Jamora J. (2014). *Variations of Sorsogon dialects as mother tongue-based medium of instruction in*

grade school mathematics. Sorsogon City: Sorsogon State College.

Estremera, M. L. (2017). The implementation of Mother Tongue-Based Multilingual Education: Viewing it from the grade III teachers' perspective. *Journal of Literature, Languages and Linguistics*, 40, 47-53.

Grimes, J. E. (1988). *Correlations between vocabulary similarity and intelligibility*. Notes on Linguistics, 41.

Holmes, J. (2013). *An introduction to sociolinguistics* (4th ed.). New York: Routledge.

Lewis, M. P., Simons, G.F., & Fennig, C.D. (Eds.). (2018). *Ethnologue: Languages of the world* (18th ed.). Dallas, Texas: SIL International. Retrieved from http://www.ethnologue.com/18/.

Lobel, J. W., & Tria, W. S. (2000). *An satuyang tataramon: A study of the Bikol language*. Naga: Lobel and Tria Partnership, Co.

Lorenzana, A. E. (2018). Phonological, lexical, and morphological analysis of the Philippine East Miraya Bikol-Pilar Dialect. *Bicol University R&D Journal*, 21 (1), 49-61.

Mann, N. W. (2005). *Syllostatistics* (Unpublished manuscript). Retrieved from http://surveywiki.info/index.php/Syllostatistics

McFarland, C. (1974). *The dialects of the Bikol area* (Doctoral dissertation). Yale University.

McFarland, C. (Ed.). (1983). *A linguistic atlas of the Philippines*. Manila: Linguistic Society of the Philippines.

Mesthrie, R. (2009). *Introducing sociolinguistics* (2nd ed.). Philadelphia: John Benjamins Publishing Inc.

Mintz, M. (1973). *Case and semantic affixes of Bikol*

verbs (Doctoral dissertation). University of Hawaii.

Nahhas, R. (2007). *The steps of language survey, an outline of practical methods*. Thailand: SIL MSEAG Survey Team.

Philippine Statistics Authority. (2015). *Population of Region V - Bicol (Based on the 2015 Census of Population)*. Retrieved 20 June 2016 from https://psa.gov.ph/content/population-region-v-bicol-based-2015-census-population

Simons, G. F. (1979). *Language variation and limits to communication*. Ithaca, NY: Cornell University, Department of Modern Languages and Linguistics. (Reissued by the SIL International, Dallas, 1983.)

Stark, C. (Ed.). (2009). *Globalization & languages: Building on our rich heritage*. Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from http://archive.unu.edu/globalization/2008/files/UNUUNESCO_2008_Globalization_and_Languages.pdf

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, *21*(2), 121-137.

Vajda, E. (2013). *Dialectology, linguistic 201 course materials*. Western Washington University. Retrieved from http://pandora.cii.wwu.edu/vajda/ling201/test3materials/dialectology.htm

Wimbish, J. (1986). The languages of the Zambales mountains: A Philippine lexicostatistic study. In Debyshire, D. C. (Ed.), *Work papers of the summer institute of linguistics, 30,* 133-142.

Appendix

Interview Questionnaire

Date: _____   Time: _____

Directions: Please complete the following questions to answer factual questions to the best of your knowledge and to reflect your opinions as accurately as possible. Your information will be kept strictly confidential.

I. **Questionnaire for Screening Language Resource Persons (LRP) based on Criteria**

The screening questions used in the present study are based on the criteria adapted from Nahhas (2007):

Criteria A: *The LRP is "from the municipality." - This is defined as growing up in the municipality, living in the municipality at present, and, if they have lived elsewhere, their time elsewhere is not a significant amount of recent time.*

1. What is your complete name?
2. Where were you born?
3. When were you born?
4. How old are you now?
5. How long did you live in here/there?
6. From what age to what age did you live here/there?
7. What language(s) do people speak in this/that municipality?
8. What do they call Bikol here/there?
9. Did you live in another place? If yes, for how long and what language(s) do the people speak there?

Criteria B: *The LRP spoke the speech variety first and currently speaks the speech variety as their best language.*

10. What language did you learn to speak first as a child?

11. What do you call your language? How do they say it in your municipality?
12. What language do you use at home and in talking to friends?
13. What other Bikol varieties do you speak?
14. What other languages aside from Bikol do you speak?
15. What is your educational attainment?
16. What is your occupation?
17. Can you understand Filipino?
18. Can you read and write in Filipino?
19. Can you understand English?
20. Can you read and write in English?

Criteria C: *The LRP has at least one parent from the municipality and that parent spoke the speech variety with them when they were a child.*

21. What language did your father learn to speak first as a child? How about your mother?
22. Is your parents' first language the same as yours?
23. What other languages do your parents speak?
24. Did your parents use their first language in talking to you at home and with their friends?

II. Questionnaire on Language Attitudes and Perception

1. Are you aware that Sorsogon Province has different speech varieties?
2. What do you call these speech varieties?
3. Do you have a specific name for each variety or do you just call them by one name?
4. What varieties in Sorsogon do you think are similar to yours?
5. What makes you think they are similar?
6. What varieties in Sorsogon are very different from yours?
7. What makes you think they are different?

8. Are there speech varieties that are more important than others? If so, which ones? Why are they considered more important?

9. What Sorsogon varieties are similar to the Standard Bikol of Naga or Legazpi?

10. Why do you think Sorsoganons call their speech as Bikol instead of specific names?

11. Is it important that you are recognized as a Bicolano rather than simply a Sorsoganon? Why?

12. Are there speech varieties which are considered NOT important or described as "bad"? Why?

13. Do you think the speech varieties in Sorsogon are purely Bikol?

14. If yes, why do you say so? If no, what other languages have influence to Sorsogon varieties?

15. What language do the students learn first at school?

16. For you, is it okay to use the students' first language as medium of instruction?

17. What other languages are taught to the students at school?

18. All languages can be written. If your language was written, would it be good for your children to be able to read and write it? Why?

19. What kinds of things would you want to have written in your language or speech variety?

20. What changes in your language or speech variety do you see 20 to 30 years from now?

III. **Word List Adapted from Swadesh (1952, 1955) and SIL South Asia Survey Word List (Blair, 1990)**

1. I (1.sg.)
2. you (2.sg)
3. he/she
4. we (exclusive)
5. they

6. this
7. that
8. here
9. there
10. who
11. what
12. where
13. when
14. how
15. not
16. many
17. some
18. few
19. other
20. one
21. two
22. three
23. four
24. five
25. big
26. long
27. thick
28. heavy
29. small
30. short
31. narrow
32. thin
33. woman
34. man
35. person (individual human)
36. child
37. wife
38. husband
39. mother
40. father
41. animal

42. fish (noun)
43. bird
44. dog
45. louse
46. snake
47. worm
48. seed
49. leaf
50. root
51. flower
52. grass
53. skin
54. blood
55. bone
56. fat
57. egg
58. horn (of bull etc.)
59. tail
60. feather
61. hair
62. head
63. ear
64. eye
65. nose
66. mouth
67. tooth
68. tongue
69. fingernail
70. foot
71. knee
72. hand
73. wing
74. belly or stomach
75. neck
76. back
77. breasts (female)

78. heart
79. liver
80. drink
81. eat
82. bite
83. suck
84. spit
85. vomit
86. blow
87. breathe
88. laugh
89. see
90. hear
91. know
92. think
93. smell
94. fear
95. sleep
96. live
97. die
98. fight
99. cut
100. split
101. stab (or stick)
102. scratch
103. dig
104. swim
105. fly
106. walk
107. come
108. lie (on side, recline)
109. sit
110. stand
111. turn
112. fall
113. give

114. hold
115. squeeze
116. rub
117. wash
118. pull
119. push
120. throw
121. tie
122. sew
123. count
124. say
125. sing
126. play
127. float
128. flow
129. swell
130. sun
131. moon
132. star
133. water
134. rain
135. river
136. sea or ocean
137. salt
138. stone
139. sand
140. dust
141. earth or soil
142. cloud
143. sky
144. wind
145. smoke
146. fire
147. ash(es)
148. burn
149. mountain

150. white
151. black
152. evening/night
153. day
154. year
155. hot
156. cold
157. full
158. new
159. old
160. good
161. bad
162. rotten
163. dirty
164. sharp (knife)
165. dull (knife)
166. wet
167. dry
168. correct
169. near
170. far
171. right (hand)
172. left (hand)
173. at or in
174. with (accompanying)
175. and
176. if
177. because
178. name
179. body
180. face
181. elbow
182. finger
183. house
184. roof
185. door

186. lightning
187. mango
188. banana
189. rice (plant)
190. chicken
191. cow
192. yesterday
193. today
194. tomorrow
195. above
196. below
197. speak
198. yes
199. youngest child
200. eldest child