Testing Intelligibility Among Sinitic Dialects

CECILIA SZETO
University of Hong Kong
kiuon@netvigator.com

1. Introduction

Intelligibility is often used to classify speech varieties as languages or dialects. In drawing the distinction, the degree of intelligibility between speakers of two different speech varieties can often indicate how close these varieties are. While talking about intelligibility as a criterion for the language-dialect classification or to group dialects of one language family, linguistic aspects like phonological and lexical factors are usually considered. Grammatical factors, on the other hand, are normally not focussed upon when intelligibility is being concerned. Grammatical divergence is an important factor in distinguishing different languages within one language family.

The Chinese dialects vary in lexical, phonological as well as grammatical aspects. Their complicated relationships with each other are often comparable to those between different languages within a family (Ramsey 1989). Even within one dialect group speech varieties may show great contrast, a well-known example being the Min supergroup where the different branches bear some grammatical differences, and also the Yue dialect group as suggested by Killingley (1993): "Yue dialects ... reveal significant differences which would and do prevent mutual understanding between speakers." This argument has added momentum to the long-standing debate of the status of the southern Chinese dialects as languages or dialects.

2. The dialects tested

To investigate the relationship between grammatical difference among dialects and the degree of intelligibility, an experiment was carried out to find out the general level of understanding by Cantonese speakers towards other southern Chinese dialects and what the major obstacles are to intelligibility. Four dialects are used: Taishan, Hakka, Chaozhou and Xiamen. They are selected because they can all be heard in Guangdong province and in Hong Kong as more people speaking these dialects have moved into the territory.

Taishan is spoken in the Siyi area in Guangdong. It belongs to the Siyi subgroup of the Yue dialects, while Cantonese belongs to the Guangfu (Guangzhou) subgroup. This sister dialect of Cantonese here is conventionally classified along with Cantonese as a dialect in the Yue group because of the many cognates, and shared phonological, and grammatical features.

For Hakka, the speech of Meixian (a city in north-eastern Guangdong) was taken, because it is widely regarded as the standard variety of Hakka. Being spoken in the mountainous areas of

Guangdong the Hakka dialects are considerably different from the Yue counterparts and they are more or less kept intact from phonological influence of Cantonese. In Hong Kong, Hakka is common only among the older generation living in villages at the outskirts of the territory.

Chaozhou is spoken in the eastern part of Guangdong close to Fujian province. It is the representative speech of the Chaoshan area under the Minnan category. Chaozhou is one of the southern Chinese dialects most commonly heard in Hong Kong as many people moved from the Chaoshan area to live in the city many generations ago. The Chaozhou clan is also one of the largest ethnic groups in Hong Kong.

Lastly Xiamen dialect is often considered the representative of the Quanzhang speech cluster of the Minnan branch. Considerable differences in grammar are found within different branches of the Min supergroup, and like all other Min dialects, Xiamen itself bears great phonological contrast and many grammatical differences when compared with Cantonese or the Yue dialects.

3. The method

The testing method was adopted from Whaley et al (1999). Four native speakers of the test dialects were invited to help with the recording of texts and sentences. These materials in each dialect, whose content do not overlap, were read in turn to a group of 20 Cantonese speakers, mostly native, selected randomly. They were asked to write down on a questionnaire any words they recognized from the tape. They could choose to write their responses either in colloquial Cantonese or in literary Chinese, or they were instructed to put a cross if they did not understand the sentence.

Subjects' responses were checked to see if they captured the meaning of the test sentences. Different representations (e.g. different vocabulary) are accepted as correct as long as they represent the same meaning.

The degree of intelligibility is then determined by the number of correct responses of the test sentences given by subjects. A high score would mean that the test sentence is highly intelligible to the subjects. It should be noted that the degree of intelligibility is determined on a comparative basis, i.e. the test sentences are compared with each other to see which one is the easiest for the subjects.

4. Linguistic background of subjects and the issue of familiarity

The questionnaire also recorded the linguistic background of subjects, including their place of origin and speech varieties used in their families. The information is useful in determining the reliability of data collected through the questionnaire, to find out whether a subject understands a certain amount of test materials because of their familiarity with the dialect or due to certain features of the dialects that the subject happens to find easy to understand. Table 1 below shows these data.

Table 1. Subjects' background

Dialects vs. No. of subjects with related background	Taishan	Hakka	Chaozhou	Xiamen	Cantonese / Others
Place of origin	11	0	2	2	5
Native relatives	6	0	2	2	10

Most subjects have Cantonese or Taishan background as Hong Kong is geographically closer to the cities of Guangzhou and Taishan. Others with Chaozhou or Xiamen background are present as the Chaozhou clan is rather large. No subject has a Hakka background, maybe because Hong Kong is not a major destination for migrating Hakka people.

However, as the results below will show, even where they do have relatives speaking non-Yue dialects, many subjects have very limited knowledge for the variety spoken by their close relatives. Such knowledge is not always sufficient to confer a significant degree of understanding of related dialects.

5. Hypothesis

A total of 43 sentences were used in the test. They were partly supplied by the informants and included simple descriptives, interrogatives, comparatives, passives, and imperatives. Although the testing focussed on grammatical influence and intelligibility, some test sentences are not specifically grammar-oriented as the testing also intended to investigate the general competence of local young Cantonese speakers' towards other southern Chinese dialects.

The general hypothesis of the test is that grammatical differences will contribute to a low degree of intelligibility. To show grammatical contrast, a sentence of a test dialect is compared with a Cantonese translation. Cognate morphemes (some grammatical but most of them lexical) of two dialects (the test dialect and Cantonese) used in different grammatical settings would have different degrees of impact on intelligibility, as illustrated by the following three cases of combinations.

(a) Cognate morphemes used in similar syntax -- when these morphemes are employed in similar grammatical structures with similar grammatical order, the degree of intelligibility would be high. As subjects already have a knowledge of grammatical structure of sentence, this might help them grasp the meaning of the test sentences.

(b) Cognate morphemes used in different syntax -- when cognate morphemes are employed in different grammatical structures, this might hinder subjects' understanding of the test dialects. Subjects may recognize similar sounding words but due to the unfamiliar grammar, the sentence might not be immediately understandable.

(c) Both morphemes and syntax are different -- this should be the most difficult case for subjects as they are presented with totally unfamiliar sounds. This situation is similar to encountering a different language.

It is predicted that intelligibility decreases from (a) to (c).

6. General results

The general results (Table 2) are that subjects showed a low degree of understanding towards the test dialects, except for Taishan. This result fits the prediction regarding traditional dialect grouping where both Taishan and Cantonese belong to the Yue dialect group. They share more common features, therefore subjects achieved a higher degree of intelligibility.

Table 2. Overall results of the experiment

	Taishan	Hakka	Chaozhou	Xiamen
A. Total number of test sentences	11	11	11	10
B. Total number of correct responses (= A x No. of subjects)	220	220	220	200
C. Actual number of correct responses given by subjects without familiarity with the test dialect	69	16	6	5
D. Relative percentage of intelligibility (= C / B x 100%)	31.3%	7.2%	2.7%	2.5%

7. The relationship between grammatical differences and intelligibility

7.1 The general case

Table 3 shows the case of comparative constructions and how structural differences influence subjects' understanding. The percentage of subjects giving completely correct responses is shown in the right column, with indication of those familiar with the test dialect in case (w F) and those who do not (w/o F). This is to give a clearer view on whether familiarity towards a

certain test dialect has affected subjects' ability to understand the sentences concerned which are set up with grammatical differences.

Table 3. A general view: the comparative structures of the test dialects

Dialect	Sentence	Complete correct response %
Taishan	$sai^1 lou^1 gai^3 su^3 leh^7 gwo^3 kwai^4 a^3 mui^2 ho^1 wu^3$ brother calculate maths clever than his sister good more 'The brother does maths better than his sister.'	65 (w F: 25 w/o F: 40)
Chaozhou	$ua^2 ma^2 li^2 loh^5 tsiah^5 pi^2 ua^2 ho^2$ I mother cook-eat than me good 'My mother cooks better than I.'	5 (w/o F: 5)
Hakka	$ngai^2 bi^3 gi^2 guo^4 tai^4$ I than he more big/old 'I'm bigger/older than him.'	0
Xiamen	$gua^1 he^1 so^1 bi^3 li^2 ka^1 ziuh^4$ I age than you more less 'My age is less than yours./ I'm younger than you.'	5 (w F: 5)

The Taishan sentence is similar to a Cantonese comparative phonologically and structurally (Subj + Adj + comparative marker /gwo^3/ + Obj); while the Chaozhou sentence is similar to a Mandarin comparative, using the '*pi*' cognate + Object of comparison + Adjective. This test sentence probably shows Mandarin influence with a more formal register.[1] The degree of intelligibility is thus lower than the Taishan sentence. Sharing the same structure, the Hakka and Xiamen sentences seem like a combination of the Cantonese /gwo^3/ cognate and the Mandarin '*pi*' marker: Subj + *pi* + Obj + /gwo^3/ cognate + Adj. This structure is relatively less familiar for Cantonese speakers. For the Xiamen sentence, the only subject to give a correct response is a Fujian native. So there is a trend showing the degree of intelligibility is decreasing from the most familiar type of grammar to the least familiar.

---

[1] There is also a relatively colloquial form of the comparative that employs 'gue' which is cognate to the Cantonese comparative marker 'gwo'. But in this context the Mandarin comparative structure was felt by the informant to be the more natural expression to use.

7.2. Case (a) of hypothesis: using cognate morphemes with similar grammar

Table 4. Cognate morphemes used in similar syntax.

Dialect	Sentence	Complete correct response %
Hakka	$sid^6 fan^4 ng^2 log^6$ eat rice not down 'I can't eat.'	70 (w F: 0 w/o F: 70)
Chaozhou	$png^7 tsiah^5 m^7 loh^5$ rice eat not down '(he) can't eat.'	20 (w F: 10 w/o F: 10)
Cantonese	$sik^8 m^4 lok^8 faan^6$ eat not down rice '(I) can't eat.'	N.A.

The Hakka sentence in Table 4 has the structure: V + Obj + resultative adverb whereas the Cantonese structure is: V + resultative adverb + Obj. In contrast to Hakka, placing the resultative adverb at phrase-final position is less typical in Cantonese, usually in more restricted environments (with pronoun as object) as in Example (1).

(1) $ngo^5 gam^1 ci^3 bong^1 lei^3 m^4 dou^2$
        I this time help you not able
        'I can't help you this time.'

However, looking back on the Hakka test sentence, there is overlap with Cantonese syntax. And with great similarity in pronunciation, subjects could easily recognize this Hakka sentence.

Some subjects also recognized the Chaozhou sentence, whose structure Obj + V + resultative adverb looks like topicalization of the Cantonese expression, yet not as many as those for the Hakka one. This was mainly due to lack of understanding of Chaozhou lexical items like /png^7/ for rice and /$tsiah^5$/for eat.

7.3 Case (b) of hypothesis: cognate morphemes in different grammatical structures

An example showing the case when grammatical difference may begin to lower the degree of intelligibility, is given below by the Xiamen pretransitive sentence.

Table 5. Xiamen pretransitive construction[*]

Dialect	Sentence	Complete correct response %
Xiamen	ka^5 mng^3 $kuan^1$ hi^8 lai^6 PRE-TR door close INCH 'Close the door.'	25 (w/o F: 25)
Cantonese (literary)	ba^2 mun^4 $gwaan^1$ hei^2 loi^4 PRE-TR door close INCH 'Close the door.'	N.A.
Cantonese (colloquial)	$saan^1$ $maai^4$ dou^6 mun^4 $keoi^5$ close V-PRT CL door it 'Close the door.'	N.A.

One fourth of the subjects got the Xiamen sentence in Table 5 completely correct and many recognized key words like 'door', 'close' and the adverb. It is not clear whether subjects recognized the grammatical significance of the pretransitive marker /ka^5/ in Xiamen, because in the colloquial Cantonese expression such marker is not employed, but its structure resembles that of the literary Cantonese expression and it might have helped subjects' understanding. It maybe said that to a lesser extent, subjects' familiarity with the Mandarin pretransitive structure also helped them understand the Xiamen sentence.

However, sometimes when cognate morphemes are used in different grammatical structures, such differences become an obstacle to understanding even when there is a phonological resemblance. The two Hakka questions in Table 6 are a case in point.

---

[*] Abbreviations: PRE-TR: Pretransitive; INCH: Inchoative; V-PRT: Verbal particle

Table 6. Two Hakka questions

Dialect	Sentence	Complete correct response %
Hakka	yiu^1 $tsan^4$ dao^4 $qian^2$ mao^2 have earn tense-marker money not 'Did you make any money?'	0
	yiu^1 $qian^2$ $tsan^4$ mao^2 have money earn not 'Did you make any money?'	0
Cantonese	ao^5 mou^5 zan^6 $tsin^2$ ah^8 have not earn money question-particle Did you make any money?'	N.A.

A typical Hakka question has the structure of Aux V + content + /mao^2/ (serving as a question marker with a rising tone). Here the Cantonese morpheme /mou^5/ however does not function as a question marker like its Hakka cognate /mao^2/, rather it simply shows a negative meaning and the sense of question is brought out more fully by the question particle /ah^8/ at the end.

It should be noted that in Cantonese another question marker /mei^6/ can be used and it is also put at the end, as in Example (2).

(2) lei^5 sik^9 zo^2 $faan^6$ mei^6 (ah^8)
        You eat Past-tense rice yet (question particle)
        'Have you eaten your meal yet?'

The focus of question (2) is to find out whether a particular event has happened or not, whereas this is not the focus in the Hakka questions in Table 6. Based on this semantic context, subjects may have tried to understand the Hakka questions in the context of the Cantonese /mei^6/ questions. Many subjects recognized key words like 'have', 'money' etc. (due to similar pronunciation in Cantonese/Mandarin), and a few of them even recognized these Hakka test sentences as questions; but no subjects gave a completely correct response.

7.4 Case (c): different morphemes in different grammatical structures

This is illustrated by a Xiamen ditransitive sentence in Table 7.

Table 7. Xiamen ditransitive

Dialect	Sentence	Complete correct response %
Xiamen	$be^3\ sen^5\ be^1\ zit^7\ dui^3\ hue^5\ ban^6\ s\tilde{a}^2\ li^2$ Mr. Be wants one pair flower-vase give you 'Mr. Be wants to give you a pair of vases.'	5 (w F: 5)
Cantonese	$baak^9\ saan^1\ soeng^2\ song^2\ (jat^7)\ deoi^3\ faa^1\ zeon^1\ bei^5\ lei^5$ Mr. Pak wants give (one) pair flower-vase give you 'Mr. Pak wants to give you a pair of vases.'	N.A.

Unlike the Cantonese structure, the Xiamen sentence differs mainly in that the modal verb 'want' is separated from the main verb 'give'. Also different morphemes are used, /be^1/ vs. /$soeng^2$/ for 'want', /$s\tilde{a}^2$/ vs. /$song^2$/ for 'give'. The lexicon /$hue^5\ ban^6$/ partly cognates with the literary Cantonese translation /$faa^1\ ping^6$/ but in different pronunciation (the term /$faa^1\ zeon^1$/ is more commonly used in daily life). So in this case where both the morphemes used and the sentence structure are completely different from those of the equivalent Cantonese expression, again no native Cantonese subjects gave a correct response.

8. Conclusion

Overall the low levels of intelligibility indicate that subjects, being native Cantonese speakers infrequently using other Chinese dialects, find understanding difficult unless they are already familiar with the dialects concerned. As expected on the basis of traditional dialect grouping, speakers of one dialect understand more readily another from the same dialect group. Thus the subjects found the Taishan text and sentences the most comprehensible. Their linguistic background, however, may not matter much. In most cases, self-reported familiarity with a certain dialect does not enable subjects to give overwhelmingly correct responses nor does it seem to contribute significantly to a high degree of intelligibility. Such discrepancy is probably because subjects have a passive knowledge of the dialect, or familiarity only arises with the particular variety or accent of the dialect spoken in their families. Given the considerable differences in phonological as well as grammatical aspects within one dialect group, knowing the speech variety of their relatives might not necessarily help subjects recognize varieties in other subgroups of the same dialect.

Speakers of one speech variety may find another variety highly intelligible primarily due to phonological similarity between the two. This is shown in 7.2 above, where most subjects

could easily understand the Hakka sentence. But the significance of grammatical difference is not to be underestimated in dialect intelligibility testing. From the three cases of the hypothesis illustrated by the examples above, grammatical differences between the test dialects and Cantonese can sometimes be viewed as impeding intelligibility, most clearly shown in Table 6 with the two Hakka questions. Because of the different question marker and the general structure of a question, subjects found these Hakka questions not readily understandable. Yet, grammatical similarity between two speech varieties may help increase intelligibility, disregarding phonological factors. In the example of the Xiamen pretransitive sentence (Table 5), although the sentence is phonologically different from Cantonese, its syntax resembles that of a literary Cantonese expression. Therefore more subjects responded correctly than to other Xiamen sentences which employ grammatical structures different from those of Cantonese.

References

Killingley, S.Y. 1993. *Cantonese*. Lincom Europa: Newcastle.

Ramsey, S. Robert. 1989. *The Languages of China*. New Jersey: Princeton University Press.

Whaley, L. J., Grenoble, L.A. & Li, F. 1999. Revisiting Tungusic classification from the bottom up: a comparison of Evenki and Oroqen. *Language* Vol.75, No.2, 286-321.