CORPORA AND OTHER SYNTACTIC ELICITATION METHODS
IN A STUDY OF THE ASPECTUAL SYSTEM IN GONG'AN DIALECT

Hong Cᴀɪ

The University of Queensland, Australia

h.cai1@uq.edu.au

Abstract

This paper is part of a larger project on Gong'an dialect, which is based on the data collected from field and aims to identify the aspect markers in Gong'an dialect and provide a systematic and comparative account of the distributional and semantic features of the aspect. More precisely, this paper engages with the data collection methods employed in this project and some of the preliminary findings of the fieldwork. The (dis)advantages of corpora and the most recently developed syntactic elicitation techniques are taken into account in the design of the fieldwork in order to lay a solid empirical basis for subsequent analysis. By combining a variety of methods, namely, acceptability judgment, spontaneous speeches, and corpora compilation, this research promises to provide insights into both the exact nature of the aspectual system in this dialect and theoretical issues of more general interest.

Keywords

Gong'an dialect, syntactic elicitation techniques, corpora, aspectual system

LOS CORPUS Y OTROS MÉTODOS DE ELICITACIÓN SINTÁCTICA EN UN ESTUDIO DEL SISTEMA ASPECTUAL DEL DIALECTO GONG'AN

Resumen

Este trabajo forma parte de un proyecto más amplio sobre el dialecto gong'an, que se basa en los datos obtenidos mediante encuestas y que tiene como objetivo identificar los marcadores de aspecto en el dialecto gong'an; de esta manera se proporciona una descripción sistemática y comparativa de los

rasgos distribucionales y semánticos del aspecto. En concreto, este trabajo se relaciona con los métodos de recogida de datos empleados en este proyecto y en algunos de los resultados preliminares del trabajo de campo. Se tienen en cuenta las (des)ventajas de los corpus y de las técnicas de obtención sintácticas más recientemente desarrolladas en el diseño del trabajo de campo con el fin de sentar una sólida base empírica para el análisis posterior. Mediante la combinación de diversos métodos, a saber, el juicio de la aceptabilidad, el habla espontánea y la recogida de corpus, esta investigación proporciona información tanto de las características del sistema aspectual en este dialecto como de cuestiones teóricas de carácter más general.

Palabras clave

Dialecto gong'an, técnicas de elicitación sintáctica, corpus, sistema aspectual

1. Introduction

This paper is the methodological part of an ongoing project on the aspectual system in Gong'an Dialect 公安话 (GAD),[1] which is based on the data collected from field and aims to identify the aspect markers in GAD and provide a systematic and comparative account of the distributional and semantic features of the aspect markers in the light of a slightly modified two-component theory of aspect.[2] More precisely, this paper is primarily concerned with a few syntactic elicitation techniques used in the data collection process and some of the preliminary findings of the fieldwork. As there was very little data available for the syntactic, semantic and (in individual cases) pragmatic purpose of the project, the main fieldwork was conducted to collect both oral and written data from September to November 2009. After preliminary analysis on the data collected from the first field trip, a supplementary fieldwork was carried out from January to February 2011 during which more data was obtained through questionnaire and spontaneous speech. Both the advantages and challenges of each method are taken into account in the design of the fieldwork and the data collection procedure in order to lay a solid empirical basis for subsequent analysis of the aspects in GAD.

---

[1] GAD is spoken in Gong'an County, Hubei Province, China, and classified as Southwestern Mandarin according to its tone values (cf. Wurm *et al.* 1987). The abbreviations used in this paper include: ACC: accomplishment; ACH: achievement; ACT: activities; ASP: aspect marker; CL: classifier; DUR: durative; GAD: Gong'an Dialect; GEN: Genitive; IPA: International Phonetic Alphabet; MC: Mandarin Chinese;
[2] The two components are 'situation aspect' and 'viewpoint aspect' (cf. Smith 1991, 1997).

Section 2 begins with a brief survey on studies on Chinese dialect grammar and the main methods adapted. [3] The motivation and justification for combining acceptability judgments and corpus data in this project are also discussed in this section. Section 3 describes the design of the elicitation questionnaire and the data collection procedure. Section 4 presents some of the preliminary findings of the data collection tasks. Section 5 briefly summarizes the implications of this study and reemphasizes the benefits of a multi-method approach in empirical dialectology.

2. Dialect grammar & research methods: some background

2.1 Studies on Chinese dialect

The study of dialects in China has a venerable tradition going back at least to Western Han dynasty (202 BC – 9 AD) when *Fang Yan*, which is considered as the first work on Chinese dialectology, was compiled by Yang Xiong (53 BC – 18 AD) (cf. L. Wang 2007: 17-24; Cen 2008: 25-26). However, previous research into Chinese dialects has concentrated primarily on salient phonological features and to some extent lexical features, and has largely ignored the grammar of dialects, which corresponds to the dialect research all over the world (cf. Cornips & Jongenburger 2001: 53). The reasons behind the traditional neglect of Chinese dialect grammar have been extensively discussed in the literature (cf. Xu 1998; Lu 2004; You 2005; Jiang 2008; Zhan 2009; *inter alia*).

Nonetheless, there has been growing interest in the domain of Chinese dialect grammar since the 1980s. In respect of the aspectual phenomena, increasing attention is being paid to them from a variety of perspectives, including synchronic, diachronic and typological studies (e.g., Mei 2000[1988]; Chappell 1992, 2001; Yue-Hashimoto 1993; Sun 1998; Wu 1999; etc.). The GAD, which is spoken in a transitional area that plays a strategic role in linking the northern and southern banks of Yangtze River and on the border of two provinces (Hubei and Hunan), is unsurprisingly of hybrid nature. GAD possesses a considerably complex aspectual system, which, however, has not

---

[3] The term 'Chinese' in this paper includes both genetically related regional dialects and the lingua franca of China, i.e. Mandarin Chinese.

been systematically examined. It is estimated that there are more than ten durative aspect markings in GAD, whereas in Mandarin Chinese (MC) the aspect marker *-zhe* plays an essential, if not exclusive, role in indicating durativity. As W. S-Y. Wang (1978: 61) remarks, "Diversity is the fundamental ingredient of change". The co-existing forms used in a single dialect potentially present different strata (Yue-Hashimoto 1991; Zhang 2000). The high diversity of aspect markers in GAD calls for a thorough investigation which may provide direction for further studies; for instance, historical approaches to the etymology of some aspect markers cross-dialectally. Furthermore, with the overwhelming influence of MC, GAD has been moving rapidly towards MC in every sense. It is therefore of great importance to study and to a certain extent "document" the idiosyncratic features of GAD.

With regard to methods used in the studies on Chinese dialect grammar, besides introspective judgment, a few data collection methods have been employed to elicit syntactic data. These include narrative recordings (e.g., Wu 1999; Zhang *et al.* 2001) and questionnaires and interviews (e.g., Yue-Hashimoto 1993; Zhang 2000; Cao 2008). A large number of synchronic studies on dialect grammar, however, have failed to make explicit how the data was obtained. A few recent studies on Chinese dialectal grammar have shown relevant information on the informants used, but made no attempt to mention either the results of the surveys or aspects of the data collection process, rather, the data elicited are directly incorporated into descriptions. Indeed, ambiguity in methods of data collection would probably cast doubt on the results presented. In other words, the reliability of the elicitation is determined by the adequacy of the methods used (cf. Greenbaum 1973: 201).

The need for application and explanation of various data collection methods in Chinese dialectology is technically obvious. In the remainder of this section, corpus and judgment data in modern dialectology are briefly discussed, followed by an introduction to and partial justification of the methodological approach of the study. Further justification is provided in Section 3.

2.2 Corpus

From the corpora compiled and used by structuralists who started with "a large collection of recorded utterances from some languages, a corpus" (Harris 1993: 27), to

today's corpora which generally refers to a collection of text or natural spoken materials that are machine-readable, representative, principled and/or balanced in terms of genres (cf. McEnery, Xiao & Tono 2005; Gilquin & Gries 2009), corpus data is playing an increasingly important role in both descriptive and theoretical linguistics (cf. Čermák 2002: 265). However, fruitful incorporation of corpus linguistics into dialectology has been seen only relatively recently, giving rise to a new branch or more specifically a new methodology, namely, "corpus-based dialectology" (cf. Anderwald & Szmrecsanyi 2009).

Both dialectologists and theoretical linguists have explored dialect corpora from a variety of viewpoints, including functional-typological, historical, generative, microvariational and quantitative (cf. Anderwald & Szmrecsanyi 2009; Nerbonne 2009; Szmrecsanyi & Kortmann 2009; Barbiers 2009; *inter alia*). The immediate benefits of well-structured dialect corpora for studies on dialect grammar include its probability to provide large amounts of data from natural contexts, whereas traditional elicitation methods may encounter difficulty in doing so. The limitations of corpora in dialect grammar study are also apparent: the existing corpora may contain very little evidence concerning the particular syntactic phenomena under investigation, or may not be valid at all. Therefore, some other methods are required to supplement the corpus data.

With respect to this project, corpus data will presumably help to identify the aspect markers in GAD, which are virtually unknown to the scholarly world and may escape the researcher's Sprachgefühl as well. In addition to that, a corpus-based approach will enable the study to gain more insights into the distributional and semantic features of each aspect marker in natural context, and therefore help to achieve the goal of providing a descriptively adequate account of the aspectual system in GAD.

Apart from the descriptive goal, this study also engages with theoretical issues of more general interest. In particular, it strives to address the question of to what extent the two-component theory of aspectuality is adequate in describing and explaining the features of the aspectual systems in MC and GAD from a comparative perspective. To ensure all the aspect markers and situation types and the interaction between them are captured in the data, an experimental method is required in the fieldwork, and this is discussed in the following section.

2.3 Judgment data

On the one hand, as the primary data source of a great deal of syntactic theory, intuitively judged (either by the linguist him/herself or by others) data has been continuously receiving comments questioning its reliability due to its susceptibility to a range of factors such as subject-related factors (e.g., individual differences, organismic factors, experiential factors) and task-related factors (e.g., procedural factors, stimulus factors) (Schütze 1996). On the other hand, instead of abandoning it entirely, considerable efforts have been devoted to improving the acceptability judgment for a number of well-known reasons (Schütze 1996: 2). [4] Techniques used in related disciplines (e.g., psychology) have been successfully adopted to diminish the artificiality of judgment data (cf. Bard *et al.* 1996; Schütze 1996; Cowart 1997). Recent developments in the study of dialect grammar have considerably improved and to certain extent standardized syntactic elicitation techniques including judgment elicitation (cf. Cornips & Jongenburger 2001; Bucheli & Glaser 2002; Cornips & Poletto 2005). It is observed that adults at least partially depend on their knowledge about the superordinate or the standard language to make acceptability judgments, and one effective way to minimize the influence is to ask for indirect acceptability judgments as described in Labov (1975) (cf. Cornips & Poletto 2005).

One of the main advantages of acceptability judgment is that particular phenomena that very rarely occur in spontaneous speech corpora or daily language can be specifically targeted by eliciting judgments which yield "a considerable degree of scientific precision and control" (Lee 2001: 141). The reason for the choice of judgment data for this project is largely due to the multi-dimensional and complex nature of the aspectual phenomena in general and the aspectual system in GAD in particular. Based on the characterization of situation aspect and the researcher's Sprachgefühl as a native speaker of GAD, it is assumed that like MC, GAD also recognizes four situation types, namely, activities, states, accomplishments, and achievements, of which activities, states and achievements are further divided into subgroups. In order to test the proposed situation types and their interactions with

---

[4] In the literature, there is a tradition of using 'grammaticality judgment' and 'acceptability judgment' interchangeably (Schütze 1996: 26).

aspect markers in GAD, judgment elicitation is used in the data collection in the form of a questionnaire.

2.4 A combined approach

Given the advantages and disadvantages of corpus and elicitation methods, an approach combining corpus data and elicitation data has been enthusiastically adopted by linguists from various sub-disciplines. A recent review of the status of using the combination of corpora and experimental methods by psycholinguists and corpus linguists has shown that 78% of the papers under review are psycholinguistically oriented, whereas only 10% are corpus-linguistically oriented (Gilquin & Gries 2009: 11). Besides providing ample evidence of inadequacy of the corpus-only approach, this review has also revealed how fruitful such a combined approach can be, and suggested that the combined approach should have been more widely used by corpus linguists.

In the area of dialect grammar, the combined approach has been adopted by dialectologists only quite recently. In a study of Lancashire dialect, Hollmann & Siewierksa (2006) have not only demonstrated the value of corpora as "an excellent methodological tool for identifying the presence of certain types of constructions in a language or dialect and gauging some idea of their distribution", but also proved that corpus-oriented methodologies fall short of exhibiting more subtle variables, and that there is a need for other methods to complement corpus data. Other studies in dialect grammar that adopted (or rather benefited from) the combined approach include, but by no means are limited to, Cornips & Jongenburger (2001), Lee (2001), Hardie & McEnery (2003), Cornips & Poletto (2005), Thráinsson *et al.* (2007).

Taking into account the advantages and disadvantages of corpus and judgment data discussed above, this project therefore takes the combined approach; that is, combining corpus data with experimental data, which promises to shed light on the methods in Chinese dialectology. As there is a lack of corpus in GAD, this project will build a corpus for its own purpose using data collected from spontaneous speech and other sources

3. The data collection procedure

In contrast to the majority of linguistic fieldwork conducted by non-native speakers of the languages or dialects under investigation, this project is in a relatively better situation since the researcher is a local GAD speaker who was born in Gong'an and lived there continuously for 18 years since birth. This brings a series of benefits to the fieldwork. Firstly, there are no significant cultural differences between the interviewer and the interviewees, and the Observer's Paradox (Labov 1972) therefore can be minimized to a large extent; secondly, the researcher can use the local dialect throughout the oral elicitation tasks which can diminish some well-known pitfalls such as 'style shifting' of the speakers, 'adjustment to the standard' (Labov 1972) or 'accommodation' (Cornips & Jongenburger 2001: 59; Hollmann & Siewierksa 2006: 26);[5] thirdly, when necessary the researcher can freely and naturally adjust the direction of the spontaneous speeches to certain topics that may stimulate the informants to produce more relevant data.

However, with all the advantages in hand, the project follows rigorous procedures in designing the fieldwork and preparing for the data collection, which will be presented in the following sections.[6]

3.1 Pre-fieldwork preparations

3.1.1 Literature study

This project starts with a comprehensive literature survey. All publications related to GAD were traced. So far there are only a few publications on GAD, which include a concise phonological description of GAD in *Report of Hubei Dialectal Survey* (Chao et al. 1972[1948]), one short article dealing with three grammar phenomena (Zhu 2005), and one monograph on GAD which is heavily involved with phonological and lexical issues yet offers an eye-opening introduction to some grammatical features

---

[5] 'Style shifting' is illustrated by Labov (1969: 19) as follows: "every speaker will show some variation in phonological and syntactic rules according to the immediate context in which he is speaking".
[6] The design of this project has drawn inspirations from the project on the *Syntactic Atlas of the Dutch Dialects* (SAND) (cf. Cornips & Jongenburger 2001), the project on the Syntactic Atlas of Swiss German Dialects (cf. Bucheli & Glaser 2002), and the Icelandic (Pilot) Project in ScanDiaSyn (cf. Thráinsson *et al.* 2007).

of GAD at both lexical and syntactic levels (Yang 1992). In addition to the above, recently Wang & Wang (2011) briefly discussed the phonetic characteristics of the dialects on the Jianghan Plain, which includes GAD.

3.1.2 Informants

Valuable suggestions on how to choose informants have been continuously proposed from the point of view of the investigator who is regarded as an 'outsider' of the language community (cf. Samarin 1967; Newman & Ratliff 2001; Crowley 2007). As an 'insider' of the language community, the researcher of this project possesses a few privileges in selecting qualified informants. By 'qualified', it means here to have the ability to cooperate with the investigator and produce natural data in GAD without apparent influence from MC or other dialects. Taking into account the education situation in China and the MC influence through mass media, this criterion tends to favor the generations who were born before 1960, raised at the survey site and had received very little education. To be specific, when identifying potential candidates, a number of variables have been considered. These include age, education, occupation, gender, birthplace, the longest time period that they have been out of Gong'an, parents' language, and the language or dialect used at home.[7] Nonetheless, reasonable adjustments on education, occupation and gender are acceptable after pilot interviews and careful scrutiny. Finally, 12 informants representing different age groups were selected to participate in judgment elicitation (JE) and spontaneous speech (SS) (see Table 1).

---

[7] The informant background checklist used for the project was adopted from Yue-Hashimoto (1993: 263-264) and slightly modified according to the requirement of this research.

Year of birth	Gender	Education	Occupation	Out of Gong'an	Tasks	Group
1932	M	A few days	Folk Artisan	6 months	JE, SS	1
1935	M	5 years	Folk Artisan	3 months	JE, SS	1
1943	M	6 years	Farmer	Less than a week	SS	4, 5
1945	M	6 years	Farmer	1 week	JE, SS	2
1946	F	1 year	Farmer	1 week	JE, SS	2
1948	F	No	Farmer	1 month	SS	4, 6
1949	M	No	Farmer	No	SS	4, 5
1950	F	1 year	Farmer	2 weeks	SS	5, 6
1952	M	8 years	Literary and art worker	6 months	JE, SS	3
1956	F	9 years	Writer	4 years	JE, SS	3
1960	F	3 years	Farmer	No	SS	6
1960	M	5 years	Farmer	2 weeks	SS	4, 5

Table 1. Informants' background.

3.1.3 Questionnaire

Based on the two-component theory of aspectuality (cf. Smith 1991, 1997) and the researcher's Sprachgefühl, a questionnaire that consisted of 536 sentences in GAD representing various situation types and aspect markers (see examples below), was constructed in the form of a combination of Chinese characters and IPA for those dialectal words without cognates in MC.[8]

(1)

他们　　　　在　　　　讲　　白话　　　　　　　　　(ACT; PRO)

tʰA⁴⁵mən　*tsai³⁴ (tai³⁴)*[9]　**tɕiaŋ²¹**　pɣ¹³xuA³⁴

they　　　ASP　　　　talk　vernacular

"They are chatting."

---

[8] Some of the Chinese characters are approximate homophones to the dialect words as their cognates in MC are unclear.

[9] tai³⁴ is possibly the phonetic variant of tsai³⁴ which marks the progressive aspect.

(2)

大门　　　关　　起在　　／　斗在　　　　　　　　(STATES, DUR)

tA^{34} mən^{13} **kuan^{45}** tɕ^{h}i^{21} tsai^{34} / tou^{21} tsai^{34}

big door　close　ASP　　／ ASP

"The gate (door) is closed."

(3)

拿 起 杯 子　找　　杯 子　　　　　　　　　(ACT, DUR)

lA^{13} tɕ^{h}i^2 pei^{45} tsɿ　tsau^{21}　pei^{45} tsɿ

take ASP cup　　look for　cup

"(She) was looking for the cup while holding the cup in hand."

(4)

碍头　　高头　　**写**　　得有　　字　　咧！(STATES, DUR)

ai^{13} t^{h}ou^{13} kau^{45} t^{h}ou　ɕiɛ^{21}　tɣ^{13} iəu^{21}　tsɿ^{34}　liɛ

stone　top　　　write　ASP　character(s)　interjection

　　　　"There are characters written on the stone!"

(5)

我 跟 她 的 伢 儿 **把** 哒　一个苹果　　　　(ACC, PER)

o^{21} kən^{45} t^{h}A^{45} ti　A^{13} ɯ^{13} **pA^{21}** -tA^{21}　i^{45} ko^{34} p^{h}in^{13} ko^{21}

　　I to　her GEN child　give　ASP one CL　apple

"I gave her child an apple."

(6)

输 哒 五 百　块　　　　　　　(ACH, PER)

su^{45} -tA^{21} vu^{21} pɣ^{24}　k^{h}uai^{21}

lose ASP　five hundred　CL

"(He) lost five hundred Yuan in gambling."

　　　The questionnaire mainly involves an acceptability task. It is arranged as follows. On the top of the first page is the background information (cf. § 3.1.2). Questions and

instructions appeared below the background section: "Have you ever encountered these utterances in Gong'an dialect? Do they sound like authentic Gong'an Dialect? If no, please tell me how you would say it."

The questionnaire is delivered orally instead of in written form due to the following reasons. Firstly, a written questionnaire is not feasible for informants who have received very little education or are completely illiterate; secondly, a written questionnaire may restrict the possible variations as the informants would concentrate on the sentences in the questionnaire only; and thirdly the written questionnaire is not reliable as it is hard for the informants to avoid interference from MC as there are no written forms for Gong'an Dialect. The advantages and disadvantages of a written questionnaire and oral elicitation is comparatively discussed in detail in the SAND project (cf. Cornips & Poletto 2005). The written questionnaire therefore is used by the researcher only to aid the researcher's memory in the fieldwork during oral elicitation.

3.1.4 The choice of the survey site

Based on the data collected from field in 1983, Yang (1992: 32) divides GAD into four parts, namely, the Northern, the Western, the Central and the Southern part according to the tone features. The GAD spoken in the current county town Douhudi 斗湖堤 area (since 1955), which is in the northeast of the central area, is considered by the natives as the "standard Gong'an accent". Taking into consideration that the tone values of GAD varies from district to district and that both the old county town Nanping 南平 and the current county town are situated in the central area according to Yang's classification, this project puts the focus on the central area. Nonetheless, the examples extracted from Gong'an ballads and folk stories do not correspond to the virtual boundaries as they are popular within the whole Gong'an area and beyond.

3.1.5 Group survey

The most natural vernacular data is probably the one obtained in group sessions, where spontaneous speech is evoked and controlled by similar factors which operate in daily conversations (Labov 1969: 64). The perceptible differences between a solo interview and group survey during a pilot interview conducted by the researcher echo

Labov's findings.[10] The group survey, therefore, is chosen to elicit both judgment data and spontaneous speech for this project. Labov goes on to claim that the group must be selected by the informants themselves, which is, however, rather indeterminate if the age and other social variables of the informants are to be taken into account. Therefore the groups are gathered by the researcher provided that the informants in the same group are familiar with each other and are of similar age, education and working background. In addition to that, the number of informants at any one time is also controlled. More precisely, two or (not more than) three informants participate in each judgment elicitation task, and two or (not more than) four in each spontaneous speech task.

3.2 Oral elicitation

Out of the 536 sentences in the questionnaire, 50 are vernacular fillers, which are randomly scattered in the questionnaire in order to distract the informant's attention from the aim of the interview and to make the dialogue as natural as possible. The questionnaire was further split up into smaller parts according to the real situation in the field.

The acceptability judgment operated as follows. First, the researcher asked one informant at a time for their background information. After finished the background section, the researcher indirectly asked the informants to express their opinion after the researcher uttered the sentence. The informants were instructed to repeat the utterance if she/he accepted it. If the informants did not agree with the utterance, they were encouraged to provide their own versions and any alternatives they were aware of. When the informants were indeterminate about the utterance, the researcher always provided a context and asked what he/she would say in such a situation. The interviews were recorded by a digital voice recorder and the differences were also immediately written down by the researcher.

---

[10] In a pilot one-to-one interview, the informant was not communicative enough to generate valid data. The situation, however, changed dramatically after the researcher invited another native speaker to join in the interview. The participation of another native speaker apparently encouraged the informant to be more active in the recording process.

3.3 Spontaneous speech

The spontaneous speech recording always started with casual talk regarding each informant's background. The informants were then encouraged to talk about anything they liked. If no proper topic was proposed by the informants, the researcher suggested topics such as "The natural disasters you went through", or "How did you celebrate the previous Chinese New Year and how are you going to prepare for the next one?" During the conversation, the researcher tried to appear as another participant rather than an observer. When the speech was not producing much of the data required, the researcher asked focus questions and successfully shifted the topic to a more productive one. Finally, eight episodes of spontaneous speeches were recorded which totaled up to 16 hours of recordings.

3.4 Data from other sources

In order to construct a corpus for this project, more data were collected from a selection of sources which reflect GAD rather faithfully in some respects. They are a book entitled *A Collection of Gong'an Ballads*《公安歌谣集》(hereafter: CGB),[11] a pamphlet entitled *An Introduction to Shuoguzi*《说鼓子简介》(hereafter: IS),[12] and stories and ballads written down in GAD in online forums by native speakers. Two local forums were also chosen to collect the relevant posts.

One forum is on "The Web of Life in Gong'an" at: http://www.gashw.com/bbs/dispbbs.asp?boardID=2&ID=67&page=1, from which the posts (dated from 24 July 2007 to 9 April 2010) dealing with Gong'an ballads were used.

The other is at: http://0716-gongan.5d6d.com/thread-235-1-1.html, from which the posts containing stories in GAD dated on 15 May 2008 and 28 May 2008 were analyzed.

---

[11] *A Collection of Gong'an Ballads*《公安歌谣集》contains popular ballads of various topics collected in the 1980s in Gong'an area and was published in 1990, by Yuan Bohua (editor in chief), Zhongguo Minjian Wenyi Chubanshe.

[12] *so⁴⁵ku²¹tsi* 说鼓子 is a folk art of storytelling and singing in dialect to drum beats and other musical instruments. It used to be popular in Gong'an and the neighboring areas such as Songzi 松滋, ShiShou 石首. The pamphlet was edited by Mao Zugui (one of the informants of this project) and printed by the House of Culture of Gong'an County in 1982.

4. Some preliminary findings of the fieldwork

4.1 On informants and group survey

In view of the data collection procedure and the results, it can be verifiably concluded that in choosing informants (at least in this project), variables such as education, age, gender and occupation do not play an equal role. The findings suggest, for example, on the condition that similarly little education was received by the informants, there are very few age-related variations among the informants involved. The observable difference between the informants with senior secondary education (educated informants) and the informants with less or almost no education lies primarily in the vocabularies they possess. The educated informants' dialectal grammar with regard to aspect marking, however, is surprisingly intact. Taking into account the fact that education is directly associated with age in China, it can be concluded that it is less risky to use informants born before the 1960s.

In terms of gender of the informants, the preferred male informants in the literature (e.g., Chambers & Trudgill 1998: 29) did not differentiate substantially from female informants in the fieldwork. It might, however, be attributed to the nature of the elicitation tasks and the native identity and gender of the investigator herself.

Occupation does play an important role; for instance, farmers' dialects have usually received less influence from MC as they are relatively less educated and less mobile. People from a working background related to folk arts and literature can also be very valuable in providing information about a decidedly under researched dialect. For example, the two folk artisans' scanty formal education and more than half a century of working experience dealing with Gong'an folk arts of talking and singing have turned them into ideal informants for dialect investigation.

The key advantage of grouping informants together lies in the natural and comfortable atmosphere it creates, which is essential for the informants to generate the most natural syntactic data. With two or more people in the group, the conversation was more like a daily gathering with neighbors or friends. During judgment elicitation, the two informants tended to compete with each other and tried their best to provide any possible variations in different contexts. When disagreements

between the two informants arose, they automatically analyzed each other's answers and pointed out the sources of regional variations.

Nonetheless, there are at least two prerequisites for employing the group survey method. First, if not a local dialect speaker him/herself, the researcher must have in-depth knowledge of the language or dialect so that the researcher can diagnose "overcorrection" and "accommodation". Secondly, the audio recording equipment must be very reliable to facilitate data transcription.

4.2 On judgment data and corpus

The oral and written data collected from field and other sources are further classified into two categories. The first such category contains only the experimental data elicited from acceptability judgment. The judgment data shows that the disagreements with the sentences in the questionnaire are predominantly on lexical items, whereas the aspectual constituents under investigation have received no objection but a few possible variations. That is to say, the proposed situation types (cf. Section 2.3) are readily recognized and the interactions between the two components of aspectual systems do exist in GAD. The second category consists of the data from narrative recordings and other sources (see § 3.5). One of the main achievements gained so far by using spontaneous narrative recordings is that it has captured the uses of two other perfective aspect markers in GAD, i.e. *-A^{21}* and *-lA^{21}*, which the oral elicitation carried out in the form of questionnaire failed to reveal due to its subtle phonetic features and its functional similarities with the more prominent aspect marker *-tA^{21}*.

The data in the ballad book and the folk art pamphlet is not ideal in terms of consistency in using aspect markers. For example, both MC aspect markers and GAD aspect markers are used in the same ballad or the same lines of one *so^{45}ku^{21}tsi* 说鼓子. This phenomenon can be in part accounted for by the constraint of the rhythmic patterns of the ballads and the uncertainty of the cognates of some GAD aspect markers. However, the co-occurrence of the aspect marks may also offer insights into the question of whether there existed a difference between literary and colloquial readings in GAD. In addition to that, there are virtually no other books or printed materials comprehensively documenting the local ballads and folk arts, it is therefore

necessary to sort out the usage of aspect markers in the above materials in order to provide more data to the study and guidance to fieldwork (e.g., ask informants to sing ballads). The seventh section of the book is not included in the frequency count as the 28 children's ballads are also available on the forums surveyed.

To sum up, there are about 3009 aspectual constructions in GAD from the data collection (see Table 2).[13] All the data are annotated and will be transcribed into IPA during analysis.

Aspect marker	Frequency					Total
	JE	Corpus				3009
		SS	CGB	IS	WEB	
$A^{21}(IA^{21})$	31	527	0	1	0	559
TA^{21}	255	1035	17	12	208	1527
ko^{34}	57	64	5	1	4	131
$tsai^{34}$ (tai^{34})	40	53	14	1	3	111
$təu^{21}$	50	32	2	0	8	92
$təu^{21}tsai^{34}$($təu^{21}tɣ$)	9	9	0	0	2	20
$tɕʰi^{21}$	64	141	47	4	4	260
$tɕʰi^{21}tsai^{34}$($tɕʰi^{21}tɣ$)	22	11	0	0	4	37
$tɕʰi^{21}lai^{13}$	21	17	17	2	4	61
$ɕiA^{34}kʰɯ^{24}$	7	2	0	0	0	9
$tsu^{21}(tsʰu^{21})$	12	8	0	0	0	20
tso	0	5	0	0	0	5
$tɣ^{24}iəu^{21}$	45	12	0	0	0	57
$tɣ^{13}$	7	5	0	0	2	14
ti^{34}	66	18	0	0	1	85
to	14	6	0	0	1	21

Table 2. The frequency of the aspect markers in the judgement data and corpus

5. Conclusion

As part of the larger project on the aspectual system in GAD, this paper focuses on the data collection methods, the design of the questionnaire and some findings of the fieldwork. Since a written questionnaire is not workable in collecting syntactic data

---

[13] The aspect marker *-tA²¹* and the sentence-final particle *tA²¹* syntactically overlap in some examples, both of which are subsumed under TA^{21} in this table. Please also note that some of the RVCs (e.g., *xau²¹*, *van¹³*, *pʰo³⁴*) denoting aspectual features in GAD but not yet fully grammaticalized are not listed in this table.

from GAD, the questionnaire has to be delivered orally. This could also be the case for collecting syntactic data of other Chinese dialects as a large number of dialect words have no corresponding written forms in Chinese characters. The findings of the data collection suggest that oral elicitation works better when informants are gathered in small groups. It is also evident that solely depending on one elicitation technique runs the risk of ignoring some important phenomena. In comparison, a multi-method approach is more reliable, which promises to provide more insights into the nature of the dialect under investigation and more general theoretical issues.

References

ANDERWALD, Lieselotte & Benedikt SZMRECSANYI (2009) "Corpus linguistics and dialectology", in A. Lüdeling & M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, Berlin, New York: Mouton de Gruyter, 1126-1139.

BARBIERS, Sjef (2009) "Locus and limits of syntactic microvariation", *Lingua*, 119 (11), 1607-1623.

BARD, Ellen G., Dan ROBERTSON & Antonella SORACE (1996) "Magnitude Estimation of Linguistic Acceptability", *Language* 72 (1), 32-68.

BUCHELI, Claudia & Elvira GLASER (2002) "The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems", retrieved 28th January, 2010, from http://www.meertens.knaw.nl/books/synmic.

CAO, Ziyun (ed.) (2008) *Linguistic Atlas of Chinese Dialects*, Beijing: the Commercial Press.

CEN, Qixiang (2008) *Yuyanxueshi gaiyao*, Beijing: Shijie tushu chuban gongsi.

ČERMÁK, František (2002) "Today's corpus linguistics: Some open questions", *International Journal of Corpus Linguistics*, *7* (2), 265-282.

CHAMBERS, Jack K. & Peter TRUDGILL (1998) *Dialectology* (2nd edition)*,* Cambridge: Cambridge University Press.

CHAO, Yuen-Ren, Ting SHENG-SHU, Yang SHI-FENG, Wu TSUNG-CHI & Tung T'UNG-HO (1972[1948]) *Hubei fangyan diaocha baogao* (Report of Hubei Dialectal Survey), Taibei: Tailian guofeng chubanshe.

CHAPPELL, Hillary (1992) "Towards a typology of aspect in Sinitic languages", *Zhongguo Jingnei Yuyan ji Yuyanxue: Hanyu Fangyan [Chinese Languages and Linguistics: Chinese dialects]*, 1, Taipei: Academia Sinica, 67-106.

CHAPPELL, Hillary (2001) "A typology of Evidential Markers in Sinitic Languages", in H. Chappell (ed.), *Sinitic grammar: synchronic and diachronic perspectives*, Oxford: Oxford University Press, 56-84.

CORNIPS, Leonie & Willy JONGENBURGER (2001) "Elicitation techniques in a Dutch syntactic dialect atlas project", *Linguistics in the Netherlands*, 18, 53-63.

CORNIPS, Leonie & Cecilia POLETTO (2005) "On standardising syntactic elicitation techniques (part 1)", *Lingua*, 115 (7), 939-957.

COWART, Wayne (1997) *Experimental syntax: Applying objective methods to sentence judgments*, Thousand Oaks: Sage Publications.

CROWLEY, Terry (2007) *Field linguistics: a beginner's guide* (edited and prepared for publication by Nick Thieberger), Oxford: Oxford University Press.

GREENBAUM, Sidney (1973) "Informant elicitation of data on syntactic variation", *Lingua*, 31 (2-3), 201-212.

GILQUIN, Gaetanelle & Stefan Th. GRIES (2009) "Corpora and Experimental Methods: A State-of-the-Art Review", *Corpus Linguistics and Linguistic Theory*, 5 (1), 1-26.

HARDIE, Andrew & Tony MCENERY (2003) "The Were -Subjunctive in British Rural Dialects: Marrying Corpus and Questionnaire Data", *Computers and the Humanities,* 37 (2), 205-228.

HARRIS, Randy A. (1993) *The Linguistics Wars*, Oxford: Oxford University Press.

HOLLMANN, Willem & Anna SIEWIERKSA (2006) "Corpora and (the need for) other methods in a study of Lancashire dialect", *Zeitschrift für Anglistik und Amerikanistik*, 54: 203-16.

JIANG, Zongxia (2008) "Hanyu fangyan fiaocha yanjiu jinbainian fazhan zhilu", *Journal of Yunan Normal University (Philosophy and Social Sciences)*, 40 (5), 46-55.

LABOV, William (1969) "The study of nonstandard English", Washington D.C.: National Council of Teachers of English, by special arrangement with the Center for Applied Linguistics.

LABOV, William (1972) *Language in the Inner City: Studies in the Black English Vernacular*, Philadelphia: University of Pennsylvania Press.

LEE, F. K. Jackie (2001) "Functions of *need* in Australian English and Hong Kong English", *World Englishes*, 20 (2), 133-143.

LU, Jianming (2004) "Remarks on the Investigation of Chinese Dialectal Grammar", *Yuyan Kexue*, 3 (2), 92-98.

MCENERY, Tony, Richard XIAO & Yukio TONO (2005) *Corpus-based Language Studies: An advanced resource book*, London: Routledge.

MEI, Tsu-Lin (2000[1988]) "The origin and function of Mandarin *zhe* 着, Wu *tsi* 仔, Southern Min *ti* 著 and their cognates in Chinese dialects", *Mei Zulin Yuyanxue lunwenji*, Beijing: Shangwu yinshuguan, 155-187.

Nᴇʀʙᴏɴɴᴇ, John (2009) "The forests behind the trees", *Lingua*, 119 (11), 1581-1588.

Nᴇᴡᴍᴀɴ, Paul & Martha Rᴀᴛʟɪғғ (eds.) (2001) *Linguistic Fieldwork*, Cambridge: Cambridge University Press.

Sᴀᴍᴀʀɪɴ, William J. (1967) *Field linguistics: a guide to linguistic fieldwork*, New York: Holt, Rinehart and Winston.

SᴄʜÜᴛᴢᴇ, Carson (1996) *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*, Chicago: University of Chicago Press,.

Sᴍɪᴛʜ, Carlota S. (1991) *The Parameter of Aspect* (1ˢᵗ ed.), Dordrecht: Kluwer.

Sᴍɪᴛʜ, Carlota S. (1997) *The Parameter of Aspect* (2ⁿᵈ ed.), Dordrecht: Kluwer.

Sᴜɴ, Caofen (1998) "Aspectual Categories That Overlap: A Historical and Dialectal Perspective of the Chinese *zhe*", *Journal of East Asian Linguistics*, 7 (2), 153-174.

Sᴢᴍʀᴇᴄsᴀɴʏɪ, Benedikt & Bernd Kᴏʀᴛᴍᴀɴɴ (2009) "The morphosyntax of varieties of English worldwide: A quantitative perspective", *Lingua*, 119 (11), 1643-1663.

Tʜʀáɪɴssᴏɴ, Höskuldur, Ásgrímur Aɴɢᴀɴᴛýssᴏɴ, Ásta Sᴠᴀᴠᴀʀsᴅóᴛᴛɪʀ, Thórhallur Eʏᴛʜóʀssᴏɴ & Jóhannes G. Jóɴssᴏɴ (2007) "The Icelandic (Pilot) Project in ScanDiaSyn", *Nordlyd,* 34, 87-124.

Wᴀɴɢ, Hongjia & Guosheng Wᴀɴɢ (2011) "Phonetic features of dialects on the Jianghan Plain", *Fang Yan*, 3, 247-251

Wᴀɴɢ, Li (2007) *Zhong Guo Yuyanxueshi*, Shanghai: Fudan University Press.

Wᴀɴɢ, William. S-Y. (1978) "The Three Scales of Diachrony", in B. B. Kachru, (ed.), *Linguistics in the Seventies: Directions and Prospects*, Department of Linguistics, University of Illinois, 63-76.

Wᴜ, Yunji (1999) *The development of aspectual systems in the Chinese-Xiang dialects*, Paris: Ecole des hautes études en sciences sociales.

Wᴜʀᴍ, Stephen A., B. T'sᴏᴜ, D. Bʀᴀᴅʟᴇʏ; (China) Li Rᴏɴɢ, Xiong Zʜᴇɴɢʜᴜɪ, Zhang Zʜᴇɴxɪɴɢ (eds.) (1987) *Language atlas of China*, Hong Kong: Longman.

Xᴜ, Liejiong (1998) "Fei duibixing de fangyanyufa yanjiu", *Fang Yan*, 3, 188-191.

Yᴀɴɢ, Lan (1992) *Gong'an fangyan yanjiu*, Shanghai: Hanyu dacidian chubanshe.

Yᴏᴜ, Rujie (2005) "Hanyu fangyanxue de xianzhuang he yuanjing", *Journal of Jinan University (Philosophy and Social Sciences)*, (118), 108-142.

Yᴜᴇ-Hᴀsʜɪᴍᴏᴛᴏ, Anne (1991) "Stratification in comparative dialectal grammar: A case in Southern Min", *Journal of Chinese Linguistics*, 19 (2), 172-201.

Yᴜᴇ-Hᴀsʜɪᴍᴏᴛᴏ, Anne (1993) *Comparative Chinese Dialectal Grammar. Handbook for Investigators* (Collection des Cahiers de Linguistique d'Asie Orientale 1), Paris: Ecole des Hautes Etudes en Sciences Sociales, Centre de Recherches Linguistiques sur l'Asie Orientale.

ZHAN, Bohui (2009) "Hanyu fangyan yanjiu 30 nian", *Journal of Yunnan Normal University (Philosophy and Social Sciences)*, 41 (2), 38-45.

ZHANG, Min (2000) "Syntactic Change in Southerastern Mandarin: How Does Geographical Distraibution Reveal a History of Diffusion?", in B. Ding, A. Yu (eds.), *Yu yan bian hua yu Han yu fang yan: Li Fang-gui xian sheng ji nian lun wen ji*, Taibei: Zhong yang yan jiu yuan yu yan xue yan jiu suo chou bei chu, Seattle: University of Washington, 197-242.

ZHANG, Yizhou, Qingyuan ZHANG & Yingshu DENG (2001) *Chengdu fang yan yu fa yan jiu*, Chengdu: Ba Shu shu she.

ZHU, Guanming (2005) "Hubei Gong'an fangyan de jige yufaxianxiang", *Fangyan*, 3, 253-257.