Eliciting Phonetic Data and Metadata during Linguistic Field Research

Katie- Louise Coxall

U1372038

MA by Research (Music, Humanities and Media)

University of Huddersfield

January 2018

Copyright Statement

i. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Abstract

Having identified a gap for research which combines the perspectives of experienced field researchers in relation to the methods used when collecting data on the phonetic and phonological aspects of language, this research will collate and explore the results of a survey of 16 experienced field linguists who have shared insight into tried and tested methods of data collection.

Through an online survey, respondents answered questions about their experiences of data collection with a particular focus on areas which were identified from the literature as having potential to cause difficulties. The areas of difficulty addressed were: the target language and informant(s), minimal pairs, affricates, connected speech, allophones, variation and metadata and, finally, recording methods. Participants were asked about the methods used to elicit data and the problems that they encountered throughout their research, followed by ways in which these problems were overcome.

Through analysis of the results and reference to the existing literature, a number of areas of methodological consensus have been identified. These areas of consensus, along with less well-known techniques, have contributed to a discussion of best practise in the collection of phonetic and phonological data. Using the results of the survey, each area has been explored and a set of recommendations have been proposed to guide future linguistic documentation projects. The suggested recommendations are also intended to serve as a prompt with additional techniques for use when problems are encountered by researchers during their research.

Whilst useful for any phonetic and phonological linguistic documentation project, the research was motivated by the diminishing rates of linguistic diversity throughout the world. By forming recommendations based on the opinions and experiences of 16 field researchers, it is hoped that the advice generated will be of assistance to future researchers in their efforts to document endangered language varieties.

Table of Contents

List of Tables and Figures

Eliciting Phonetic and Phonological Data during Field Research

1. Introduction

Field work is a vital part of linguistic research and enables researchers to gain a valuable insight into language varieties and their use. Language endangerment in particular is an issue that has been gathering an increasing amount of attention from field linguists as a result of the rising numbers of vulnerable languages. Whilst assessments of the number of languages at risk vary, it has been suggested that at least 3,500 of the estimated 7,000 languages spoken worldwide could become extinct within a few more generations (Austin & Sallabank, 2011, p.1). With the decline of linguistic diversity, it has been argued, comes a reduction in cultural diversity and the loss of traditional knowledge systems (Hirsh, 2013). This loss of identity can not only have a profound effect on those communities affected, but also on the linguistic discipline. As a consequence of the vast number of 'at risk' languages, linguists are faced with a large number of languages which, without documentation, may cease to exist. Therefore, documenting languages in an efficient and timely manner has become increasingly important in the preservation of linguistic diversity. There are currently a number of linguists who conduct research in the areas of both language documentation and the endangerment of languages. Some of these linguists have produced literature which aims to guide researchers who are embarking on language documentation projects. Amongst these works, the following are considered to be some of the most influential sources of information in this linguistic field and include: Crowley (2007), Bowern (2008) and Chelliah & Reuse (2011).

1.1 Research Motivation

The motivation of this thesis is to assist field linguists in overcoming some of the common problems that are encountered when undertaking linguistic fieldwork. Whilst such guides do already exist, these are written from the perspective of just one or two researchers. However, through the use of a survey, this research fills a gap in the literature by collating the perceptions and experience of a

total of 16 field researchers in relation to specific areas of difficulty in the collection of phonetic and phonological data during field research. The techniques and suggestions given by respondents have been analysed and synthesised so that well informed recommendations can be made. The resulting work offers solutions to common problems that are experienced whilst collecting data of this type. As a result of the guidance obtained from participants, it is intended that future field linguists can improve the efficiency of their research.

1.2 The Research Area

The research focuses specifically on the collection of phonetic and phonological data. This area was chosen as understanding the phonetic and phonological system is considered to be one of the initial priorities when beginning to work on a language (Bowern, 2008, p.34). This research fills a gap in the literature for work which collates the perspectives of a number of field linguists in relation to the challenges faced when collecting phonetic and phonological data during field research. Whilst a number of guides to linguistic fieldwork exist, these are based primarily upon the opinions of one field linguist. However, by combining the experience of a number of field linguists, who all provide answers to the same survey questions, the conclusions drawn from the research become more objective and relate to a broader range of experiences and language types.

1.3 Areas of Difficulty in the Collection of Phonetic and Phonological Data

A number of areas of difficulty in the collection of phonetic and phonological data have been identified and form the basis of the research: field research, informants and elicitation methods; minimal pairs; affricates; connected speech; allophones; variation and metadata and, finally, recording methods.

1.4 Structure

In Chapter (2), the Literature Review, I present an overview of the existing works which are most significant to the research. This places the study into context within the linguistic field and helps to clarify how this research contributes to the existing body of work. This is followed by Chapter (3),

the Methodology, in which the research objectives are outlined along with the process that was followed throughout the planning and implementation of the study. In Chapter (4), the results of the research are presented. This is followed by Chapter (5), the Discussion, which examines the results in further detail and provides recommendations for future research based on the experiences of survey participants. Following the discussion a number of recommendations have been proposed as a guideline for future research. The thesis concludes in Chapter (6), the Conclusion, where the findings of the research have been summarised and explained in the context of how they have met with the research motivations.

2. Literature Review

This chapter provides an overview of the existing literature that is most significant to the preservation and documentation of languages through field linguistic research. In order to put the research into context, and justify the importance of both this thesis and the preservation of linguistic diversity as a whole, the chapter begins with a discussion of the perspectives through which languages can be viewed and defined in relation to one another. This is followed with a discussion of the reasons for language endangerment and the importance of language documentation: both to speech communities and the linguistic discipline itself. Next, we address the assessment of language vitality and the impact of these observations on the prioritisation of languages for documentation. Finally, the more specific areas of phonetics and phonology are defined and explored, with an emphasis on the difficulties that can be faced when collecting linguistic data of this type.

2.1 The relationship between languages

Whilst it is often taken for granted that the term 'language' refers to a solid and established linguistic variety, there is some controversy as to how this is defined. According to Weber & Horner (2012), in general there are considered to be two 'models' of language: the 'expert' model and the 'popular' model. When viewed from the perspective of the popular model, languages are differentiated from dialects, with languages taking a higher status in a perceived hierarchy between

the two. Within this view, the standard variety of a language tends to be consistent with the version that appears in grammars and dictionaries and is therefore afforded the most prestige within society. On the other hand, proponents of the expert model recognise no linguistic or hierarchical differences in status between languages and dialects and, as a result, consider all varieties to be equally as important (p.27). These perspectives are important in understanding attitudes towards varieties of language and can, in some cases, help to explain why the use of certain, often less prestigious, varieties decline.

2.2 Why do Languages Become Endangered?

Language endangerment is a complex issue which cannot be adequately explained by any one cause, but by a number of contributing factors. In general, the factors which lead to language endangerment fall into one of two categories: causes which encourage language shift and those which cause abrupt language loss. In order to help to explain incidences of language loss, Campbell & Muntzel (1989, p.183-186) suggest a taxonomy which categorises the reasons for language loss alongside providing an indication of the speed at which the loss occurs in each case. These categories comprise: sudden attrition, radical attrition, gradual attrition and bottom-to-top attrition.

2.2.1 Campbell & Muntzel's (1989) Taxonomy

In cases of sudden attrition, language loss occurs abruptly due to the widespread loss of a speaker population as a result of an outbreak of disease, war or natural catastrophe. An example of sudden attrition can be seen during the 16th century in Hispanic South America where, according to Adelaar (2007), the spread of epidemic diseases were introduced by Spanish conquerors. The resulting disappearance of many coastal populations led to the loss of a number of languages and ethnically diverse communities. One example was the Quingnam language whose survivors eventually adopted the Spanish language of the new settlers (p.9).

Radical attrition occurs as a means of self-defence due to political issues such as repression of a population or genocide. Speakers therefore switch to what is seen as a more desirable variety in

order to distance themselves from their ethnic group and protect themselves from persecution. One example of radical attrition is in the case of Maori in New Zealand which, according to Weber & Horner (2012), became severely endangered following an extended period of colonial oppression and cultural integration policies in education. However, following a successful revitalisation, the Maori language now enjoys its status as official state language of Aotearoa (p.54).

As its name suggests, gradual attrition occurs more slowly and, as a result, is less likely to cause concern in its initial stages. This kind of shift occurs when a population begins to make the transition from an indigenous language to a more dominant language of wider communication. Due to the gradual nature of this change, it could be argued that by the time the language no longer exists, the community hold less of an attachment to the language and are often enjoying the economic benefits of switching to a more dominant variety. Therefore, instances of gradual attrition do not appear to be of significant focus in the literature and, as such, is an area that could benefit from further research.

Bottom-to-top attrition denotes a situation in which the language is no longer used in most settings such as the family home, but is retained in religious practices due to its significance to the history of the religion or ritual. As Grenoble and Whaley (2006) point out, religion is often one of the last domains in which a vulnerable language is used (p.41). Though it has now experienced a revitalisation, one example of bottom-to-top attrition would be in the case of Hebrew which was once critically endangered and was mostly confined to use in a religious capacity (Weber & Horner, 2012, p.54).

2.2.1.1 Evaluation of Campbell & Muntzel's taxonomy

Campbell and Muntzel's (1989) taxonomy is of great help in explaining the reasons why languages become endangered and works of this type not only help with the understanding of language endangerment once it has occurred, but also help to identify languages which may be at risk of endangerment by highlighting the circumstances which have, historically, caused a decline in language vitality. However, due to the time that has passed since this taxonomy was proposed, this

work would benefit from modernisation in order to account for advances in technology which have undoubtedly had an effect on language use in recent times.

2.2.2 Austin & Sallabank's (2011) Taxonomy

Offering an alternative, and more recent, explanation of the factors that led to language endangerment, Austin & Sallabank (2011) suggest that the causes can be split into four main factors: natural catastrophes, famine and disease, war and genocide and, lastly, overt repression and cultural/ political/ economic dominance (p.5). These factors can overlap or occur together and the fourth category can be further divided into five additional factors: economic, cultural dominance, political, historic and attitudinal (p.5). Each of these factors can eradicate large sections of a language community and leave an inadequate number of speakers to produce new generations of speakers large enough to keep a language variety alive. Issues of cultural, political and economic dominance can cause communities to become marginalised and forced to adopt a more prestigious variety of language in order to access essentials such as education and jobs. Whilst these reasons help to account for some of the causes of endangered languages, perhaps due to their age, it could be argued that they are over-simplified and discount the importance of advances in technology which have revolutionised the way that many modern societies communicate.

2.2.2.1 Evaluation of Austin & Sallabank's (2011) Taxonomy

Whilst Austin and Sallabank's reasoning suggests that languages become endangered through factors which are out of a community's control, Maffi (2002) acknowledges that whilst this is sometimes the case, languages also cease to exist when speakers voluntarily shift to another, often more prestigious, variety (p.385). Similarly, the loss of a language is usually described in negative terms, yet for communities themselves the shift to a more common variety often brings with it increased employment and trade opportunities. As a result, communities can view the shift to a prestigious language as a positive due to the economic benefits that a new variety brings.

2.2.3 Factors Which Lead to Language Shift

Language shift is the primary cause of language loss and occurs when speakers switch to a variety of language other than their native tongue (Grenoble, 2011, p.27). Although the process of language shift can sometimes happen quickly, the process tends to be gradual. In the following section I examine some of the factors that encourage the process of language shift. These factors include globalisation, which incorporates advances in technology and an increase in travel, along with linguistic imperialism.

2.2.3.1 Globalisation

In recent times, globalisation can be said to account for many of the shifts in language behaviour due to an increase in contact between different language communities. While advances in transportation have made international travel for leisure, migration and trade an option for many; technologies such as the internet have further reduced the boundaries that once existed between language communities. As a result, exposure to different languages and cultures has, in many cases, become commonplace. Consequently, some minority language communities are no longer teaching children the indigenous variety as a first language resulting in language endangerment (Austin & Sallabank, 2011, p1). Whilst the effect of globalisation on language endangerment has become clear in recent years, the literature on the topic is not extensive and requires further investigation.

2.2.3.1.1 Advances in technology

In terms of technological advances, the internet has made access to information easier than ever. This means that communities which previously had little contact with outsiders are now able to interact with online content which exposes them to different ways of life and varieties of language. With endangered languages not widely used in online content, users may feel the need to adapt to a different variety of language in order to become part of the digital community and access the content on offer. Therefore, failure of a language to venture into modern technological domains can increase the threat to its stability (Brezinger, 2007, p.4) with young people in particular encouraged to transition to other languages in order to gain greater accessibility to modern life, putting the indigenous language at risk.

On the other hand, advances in technology can be a positive for minority languages whereby language users adapt the variety to modern life which, as a result, affords the language a higher level of prestige within rural communities (Brezinger, 2007, p.152). This increase in prestige is likely to invoke pride in the language and, therefore, offer the language a level of protection from endangerment. Cocq (2015) suggests that social media opens up new sources of knowledge and modes of communication for indigenous communities, allowing them to express their culture and enabling both revitalisation of language and empowerment of indigenous communities (p.274). Adaptation to modern life is also likely to mean that the language remains relevant and therefore encourages the younger generation to continue using the language to fulfil both traditional and modern roles.

2.2.3.1.2 Increased travel

With international travel now commonplace, tourism has extended beyond the traditional holiday destinations to incorporate options which take travellers to more remote communities thus exposing these communities to other languages. As a result, in some cases communities may adapt to the language of tourists in order to facilitate further tourism and provide a boost to the economy through *linguistic capital* (Bourdieu, 1992) whilst retaining some elements of the traditional language for the purposes of encouraging tourism through *cultural capital* (Bourdieu, 1992).

2.2.3.2 Linguistic imperialism

With power usually in the hands of speakers of the majority languages, minority language speakers often find themselves under pressure to shift to the more influential language variety (Austin & Sallabank, 2011, p.1). Governmental language planning policies often further encourage this transition through the implementation of the dominant language in education, politics and other official capacities thus making knowledge of the language advantageous for success in both schooling and careers. These more prestigious, and often more widely spoken, languages are considered to have *linguistic* and *economic capital* (Bourdieu, 1992). Over generations, this results in a decreased fluency

in the minority language and therefore, with fluency only existing in the older generation, the language gradually becomes endangered.

2.2.3.2.1 Linguistic imperialism in Singapore

Whilst the language situation in Singapore has not led to the endangerment of a language, its complexity showcases a situation in which the language planning policies of a country can be seen to undermine some varieties of language in favour of others and demonstrates ways in which language use can decline due to political influence.

Due to the complex heritage of the population of Singapore, the country presents a varied linguistic landscape with four co-official languages: Mandarin, Malay, English and Tamil assigned to the country. With the population consisting primarily of Indian, Malay and Chinese, the co-official languages are designed to maintain the heritage of the different ethnic groups with the allocation of a mother tongue for each, whilst encouraging a united identity for all ethnicities through the use of English. However, despite the co-official status of the official languages, in practise the languages are not given equal status. One example of which is the educational policy of 'English plus mother tongue bilingualism', where students are taught English as a first language, and their government assigned mother tongue as a second language (Weber & Horner, 2013, p.72; Ng, 2014, p.364-365). This demonstrates linguistic imperialism in action and shows that English is seen to have both *linguistic* and *economic capital* (Bourdieu, 1992) to the detriment of less prestigious varieties of language.

2.3 The classification of degrees of language endangerment

In order to help prioritise their documentation efforts, linguists must be able to gauge the vulnerability of languages. However, according to Brezinger (2007), this process is challenging due to the complex nature of speech communities and the wide variety of patterns of language use that communities exhibit. Furthermore, it is not possible to rely on speaker numbers to accurately assess the vitality of a language, as it can be the case that a language which is spoken by thousands of people is endangered; whereas a language with a smaller number of speakers is stable (p.x).

In order to assist in the process of assessing language vitality, Krauss (2007) suggests a framework which classifies languages on a spectrum from safe to extinct.

'safe'			a+	
e n d a n g e r e d		stable	a−	all speak, children & up
	i n d e c l i n e	instable; eroded	a	some children speak; all children speak in some places
		definitively endangered	b	spoken only by parental generation and up
		severely endangered	c	spoken only by grandparental generation and up
		critically endangered	d	spoken only by very few, of great-grandparental generation
extinct			e	no speakers

Figure (1): Krauss (2007, p.1) Schema for classification of languages

Krauss's schema is valuable as it provides a means of classifying and comparing differing language situations and, subsequently, prioritising their need for documentation. However, it could be argued that the categories are not detailed enough and lack consistency. For example, for a language to be classed as 'definitively endangered', the only stipulation is that the language is 'spoken only by the parental generation and up'. Therefore, a language where only 20% of the adult population speak the language in question and a language where 99% of the adult population speak the language would fall into the same category, despite the levels of vitality being very different.

However, several other schemas for identifying endangered languages have also been suggested. UNESCO devised a language vitality index which uses nine factors of equal importance (Lee & Van Way, 2016, p.273-274).

Factor number	Factor
1	Intergenerational language transmission
2	Absolute number of speakers
3	Proportion of speakers within the total population
4	Trends in existing language domains
5	Response to new domains and media
6	Materials for language education and literacy
7	Governmental and institutional language attitudes and policies, including official status and use
8	Community members' attitudes toward their own language
9	Amount and quality of documentation

Figure (2): UNESCO's nine factors. Adapted from Lee & Van Way (2016)

Each of the factors are rated from 1-5, with 5 depicting the most favourable situation. These factors are then used to assess the requirements of a language and identify areas where support would be beneficial (Lee & Van Way, 2016, p.273). One of the benefits of this scale is that it acts as a scoring system which enables quick comparison between a large number of languages in order to understand which varieties are the most at risk.

Still, despite the emergence of a number of scales and schemas for the classification of endangered languages, it is widely recognised that a single, standardised scale would be beneficial (Lee & Van Way, 2016, p.273). This would help to ensure that languages are being compared according to the same criteria and the most endangered languages receive priority in terms of documentation.

2.4 Field linguistics and the significance of language documentation

Field linguists are involved in the study and description of language varieties and, with linguists predicting that by the end of the 21st century only half of the world's estimated 6912 languages will survive (Harrison, 2008, p.3), preserving the world's linguistic diversity is vital: both to the wider discipline of linguistics and to language communities themselves.

2.4.1 Why is the documentation of language important?

Language documentation is important, not only for those language communities whose variety is at risk of extinction, but also for the discipline of linguistics as whole.

2.4.1.1 Importance for the linguistic discipline

Language data is vital for the linguistic discipline which relies on the documentation of languages to enrich understanding of the subject area. Creating written records of language data contributes not only to the understanding of one particular variety but, with every addition of language data to the catalogue, also strengthens the empirical foundations of linguistics and provides the opportunity for further study (Gippert, Himmelman & Mosel, 2006, p.1). The retention of language data through documentation also improves the ability of linguists to conceptualise the extent of what is possible in both human language and cognition (Lee & Van Way, 2016, p.272). The greater the inventory of language data, the more evidence is available for linguists to examine and use to reassess their findings, resulting in more accurate research and conclusions.

2.4.1.2 Importance to language communities

For language communities, particularly those in which a language variety is facing extinction, language documentation can be beneficial on many levels.

2.4.1.2.1 Relationship between language, thought, culture and identity

Evidence of a relationship between language and thought, and thus a community's way of life, adds further motivation for the documentation of vulnerable languages. Through Benjamin Whorf's (1956) study of the Hopi community, and his subsequent proposal of the Sapir-Whorf Hypothesis, important differences were drawn between the way that the Hopi community viewed and referred to time, space and matter and the way that 'Standard Average European' languages viewed and referred to them. These findings indicate that language and thought are significantly interlinked and thus the loss of a language could have a huge impact on a community's way of life. Furthermore, the way in which people think and refer to things is also a reflection of the cultural background of a person and suggests that the loss of language could also impact upon the vitality of an entire culture.

More recently, Hirsh (2013) has elaborated upon the research into the relationship between language, thought and culture and found that the attitude towards languages is changing, with many communities appreciating the cultural knowledge and belief systems that are embedded within their languages, resulting in an increased sense of pride and unity being attached to their use (p.11)

2.4.1.2.2 Increases the prestige of a language variety

In addition, the documentation of language can be seen to increase the status of a language variety (Chelliah & Reuse, 2011, p.83; Crystal, 2000, p.151). The formal process of documentation often leads to the formation of a dictionary, along with the implementation of grammatical rules. This process often increases pride amongst a language community and can encourage revitalisation.

2.4.1.2.3 Knowledge systems

In a similar way to the way in which language and thought are intertwined, knowledge systems and are also connected with a language. Hirsh (2013) suggests that these knowledge systems are valuable sources of information about the medicinal qualities of plants and the natural world (p.11). Crystal (2000) argues that the knowledge of indigenous people is unrivalled in the western world, with communities demonstrating a profound awareness of the ecology of their environment (p.46-47). The documentation of language therefore ensures the retention of knowledge which has been identified and passed down through generations of language speakers.

2.4.1.2.4 Preserves linguistic diversity

Furthermore, in a world where linguistic diversity is threatened, linguistic fieldwork encourages the preservation of diversity (Crowley, 2007, p.1), not only in the linguistic sense but, as we have seen, also diversity of beliefs, knowledge and ways of life.

2.5 Choosing a language to document

Whilst the preservation of linguistic diversity may be a factor in deciding which language to document, there are also a number of other considerations to be made. Some of these factors will be outlined below.

Firstly, the language capabilities of the researcher are likely to have a significant influence. Crowley (2007) suggests that, in many cases, fieldwork will require an advanced knowledge of a language other than English (p.58). Therefore, it would be advisable for a researcher to seek out a field site where communication can initially take place in a lingua franca which is known to both researcher and informant.

Secondly, prior knowledge of the endangered language itself may play a role. For example, if a researcher has studied a particular variety of endangered language and already has an understanding of some of its elements prior to embarking on fieldwork, it would make sense to use this background knowledge to help with the documentation process (Crowley, 2007, p.59).

Whilst estimations of language vitality are likely to play a role when deciding which language variety to document, it is also important to take into consideration whether any documentation work has previously taken place on the language in question. According to Chelliah & Reuse (2011), grant applications can be declined on the basis that someone else has already documented elements of a language (p.84). Therefore, it is advisable to check if there are any documentation projects underway and, if there are, it may be wise to either choose another language or to suggest working together with the existing field linguists in order to prevent duplication of effort.

Finally, personal factors may play a role such as religion, sexual orientation and gender. It would be unwise to conduct fieldwork in a community where individuals are harassed or demeaned due to their beliefs or way of life (Chelliah & Reuse, 2011, p.84-84). Not only would this impair a researcher's ability to conduct their fieldwork efficiently, but this could place a researcher in a dangerous situation.

2.6 Elicitation

Whilst researchers approach linguistic fieldwork with a variety of agendas and research questions, for those that need to collect linguistic data they are almost always going to interact with phonetic/ phonological data at some level. Therefore, the elicitation of phonetic and phonological data is of relevance to a broad range of linguistic field research projects.

2.6.1 Elicitation of phonetic data

One of the first priorities when beginning to work on a language during linguistic fieldwork should be establishing the basic phonetics and phonology of the language in question (Bowern, 2008, p34; Crowley & Thieberger, 2007, p96). In order to document and understand these aspects of the language, researchers will need to work alongside language informants to elicit data (Crowley, 2007, p.85). However, there are a number of considerations to be made before commencing elicitation work with informants including the influence of the researcher and the most suitable methods of elicitation.

2.6.2 The influence of the researcher

There are a several ways in which a researcher can unintentionally influence the elicitation process.

First of all, it is important to note that the instructions given to informants at the beginning of any elicitation sessions can influence the way in which the data is presented by the informant. Chelliah & Reuse (2011) agree that it is important to avoid prompting an informant to produce a particular construction through the wording of a question as, they argue, it is well-known that speakers are likely to recall recently encountered words in an effect known as priming (p.206). Therefore, instructions should be delivered in an objective manner, which in no way coerces a participant towards a particular response. Bowern (2008) suggests using a script to introduce word list elicitation (p.66). This script can then be read to each informant, ensuring that the same instructions are given to each participant. Furthermore, the idea of using a script can also be used to introduce and explain other elicitation tasks in order to ensure that the information received by participants in each comparable task is consistent.

Giving instructions in this way helps to eliminate bias and helps to make the results more easily comparable.

Secondly, when attempting to elicit a word it is unwise for a researcher to say the word in the target language as this can, again, prime the participant to speak in a particular way (Bowern, 2008, p.67; Chelliah & Reuse, 2011, p.206). Instead, the researcher should ask for the target word in another, common language. The informant can then relay the equivalent word in the target language.

In addition, it is important to consider *The Observers Paradox* (Labov, 1972) which could cause an informant to speak differently due to the fact they are being observed. Bowern (2008) suggests that informants should be asked to speak clearly in order to avoid citation tone or quick repetition (p.66); however, this may lead the informant to make a conscious effort to pronounce things 'correctly', rather than in the way that the word would usually be spoken, which could therefore impact upon the results. In a similar example, Crowley (2007) suggests that native speakers can have prescriptive views of the way that their language should be used and, therefore, wish to represent words in a particular way which may not always be reflective of use (p.140). Thus, it is important to decide whether you are trying to achieve a representation of how people actually speak, how they feel that they 'should' speak or examples of both. On the other hand, in languages where there is no dictionary or grammatical rules, there may be a lack of consistency in the pronunciation of words and no 'standard' variety.

Another consideration would be the number of times to ask for the word elicited to be repeated. Efforts should be made to ensure that each word is repeated the same number of times which will help to control bias and also ensure that speaker fatigue does not impact the way that words are pronounced. According to Bowern (2008), three repetitions of each word is desirable, however warns that this is not always possible (p.65). Moreover, if the informant has undertaken a particularly long elicitation session, the words at the end may be produced by a more impatient informant than those words elicited at the start as, according to Chelliah & Reuse (2011), a fatigued informant is unlikely to be as careful with pronunciation as an alert informant (p.201).

2.6.3 Methods of elicitation

There are a number of possible methods of elicitation which can be used during data collection, each with different benefits and suitability for eliciting phonetic and phonological data.

2.6.3.1 Readings

In situations where an informant is able to read, asking them to read a passage can be a useful technique when eliciting phonetic and phonological data. As the text contains complete sentences, parsing the speech stream and providing a gloss is less of an issue making the analysis easier to conduct (Chelliah & Reuse, 2011, p.427). In addition, the exact same reading can also be given to several different informants making each response comparable. This method gives a researcher the opportunity to isolate and compare the pronunciation of phonemes, as well as identify any potential instances of free variation without additional analytical complications.

However, as reading tone can be different from the way that people typically speak this is likely to result in less natural linguistic data, than if the word was simply overheard during conversation. Prescriptivism can also effect speaker pronunciation, whereby a speaker tries to pronounce words in the way that they feel is 'correct' rather than the way that they would naturally speak (Chelliah & Reuse, 2011, p.173). This is more likely to occur in non-spontaneous methods of elicitation which feel more artificial to an informant. In addition, it should be noted that many communities may have few members who are literate (Crowley, 2007, p.28), therefore this should be considered prior to suggesting a reading as a method of elicitation as this could cause offence or embarrassment.

2.6.3.2 Spontaneous speech

Field linguists are often advised to utilise naturalistic data as much as elicited data (Bowern, 2008, p.115), with one of the reasons being that spontaneous speech is likely to be the most natural way of capturing data and would reduce the effects of observer's paradox. This method would also be the most likely to capture words as they are most likely to be spoken in everyday conversation and,

also, features of language which would not generally be used in elicited speech such as code-switching (Bowern, 2008, p.123). Whilst these features may make analysis more difficult, natural speech is likely to offer insights into language use and pronunciation that cannot be gained through elicitation alone. However, taking into account the complications of parsing and translating spontaneous speech, it is important to consider the additional time that an accurate transcription would take. For this reason, it would be preferable to use a combination of both spontaneous and non-spontaneous methods of data collection.

2.6.3.3 Tasks

Asking participants to complete tasks can encourage the elicitation of words in a particular semantic area (Bowern, 2008, p.82) and, therefore, produce a range of sounds for analysis. For example, suggesting a task which involves directing a participant to a particular place would promote the use of directional language. Once participants become absorbed in the task itself, this can also help to reduce observer's paradox and produce the necessary repetitions of a word. Picture naming tasks can cover a number of different semantic fields and word categories and, being a relatively simple task, can make a welcome change to the more intensive elicitation sessions. However, when conducting picture naming it is important to frame questions so that the informant understands what information is required. Chelliah & Reuse (2011) provide the example of a picture of a horse which, when shown to an informant could trigger an answer which provides the general name for 'horse' in the target language or an informant could give the name of the specific breed of horse, therefore it is important to be clear about the information that a researcher is hoping to elicit (p.229).

2.6.3.4 Questions/ Wordlists

Wordlists are often an essential part of the elicitation process as they allow for quick elicitation of a large number of words and are a useful way of illustrating sound contrasts (Chelliah & Reuse, 2011, p.258). Offering a structured elicitation process, word lists can be prepared ahead of elicitation sessions, helping to ensure efficiency for both the researcher and language informant. The existence of commonly used wordlists such as the Swadesh List (Samarin, 1967, p.220) reduces the

preparation that a fieldworker would have to put in to the elicitation process, only needing to build on the existing 200 word list, with other more culture-specific words. In addition, word lists offer a way of building up a researcher's knowledge on an unfamiliar language prior to the elicitation of more complicated data such as connected speech.

On the other hand, eliciting words using a wordlist could become tiresome for an informant, particularly when asked to repeat words for checking. Therefore, Bowern (2008) recommends eliciting three tokens of each word which is enough to produce statistically significant results, without boring the informant (p.65). Furthermore, when eliciting for the purposes of phonetic and phonological data collection, it may be necessary to use a frame such as 'this is X' which will help to prevent speakers using citation tone (Bowern, 2008, p.66 & p.70).

2.7 Areas of focus and difficulties associated with the collection of phonetic data

Through reviewing the literature around the topic of phonetic and phonological data collection a number of areas difficulty have been identified, including minimal pairs, affricates, connected speech, allophones, variation and metadata and, finally, recording methods. These areas and the associated difficulties will be explored further in the following section.

2.7.1 Minimal pairs

According to Crystal (2008), a minimal pair can be defined as "two words which differ in meaning when only one sound is changed" (p.219). For example, the words mat and met or cat and hat.

As the corpus of language data grows, it is useful to begin to check for minimal pairs. The accuracy can then be checked by verifying the pronunciation with the informant; however, even if the words are said incorrectly by the researcher, an informant could recognise what the researcher is trying to say and confirm that this as correct. Therefore, it is often better to ask the informant to repeat the word that the researcher was trying to pronounce in order to confirm its accuracy (Crowley, 2007, p.97).

According to Crowley (2007), minimal pairs can also be used to check for other words through sound substitution. For example, if a word such as pari exists in the language, by substituting sounds, other words can be suggested. Eg. padi, pabi or bari, badi (p.110).

2.7.2 Affricates

An affricate is the name given to a consonant sound which is produced "when the air-pressure behind a complete closure in the vocal tract is gradually released" (Crystal, 2008, p.12).

The identification of affricate sounds can be complicated due to the similarity of affricates to other sounds. For example an affricate can appear similar to a consonant pair, or palatal affricates can be confused with palatal stops due to the palatal stop displaying some affrication (Bowern, 2008, p41).

2.7.3 Connected speech

Depending upon the method of elicitation, connected speech can pose an additional problem for a researcher as the speech stream must be parsed in order to identify the boundaries of words and their individual phonological make up. Furthermore, when transcribing conversational data there may be overlaps and interruptions to deal with, as well as slang language (Bowern, 2008, p119). Parsing the speech stream is particularly challenging at the beginning of the process of linguistic fieldwork due to a lack of familiarity with the language. As exposure to the language variety increases, a researcher will be better equipped to begin to analyse this type of linguistic data.

2.7.4 Allophones

An allophone refers to phoneme variations which do not alter the meaning. For example, the English phoneme /t/ can be pronounced in a number of different ways (Crystal, 2008, p.14-15).

Deciding whether sounds are separate phonemes or allophonic variants can be problematic, particularly because sounds that are allophonic in a researcher's native language may be phonemic in the target language (Bowern, 2008, p.40). Therefore a researcher must step outside of the assumptions

that govern their native language and approach data collection with an open mind as to what is possible in the target language.

2.8 Variation and metadata

Variation can relate to both speakers and the data itself and can have an affect on the outcome of data collection. Aspects to consider relating to the background of the speaker are, regional/where an informant has lived and socialised, their age, gender, socioeconomic status and educational background. Each of these variables and the impact that they can have upon both a speaker and the linguistic data that they produce will be discussed in the following section.

2.8.1 Regional variation

Regional variation can greatly impact upon language use, particularly from a phonetic and phonological perspective. Although not always clear to a non-native speaker, native speakers are likely to display differing accents which has an effect upon pronunciation. Whilst semantic regional variation is covered in the literature, phonetic and phonological variation by region appears to be a less explored area which would benefit from further research.

2.8.2 Age

The age of a speaker is an important variable to consider, with different age groups generally contributing to elicitation tasks in different ways. Those from the older generation are more likely to have retained fluency of an endangered language and have maintained use of a particular speech sound that has not been passed down to the younger generation (Chelliah & Reuse, 2011, p.169). In addition, prescriptivism is more common amongst older speakers which may lead them to provide more traditional versions of the language in question, rather than the version that is used in everyday speech (Chelliah & Reuse, 2011, p.170). However, when using older speakers for phonetic purposes, it is important to consider that pronunciation may be effected by a lack of teeth and unclear articulation (Crowley, 2007, p.91). On the other hand, younger speakers are, in general, better at paradigmatic elicitation (Crowley, 2007, p.91). Therefore, according to Bowern (2008), for phonetic

purposes, a speaker that is "not too old and not too young" and retains a full set of teeth is preferable (p.65).

2.8.3 Gender

Gender can impact greatly upon an informant's experiences of the world, their education and literacy and, as a result, their use of language. In many traditional communities, females and males have very different ways of life. In some communities males will receive an education, whereas females will not. Furthermore, as the children grow older, men will go out to work and women will stay at home to raise the family, cook and clean. As a result, women tend to be more conservative speakers (Bowern, 2008, p.134). Conversely, in traditional societies, men are likely to have a much broader network of contacts and experiences, which means that they develop a wider vocabulary and therefore often make good language informants (Chelliah & Reuse, 2011, p.171). As a result of the different roles and experiences that men and women have, it may be useful, where possible, to incorporate both genders into the elicitation process.

2.8.4 Socio-economic status/ Educational background

Socio-economic status and educational background can often be linked. With the provision of a financially or educationally wealthy background, informants are likely to have a broader vocabulary and possibly differences in pronunciation to those who have received no formal education. Literate informants bring a number of benefits to the data collection process, including a broader awareness of the technical aspects of the language and often a wider vocabulary. However, according to Chelliah & Reuse (2011) knowledge of the written language can result in prescriptive views of language use and therefore influence pronunciation (Chelliah & Reuse, 2011, p.173). On the other hand, according to Bowern (2008) uneducated speakers may have a better knowledge of the oral history of the language due to the time spent within the community, rather than at school (p.131). Therefore, informants of both socio-economic/educational backgrounds offer different contributions to the data collection process.

2.9 Recording methods

The choice of recording method is vital as, after an elicitation session has ended, this contains the only record through which further analysis can take place. There are a number of possible recording methods including electronic and non-electronic methods.

As data is likely to be saved for future linguistic analysis, it is preferable to record the sessions electronically, either through an audio recorder or video recorder, so that sessions can be played back for analysis. Whilst video recordings can offer insight into extra-linguistic aspects of conversation, for phonetic and phonological data collection an audio recorder is preferable as it offers higher quality sound playback. Whilst there are a number of types of audio recorder, Bowern (2008) recommends using a digital recorder due to their ability to produce high-quality recordings and versatility (p.17). Clear recordings are essential for the purposes of analysis, particularly when transcribing important phonetic distinctions which may be difficult to perceive.

In addition, note taking, which can be typed or handwritten, can be used alongside recording equipment. Handwritten notes would be preferable if used alongside an electronic recording method as this prevents a researcher from relying solely on electronic equipment which can encounter failures such as loss of battery or equipment malfunction due to the extreme conditions (Bowern, 2008, p.26-27). Not only is this method useful in that it does not rely on technology or electricity, but it can be easily transported and is often considered less intrusive or distracting than electronic devices, particularly in communities where electronic equipment is not the norm (Bowern, 2008, p.26-27).

2.10 Limitations of the literature

As demonstrated within this chapter, there are a number of sources of informative literature in the area of field linguistics and the collection of data whilst undertaking research; however, these ideas and opinions originate from individual scholars and are therefore in no way comprehensive. Through my research I intend to increase the breadth of the existing recommendations and also

confirm which techniques are the most successful so that future field linguists can benefit from the combined experience of others before them.

3. Methodology

This chapter describes the processes involved in carrying out the research. It begins with an explanation of the motivations behind the choice of topic and how these justify the research method used. This is followed by further description of how the research was conducted.

3.1 Research Motivation

The literature discussed in chapter 2 demonstrates that linguists are seeing a sharp decline in the number of languages which are spoken throughout the world. The impacts of this decline are broad and affect both communities themselves and the wider linguistic field. Due to the time sensitivity of the issue of language endangerment, more research is needed as a priority so that languages can be preserved before they become extinct.

Furthermore, whilst there is extensive literature around the subject of field linguistics and language endangerment, there is a lack of literature which combines the advice of experienced field linguists. Past work that has combined the opinions of a number of field linguists has been in the form of individual chapters in an edited book. The advice or opinions expressed in these editions, therefore, still only reflect the opinion of the one or two field linguists that wrote the chapter. Consequently, a gap exists in the literature for advice which is derived from the collective knowledge of experienced field linguists and, for this reason, research of this kind would make a significant contribution to the linguistic field.

In order to address this gap, field linguists were invited to share their experiences of language documentation in relation to the collection of phonetic data and metadata during field research. The findings were then analysed and any patterns identified were used to create recommendations based on the knowledge and feedback of experienced field linguists. These recommendations can then be

used to guide and assist future field linguists in their language documentation projects. The specific areas of phonetic data and metadata were chosen due to the lack of focus on these areas in the existing literature.

Having already completed a research project which explored the elicitation of lexical and semantic data during field research, I expressed an interest in extending my research to cover the collection of phonetic data with a view to contributing to my goal of producing a series of guides which provide recommendations for the elicitation of linguistic data during field work. The recommendations included in these guides would be derived from the feedback and knowledge of a number of experienced field linguists which would be obtained through the use of an online survey. For this reason, I have chosen to build upon my past research at Masters Level and address a further area of linguistic data collection – phonetics and phonology.

In keeping with my past research, this project fills a gap in the research area as, through my reading around the topic area, research and advice relating to the collection of phonetic and phonological data during field research is minimal. Furthermore, whilst there are many useful guides to linguistic fieldwork these have primarily focused on the views of one or two field linguists; whereas my project differs due to the collaborative element whereby a number of experienced field linguists have been invited to take part.

In order to fulfil these motivations, an online survey method was chosen. Not only does this method suit the time sensitive nature of the research by creating the potential for a quick data collection process, but it also fulfils the need for the survey to be easily distributed to a large number of participants in order to bring together multiple perspectives on the topic. In addition, the online survey method helped to ensure a broad geographic scope of participants, which was intended to enhance the range of the languages documented by respondents.

3.2 Creating the survey and formulating questions

After deciding on the topic area, I began to devise questions which, based upon my review of the literature, would address areas of difficulty in the collection of phonetic and phonological data during field research.

3.2.1 Question Types

A combination of open and closed questions were used in order to vary the structure of the survey and provide a more interesting experience for the respondent; thus increasing the likelihood of obtaining completed responses from participants. Open questions were chosen to allow for detailed answers and qualitative information; however can also lead to an excess of irrelevant information (Bloomer, Wray & Trott, 2006, p.155). On the other hand, closed questions allow for precise answers which can be easily analysed in a quantitative manner, but limit the respondent's opportunity to expand and offer extra information (Bloomer, Wray & Trott, p.155). With closed questions it could also be argued that the provision of suggested answers may limit the accuracy of a response, with the respondent choosing the closest answer, rather than the most precise answer.

3.2.2 Question Content

The questions were chosen in order to address the early concerns that a field linguist may encounter when collecting data on the phonetics and phonology of a language. These specific areas were identified from the literature as potential problem areas for field linguists and include: general questions about the language and informants, elicitation, identifying the sound inventory of the language, affricates, connected speech, allophones, variation and metadata and, finally, recording methods.

3.2.3 Creating the Survey

Once my survey was finalised and had received ethical approval from my research supervisor and the university's ethics board, I began to transfer the information to my chosen online free survey software, Survey Gizmo.

The questions which were included in my survey are listed below:

1) Which language(s) were you documenting/analysing in your field work/research?

2) Where in the world is/ were this/these language(s) spoken? By which community?

3) Has your data collection focused on:

   (choose one)

   a. Your native language   *b. A non-native language*   *c. Both*

4) Did you have a language in common with your informants? For example did you both speak English, and could therefore use this language to communicate.

         a. Yes   *b. No*

- *Elicitation*

5. What kind of speech do you elicit: (choose as many as apply)

         a. Readings (eg. reading aloud from a book)

         b. spontaneous speech (eg. conversations which happen around you naturally)

         c. Conduct a task which requires conversation/ explanation

         d. Ask questions

         e. Word lists

         f. Other (please specify)

         Identifying the Sound Inventory of the Language

- *Minimal Pairs*

6. What techniques have you used to identify minimal pairs?

         a. Inventing minimal pairs to test

         b. Trying free variation of sounds

c. Ask for words that rhyme or sound similar

d. Other

7. Have you encountered any problems with the identification of minimal pairs in your past research?

a. Yes   b. No

8. If yes, what were they?

9. How did you overcome these problems?

- *Affricates*

    10. What techniques have you used to determine whether a stop and fricative combine to make an affricate, or whether they are a consonant pair?

    11. Have you ever misrepresented an affricate as a consonant pair and vice versa?

    a. Yes   b. No

    12. If yes, why did this happen?

    13. How did you realise that this was the case?

    14. Have you encountered any other problems with the identification of affricates in your past research?

    a. Yes   b. No

    15. If yes, what were they?

    16. How did you overcome these problems?

- *Connected speech*

    17. What techniques have you used to separate the speech stream?

18. Have you ever misrepresented a number of separate words as one word? Or incorrectly segmented a stream of sound into separate phonological units?

a. Yes   b. No

19. If yes, why did this happen?

20. How did you realise that this was the case?

21. Have you encountered any other problems with parsing the speech stream?

a. Yes   b. No

22. If yes, what were they?

- *Allophones*

23. What techniques did you use to identify the allophones of a particular phoneme?

24. Have you ever misrepresented an allophone as a phoneme?

   a. Yes   b. No

25. If yes, why did this happen?

26. How did you realise that this was the case?

27. How do you establish whether two sounds are separate phonemes, or allophonic variants?

28. Have you encountered any other problems with the identification of allophones in your past research?

a. Yes   b. No

29. If yes, what were they?

30. How did you overcome these problems?

Variation and Metadata

- *Variation and metadata*

  31. How much data did you collect on the language?

  32 How many speakers did you collect data from?

  33. If you had more than one informant, did you obtain an equal amount of data from each?

  34. Did you have controls in place for variation (tick as many as apply):

  Regional

  If yes, give details

  Socio-economic

  If yes, give details

  Gender

  If yes, give details

  Age

  If yes, give details

  Education

  If yes, give details

  Other

  If yes, give details

- *Recording Methods*

  35. What methods do you use to record your language data? (tick as many as apply)

  Hand-written notes

  Typed notes

  Using an audio recorder

  Using a video recorder

Other (please specify)

3.3 Selecting participants and distributing the survey

Participants were academics who worked at universities and research institutions. I identified these professionals using Google searches for 'field linguistics research' which led me to the websites of universities and research institutions involved in this type of research. The individuals were then selected based on their interest in field linguistics and linguistic research as identified through the research areas of interest detailed on their academic profiles. In order to invite the selected participants to take part in the survey, I emailed each of them with the link to the survey alongside an introduction to myself and my research:

Good Afternoon (insert name),
I am a student at the University of Huddersfield currently undertaking a Masters of Research in Linguistics. My project focuses on the problems and pitfalls encountered when collecting phonetic data during field research. Through my survey, I aim to collate the advice of experienced field linguists in order to provide a guide for future language documentation projects of this kind.
As an individual who may have undertaken field research of this type, I would be grateful if you could spare 10 minutes of your time to participate in this short survey which will form the basis of my research project.

Below is the link to my survey:
http://www.surveygizmo.com/s3/3467978/Eliciting-Phonetic-Data-During-Field-Research

Thank you for your time.

Kind Regards,

Katie Coxall

By clicking the link on the email, participants were directed to an introductory page which explained the aims of my research and contained an ethics disclaimer.

In total 95 participants were invited to take part in the survey, and a total of 16 of these individuals completed the survey which is a completion rate of 16.8%.

3.4 Ethics

In order to comply with ethical considerations, prior to administering my survey I completed the ethical declaration documents and submitted these to the university for approval. The ethics documents confirmed that the nature of my study would not cause any harm to participants and outlined that I would handle any data in line with the Data Protection Act. Once approval was received, my survey was administered to participants. In order to ensure that participants were fully aware of the aims and outcomes of my survey I included an ethical declaration at the very start of the survey, along with an explanation of how the data would be stored and the participant's rights to withdraw at any time. A brief introduction to me and the aims of my research were also included in the initial email inviting participants to take part in the survey.

Ethical declaration:

Thank you for agreeing to participate in this survey on the elicitation of phonetic data during field research.

The aim of this survey is to collate the experiences of field linguists in the collection of phonetic data and the problems associated with data collection of this type in order to inform future field linguists of tried and tested methods. I hope to contribute strategies which will assist in simplifying the process of linguistic data collection and documentation so that languages, especially endangered languages, can be documented more easily.

All answers will remain anonymous.

For this study, you will be completing a short survey about your previous experiences of data collection whilst undertaking linguistic field work. If you have any questions at any point before you complete (or during) this survey, please email me, Katie Coxall, at Katie-louise.Coxall@hud.ac.uk. All responses provided for this study will be kept completely confidential, and all data will be held in accordance with the Data Protection Act (UK). When the results of the study are reported, you will NOT be identified by name or by any other information that could be used to infer your identity.

By clicking 'Next' below you acknowledge that you have read and understand that:

- *Your participation in this survey is voluntary. You may withdraw your consent and discontinue participation in the project at any time, for any reason, and without having to state a reason.*
- *You can contact the researcher (Katie Coxall) with questions at any point in the survey and your questions will be answered.*
- *The information you provide will be held in confidence by the researcher.*
- *You are giving your consent to be a subject of this research.*

3.5 Evaluation of the Survey Method

The survey method was advantageous in that it allowed me to target a number of participants in a short space of time, as well as being relatively geographically unrestrictive. This allowed me to invite participants from different countries to take part. The higher the number of participants and the wider the geographic spread of respondents, the more representative the research becomes.

Using an online survey, rather than a face to face method, also enables a researcher to minimise the likelihood of bias by delivering the same messaging to each respondent. The absence of face-to-face communication excludes the possibility that the presence of a researcher could have influenced the results through body language, changes in intonation or varying ways of expressing the questions or guidance. Furthermore, if a researcher is present, a respondent may be more inclined to ask the researcher for further clarification which may then influence their response. Therefore, being able to control these variables by removing any face-to-face contact between researcher and respondent increases the reliability of the results. In terms of the advantages for respondents, the survey method is unobtrusive in that it gives respondents the opportunity to complete the survey at a time that is convenient to them and also makes the research easier to decline which could be seen as beneficial from the perspective of a respondent.

Conversely, the online survey method also has some disadvantages. From the point of the researcher, the method can restrict the number of responses received in that there is no face to face contact from the researcher and as such the participant may find the research easier to decline. This can, in comparison to face-to-face methods of data collection, result in low response rates which are

considered to be typical for data collection of this type. My survey was distributed to 95 respondents and received 16 completed responses which is a response rate of 16.8%. Mcneill, Chapman and Chapman (2005) suggest that the response rate for a postal survey is 30-40%, whereas face-to-face research typically receives a response rate of around 70-80% (p.52). Furthermore, it can be argued that the method of data collection predisposes a certain type of respondent, as some people are more likely than others to respond to a request for this kind of information. The survey method also restricts the opportunity to request further information from a respondent and can, as a result, leave unanswered questions.

3.6 Process of analysing the survey results

Once the survey had been completed by all 16 respondents, the results were exported to an Excel worksheet. This allowed for all results to be viewed together in the same document making it easier to spot trends, rather than switching between pages on the survey platform. These responses were then presented in Chapter 4, the Results.

When analysing the results, where appropriate, these were grouped into more specific categories. For example, if a respondent indicated an answer of 'other', but then in their explanation suggested that they had used one of the methods already listed in the suggested response list, this response would be re-categorised as appropriate. Furthermore, where a number of respondents gave similar answers, their responses were grouped under one heading. The results were in some cases presented in charts or tables and in others were explained in text. As some of the questions allowed for multiple selections, whereas others allowed only one selection, results were presented in percentage values to make the results more easily comparable.

4. Results

In the following section, the results of the research are presented. The results have been structured to follow the order in which the sections appeared in the survey: field research, informants and elicitation methods; minimal pairs; affricates; separating the speech stream; allophones; variation

and metadata and, finally, recording methods. Where an answer was given by a respondent which did not directly answer the question, this was omitted from the results section. The results are to be followed by a more thorough discussion in chapter 5 where the significant results will be explained in further detail.

4.1 Field Research, Informants and Elicitation Methods

In this section, participants answered questions relating to some of the broader concerns of linguistic fieldwork, beginning with questions surrounding the language variety and location in which the variety is spoken, followed by issues relating to the communication between researcher and informant and, lastly, the type of speech that was elicited.

Through this research, data was analysed from the responses of 16 researchers who had collectively worked on, in excess of, 39 languages from a total of 6 continents. The languages are organised into columns according to continent. The countries in which the languages are spoken, or were recorded, are then listed followed by the name of the language variety.

Table (1): Distribution of languages documented by survey participants

Continent					
Africa	**North America**	**South America**	**Asia**	**Oceania**	**Europe**
Cameroon: Kom, various Bantoid languages, Medumba (Grassfields Bantu),	California: Babanki	Bolivia: Quechua	Taiwan: Tsou, Amis, Bunun, Atayal, Kavalan	Australia: Arabana, Iwaidja and Wubuy	Austria: Halkomelem, Blackfoot, Kutenai and Upper Austrian
Ghana:	Southern	Northwest	China: Uyghur		

Nafaanra	Mexico: Teotitlan del Valle Zapotec,	Amazonia: Desano			
Swaziland: SiSwati	Canada: Witsuwit'en, Tlingit (Na-Dene), Plains Cree (Algonquian), ʔayʔaǰuθəm (Comox-Sliammon), Kutenai (aka Ktunaxa, Kootenay, Ksanka.)	Brazil: Siriano	Turkmenistan: Turkmen		
CÃ´te d'Ivoire: Guébie (Kru)	Alaska: Tlingit (Na-Dene)	Columbia: Siriano, Gitksan			
	USA: Cherokee, Deg Xinag, English, Plains Cree (Algonquian)	Northern Peruvian Amazonia: Máíhɨ̃ki			
	North America: Halkomelem, Blackfoot,	Mato Grosso: Panará, Mebêngôkre			

	Kutenai and Upper Austrian				
		Para: Panará, Mebêngôkre			

When asked whether their data collection focused on their native language, of the 16 respondents, 14 respondents (87.5%) indicated that they had documented a non-native language, whereas 2 respondents (12.5%) documented both a native and non-native language.

Of these 16 respondents, 14 (87.5%) initially confirmed that they had a language in common with their informants; whereas the remaining 2 respondents (12.5%) selected *other*. After reviewing the two *other* responses, one explanation indicated that the respondent did have a language in common with their informants, taking the total number of respondents for this category to 15 (93.75%). The remaining *other* response was an answer of 'both', suggesting that the respondent had worked on at least one language where there had been a language in common between researcher and informant, and at least one language where there had been no language in common.

Respondents were asked about the methods of elicitation that they used during their field research. In this question, participants were able to choose as many options as they felt applied. The answers are represented in *Figure (3)*.

Figure (3): Methods of Elicitation
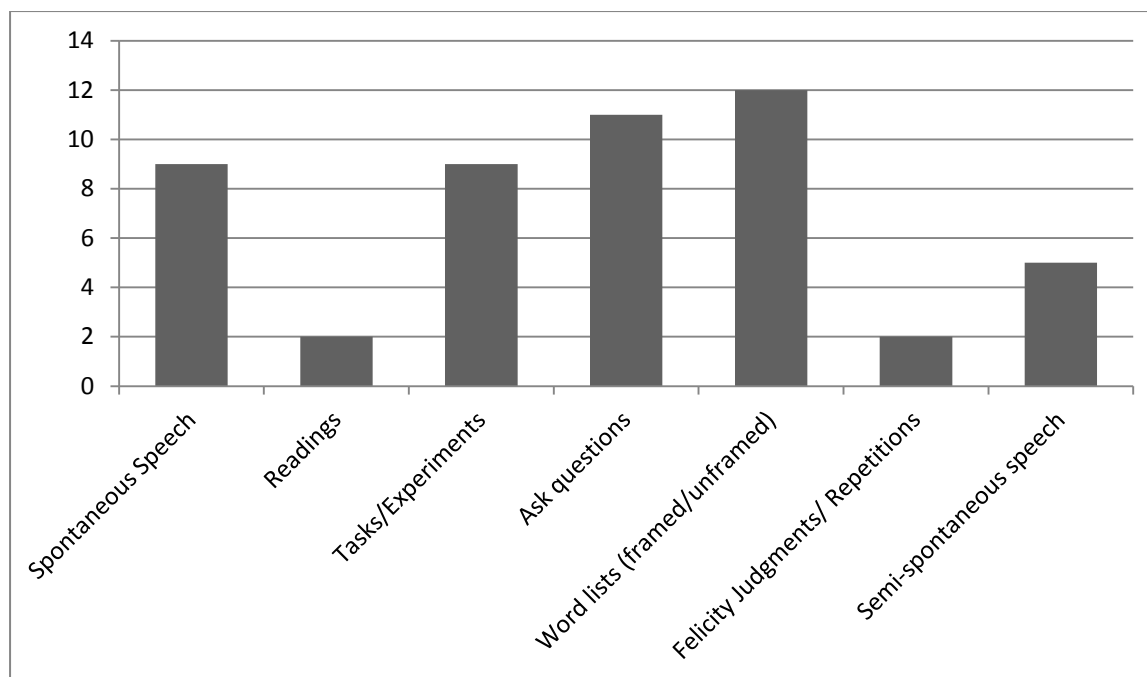
When asked about methods of elicitation, *other* was the most commonly selected option with 12 selections. The information given by the *other* respondents has been re-categorised and represented within *Figure (3)*. Taking these responses into account, *word lists* was the most popular category with 12 selections. *Ask questions* was the second most popular category with 11 selections, followed by *tasks/experiments* and *spontaneous speech* which each had 9 selections. *Semi-spontaneous speech* which represented picture tasks, storyboards and narratives, received a further 5 selections. The categories chosen the least were *readings* and *felicity judgments* which attracted 2 selections each.

4.2 Minimal Pairs

In this section, participants were asked about their experiences of eliciting minimal pairs, including the techniques used to elicit this kind of data, the problems encountered and the ways in which these problems were overcome.

Participants were asked which techniques they used to identify minimal pairs and were able to choose as many options as they felt applied. Responses are represented in *Figure (4)*.

Figure (4): Techniques used to identify minimal pairs
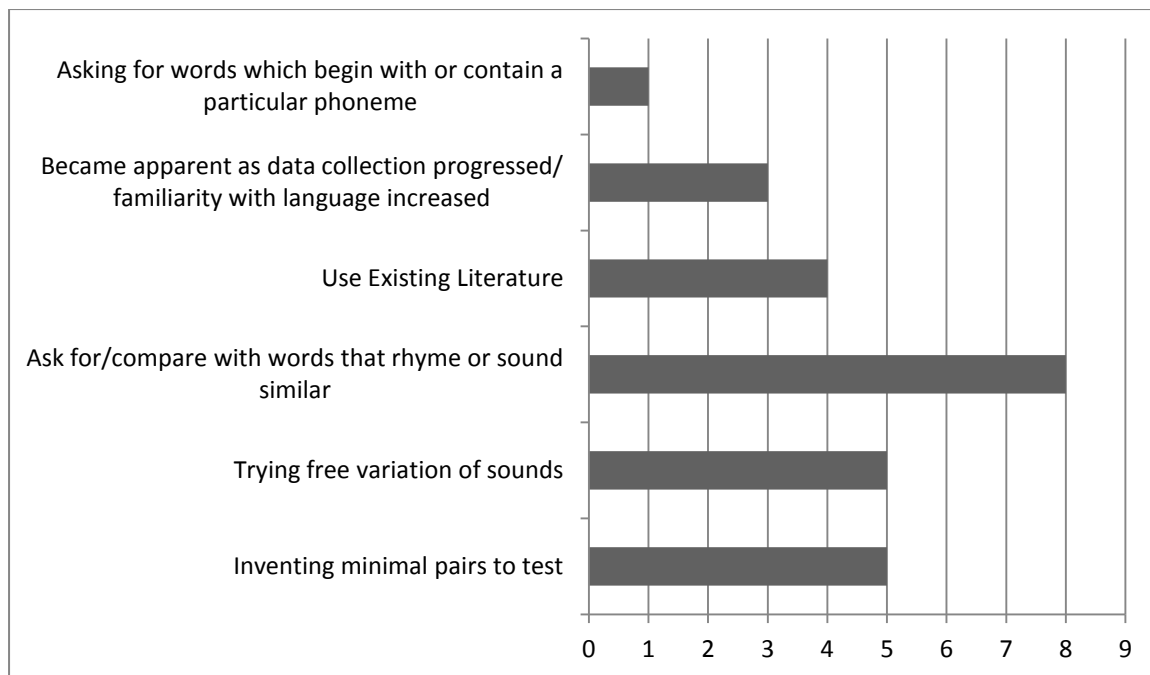
Figure (4) shows the techniques used by respondents when identifying minimal pairs. The most popular category was again *other* with 9 selections; however, in order to represent these responses more clearly, they have been re-categorised and included within *Figure (4)*. Taking the *other* responses into account, the most popular category is, *ask for/compare with words that rhyme or sound similar*, as indicated by 8 respondents. The second most popular methods for the identification of minimal pairs were *inventing minimal pairs to test* and *trying free variation of sounds* with 5 selections each. *Use existing literature* received 4 responses, and *became apparent as data collection progressed/ familiarity with language increased* received a further 3 selections. Finally, *asking for words which begin with or contain a particular phoneme* received 1 selection.

8 out of a possible 16 respondents indicated that they had experienced problems when trying to identify minimal pairs in their research. This equates to 50% of respondents. 5 respondents (31%) had no problems in this area; whereas another 2 respondents (12.5%) selected *other*. 1 respondent (6.25%) chose not to answer this question.

When taking into account and re-categorising the *other* responses, a further respondent had in fact indicated that they had experienced challenges when identifying minimal pairs; therefore, those that experienced a problem during their research can be increased to 9 respondents (56.25%). The

47

second *other* respondent indicated that they did not require minimal pairs for their research and therefore this response is most accurately categorised as *other,* taking the total number of responses for this category to 1 selection (6.25%).

When asked about the problems experienced with the identification of minimal pairs 11 responses were received. Of these responses, the complexity of language tone systems and suprasegmentals were issues which caused difficulty for 3 respondents (27%), vowel distinctions posed problems for 2 respondents (18%). A further 2 respondents (18%) had difficulty gaining assistance from untrained native speakers; whereas each of the following issues were suggested as problematic by 1 respondent each (9%): sociolinguistic variation, morphophonological complexity, few minimal pairs and difficulty finding minimal pairs using a dictionary.

Participants were asked how they overcame problems with the identification of minimal pairs. Respondents were able to provide as many techniques as they felt applied. These responses have been represented in *Figure (5)*.

Figure (5): How did researchers overcome problems with the identification of minimal pairs
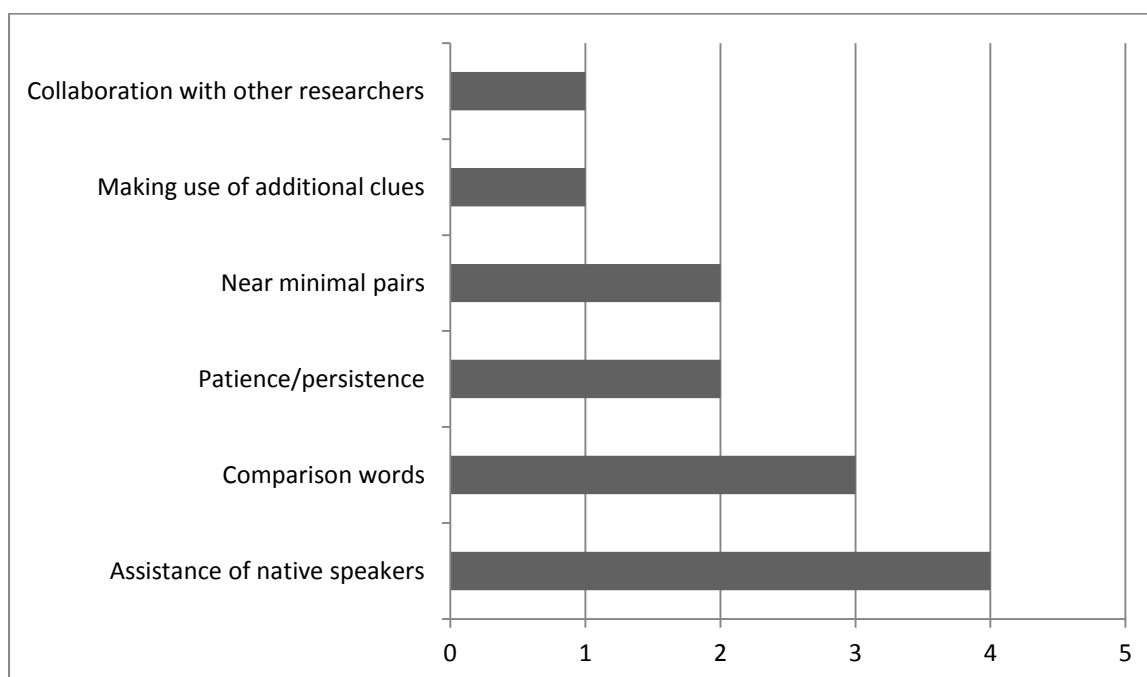
Together, 8 respondents suggested 13 ways in which they overcame the problems experienced with the identification of minimal pairs. 4 responses (31%) suggested the *assistance of native*

speakers; 3 responses (23%) suggested making use of *comparison words*; 2 respondents suggested *patience and persistence* (15%) and also the use of *near minimal pairs* (15%); whereas 1 suggestion was made for both *making use of additional clues* (8%), and *collaboration with other researchers* (8%).

4.3 Affricates

In this section, participants were asked about the techniques used to elicit data relating to affricates, the types of problems that were encountered whilst eliciting this type of data and also how these issues were overcome.

When asked which techniques were used to determine whether a stop and fricative combine to make an affricate, or whether they are a consonant pair, 12 respondents provided feedback to this question. However, 6 of these responses were disregarded from this section as they were not deemed to answer the question. Of the 6 remaining responses, several contained a number of suggestions and have therefore been re-categorised to reflect each separate suggestion. Responses are represented in *Figure (6).*

Figure (6): Techniques used to determine whether a stop and fricative combine to make an affricate, or a consonant pair
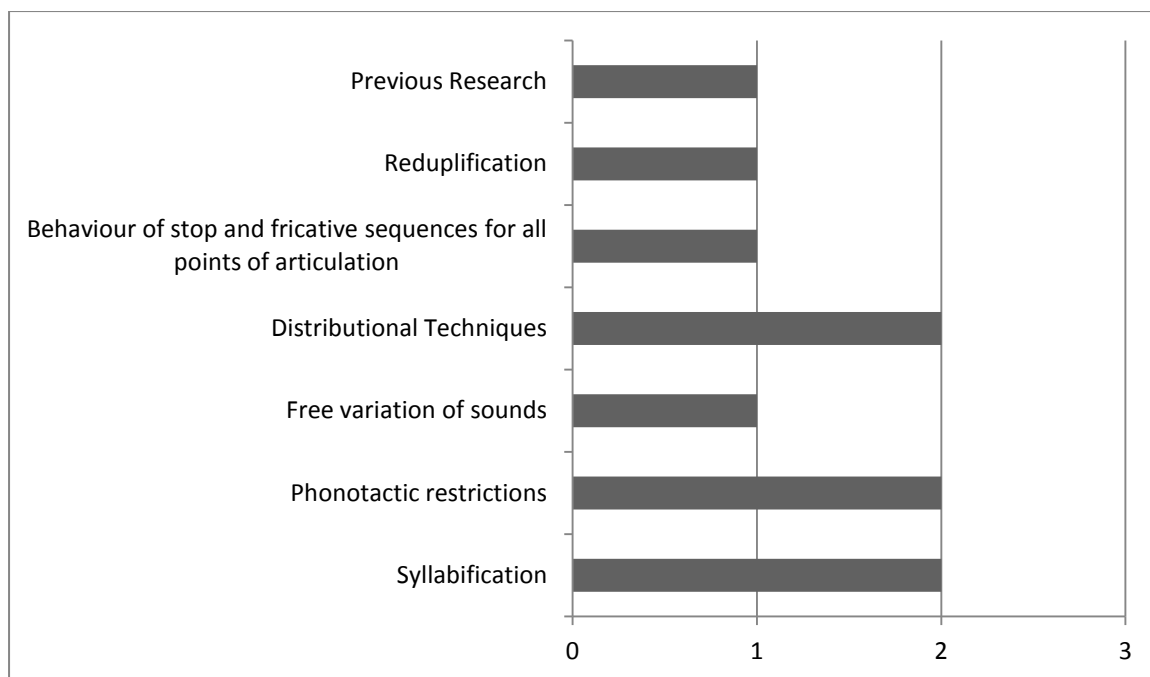
Figure (6) shows that the most popular techniques for the identification of affricates were *distributional techniques, phonotactic restrictions* and *syllabification* which each received 2 suggestions out of a total of 10 suggestions (20% each). Each of the remaining techniques were suggested once each (10%).

When asked whether they had misrepresented an affricate as a consonant pair, of the 16 potential respondents, the highest proportion of respondents (7 or 44%) indicated that they had not misrepresented an affricate as a consonant pair or vice versa; whereas 3 respondents (18.75%) indicated that they had. 4 respondents (25%) selected *other*. 2 respondents (12.5%) chose not to answer this question.

Taking into account the *other* responses, and having re-categorised these to fit in to the most appropriate category, 10 respondents (62.5%) indicated that they had not misrepresented an affricate as a consonant pair and vice versa, whereas 4 respondents (25%) said that they had. 2 respondents (12.5%) did not answer the question.

Indicating why they misrepresented an affricate as a consonant pair, or vice versa, 1 respondent commented, "ʔayʔaǰuθəm has a lateral affricate, but also has an alveolar stop and lateral

fricative, so it is possible (to non-native listeners) to mistake the sequence as two phonemes instead of one." Although 4 responses were offered to this question, 3 were discounted from this section as they were not considered to be relevant to the question.

Respondents were asked how they realised that they had misrepresented their data. Of the 4 responses given, 2 were discounted from this section as they were not deemed to be relevant. Of the remaining 2 responses, 1 respondent indicated that they had identified that they had misrepresented their data "based on analysis"; whereas the second respondent said, "The easiest way to check whether the sound is one phoneme or two is to pay attention to where it occurs (there are no complex onsets of the form stop-fricative, so in onset position the sequence must be an affricate) and to check how it behaves when different morphological processes apply (does it fill a single consonant slot in reduplication, for instance)."

11 of a possible 16 respondents (69%) indicated that they had encountered no other problems with affricates in their past research; whereas 3 respondents (19%) said that they had experienced issues. 2 respondents (12.5%) did not answer this question.

Despite only 3 respondents having detailed that they had encountered further issues with affricates, 4 respondents went on to detail further issues experienced with affricates in their previous research. However, 1 of these responses was disregarded due to its lack of relevance to the question. The problems experienced by the remaining 3 respondents have been summarised as follows: sociolinguistic variation; germinate consonants which had not previously been identified within the literature and stops which are affricated in certain prosodic positions or contexts.

In order to overcome these problems, respondents gave the following suggestions: *collecting more data made it clear that the variation was sociolinguistic*; *the help of other researchers* and the final respondent *elicited consonants in a variety of prosodic and vowel environments rather than using a standard list*, the same respondent also *listened to lots more conversational speech*. Although 4 respondents completed this question, 1 response was disregarded as it did not answer the question.

4.4 Separating the Speech Stream

In this section, respondents were asked about the techniques which they had used to separate the speech stream, along with further problems which were encountered in this area. Participants were able to choose as many techniques as they felt applied.

11 respondents suggested techniques which were used to separate the speech stream; however, 3 were disregarded from this section as the answers were not deemed to contribute to the understanding of this area. Of the remaining 8 responses, 2 of the responses offered more than one technique, each are represented in *Figure (7)* below.

Figure (7): Techniques used to separate the speech stream
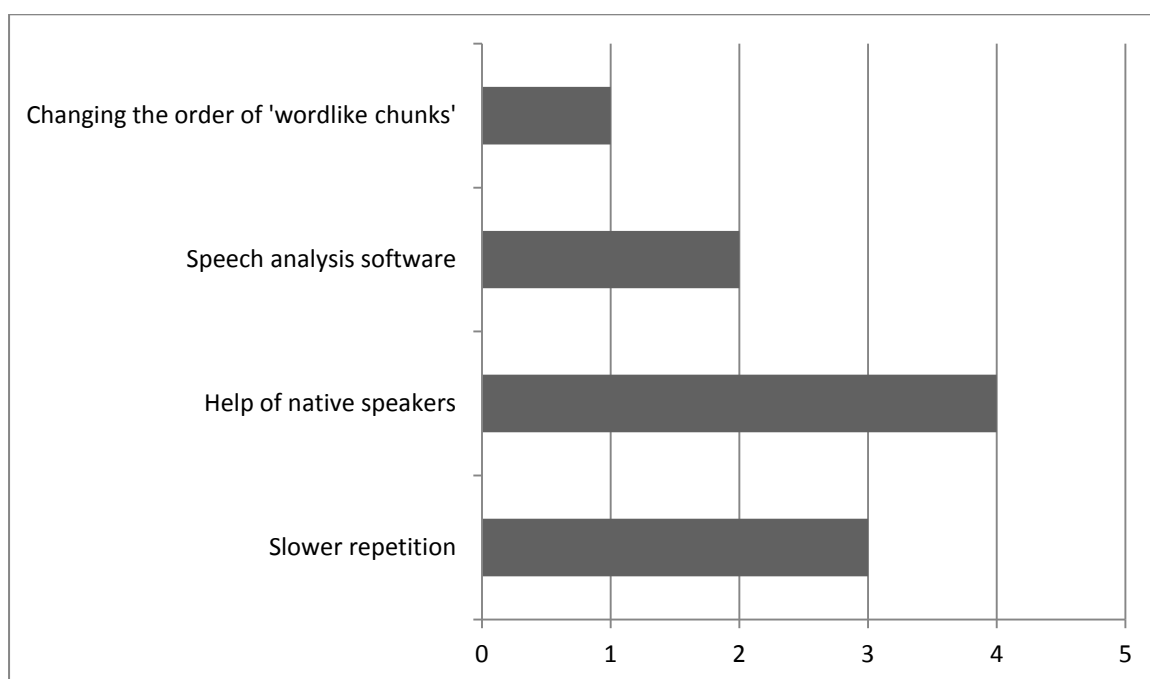
Help of native speakers was the technique enlisted most frequently, with 4 out of a total of 10 suggestions (40%), *slower repetition* was the second most popular category with 3 suggestions (30%), *speech analysis software* was suggested by 2 respondents (20%), and *changing the order of wordlike chunks* was suggested once (10%).

Out of a possible 16 respondents, 13 respondents (81%) indicated that they had experienced problems with the segmentation of the speech stream; whereas 1 respondent (6%) experienced no

issues. 1 respondent (6%) selected the *other* option and a further respondent (6%) chose not to answer this question. However, after taking into account and re-categorising, the feedback that was given from the *other* respondent, the number of respondents indicating that they had experienced no issues can be increased to 2 selections (13%).

The issues experienced with the segmentation of the speech stream are represented in Figure (8) below.

Figure (8): Issues experienced with the segmentation of the speech stream
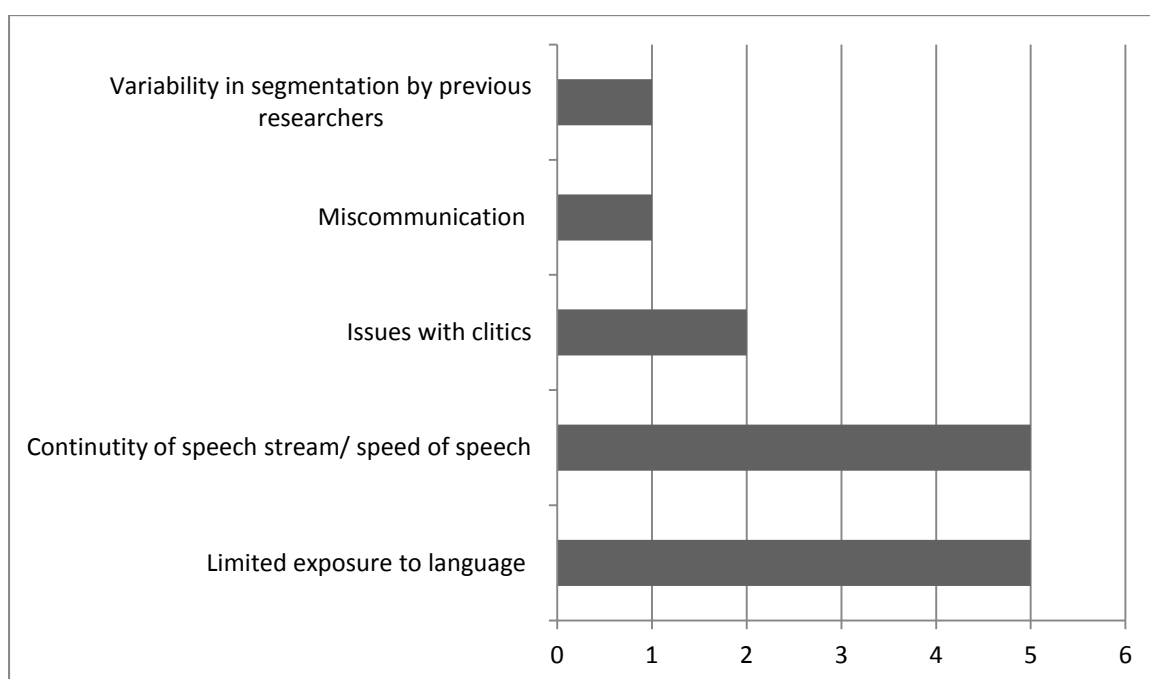
13 respondents suggested reasons why they had encountered issues with the segmentation of the speech stream. As respondents were able to suggest more than one issue, between the 13 respondents, 14 suggestions were made. Of the 14 suggestions, both *limited exposure to language* and *continuity of the speech stream* accounted for 5 selections each (36% each); *issues with clitics* were cited in 2 responses (14%); whereas *miscommunication* and *variability in segmentation by previous researchers* were each suggested once (7% each).

Respondents realised that they had misrepresented elements of the speech stream in various ways. Out of a total of 12 responses, *Native speaker judgements* received 4 selections (33%); whereas

increasing familiarity with the language in question was suggested twice (17%). *Clitics* were cited in a further 2 responses (17%); whereas *back translation*, *use of a dictionary* and *morphological analysis* contributed to 1 response each (8% each). Of the 12 responses received to this question, 1 response (8%) was disregarded as it wasn't deemed to contribute to the understanding of the area.

When asked whether they had experienced further problems when parsing the speech stream, out of the 13 responses, 6 respondents (46%) indicated that they had experienced further problems when parsing the speech stream. An equal number of respondents indicated that this caused them no further problems during their research (46%). 1 respondent (8%) selected *other*. Taking into account the *other* responses, the feedback suggests that the answer can be re-categorised as *yes*. This increases the number of *yes* answers to 7 (54%).

Of those that experienced further problems with parsing the speech stream, the most frequent problem was the reduction of words, or elision, in fast speech. 1 respondent also indicated difficulty transcribing tone and a further respondent had problems with applying the concept of a word in languages where separate words are not widely recognised. Of the 6 responses, 1 was disregarded as it was not deemed to directly answer the question.

4.5 Allophones

In this section, participants were asked about the techniques used to identify allophones, along with the issues experienced and how these problems were overcome.

Respondents suggested the techniques which they had used in order to identify allophones. Of the 12 responses, 2 were discounted as their answers did not directly answer the question. Within the remaining 10 responses, some respondents detailed more than one technique which resulted in a total of 17 suggestions. These responses were re-categorised in order to collate the results. *Distributional analysis* was the most suggested technique with a total of 4 suggestions (24%), followed by *comparison with neighbouring language data*, *previous research* and *native speaker judgements*, along with *collecting target sounds in different environments* which each received 2 suggestions

(12%). 1 respondent (6%) suggested *extended exposure and use of the language*; whilst a further respondent suggested *using* the *standard approach*. Another respondent suggested *collecting paradigms of word forms*; whereas *suffixation* and *reduplification* were also proposed.

When asked whether they had ever misrepresented an allophone as a phoneme, 14 responses were received. Of these responses, the highest proportion of respondents (7 or 50%) indicated that they had; whereas 4 respondents (or 29%) indicated that they had not. 3 respondents (21%) selected other.

When taking into account the detail offered by the *other* respondents, one of the *other* responses can be re-categorised as *no*, which gives 5 *no* responses (36%); whereas another response can be re-categorised as *yes*, taking the total number of responses to 8 (57%). The remaining *other* answer is correctly categorised, taking *other* responses to 1 (7%).

Respondents were asked why they had misrepresented allophones as phonemes and vice versa. Whilst 10 responses were received, 2 were discounted due to a lack of relevance. The remaining 8 responses are represented in Figure (9) below.

Figure (9): Issues causing allophones to be misrepresented as phonemes and vice versa
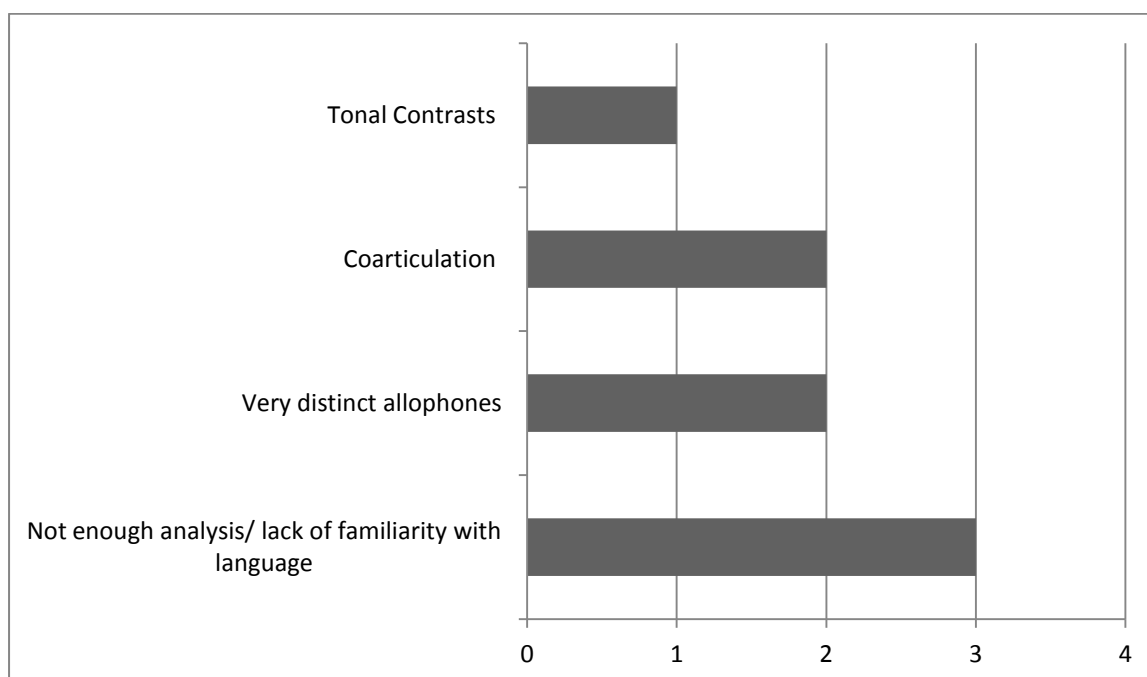
As shown in Figure (9), of the remaining 8 responses the highest proportion of 3 respondents (37.5%) suggested that *not enough analysis/ lack of familiarity with the language* were factors in them misrepresenting allophones as phonemes; whereas *very distinct allophones* and *coarticulation* were indicated by 2 respondents each (25%). 1 respondent (12.5%) said that *tonal contrasts* had caused difficulty.

When asked how they realised that they had made an error, 9 responses were offered. Of these 9 responses, 3 respondents (33%) realised that they had made an error due to *co-variation*. A further 2 respondents (22%) enlisted *distributional analysis* which highlighted inconsistencies, whereas 2 respondents (22%) realised their error through a *review of previous research*. The final respondent (11%) discovered their error through *native speaker judgement*.

The techniques used by respondents to establish whether two sounds are separate phonemes or allophonic variants are represented in Figure (10) below.

Figure (10): Techniques used to establish whether two sounds are separate phonemes or allophonic variants
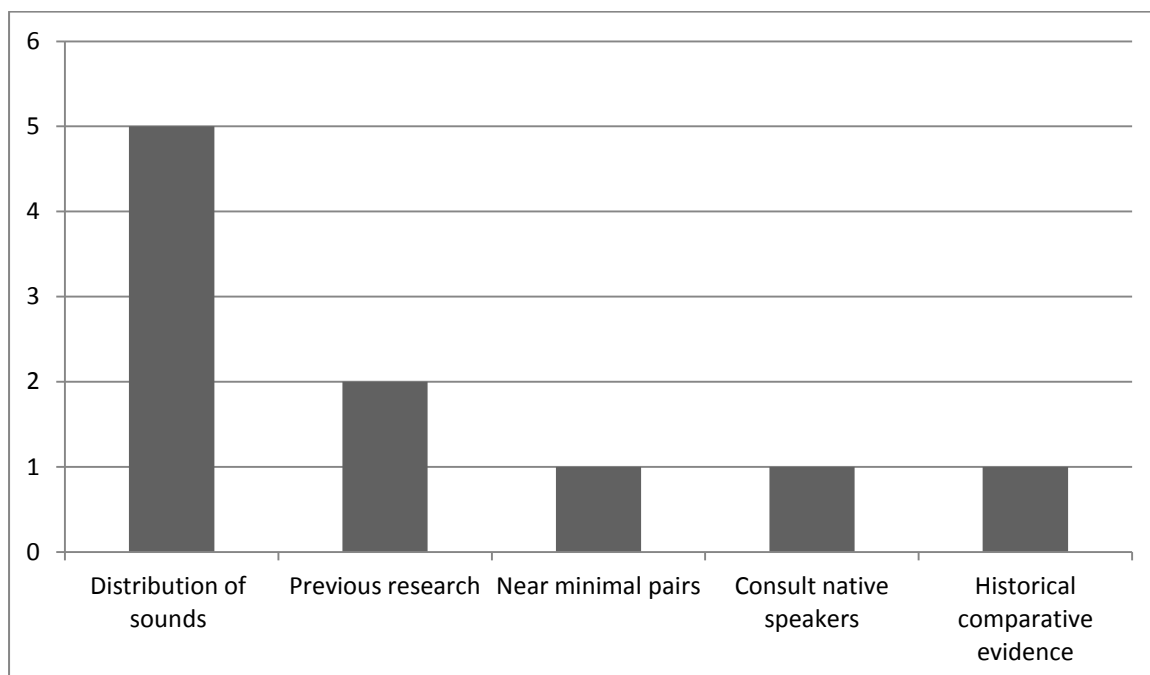
9 respondents suggested techniques to establish whether two sounds are separate phonemes or allophonic variants. 1 of the 9 responses was disregarded as it did not directly answer the question. Within the remaining 9 responses, 10 suggestions were made as seen in *Figure (10)*. Of the 10 suggestions, the most popular technique was *distribution of sounds* which received 5 suggestions (50%). *Previous research* received 2 selections (20%), whereas *near minimal pairs*, *consult native speakers* and *historical comparative evidence* received 1 suggestion each (10%).

Of the 16 potential respondents, 8 respondents (50%) did not experience any further problems with the identification of allophones in their research. However, 3 respondents (19%) did experience problems. 1 respondent (6%) selected *other* and indicated that *glides/off glides* were a problem for them. 4 respondents (25%) chose not to answer this question.

Although only 3 respondents said that they had experienced further problems with the identification of allophones in their research, 4 respondents stated further issues. Of the 4 respondents, 1 respondent experienced *problems with glides*; 1 respondent indicated that *fast speech* was an issue; 1 respondent had difficulties establishing *which allophone is underlying*; whereas the final respondent identified *rare sounds* as causing a problem for them during their research.

The ways in which participants overcame problems with allophones are represented in Figure (11) below.

Figure (11): How were problems with the identification of allophones overcome
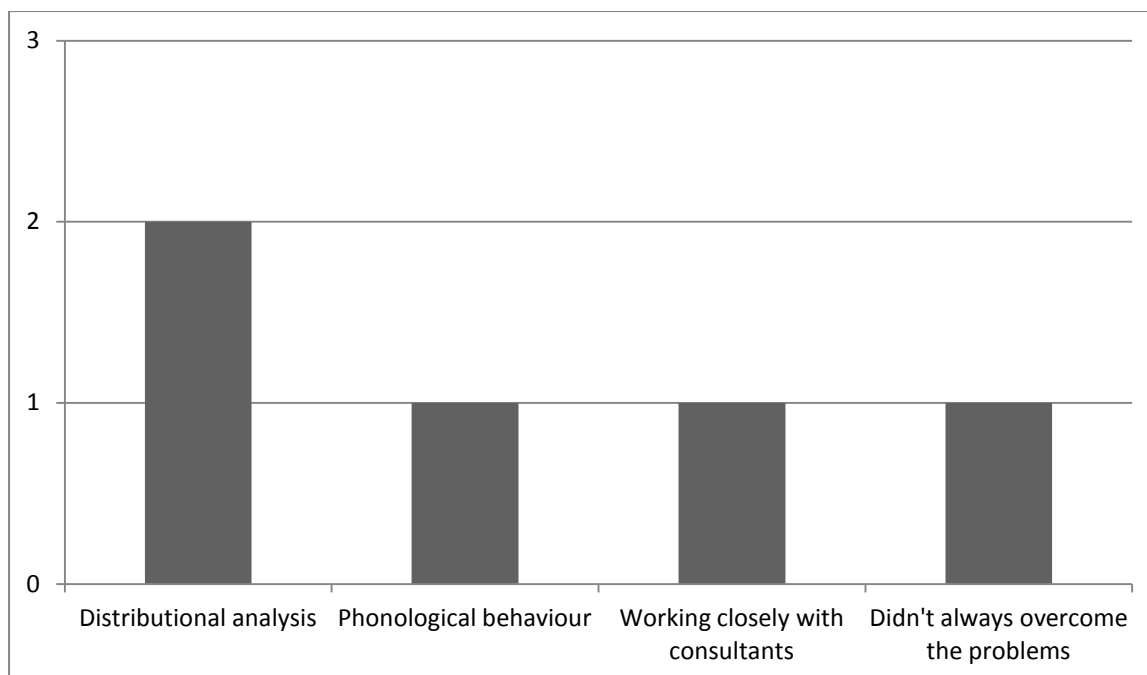
When asked how problems relating to the identification of allophones were overcome, 5 participants responded. 2 answers were disregarded as they were not deemed to contribute to the understanding of the area. Of the remaining 3 responses, 5 suggestions were made. Of these 5 suggestions, *Distributional analysis* was the most popular answer with 2 suggestions (40%). *Phonological behaviour, working closely with consultants* and *didn't always overcome the problems* were suggested once each (20%).

4.6 Variation and Metadata

In this section, participants were asked questions about the quantity of data collected, as well as the numbers of informants that were used. Participants were also asked about the controls that they had in place for speaker variation.

The quantity of data collected by survey participants varies broadly. As the wording of the question was extremely open, answers were also quantified differently by survey participants with some responding with answers which referred to how many gigabytes of data were collected, others referring to the number of speakers used and further respondents quantifying the data by number of hours of recorded speech. This data has been displayed in Table (2) below.

Table (2): Quantity of data collected by respondents

Respondent	Quantity of Data Collected			
	Gigabytes	Hours	Speakers	Tokens
1a[1]	50GB		5	
1b	10-20GB		1	5 per word elicited
2a		12		
2b		130		
3		Hundreds		
4a		50		
4b		5		
5a			12	
5b			24	
5c			24	
5d			24	
5e			42	
5f			12	
6		1-2 hours per day, per consultant		
7		100		
8				380
9	Several GBs			
10	Depends on the type of research			
11	Depends on the type of research			
12	Depends on the type of research			

[1] Instances of the same number indicate the responses of one participant. Letters are added to suggest responses which relate to further languages documented by the same respondent.

13	As much as possible
14	As much as possible
15	As much as possible
16	As much as possible

When asked how many speakers data was collected from, respondents were able to provide multiple responses to account for documentation of more than one language variety. 17 responses were received and are represented in Figure (12) below.

Figure (12): How many speakers was data collected from
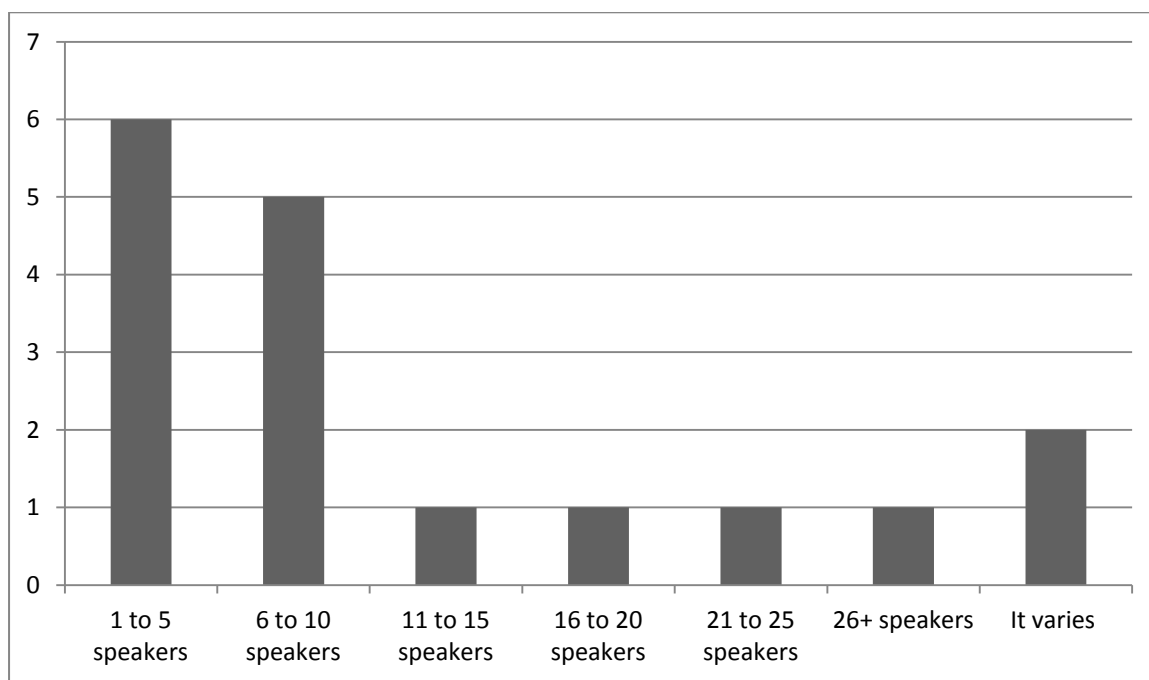
Figure (12) shows the number of native speakers that a researcher collected language data from. Participants could give multiple answers to account for each language that they had worked on. The most popular answer was 1-5 speakers, which was selected 6 times. This was closely followed with the 6-10 speaker category which was selected 5 times.

When collecting data from informants, 12 respondents (75%) indicated that they did not collect an equal amount of data from their informants; whereas 2 respondents (12.5%) said that they did. A further 2 respondents (12.5%) chose the *other* option. Of the 2 respondents that selected *other*,

1 participant explained that they collected an equal amount of data from each respondent "when possible", whereas the second participant said that, "For structured experiments, I collect the same amount of data from everyone. For elicitation and spontaneous data, I collect more data from people who interact with me more regularly than other members of the community".

Respondents were asked whether they had controls in place for speaker variation. Participants were able to select as many responses as they felt applied. Responses can be seen in *Figure (13)*.

Figure (13): Were there controls in place for speaker variation
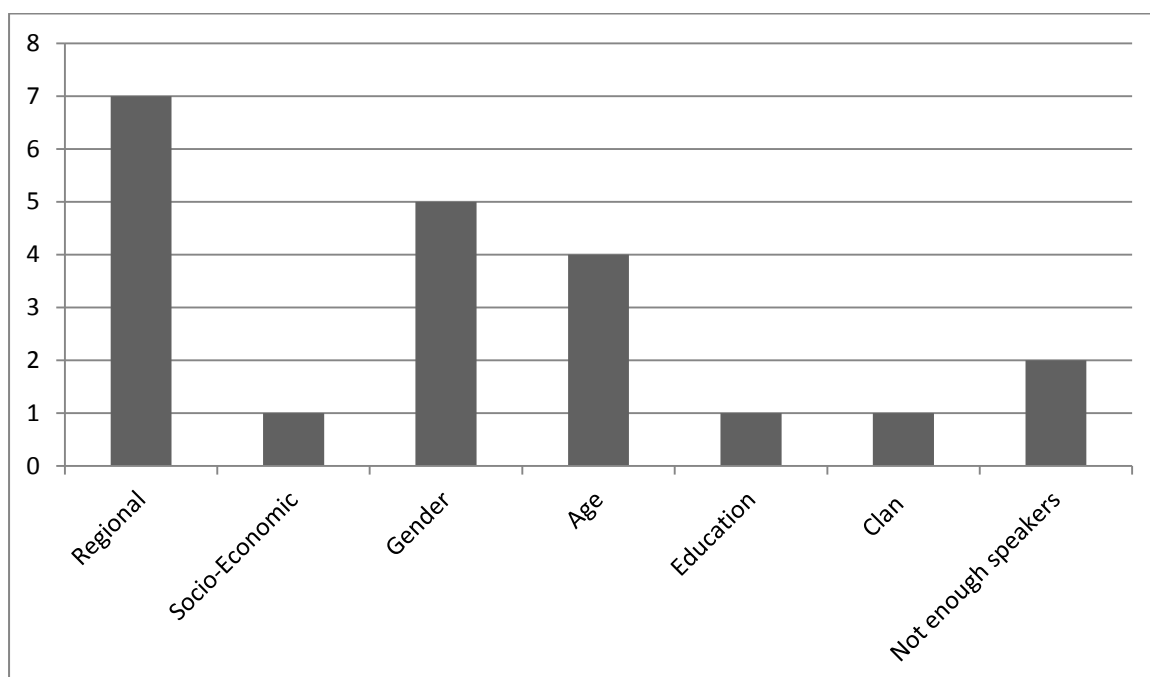
Figure (13) shows the areas in which respondents had controls in place for speaker variation. Respondents were able to select as many areas as applied. *Regional* was the most popular category with 7 out of a total of 21 selections (33%). This is followed by *gender* which has 5 selections (24%). *Age* received 4 selections (19%); both *education* and *socio-economic* received 1 selection each (5% each). This questioned allowed for further comments on each selection. 4 participants selected *other*, however their responses were re-categorised and are incorporated in *Figure (13)*.

Participants were given the option to offer additional feedback on the controls that they had in place for speaker variation. In terms of socio-economic controls, it was indicated that these were put

in place "when it made sense"; whereas for gender, comments were: "they were all male", "when possible" and "male vs. Female". Age was suggested to be "one of the hardest" variables to control.

4.7 Recording Methods

Respondents were asked which recording techniques they favoured in their research. Participants were able to offer as many techniques as they felt appropriate. Answers are presented in *Figure (14)*.

Figure (14): Recording Methods
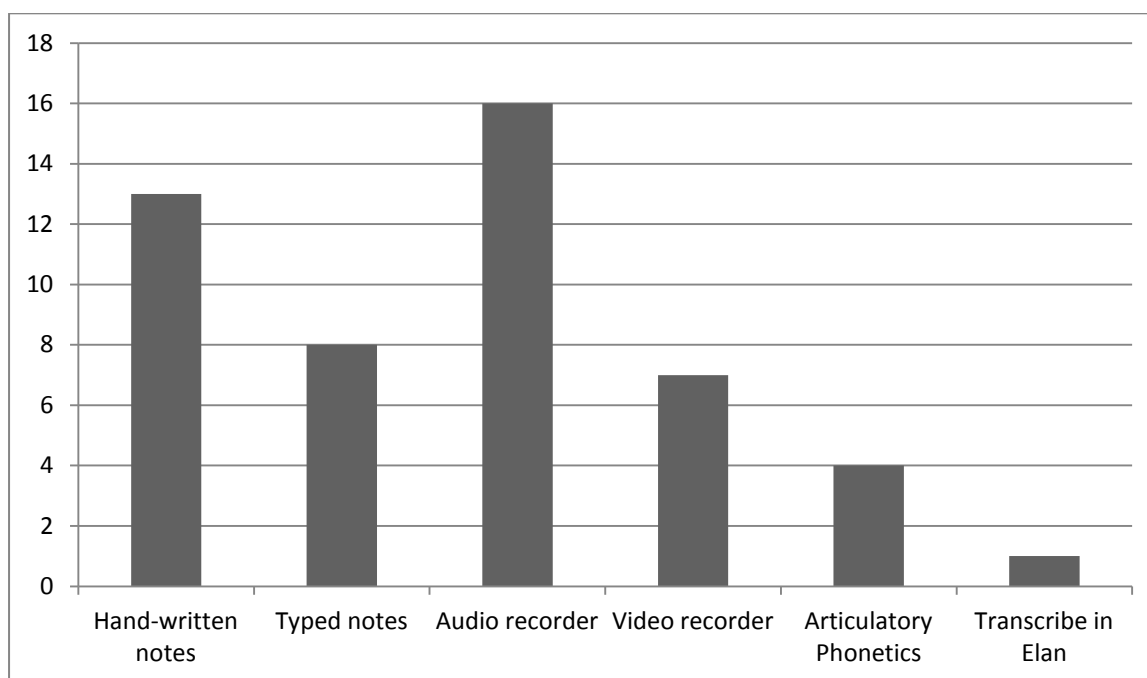
The most commonly selected recording method in *Figure (14)* is *audio recorder* which received 16 (33%) out of a total of 48 selections. This is followed by *hand-written notes* which received 13 selections (27%); *typed notes* received 8 selections (17%); *video-recorder* received 7 selections (15%) *articulatory phonetics* received 3 selections (6.25%) and *Transcribe in Elan* received 1 selection (2%).

Having analysed and presented the results, it is possible to identify a number of trends which can be used to guide field linguists in their documentation projects. These findings will now be explained in further detail in Chapter 5, the Discussion section, which is to follow.

5. Discussion

This chapter begins with an introduction explaining how the research has contributed to the field, where it fits in with the existing literature and the research gap that it addresses. This is followed with a discussion of the results, with a particular focus on the findings that are significant in helping to form guidance for future field research. Where possible, the results have been compared against the findings and opinions of the authors of some of the most influential guides to fieldwork including: Bowern (2008), Crowley (2007*)* and Chelliah & Reuse (2011). Using the insight offered by survey respondents, recommendations have been made in order to assist future field linguists in their documentation projects. This chapter is organised in sections which, for clarity, will follow a similar order and headings as both the survey and the results section.

5.1 Contribution to the field

Having identified a number of potential areas of difficulty for field researchers when collecting phonetic and phonological data, analysis of the survey results provided by 16 experienced field researchers has offered an insight into the methods that have been used to collect this type of data, the difficulties that are faced during data collection and the techniques that can be used to overcome these issues. Although multiple guides to linguistic fieldwork exist, these are primarily based upon the individual opinions and experiences of one researcher. Where several opinions have been collated, this has been in the form of individual chapters in an edited book whereby only one or two researcher's perspectives are presented. Therefore, this research addresses a gap in the existing literature for work which combines multiple perspectives on the collection of phonetic and phonological data by gathering the opinions of researchers in relation to identical survey questions. This allows for areas of both consensus and disparity to be identified which will help to guide future field linguists in their research.

5.2 Language, Informants and Elicitation Methods

Through the use of a survey, data was collected from 16 field researchers who had collectively worked on, in excess of, 39 languages. The survey method meant that the geographic scope of the research was vast, with data collected on languages from 6 continents; thus making the results as representative as possible. With the results also incorporating comments relating to the less explored area of tonal languages, the research fills a further gap in the literature for advice relating to the collection of data in this category of languages.

While 87.5% of respondents indicated that they had documented a non-native language; 93.8% stated that they had a language in common with their informants. This suggests a trend for documenting a non-native language and also for selecting a language where the researcher has a language in common with their informants. This could be seen to imply that, despite the scales of language vitality which are intended to indicate which languages are the most vulnerable (see chapter 2), the languages with the least speakers are less likely to get documented due to the likelihood of a language barrier between researcher and informant. Bowern (2008) recognises the benefits of having a language in common with the community that a researcher is studying, suggesting that not only does this help with elicitation between researcher and informant, but it also helps with interaction between researcher and the wider community (p133). Whilst the wider community may not be actively involved in the elicitation or data collection process, having an understanding of the general conversation that is taking place around the community will enrich a researcher's understanding of both the culture and traditions of the society, each of which have a significant impact upon the language itself. Knowing the language also affords a researcher greater control over the data, given that the researcher will have a better understanding and intuition as to what is feasible within the language (Bowern, 2008, p.9). Furthermore, Crowley (2007) is of the opinion that true ethnographic insight can only come through communication in the local language, not through a lingua franca or translator (p.preface), yet if a lingua franca or translator is the only option, then it would be preferable for a researcher to switch to monolingual elicitation as soon as they had learnt enough of the local language that they could communicate directly with the informant (p.93-94).

In terms of methods of elicitation, Figure (3), chapter 4, shows that wordlists were the most popular, closely followed by asking questions. This supports the observations made in, section 2.6.3.4 regarding the efficiency of word list elicitation and suggests a preference for structured techniques which can be planned in advance. In addition, planning to elicit data using a wordlist is made easier through the existence of pre-prepared wordlists such as the Swadesh list of basic vocabulary (Samarin, 1967, p.220), or other regionally specific lists. These pre-existing wordlists are useful resources which save researchers time when beginning their fieldwork. Furthermore, whilst useful as a starting point, the process of elicitation using a wordlist can also attract the interest of informants and open up opportunities for further data collection in semantic areas which are of particular interest. This type of approach can be useful when collecting large amounts of data as having a structured, pre-planned way of working can help to minimise distractions and the risk of going off task which, as a result, reduces the time that the elicitation process takes., Bowern (2008) agrees that wordlists are a good place to start when eliciting data on an unfamiliar language. Not only are word lists useful for providing information relating to the phonetic and phonological aspects of a language, but they are helpful in establishing word boundaries and familiarising a researcher with how best to transcribe the language. This also gives the researcher an opportunity to introduce informants to the elicitation process with a fairly simple task (p.35). Crowley (2007) also suggests wordlists as a useful starting point for collecting data on a language, though he suggests that veering away from the list can also be a positive if informants show particular interest in particular semantic domains (p.95). However, although there are many benefits to elicitation, this method can present issues of bias due to the choices that a researcher makes with regards to the topics and word categories that are to be asked about. For this reason, a comprehensive account of a language should be based upon both elicited and spontaneously produced speech (Bowern, 2008, p.115).

Although less popular than structured methods, the results of the survey show that spontaneous and semi-spontaneous methods were also useful. As discussed in section 2.6.3.2, this category of methods are likely to provide a more realistic representation of language use in the community and help to expose a researcher to common phrases or categories of words that occur

more naturally during speech. Furthermore, Bowern (2008) recommends making observations about the way that language is used outside of elicitation sessions as this may expose data that, for a number of reasons, would not have been picked up during elicitation (p.115). However, the lack of structure in spontaneous and semi-spontaneous data collection can make the process more time consuming, with no guarantee of much relevant data. The analysis of data collected in this way is also likely to take much more time which could help to explain why structured methods are a more popular choice amongst participants of this research.

5.2.2 Minimal Pairs

Minimal pairs were identified in section 2.7.1 as an area of difficulty in the collection of phonetic and phonological data during field research and this was confirmed based on the results of this research which showed that 50% of respondents had, at some point in their research, encountered difficulties with the identification of minimal pairs. In order to assist in preventing some of these issues, the methods for elicitation of this type of data and the problems encountered were explored along with the techniques that can be used to overcome these difficulties.

5.2.2.1 Methods of Elicitation

Figure (4), in chapter 4, demonstrates the techniques used by survey participants when identifying minimal pairs. The most popular technique was *asking for and comparing with words that sound similar*, followed by *trying free variation of sounds* and *inventing minimal pairs to test*. Each of these techniques incorporate using the judgement of native speakers, which is a technique also supported by Crowley (2007), who suggests pronouncing the pairs of words and asking the informant to verify whether the words were pronounced correctly. However, he points out, it is important to consider that informants may provide a positive response to indicate that they have understood what a researcher is trying to say; despite the pronunciation of the word being incorrect (p.97). Chelliah & Reuse (2011) recommend that native speakers are enlisted to assign meaning to suspected minimal pairs in order to verify whether they are indeed minimal pairs, or whether the sounds are in free variation (p.258).

5.2.2.2 Problems Encountered and Techniques Used to overcome issues with Minimal Pairs

In terms of the problems encountered whilst identifying minimal pairs, the complexity of language tone systems and suprasegmentals were the most frequently given answers. The term suprasegmental denotes a phonological feature that occurs alongside an utterance, such as stress, pitch or tone and, as these features effect the pronunciation of a phoneme, they can complicate the process of data collection.

As shown in Figure (5), chapter 4, in order to overcome the issues experienced with the identification of minimal pairs, the most suggested solution was, again, enlisting the assistance of native speakers, followed by the use of comparison words. However, more specifically, when overcoming the problems associated with the complexity of tonal language systems one of the suggestions was collaboration with experts in the research area. Calling on the expertise of another researcher can save a significant amount of time and also bring a new perspective to the data collection process. This approach is supported by Crowley (2007), who suggests that it is wise to consult an expert when dealing with topics in which a researcher is not confident (p.46). Furthermore, as mentioned, comparison words were also suggested as a technique for overcoming issues with identifying minimal pairs in tonal languages. The suggestion of one survey participant involved choosing various comparison words at different tone levels as a baseline for the tonal transcription; whilst whistling the tone of words was also suggested as a method to detect tonal differences. This approach is also proposed by Chelliah & Reuse (2011) however they warn that some informants will hum or whistle in a random fashion, which does not aid a researcher in detecting tonal differences (p.269).

With regards to overcoming issues with suprasegmentals, one respondent suggested paying particular attention to the way that native speakers corrected attempts at pronunciation, once again reaffirming the crucial role of native speaker judgements.

5.2.3 Affricates

Affricates were identified in section 2.7.2 as a potential area of difficulty during linguistic fieldwork and this was confirmed by 25% of survey respondents who indicated that they had, at some point during their research, misrepresented an affricate.

5.2.3.1 Anticipated Problems

The anticipated problem with affricates was the misrepresentation of an affricate as a consonant pair and vice versa. Respondents were asked whether this had been a problem for them during their research. As seen in section 4.3, 25% stated that this had been an issue during their research.

5.2.3.2 Methods of Elicitation

Respondents were asked which techniques they had used to determine whether a stop and fricative combined to make an affricate or whether they were a consonant pair. As shown in Figure (6), chapter 4, participants indicated several techniques. These techniques include syllabification, which involves breaking words down into syllables; phonotactic restrictions, which involves the identification of the patterns which govern the possible combinations of phonemes in a language; and distributional techniques, which involve checking "*like* consonants and vowels in a number of phonetic environments" (Chelliah & Reuse, 2011, p.259). Each of these techniques received two selections each.

5.2.3.3 Further Problems Encountered

19% of respondents indicated that they had encountered additional problems with the identification of affricates during their research.

The first issue indicated was that phonetically, stops can appear to be affricated in certain prosodic positions or in other, unanticipated, triggering contexts. Bowern (2008) agrees that stops can be easily confused with affricates, suggesting that palatal stops in particular can be problematic. Aspiration can also be an area of difficulty when dealing with affricates and is a feature that is often

missed when working on a language (p.41). A further issue was sociolinguistic variation in pronunciation of phonemes. Finally, germinate, or elongated, consonants were a problem for one researcher as this was not a known feature of the target language.

5.2.3.4 Overcoming the Issues

In order to overcome issues with affricates, a number of suggestions were made. Firstly, checking where the sound occurs as, as per the respondent's example, there are no complex onsets of the form stop-fricative, so in onset position the sequence must be an affricate. How the sound behaves when different morphological processes apply, for example, does the sound fill a single consonant slot in reduplication, was proposed as a method of checking whether the sound was an affricate. In addition, collecting more data may resolve the issue in some cases, as additional data may help to corroborate a theory, such as that the variation is sociolinguistic in nature for example. The help of other researchers is also suggested as a valuable method, along with eliciting consonants in a variety of prosodic and vowel environments rather than using a standard list.

5.2.4 Connected Speech and Parsing the Speech Stream

As identified in section 2.7.3, connected speech is likely to cause problems in the early stages of elicitation until a researcher becomes familiar with aspects of a language (Bowern, 2008, p.74); however, this type of data is valuable due to its more naturalistic representation of the way that language is used. Even when a linguist becomes more familiar with a language, fast speech, overlaps between speakers and slang words make this type of data particularly difficult to parse and transcribe (Bowern, 2008, p.119). As seen in section 4.4, 81% of respondents indicated that they had experienced a problem in this area during their research.

5.2.4.1 Methods of Elicitation

Respondents were asked which techniques they had used to separate the speech stream. As shown in Figure (7), chapter 4, the help of native speakers was the most suggested method which, again, highlights the importance of working with an informant with whom you are able to

communicate well with. Slower repetition, which also enlists the assistance of native speakers, was the second most popular option. With the help of native speakers and slower repetition accounting for seven selections (70%), this clarifies the importance of language informants in making the elicitation process a success.

5.2.4.2 Problems Encountered and Techniques Used to Overcome the Issues

As seen in Figure (8), chapter 4, the most frequent issues reported when parsing the speech stream relate to the continuity/ speed of speech along with limited exposure to the language, with each receiving five selections. In order to avoid some of the issues with connected speech, Chelliah and Reuse (2011) suggest asking informants to "read or paraphrase a written story", this method makes parsing the speech easier as there is a certain level of predictability (p.427). Taking this approach prior to attempting to parse spontaneous connected speech is likely to help a researcher to gain a deeper understanding of the language and make parsing entirely spontaneous connected speech more successful.

5.2.5 Allophones

As discussed in section 2.7.4, determining whether two sounds are phonemes, or whether one is an allophonic variant can be problematic (Bowern, 2008, p.38). Consequently, the techniques used to elicit this kind of data are important.

5.2.5.1 Methods of Elicitation

The most suggested method of elicitation was *distribution of sounds* which was put forward a total of five times, with the next most popular technique being using *previous research* which received two suggestions.

5.2.5.2 Problems Encountered and Techniques Used to Overcome the Issues

Four respondents indicated that they had experienced further problems with allophones. As shown in Figure (10), chapter 4, *distributional analysis* was the most popular method by which

researchers overcame issues with allophones, followed by *comparison with neighbouring language data*, *previous research* and *native speaker judgements*, along with *collecting target sounds in different environments* which were each suggested twice.

5.2.6 Variation and Metadata

As addressed in section 2.8, whilst collecting data on the phonetic and phonological aspects of a language a number of considerations have to be made with regards to both the selection of the informants themselves and also the quantity of data to be collected.

5.2.6.1 Informants and Data

As shown in Figure (12), chapter 4, using 1-5 speakers was the most common response followed by 6-10 speakers. The results of the survey demonstrate that the number of speakers used in the collection of data varies broadly, with answers ranging from one speaker to over 50 speakers. This does not appear to be an area of agreement for a number of reasons. There are many factors to consider during the selection of informants for field research including candidate suitability and how many speakers will be required, the age, gender and sociolinguistic status of informants; how many speakers exist in the location, how many will be suitable for the task and then how many of the suitable candidates will be asked to help in the collection of language data.

Whilst, ideally, the target may be to collect data from as many informants as possible, in communities where the language variety is endangered there are a number of factors which prevent this from being possible. Firstly, there may only be a handful of speakers of an endangered language left throughout the world or in that particular community. Furthermore, there may be restrictions on working with community members of a particular age or gender (Bowern, 2008, p.134; Crowley, 2007, p.90; Chelliah & Reuse, 2011, p.170-171), which further reduces the number of eligible informants. Or, where there are more speakers, there may be hostility towards outsiders and unwillingness from some members of the community to share their language and knowledge. Lack of literacy is also an issue in some areas, whilst an ageing population may also cause difficulties. If the

only informants available are old aged, they may be hard of hearing, unclear articulators due to a lack of teeth or simply not well enough to sit with a researcher for long periods of time (Chelliah & Reuse, 2011, p.169-170).

From the results of the current survey, another factor that influenced the number of informants enlisted to help with the field research was the area and method of data collection. For example, one researcher specified that they had used five speakers for phonetic purposes and 25 for documentation purposes; whereas another researcher said that they had used four speakers for general data collection and 10 speakers for a task on colour naming. Bowern (2008) agrees that the number of speakers required varies by task. For phonetic/phonological elicitation it is preferable to conduct elicitation one-on-one, where noise and distractions can be kept to a minimum. However, Bowern (2008) concedes that, if there are only a small total number of speakers of the language remaining, it is advisable to work with as many of those as possible (p.135). Chelliah & Reuse (2011), argue that, for phonetic fieldwork, multiple speakers of both genders should be recorded (p.175).

Crowley (2007) acknowledges that for ease, a researcher may wish to choose just one informant to work with; however, he points out that it is important to keep in mind that this person will have other commitments such as family and community roles, so they may not always be available when required. For this reason, he suggests arranging to work with several informants so that this does not become a problem (p.89).

The data collected can be measured in a number of different ways, dependent upon the recording method. Some methods, such as audio or video recordings can be measured in hours of gigabytes, whereas word lists can be measured in the number of tokens elicited. For this reason, the answers relating to the amount of data recorded by participants of the research varies and, if I were to conduct this research again, the question would be reworded to be more specific. This would have allowed me to obtain more comparable data in this area.

When asked whether they had collected an equal amount of data from each informant, 75% of respondents indicated that they did not. Whilst collecting an equal amount of data from each

informant would be preferable, this is not usually possible for a number of reasons. Firstly, if using unstructured techniques, like conversation, it is difficult to elicit the same quantities of data from person to person due to the unique nature and spontaneity of conversation. On the other hand, when using word lists or more structured techniques this would be easier to achieve. Other factors influencing the collection of data include how much time and patience an informant has. Whereas one informant may be happy to devote hours of their time to helping a researcher, others may not be as willing and may become bored or disillusioned within only a few hours. Bowern (2008) suggests that with wordlists, researchers need to collect enough tokens of each word that the results are statistically significant, but not so many that informants become bored (p.65).

5.2.6.2 Controls for Variation

Whilst collecting language data it is important to be aware of factors which may influence that data such as the sociolinguistic background of an informant, regional variations, gender, age, education and levels of literacy along with membership of groups such as different clans. As a result of the survey, it was shown that regional was the most popular area for a researcher to have controls in place with 33% of respondents saying that they had accounted for this. This was followed by gender which received 24% of selections. This was closely followed by age with 19% of selections. This suggests that these are categories whereby researchers feel that speaker variation has the most impact.

As discussed in section 2.8.1, regional variation can have a profound effect upon data. For example, if two informants have grown up in two different parts of a country, yet speak the same language, there may still be some differences. Pronunciation is likely to differ, along with some elements of vocabulary. Bowern (2008) notes that semantic domains are often specific to a region, particularly with regards to flora and fauna (p.109). Therefore, differences in regional background should be noted when collecting data from more than one informant.

Further to the discussion in section 2.8.3, gender can also have a significant effect upon results, with women and men often having very different life experiences and, therefore, differing

ways of using a language. In countries where men typically travel for work and women remain within the community raising the family it is often the case that women are much more conservative speakers (Bowern, 2008, p.134). In this situation men are likely to be exposed to a much broader vocabulary which covers different semantic fields from women, for example, trade (Chelliah & Reuse, 2011, p.171). Also, with increased travel comes exposure to other communities and languages, which may have an impact on pronunciation. Even in situations where men don't travel, they are likely to have different knowledge and tasks to women which may increase their linguistic awareness in semantic domains such as hunting or farming. In addition, in some communities men may be more likely to receive an education which would therefore increase their literacy and, as a consequence, their vocabulary. Education and literacy levels are likely to impact upon both the way that informants speak and pronounce words, possibly their accent and also their understanding of the rules of that govern their language and also the instructions and terminology used by the researcher. In addition, the vocabulary of a person who has received an education is likely to be wider, with an awareness of words in different semantic domains.

As discussed in section 2.8.2 age is another factor which impacts upon vocabulary use. Whilst elders are often the most respected within a community due to the wisdom that they have generated, their health may prevent them from being suitable informants – particularly when the goal is to elicit phonetic and phonological data. The loss of teeth and hearing amongst the older generation could lead to confusion with regards to the task at hand and also difficulties with clear pronunciation (Chelliah & Reuse, 2011, p.164). On the other hand, the younger generation are more likely to be healthy, with competent hearing and a full set of teeth which enhances their linguistic control and improves articulation. This makes differentiation between phonemes easier for a researcher. Using a combination of informants from different age groups may help a researcher to garner the benefits from each group however, for the purposes of phonetic and phonological data collection, a younger adult may be preferable.

5.2.7 Recording Methods

An audio recorder was the most popular choice of recording method with 16 out a total of 49 selections (32%). Having a recording of language data allows a researcher to return to the data an unlimited amount of times to verify its accuracy and also, providing that the correct ethical consent has been obtained, creates the possibility of retaining and backing up the recordings for future generations. Hand-written notes were the second most popular method with 13 selections (26.5%). This method allows for notes to be made with little reliance on technology. However, if notes alone are used to record language data it is difficult to verify the accuracy of the work at a later date. Notes are also difficult for a different researcher to understand and analyse in the future.

5.3 Recommendations for future linguistic fieldwork

In keeping with the motivation of this research, which was to assist those undertaking future phonetic and phonological linguistic documentation projects, the survey feedback has been analysed and the following recommendations have been derived from the results. These recommendations will then be discussed in further detail in the following section.

Recommendation 1: Data collection will be easiest when working on a language where there is a common language between researcher and informant. However, if working on a variety where there is no common language, it is important for a researcher to be transparent about the limitations of their work.

Recommendation 2: Suggested elicitation techniques in order of most popular:

- Ask questions

- Spontaneous speech

- World lists

- Conduct a task

- Storyboards

- Readings

- Words embedded in carrier phrase/ sentence frame

- Single tokens for palatography

- Felicity judgments

- Showing photos with items to name

- Psycholinguistic experiments

- Narratives/stories

Recommendation 3: Use a combination of spontaneous and non-spontaneous methods of elicitation.

Recommendation 4: Suggested techniques for the identification of minimal pairs:

- Ask native speakers for suggestions words that rhyme or sound similar

- Try free variation of sounds

- Invent minimal pairs to test

- Use existing resources such as research and dictionaries

- Comparison of growing lexicon

- Explore words which begin with or contain a particular phoneme

Recommendation 5: Suggested techniques for overcoming difficulties with the identification of minimal pairs:

- Practice and habituation.

- Pay attention for cues other than general formant values, e.g. internal dynamicity of one vowel and not the other.

- Collaborate with researchers who specialise in this area.

- Find tonal minimal pairs by choosing comparison words at different tone levels as a baseline for the tonal transcription.

- Whistle and/or ask speakers to whistle the tone of a word to detect differences.

- For vowels, use reference words whose vowels are established to help.

- Native speaker judgements, including corrections made.

- Train consultants in basic linguistic analysis.

- Use near-minimal pairs.

Recommendation 6: Techniques which can be used to determine whether a stop and fricative combine to make an affricate or whether they are a consonant pair:

- Observe the behaviour of stop and fricative sequences for all points of articulation, i.e., for the entire natural class of stops as well as the entire natural class of fricatives.
- Use distributional techniques including free variation.
- Reduplification.
- Syllabification tasks.
- Phonological behaviour.
- Use previous research.

Recommendation 7: Techniques which can be used to separate the speech stream:

- Record the speech, then replay to the informant and ask them to repeat the speech slowly.
- Annotate the speech with the help of native speakers and ask about word boundaries.
- Use software such as ELAN, Praat and FAVE to assist.
- Using the morphology of the language.
- Change the order that 'wordlike' chunks appear in and use intervening words to see if they are interruptible.
- Use prosodic patterns to help in identifying words.
- Use a dictionary if available.

Recommendation 8: Techniques to identify the allophones of a phoneme:

- Phonological distributional analysis
- Similarity of allophones to one another
- Previous research about the language in question and related languages/ comparison with neighbouring languages.
- Extended exposure to the language.

- Speaking to native speakers using their language

- Collect target sounds in as many environments as possible. Eg. word position, prosodic

  position, before and after a variety of segments.

- Collecting paradigms of typical word formation eg. nouns, verbs etc.

Recommendation 9: Suggested methods for recording language data:

- Audio recorder

- Hand-written notes

- Typed notes

- Video recorder

- Articulatory phonetics - Static palatography/ Ultrasound/ EMA

- Speech analysis software such as ELAN

5.4 Discussion of recommendations

In this section each recommendation will be discussed, along with the survey feedback that led to its inclusion. When discussing a recommendation which consists of a list of suggested methods, all relevant methods put forward by survey respondents have been included. This way, the collective experience of those who have undertaken linguistic fieldwork can be combined and used as a guide for future field linguists. Whilst the methods suggested have both strengths and limitations, by presenting all methods together, a researcher can make a decision as to which technique is best suited to their situation. In addition, where a technique does not produce the desired results, a researcher can revert to the list for advice on an additional method.

5.4.1 Recommendation 1

As seen in section 5.3, recommendation (1) relates to the selection of an informant and the frequency with which the researchers and informants have a common lingua franca. In section 4.1 the proportion of respondents that documented a language where there was a language in common between researcher and informant was discussed. Although 14 respondents indicated that they had a

language in common with their informant, when taking into account the comments relating to this question, this figure is in fact 15 of the 16 respondents (94%). This result suggests a trend for researchers working with communities where communication does not pose an additional barrier to their data collection and, for this reason, is an important factor to be considered by a researcher when deciding upon a destination in which to undertake linguistic fieldwork (Crowley, 2007, p.58). The frequency with which researchers selected a language where they had a language in common with the informant indicates that the documentation process is easier this way and is likely to result in a quicker and more accurate documentation process. However, on the other hand, this suggests that those languages which have the smallest number of speakers, are the least likely to be documented and puts these already vulnerable varieties at increased risk.

5.4.2 Recommendation 2

The results of the survey revealed a number of methods for the elicitation of data during field research. Whilst some suggested methods were provided in the survey, respondents also put forward their own suggestions resulting in a bank of methods which will serve as a help to future field linguists. The number of different methods selected by respondents suggests that it often takes a combination of techniques to elicit the data required and, furthermore, the most appropriate method varies both according to the type of data that is sought, and the strengths and preferences of both the researcher and informant. Having in place an inventory of suggested methods will help to serve as a prompt for a researcher when one particular method is not producing the desired results.

5.4.3 Recommendation 3

Using a combination of spontaneous and non-spontaneous techniques has a number of advantages. Whilst both techniques are useful for eliciting large amounts of data, non spontaneous techniques such as word lists tend to offer the ability to gather more precise data in a shorter space of time. Spontaneous techniques on the other hand offer opportunities to capture detailed data, which offers a more in depth insight into the language in question. These techniques can also present unexpected data which could have been missed if relying solely on non-spontaneous techniques.

However, whilst the additional data generated by spontaneous techniques is useful in establishing a more in depth understanding of the language, this additional data is likely to be time consuming to analyse. Non spontaneous techniques on the other hand are, in general, quicker and easier to analyse, as the data collected relates to a particular semantic domain, or is in answer to a specified request for information.

5.4.4 Recommendation 4

The survey showed that a number of techniques were successfully enlisted whilst identifying minimal pairs and suggests that, again, it may take a combination of methods to elicit the required data. Combining the input of native speakers with the findings of previous research and other recognised techniques of elicitation offers a number of options for a researcher when eliciting data of this kind. Using more than one method can also provide a way of checking the data. For example, using the judgement of native speakers to verify the accuracy of a researcher's intuitions can help to ensure that the results of the research are as accurate as possible.

Asking for words that rhyme or sound similar was suggested by respondents as a method for identifying minimal pairs. This makes use of the knowledge of native speakers, without the informants needing to possess a high level of linguistic knowledge and is likely to generate a number of minimal pairs relatively quickly. However, the minimal pairs produced are most likely to be those that contrast in word-initial position, as these tend to be the most easily recognised by speakers (Chelliah & Reuse, 2011, p.258).

Free variation of sounds entails checking whether two sounds can be used interchangeably without either altering the meaning or making the word incorrect. If they can be used interchangeably without a change in meaning, this would indicate that these sounds are allophones of the same phoneme; whereas, if they can't, this likely denotes a minimal pair.

Another recommended technique is inventing minimal pairs to test. When a researcher has become familiar enough with a language, it is likely that possible minimal pairs can be predicted.

These combinations can then be verified with the informant to establish which are valid in the language.

If previous research has already been conducted on the language, it is likely that some information on minimal pairs already exists. Utilising this research along with any existing dictionaries can save a researcher valuable time during the data collection process.

As research progresses, comparison of the growing lexicon can reveal possible minimal pairs which can then be tested. Exploring words which begin with or contain a particular phoneme is a systematic way of approaching this.

5.4.5 Recommendation 5

The recommended techniques for overcoming problems encountered with the identification of minimal pairs incorporates enlisting the help of other researchers and native speakers, suggesting that it may be wise to seek the opinions of native speakers following the initial elicitation in order to prevent problems occurring further into the research process. However, if problems do occur, there are a number of suggested ways to overcome these.

Practice and habituation is, in a variety of scenarios, one way of overcoming problems. The more familiar a researcher becomes with a both a language and the techniques used to elicit data, the easier it is to identify combinations which are likely to be correct and, also, to recognise those which do not follow with the other patterns in the language.

If conducting acoustic analysis, paying attention for cues other than general formant values can be useful. For example, internal dynamicity of one vowel and not the other would suggest that two words could be minimal pairs.

Collaboration with other researchers who specialise in an area is always a good idea, particularly when encountering issues. Not only does this bring a broader range of expertise to the data collection process, but this also offers a second perspective on the data meaning that errors are likely to be identified more promptly.

Once research has resulted in the identification of some of the vowels/minimal pairs in the language, these vowels/minimal pairs can be used as a reference point for the comparison of new suspected vowels/minimal pairs. For example, using the minimal pairs 'pin' and 'pan', if these words have been identified as minimal pairs in the language and the vowels 'i' and 'a' have been verified as separate phonemes, these examples can then be used as a comparison to try and establish new vowels. If the researcher suspected 'e' to be a vowel, this can be compared to the existing vowels 'i' and 'a'. Furthermore, the newly identified vowel can then be entered into the vowel position in the existing minimal pair to make 'pen'. The researcher can then verify the new word and vowel with their informant.

Again, native speaker judgements are deemed to be extremely valuable in the correction of errors. This is often a quick way of checking data as a native speaker will usually recognise intuitively if something is not quite right and, even if they can't explain this using linguistic terminology, they will usually be able to offer some suggestions as the how a word should be correctly represented.

Taking into account the value of native speakers in the data collection process, it may be advisable to train consultants in basic linguistic analysis. This way, the important role of native speakers can be further enhanced. Not only can informants understand the reasoning for certain questions, but they can also be more informative when explaining why a linguistic item is incorrect.

Where minimal pairs cannot be found, the use of near-minimal pairs can be used to establish contrasts within the language.

For tonal minimal pairs, researchers may come across challenges which require a slightly different approach. In this situation, researchers can find tonal minimal pairs by choosing comparison words at different tone levels as a baseline for the tonal transcription. It is also possible to detect tonal differences by whistling and/or asking speakers to whistle the tone of a word.

5.4.6 Recommendation 6

In order to establish whether a stop and fricative combine to make an affricate or whether they are a consonant pair, respondents made a number of recommendations.

Firstly, observing the behaviour of stop and fricative sequences for all points of articulation, i.e., for the entire natural class of stops as well as the entire natural class of fricatives. By exploring how the sounds work individually it should then be easier for a researcher to ascertain whether the sound in question is a consonant pair or an affricate.

Using distributional techniques can be useful in determining how to categorise sounds. For example, if a phonetic stop and fricative appear in an onset, it may be assumed that the phonemes form part of an affricate. In addition, observing free variation in the production of sounds can be a helpful technique. If two sounds can occur in the same environment without altering the meaning of a word, they can be considered to be one phoneme which is in free variation; whereas if the substitution of one phoneme by the other changes the meaning or sense of the word, the phonemes are a consonant pair.

From a morphological perspective, looking at instances of how the sound behaves when different morphological processes apply can be insightful. For example does it fill a single consonant slot in reduplification - a morphological process whereby a prefix or suffix are in some way reflective of the root word due to repetition (Crystal, 2008, p.293). This can help to ascertain whether the sounds always occur together and form an affricate or, whether they are actually a consonant pair.

Syllabification tasks involve the process of breaking down words into syllables (Crystal, 2008, p.338), and was suggested as a technique through which a researcher can determine whether a consonant pair and fricative combine to make an affricate or whether they belong to a consonant pair. If the stop and fricative can be broken apart and occur in separate syllables, the sounds can therefore be classified as a consonant pair rather than an affricate.

Paying close attention to phonological behaviour can be insightful. For example, fricatives cannot be glottalised but affricates can. Therefore, when a stop-fricative combine and can be glottalised, we

can identify them as an affricate. In addition, observing the phonotactic restrictions within a language offers insight into permissible constructions.

Previous research may have explored the topic and may help to inform the current research. If previous research doesn't exist, there may have been work conducted on related languages which could offer some insight.

5.4.7 Recommendation 7

Parsing the speech stream of an unfamiliar language can be one of the most challenging aspects of the speech analysis process which can become further complicated by fast or unclear articulation. However, there are a number of techniques which can be adopted to make the process easier.

In order to counteract the effects of fast speech, the speech can be recorded and played back allowing a researcher to pause and repeat speech an infinite number of times. The recording can also be played back to an informant, so that they can then be asked to repeat the speech more slowly.

Another technique which enlists native speaker judgements is the annotation of speech where informants can help to identify where they believe word boundaries to be. Using the valuable instincts of an informant in this way, can greatly reduce the time that a researcher takes to identify possible boundaries.

Speech analysis software such as ELAN, Praat or FAVE can assist in identifying word boundaries and with speech annotation and analysis.

Using the morphology, or typical word structures, of the language can help in identifying a likely beginning, middle and end of a word, therefore helping to establish possible word boundaries.

Similarly, using prosodic patterns can be of help in identifying word boundaries. It may be that words tend to end in a particular way, perhaps with a pause or a change in stress patterns which alerts a researcher to a likely transition between words.

Where words boundaries are suspected, changing the order that the 'wordlike' chunks appear in can help to establish whether word boundaries have been identified correctly and, in addition, the use of intervening words identifies whether the suspected words are interruptible.

If a dictionary is available, this can be used to look up suspected words and clarify where boundaries occur.

5.4.8 Recommendation 8

Looking at phonological distribution can help in identifying the allophones of a phoneme. If the various phonemes can occur in the same place in a word without changing the meaning, this would suggest that the sounds are allophones of the same phoneme. On the other hand, if the words cannot appear in the same place in the word without changing the meaning, the sounds are separate phonemes which are in contrastive distribution.

Whilst not a technique or method as such, researcher instincts about similarities of sounds are often a good indication of sounds which may be allophonic and require further confirmation. Extended exposure to the language is invaluable in gaining an accurate picture of aspects of a language, including allophones.

Consulting previous research is always wise when working on a language. As the phonetics and phonology of a language are one of the first things that tend to be investigated, the likelihood is that if previous research exists it will contain some useful information in regards to the language's sound inventory. Similarly, information on neighbouring languages may be of use as due to the proximity between the languages it is likely that some features are shared between the languages.

Speaking to native speakers using their language requires practice; however, this is a good way to gain insight into the language. Where mistakes are made, the feedback and corrections from native speakers can offer valuable information about the language and enable a researcher to achieve a much deeper understanding of the language and the way it is used. In relation to allophones, a speaker

would become familiar with the sounds which are used interchangeably and have a greater instinct as to which sounds are allophonic.

Collecting target sounds in as many environments as possible, for example word position, prosodic position, before and after a variety of segments, can help to show how phonemes change based on their environment and can help to identify allophonic variants.

Collecting paradigms of typical word formation eg. nouns, verbs etc can help to highlight similar phonemes.

5.4.9 Recommendation 9

Survey respondents suggested a number of methods for recording language data. Whilst the recommendations are useful as a quick reference, there are a number of additional factors to take into consideration when determining which method is most suitable including the nature of the project, the budget and timescale. The methods suggested include electronic and pen and paper methods. With electronic data, there are a number of considerations to be made. Firstly, the budget will help to determine what level of equipment can be used. If the budget is tight, note-taking may be the only viable option. Secondly, future access to the data must be considered. Bowern (2008) points out that technology can pose problems as, due to the speed at which technology advances, some mediums can quickly become obsolete making accessing the data more difficult in the future (p.16-17). Therefore, it is important to consider how likely the technology is to be accessible to others in years to come. In addition, the recording equipment must be sturdy enough to function under the conditions of the field site, for example extreme weather conditions. Adequate protective equipment should also be considered, as well as back-up devices (Bowern, 2008, p.18-19). Furthermore, electricity may not be available to charge electronic devices, so spare batteries and alternative methods of recording data must also be available.

Audio recordings are beneficial as the recordings can be revisited for analysis, helping to ensure that transcription is as accurate as possible. Providing that the correct ethical procedures are

followed, the recordings can also be accessed by others wishing to verify the accuracy of the transcriptions or wishing to study the language themselves. Bowern (2008) recommends using a digital recorder, as the quality of the recording is superior to analogue recording equipment (Bowern, 2008, p.17; Crowley, 2007, p.74).

Another suggestion was using a video recorder. Video recorders are useful for capturing extra-linguistic cues such as hand gestures, facial expressions and deictic markers (Bowern, 2008, p.24). These additional non-linguistic clues can help to enhance a researcher's understanding of a situation or story and make the transcription process quicker, simpler and more accurate. However, one of the limitations of video recordings is often a reduction in the quality of sound which, when recording data for phonetic and phonological purposes, would make video recordings alone unsuitable. For this reason, it is advisable to use either a microphone attachment to enhance the sound quality or to combine the video recording with another complementary method of data collection.

Typed notes can be used as a method of recording language data and are often quicker to make than the written variety. This type of record can be made using a laptop or hand-held personal digital assistant device (PDA) which is more easily transportable.

Despite modern technology offering many advantages to the fieldworker, hand-written notes can also be invaluable. Hand-written notes don't rely on batteries or other sources of power, are easily transportable and are often less intimidating for informants who may not be familiar with electronic equipment. Bowern (2008) also points out that it is easier to draw diagrams or illustrations using this method, rather than electronically (p.27). If an informant and researcher are struggling with communicating a concept, a drawing can be a good way of bridging the language barrier.

When conducting research into the phonetics and phonology of a language, it may be useful to find out more about the way that the phonemes are articulated. This can be achieved in a number of ways: using static palatography – a technique which determines where contact occurs between the tongue and palate (Chelliah & Reuse, 2011, p.262); ultrasound, which can be used to record tongue movements during articulation (Chelliah & Reuse, 2011, p.254) or electromagnetic articulography is a

further method by which the movement of articulators can be tracked using sensors. By looking at the way that phonemes are articulated, this may help in differentiating similar sounding phonemes.

Whilst not strictly a method of collecting data, speech analysis software is an important part of the transcription, organisation and processing of linguistic data. By combining the data in a searchable database, the information becomes easier to categorise and manage. There are a number of programmes, each with differing functions, to choose from including ELAN, PRAAT, TRANSCRIBER and EXMARaLDA (Chelliah & Reuse, 2011, p.219).

5.5 Significant observations

While a host of valuable recommendations and advice have been presented in this section, some consistent observations have been identified. Throughout all of the areas of phonetic and phonological elicitation that were explored, native speakers were suggested to be of utmost importance. As a small number of native speakers are, in many cases, the only source of information on a language, maintaining a good working relationship between researcher and informant is vital. As a result, the process and decisions that are made when selecting an informant, along with the ways in which informants are retained, becomes a crucial consideration.

It also becomes apparent from analysis of the results that, whilst there are a number of suggestions and opinions on best practise in the field, these ideals are not always possible to obtain. For example, when recruiting a language informant for phonetic and phonological fieldwork, the ideal candidate would have a full set of teeth; however, in some communities the only remaining speakers are old aged and are unlikely to have retained all of their teeth. Furthermore, collecting equal amounts of data for each informant, although desirable, is not always possible. Therefore, in some circumstances, a researcher must be prepared to work with what is available in order to develop as much knowledge as is feasible within the constraints of a fieldwork situation.

Furthermore, observing the number of suggested methods it is apparent that the successful elicitation of phonetic and phonological data often takes a combination of methods, with the success

of each depending on the appropriateness of the method and also the strengths of a language informant.

6. Conclusion

The aim of this research has been to bring together the experience of field researchers who have collected and analysed data relating to the phonetic and phonological aspects of language during linguistic research. Through doing so, a series of recommendations have been formulated to assist future field linguists in the efficient, and timely, completion of their linguistic fieldwork. Most importantly, the advice and recommendations suggested in this research are intended to help in the preservation of endangered language varieties and thus facilitate continued linguistic diversity through the provision of lasting records of language data. This research complements existing guides to linguistic fieldwork, by filling a research gap for recommendations and advice which are not just based on one point of view, but combine the opinions of experienced field researchers in relation to phonetic and phonological data collection.

Chapter 1 introduced the research agenda and provided a brief overview of the research motivations, including the situation of language endangerment and the diminishing linguistic diversity that is of current issue throughout the world. Anticipated areas of difficulty in the collection of phonetic and phonological data collection were also outlined and the structure of the thesis was explained.

In chapter 2, language endangerment, the importance of language documentation and phonetic and phonological research were examined in further detail through a review of the existing literature. By exploring existing research, areas of both consensus and disagreement were identified, along with areas of difficulty. This review helped to inform the current research by highlighting areas of difficulty within phonetic and phonological data collection, and helped to establish those areas where researchers would benefit from advice. Reviewing the existing literature also enabled me to establish a gap where this research could be of particular value to the linguistic field.

Chapter 3 described the process through which the research evolved including how the motivations for the research influenced the method used, the process of obtaining ethical approval for the project and the progression from designing through to conducting the survey. Finally, the methods used to analyse the results were described.

The results of the survey were presented in chapter 4, along with a brief description of what those results showed in preparation for a more thorough discussion in chapter 5.

In chapter 5, the discussion, the results were explored in further detail and a number of trends were discussed. Following this discussion, a set of recommendations were suggested to help inform future linguistic documentation projects. Finally, each of these recommendations were explored in detail in order to justify their inclusion.

Serving as a guide of tried and tested methods, future field linguists can refer to the recommendations which have been made within this research with the knowledge that they are utilising the advice of 16 researchers who are knowledgeable in the collection of phonetic and phonological data collection. As a result, linguists have a pre-prepared list of methods and advice which will help to keep research projects moving forward, even when encountering difficulties. These recommendations will help to serve as a prompt where difficulties do occur, by offering guidance and alternative methods which will enable fieldworkers to complete their project. As a result of the guidance offered, it is hoped that an increasing number of languages can be successfully documented.

In terms of suggestions for future research, this project would benefit from replication on an even larger scale to incorporate the opinions of a broader number of researchers with experience of working on additional language varieties. Further research could also act as an extension of this work and cover other linguistic sub-fields such as morphology and syntax in order to work towards the creation of a comprehensive guide to linguistic fieldwork.

7. Reference List

Adelaar, W (2007). Threatened Languages in Hispanic South America. In M. Brezinger (Eds.), *Language diversity endangered* (pp9-28). Berlin: Mouton de Gruyter.

Austin, P., & Sallabank, J. (2011). *The cambridge handbook of endangered languages*. Cambridge: Cambridge University Press.

Bloomer, A., Wray, A., & Trott, K. (2006). *Projects in linguistics* (2nd ed.). London: Hodder Education.

Bourdieu, P. (1992). *Language and Symbolic Power*. Cambridge: Polity Press.

Bowern, C. (2008). *Linguistic fieldwork: A practical guide*. Basingstoke: Palgrave Macmillan.

Brenzinger, M.. (2007). *Language diversity endangered*. Berlin: Mouton de Gruyter.

Campbell, L., & Muntzel, M. (1989). *The structural consequences of language death*. Cambridge: Cambridge University Press.

Chelliah, S. L., & Reuse, W. J. d. (2011). *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer. doi:10.1007/978-90-481-9026-3

Cocq, C (2015). 'Indigenous voices on the web: Folksonomies and endangered languages'. *The Journal of American Folklore*, *128*(509), 273-285. doi:10.5406/jamerfolk.128.509.0273

Crowley, T., & Thieberger, N. (2007). *Field linguistics: A beginner's guide*. Oxford: Oxford University Press.

Crystal, D. (2002). *Language death* (Canto ed.). Cambridge: Cambridge University Press.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Oxford: Blackwell.

Gippert, J., Himmelmann, N. P., & Mosel, U. (2006). *Essentials of language documentation*. Berlin: Walter de Gruyter.

Grenoble, L (2011). Language ecology and endangerment. In P. Austin & J. Sallabank (Eds.), *The cambridge handbook of endangered languages* (pp27-44). Cambridge: Cambridge University Press

Grenoble, L. A., & Whaley, L. J. (2006). *Saving languages: An introduction to language revitalization*. Cambridge: Cambridge University Press.

Harrison, D. K. (2008). *When languages die: The extinction of the world's languages and the erosion of human knowledge*. Oxford: Oxford University Press.

Hirsh, D. (2013). 'Endangered languages, knowledge systems and belief systems'. *Linguistic Insights - Studies in Language and Communication*, *163*, 1-153.

Krauss, M. (2007). Classification and terminology for degrees of language endangerment. In M. Brezinger (Eds.), *Language diversity endangered* (pp.1-8). Berlin: Mouton de Gruyter.

Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell

Lee, N.H., & Van Way, J.(2016). Assessing levels of endangerment in the Catalogue of Endangered Languages (ELCat) using the Language Endangerment Index (LEI). *Language in Society*, 45(2),271-292. doi:10.1017/S0047404515000962

Maffi, L. (2002). Endangered languages, endangered knowledge. *International Social Science Journal*, 54(3), 385-393. doi:10.1111/1468-2451.00390

McNeill, P., & Chapman, S., & Chapman, S.R. (2005). *Research methods* (3rd ed.). London: Routledge. doi:10.4324/9780203463000

Ng, C. (2014). Mother tongue education in singapore: Concerns, issues and controversies. *Current Issues in Language Planning*, 15(4), 361-375. Retrieved from http://search.proquest.com/docview/1650526178?accountid=11526

Samarin, W. J. (1967). *Field linguistics: A guide to linguistic field work*. New York: Holt, Rinehart and Winston.

Weber, J. J., & Horner, K. (2012). *Introducing multilingualism: A social approach*. London: Routledge. doi:10.4324/9780203149713

Whorf, B. L. (1956). *Language, thought, and reality; selected writings*. United States/ London: Published jointly by The Technology Press of Massachusetts Institute of Technology and John Wiley & Sons New York/ Chapman & Hall Ltd London.

8. Appendix 1 – Ethics Documents

POSTGRADATE STUDENT / STAFF RESEARCH ETHICAL REVIEW

SECTION A: TO BE COMPLETED BY THE APPLICANT

Before completing this section please refer to the School Research Ethics web pages which can be found at https://www.hud.ac.uk/mhm/researchgovernanceandethics/. Applicants should consult the appropriate ethical guidelines.

Please ensure that the statements in Section C are completed by the applicant (and supervisor for PGR students) prior to submission.

Project Title	Eliciting Phonetic Data During Field Research
Applicant	Katie Coxall
Supervisor (where applicable)	Erica Gold
Award (where applicable)	Masters by Research (Linguistics)
Project start / end date	September 2016 – January 2017

Mark 'X' in one or more of the following boxes if your research involves:

- ☐ direct contact with human/animal participants
- ☐ access to identifiable personal data for living individuals not already in the public domain
- ☐ increased danger of physical or psychological harm for researcher(s) or subject(s)
- ☐ research into potentially sensitive areas
- ☐ use of students as research assistants
- ☐ covert information gathering or deception
- ☐ children under 18 or subjects who may unable to give fully informed consent
- ☐ prisoners or others in custodial care (e.g. young offenders)
- ☐ significantly increased danger of physical or psychological harm for researcher(s) or subject(s), either from the research process or from publication of research findings
- ☐ joint responsibility for the project with researchers external to the University.

Please note that if you provide sufficient information about the research (what you intend to do, how it will be carried out and how you intend to minimise any risks), this will help the ethics reviewers to make an informed judgement quickly without having to ask for further details.

SECTION B: PROJECT OUTLINE (TO BE COMPLETED IN FULL BY THE APPLICANT)

Issue	Please provide sufficient detail to allow appropriate consideration of any ethical issues. Forms with insufficient detail will need to be resubmitted.
Aims and objectives of the study. Please state the aims and objectives of the study.	To collate information relating to the collection of phonetic data in field research so that recommendations can be made based on tried and tested methods of data collection which will assist future field linguists in their documentation efforts.
Brief overview of research methodology The methodology only needs to be explained in	Data will be collected using a survey which will be created using online software. This will be distributed to

sufficient detail to show the approach used (e.g. survey) and explain the research methods to be used during the study.	field linguists who will be identified using their profiles on the websites of the institutions at which they work.
Does your study require any permissions for study? If so, please give details	NA.
Participants Please outline who will participate in your research. Might any of the participants be considered 'vulnerable' (e.g. children)	Participants will be experienced field linguists who work at research institutions/ universities.
Access to participants Please give details about how participants will be identified and contacted.	Participants will be identified based on their research interests listed on the website of the institutions at which they work. They will be contacted via the email address listed on their profile.
How will your data be recorded and stored?	My data will be held in accordance with the Data Protection Act (UK). When the results of the study are reported, participants will not be identified by name or by any other information that could be used to infer their identity.
Informed consent. Please outline how you will obtain informed consent. **If informed consent or consent is NOT to be obtained please explain why.**	A disclaimer will be included on the first page of the survey, before any questions. This will inform the participant of the aims of the research and advise them that by clicking 'next' and completing the survey they have provided their consent to become a subject of my research. I will include my email address which will allow researchers to contact me directly with any questions or concerns and will outline that the participant is taking part on a voluntary basis.
Confidentiality Please outline the level of confidentiality you will offer respondents and how this will be respected. You should also outline about who will have access to the data and how it will be stored. (This information should be included on Information your information sheet.)	All responses provided for this study will be kept completely confidential, and all data will be held in accordance with the Data Protection Act (UK). When the results of the study are reported, participants will not be identified by name or by any other information that could be used to infer their identity. I will be the only person with access to the data.
Recorded Media Will the research involve the production of recorded media such as audio and/or video recordings? If so how will you ensure that there is a clear agreement with participants as to how these recorded media may be stored, used and (if appropriate) destroyed?	No audio or video recordings will be taken.
Anonymity If you offer your participants anonymity, please indicate how this will be achieved.	Participants will not be referred to by name, or by any other means in which their identity could be inferred.
Harm Please outline your assessment of the extent to which your research might induce psychological stress, anxiety, cause harm or negative consequences for the participants (beyond the risks encountered in normal life). If more than minimal risk, you should outline what support there will be for participants. If you believe that that there is minimal likely harm, please articulate why you believe this to be	As the respondents will be anonymous there is no risk of harm.

so.	
Does the project include any security sensitive information? Please explain how processing of all security sensitive information will be in full compliance with the "Oversight of security - sensitive research material in UK universities: guidance (October 2012)" (Universities UK, recommended by the Association of Chief Police Officers)	No security sensitive information.

Retrospective applications. If your application for Ethics approval is retrospective, please explain why this has arisen.

SECTION C – SUMMARY OF ETHICAL ISSUES (TO BE COMPLETED BY THE APPLICANT)

Please give a summary of the ethical issues and any action that will be taken to address the issue(s).

Informed consent will be obtained by including a disclaimer in the first page of the survey which states that by completing the survey, the participant agrees to become a subject of the research. The participant will be provided with my email address and advised that if they wish to withdraw, they may do so at any time.

Information will be stored in accordance with the Data Protection Act.

Participants will remain anonymous.

SECTION D – ADDITIONAL DOCUMENTS CHECKLIST (TO BE COMPLETED BY THE APPLICANT)
Please supply copies of all relevant supporting documentation electronically. If this is not available electronically, please provide explanation and supply hard copy.

I have included the following documents

Information sheet   Yes   X   Not applicable ☐

Consent form   Yes   ☐   Not applicable X (included on
                                                               first page of survey)

Letters   Yes   ☐   Not applicable X

Questionnaire   Yes   X   Not applicable ☐

Interview schedule   Yes   ☐   Not applicable X

SECTION E – STATEMENT BY APPLICANT

I confirm that the information I have given in this form on ethical issues is correct. (Electronic confirmation is sufficient).

Applicant name: Katie-Louise Coxall

Applicant Signature: K L Coxall

Date: 09/06/2017

Affirmation by Supervisor (where applicable)
I can confirm that, to the best of my understanding, the information presented by the applicant is correct and appropriate to allow an informed judgement on whether further ethical approval is required

Supervisor name: Erica Gold

Supervisor Signature: Erica Gold

Date: 09/06/2017

Participant Information Sheet

Research Project Title: Eliciting Phonetic Data during Field Research

Name of Researcher: Katie Coxall

Contact Details of Researcher: Katie-louise.Coxall@hud.ac.uk

You are being invited to take part in a research project. Before you decide, it is important for you to understand why this research is being done and what it will involve. Please take time to read the following information and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. May I take this opportunity to thank you for taking time to read this.

1. What is the purpose of the project?
The research project is intended to provide the research focus for a module which forms part of my degree.

The aim is to gather information relating to the experiences of field researchers in the collection of phonetic data during field research. This information will be collated and recommendations/ strategies will be suggested based on the results. These suggested strategies are intended to assist future field linguists in their documentation efforts.

2. Why have I been chosen?
You have been selected based on the research interests which are listed on your academic profile, and your previous field linguistics experience.

3. Do I have to take part?
Participation on this study is entirely voluntary, so please do not feel obliged to take part. Refusal will involve no penalty whatsoever and you may withdraw from the study at any stage without giving an explanation to the researcher.

4. What do I have to do?
You will be invited to take part in a survey which should take no more than 10-15 minutes of your time.

5. Are there any disadvantages to taking part?
There should be no foreseeable disadvantages to your participation. If you are unhappy or have further questions at any stage in the process, please address your concerns initially to the researcher if this is appropriate. Alternatively, please contact Professor M. Adkins (**m.adkins@hud.ac.uk**) at the School of Music, Humanities and Media, University of Huddersfield.

6. Will all my details be kept confidential?
All information which is collected will be strictly confidential and anonymised before the data is presented in any work, in compliance with the Data Protection Act and ethical research guidelines and principles.

7. What will happen to the results of the research study?
The results of this research will be written up for my Masters of Research project. If you would like a copy please contact the researcher.

8. What happens to the data collected?

The data collected will be collated any analysed in order to deduce advice relating to the collection of phonetic data during linguistic fieldwork. This advice will form the basis of some recommendations for future field linguists in the collection of data of this type.

9. Will I be paid for participating in the research?
Your participation in this research is entirely voluntary, and no payments will be made.

10. Where will the research be conducted?
The research will be conducted via an online survey. The link will be emailed to you.

11. Criminal Records check (if applicable)
NA.

12. Who has reviewed and approved the study, and who can be contacted for further information?
You can contact the researcher, Katie Coxall, for further information via email Katie-louise.Coxall@hud.ac.uk. Alternatively, you can contact the research supervisor, Dr. Erica Gold, via email e.gold@hud.ac.uk .