Recruitment techniques for minority language speech databases: some observations

Rhys James Jones*, John S. Mason*, Louise Helliker †, Mark Pawlewski†

* Speech and Image Processing Research Group, University of Wales Singleton Park, Swansea SA2 8PP, Wales, UK {R.J.Jones, J.S.D.Mason}@swansea.ac.uk

[†]BT Advanced Communications Engineering, Adastral Park Martlesham Heath, Ipswich, Suffolk IP5 3RE, UK {louise.helliker, mark.pawlewski}@bt.com

Abstract

This paper describes the collection efforts for SpeechDat Cymru, a 2000-speaker database for Welsh, a minority language spoken by about 500,000 of the Welsh population. The database is part of the SpeechDat(II) project. General database details are discussed insofar as they affect recruitment strategies, and likely differences between minority language spoken language resource (SLR) and general SLR collection are noted. Individual recruitment techniques are then detailed, with an indication of their relative successes and relevance to minority language SLR collection generally.

It is observed that no one technique was sufficient to collect the entire database, and that those techniques involving face-to-face recruitment by an individual closely involved with the database collection produced the best yields for effort expended. More traditional postal recruitment techniques were less successful. The experiences during collection underlined the importance of utilising enthusiastic recruiters, and taking advantage of the speaker networks present in the community.

1. Background

As international demand for speech technology grows, it is inevitable that it will encompass languages and domains for which such technology does not currently exist. A need arises to collect new spoken language resources (SLRs) in languages for which there are currently no speech databases. For large-scale SLRs, speaker recruitment techniques must be identified. Techniques for collecting SLRs for major languages are well-known and documented (van den Heuvel et al., 1998). However, the process of collecting a minority language SLR is less well-understood.

Little previous collection of Welsh has been undertaken. As noted by Williams (1999), most of the corpora currently available are not machine-readable. To the best of the authors' knowledge, only one other significant machine-readable corpus of Welsh exists: a collection of six speakers, recorded in a clean recording studio environment, intended for use in basic Welsh phonetic research (Williams, 1999). The collection described here has very different aims. The objective of SpeechDat(II), of which SpeechDat Cymru is part, was to record between 500 and 5000 speakers in each of 28 languages and dialectical variants, over fixed and mobile telephone networks, for speech recognition and speaker verification (van den Heuvel et al., 1998). In the case of SpeechDat Cymru, the target was to record 2000 Welsh speakers – defined as those who had spoken Welsh since they were at least 5 years old – over a fixed telephony network (PSTN).

Contributors to SpeechDat Cymru were required to be demographically balanced in terms of gender (45-55% male, 45-55% female) and age (an approximately uniform distribution of callers across four age ranges stretching from 16-61+, with a recommendation that a small number of children under 16 also be recorded).

SpeechDat databases contain mostly read speech, with some items being elicited spontaneously. Each volunteer is hence given a unique prompt sheet. Details of the items to be elicited from each caller are shown in Table 1. In order to record the required number of callers, prompt sheets were tracked, and calls made to the recording platform were continuously monitored. Any prompt sheet used was immediately withdrawn from circulation, ensuring that no prompt sheet is recorded more than once.

Description		Number per call
application words		6
sequence of 10 isolated digits		1
	sheet number (5+ digits)	1
4 connected digits	telephone number (9-11 digits)	1
	credit card number (14-16 digits)	1
	PIN code (6 digits)	1
	spontaneous date, e.g. birthday	1
3 dates	prompted date, word style	1
	relative and general date expression	1
word spotting phrase using an embedded application word		1
isolated digit		1
	spontaneous, e.g. own forename	1
3 spelled word	spelling of direct. city name	1
(letter sequences)	real/artificial for coverage	1
currency money amount		1
natural number		1
	spontaneous, e.g. own forename	1
	city of birth / growing up (spontaneous)	1
5 directory assistance	most frequent cities (set of 500)	1
names	most frequent company/agency (set of 500)	1
	"forename surname"	1
2 questions, including	predominantly "yes" question	1
"fuzzy" yes/no	predominantly "no" question	1
	time of day (spontaneous)	1
2 time phrases	time phrase (word style)	1
phonetically rich words		4
phonetically rich sentences		9
additional application word (specific to SDCymru)		1

Table 1: Contents of one call

2. Recruitment strategies

Speakers of minority languages seldom exist in a vacuum, but have contact with other speakers of the language. An obvious instance of such a speaker network is a neighbourhood or local community, but others also exist: schools teaching through the medium of the minority language; clubs or organisations devoted to that language or operating within the medium of that language; Internet discussion forums aimed at bringing together the language's speakers, and so on. In collecting SpeechDat Cymru, use was made of Welsh speaker networks. It was found that they gave rise to all the most effective recruitment techniques.

The initial recruitment plan was to contact the 60 or so bilingual secondary schools within Wales. These schools educate 11 to 18-year old children through the medium of Welsh. It was believed that they would give easy access to a network of parents supportive of the Welsh language.

Each school was contacted by letter, and informed of the collection and its aims. They were advised to contact those administering the collection should they not want to take part. If a reply was not received within a week, they were then sent a set of 200 slips to forward to parents. The parents would then provide names and addresses of volunteers on these slips, which the school would collect and in due course send back to the recruiters. Prompt sheets were subsequently sent to the addresses provided, together with detailed instructions on how to make the call.

Incentives are important when collecting any spoken language resource (SLR). Some volunteers, especially for minority language SLRs, may be enthusiastic to do something "for the good of the language", and may not need any incentives to take part. However, for the majority of volunteers a tangible reward is needed. In the case of the work described here, £2 was given to a charity or school nominated by the volunteer, on completion of his/her call. For school collections, the £2 donations were given to the school itself.

The initial anticipation was that something in the order of 100 responses might be received from each school, and from these perhaps 50 people per school would make a call to the recording platform. To avoid potential overloading of the platform, a staggered recruitment programme was set up, and care was taken to avoid sending out a large number of prompt sheets simultaneously.

In three months, recruitment through the 60 schools yielded 900 calls: an average of 15 per school, far less than expected. Returns from schools varied enormously even when allowing for different school sizes, from under 10 at some to over 200 at others.



Graph 1: Increase in calls made during collection

The main merit of the secondary schools approach is the low administrative overhead and the comparatively high yield of calls by this method. Were the number of secondary schools larger, recruitment could have been completed by this method alone. In the event, numerous other techniques were needed for recruitment, after exhausting the 60 schools.

School recruitment used a direct mail approach to recruitment. Subsequent recruitment used a mixture of direct mail and face-to-face approaches. The use of faceto-face recruitment enabled any queries or reservations potential volunteers might have to be discussed, and hopefully overcome.

In particular, many minority languages have a large base of speakers who use the language primarily in conversational mode, rather than a formal literary or written mode. Formal prompt sheets could prove daunting to such people. Such a format was required to ensure a full phonetic coverage as specified by SpeechDat. Face-toface recruitment enabled the recruiter to explain the general recording process and any unfamiliar terms or phrases used in the prompt sheets to the potential volunteer. In the vast majority of cases, the volunteer was reassured by this process and was then eager to make the call.

Face-to-face recruitment for SpeechDat Cymru took place in the National Eisteddfod of Wales - a week-long Welsh cultural festival. Passers-by were approached, and the collection process and reasons behind it were briefly explained. During the opening conversations, particular emphasis was placed on the potential benefits the collection would have for the Welsh language. If they were willing to do so, volunteers were asked to make a recording there and then. The potential caller's questions or apprehensions were answered and addressed in advance of the recording being made.

Face-to-face recruitment also occurred at organisations and businesses employing a significant number of Welsh speakers, where similar techniques were employed. Lindberg et al. (1998) state that in the context of delivering a finished database, 'speaker recruitment was the most difficult task for all partners, and it took longer than expected for most partners'. This was especially true for Welsh, where recruitment stretched from April 1997 to October 1998 - a total of a year and a half, nearly double the original estimate. The increase in calls recorded with time can be seen in Graph 1. As can be seen from this graph, slightly more than 2000 calls were actually recorded. This was due to a small number of calls that had to be re-recorded, due to either the poor quality of the line or confusion on behalf of the caller resulting in too many invalid items being recorded.

The trends in the graph are as follows:

• from the end of March 1997 to end July 1997: This was the extent of the school collection programme. The number of calls made per day increased gradually as more and more schools were contacted. During the peak period for calls (between the beginning of June and the middle of July) an average of about 10 calls were made per day.

from the beginning of August 1997 to the middle of June 1998: the collection efforts during this time concentrated mainly on snowball recruitment (see next Section), and contacting a wider range of smaller primary schools and charities. Regardless of the collection efforts during this time, the rate of calls received remained remarkably constant at about 2 per day. The brief increase in call rate during early September was due to further slips being returned from schools that had recently finished their summer holidays. That in February 1998 resulted from the response to a second round of publicity. All other brief increases were due to a large number of slips being returned from a particular charity, and the prompt sheets that were then sent out resulting in a large number of calls being made.

• from the middle of June 1998 to the end of September 1998: an emphasis on face-to-face recruitment, together with a continuation on snowball recruitment, resulted in a slightly higher overall rate of calls received (about 3 per day) coupled with larger individual increases on certain days, as particular visits to companies and organisations were made. In particular, the large increase at the beginning of August coincided with the week-long visit to the National Eisteddfod of Wales, where about 25 calls were made per day.

• from the end of September 1998 to the end of collection: occasional calls were made, mostly to compensate for those calls which were discovered to be poorly or inadequately recorded.

It is worth pointing out the scale of the collection. Welsh, according to the most recent UK census, has just over 500,000 speakers. To recruit 2,000 of these corresponds to recording 1 in every 250 Welsh speakers. To put this figure in perspective, recording 1 of every 250 British English speakers within the UK would result in a database of 250,000 speakers.

3. Comparisons

The following recruitment schemes are discussed in Lindberg et al. (1998). Their relevance to Welsh is as follows:

Hierarchical recruitment, defined as the process whereby a recruiter distributes prompt sheets among colleagues and employees. For any minority language utilising this technique, care must be taken to choose a group of people among which there is a high incidence of speakers of the language. Even among such groups of people, an immediate face-to-face recruitment technique would be preferable: here, an individual closely associated with the overall collection visited the group, explaining the aims of the collection and the recording process, and giving people a chance to call the recording platform there and then. It was felt that such a technique, though more timeconsuming, would result in a greater yield of calls, as it gave the recruiter a chance to answer any queries or apprehensions potential volunteers might have about the recording process, making the potential volunteer more likely to make the call.

• **Snowball recruitment** occurs when an existing volunteer is asked to nominate further potential volunteers – typically their friends and relatives. In the case of Welsh, prompt sheets were sent out with a sheet with room for up to five names, and a stamped addressed envelope for it to be sent back to those adminstering the collection. Prompt sheets were then sent to the addresses provided, with an explanation that their details had been provided by a friend or relative. These prompt sheets were also sent together with a sheet with room for up to five names, hence the 'snowball' analogy.

• **Direct mail** was used only in a limited context for Welsh, due to the difficulty in identifying the Welsh speakers in any lists that

were publicly available (or that could be bought). A list was, however, available of all the Welshspeaking staff at the University of Wales Swansea, who administered the collection. Prompt sheets were thus sent to every member of staff (about 100) on this list, along with instructions on how to make the call to the recording platform. This resulted in about 25 calls being made.

• Newspaper and magazine advertisements were not used as such. Use was made, however, of local newspaper and radio calls for participation (see below) and information about the project was also placed on Internet mailing lists and Usenet newsgroups relating to Wales and the Welsh language.

• Local newspaper, radio and television calls for participation were carried out in two main rounds of publicity, resulting in a substantial number of articles. Most of these articles also mentioned a free phone number – this enabled volunteers to request a prompt sheet which was subsequently sent to them. Approximately 100 calls were made to the recording platform as a result of this publicity. Although it was not possible to precisely determine the most popular sources of calls for participation, casual observations seemed to suggest that publicity worked best when placed in specifically Welshlanguage media.

In addition to these techniques, some partners also used a market research company to recruit all or part of their databases. Such an approach was not used for Welsh, for a number of reasons. As other partners discovered, market research companies only guarantee to contact a given number of people, not to ensure the correct number of calls are made. Further, it is not guaranteed that the recruiting staff will be familiar enough with the recording process to answer a potential volunteer's detailed questions. An additional problem for Welsh is that of having to recruit a subsection of the geographical population of the country. With the percentage of Welsh speakers below 10% in many urban areas of Wales, this would present a challenge and greatly increase the cost of collection via a market research company. Using a market research company for collection of a minority language database would, in short, be a very expensive process.

4. Findings

In common with other partners (Lindberg et al., 1998), it was found that no one recruitment technique was adequate to collect the 2000 speakers. For collection of minority language SLRs, as with collection of SLRs generally, many different recruitment techniques should be initiated, and those with the greatest impact continued. Often there will be a trade-off between the effort involved in a recruitment technique and its yield. In order of greatest yield for effort, the most successful recruitment techniques for Welsh were:

• face-to-face recruitment at the Eisteddfod, yielding about 180 calls in 6 days. This required 6 whole days of one person's time (to recruit on the field, and supervise recordings).

• other face-to-face recruitment (at schools and organisations), with slightly smaller yields for the amount of effort expended.

• School postal recruitment, yielding about 900 calls in 3 months. This required a great deal of effort in prompt sheet distribution and dealing with the returned volunteer slips from schools. Nevertheless, this method alone might have been sufficient had a greater number of secondary schools existed.

• 'Snowball'/word-of-mouth recruitment, where a sheet with room for five names was sent out with every prompt sheet enabling the callers to recruit up to five more volunteers.

• General publicity (printed media, radio, television). If the minority language under collection is sparsely distributed within a country, publicity must be directed at a subsection of the country's population.

• Direct mail to individuals.

A general point arising from the collection is the importance of enthusiastic recruiters. The reason behind the large variation in the number of slips returned and the number of calls subsequently made between one school and another was largely the degree of enthusiasm of the teacher in charge of co-ordinating recruitment within that particular school. In the school which produced the best return (nearly 220 potential volunteer names, resulting in over 100 calls being made), the headmaster had appointed the head of the Welsh department to be in charge of the collection, on the understanding that the £2 per caller would be given directly to her department. This resulted, predictably, in a very enthusiastic and valuable recruiter who produced an excellent return. At other schools, response slips, if they were distributed at all, were distributed sparsely - often to just one age group - or poorly, where sometimes the pupils were not compelled to return them. Approximately half of the schools to which the response slips were distributed returned none, despite repeated reminders and phone calls to those schools. Of those who did respond, the average return of slips was 30 - given that the best return was 220, this underlines the difference an enthusiastic recruiter can make.

5. References

Lindberg, Borge et al. (1998). Speaker recruitment methods and speaker coverage - experiences from a large multilingual speech database collection. In Proc. ICSLP 1998.

van den Heuvel, Henk et al. (1998). SpeechDat experiences in creating large multilingual speech databases for teleservices. In Proceedings of the First International Conference on Language Resources and Evaluation, vol. 1, pp. 361-366.

Villarrubia, L. et al. (1998). VOCATEL and VOGATEL: Two telephone speech databases of Spanish minority languages (Catalan and Galician). In Proceedings of the Workshop on Language Resources for European Minority Languages, LREC 1998.

Williams, Briony (1999). A Welsh speech database: preliminary results. In Proc. EuroSpeech 1999, vol. 5, pp. 2283-2286.