

Philip Carr, Jacques Durand and Monika Pukli

*The PAC project : Principles and Methods**

NB This is the prefinal version of an article published under the same title in *La Tribune Internationale des Langues Vivantes* (TILV), Vol. 36, 2004, p. 24-35. If you wish to quote this article, please use the final version in TILV.

0. Introduction

PAC (*La Phonologie de l'Anglais Contemporain: usages, variétés et structure / The Phonology of Contemporary English: usage, varieties and structure*) is a project coordinated by Philip Carr (Montpellier III & ERSS-CNRS) and Jacques Durand (Toulouse II & ERSS-CNRS). After some pilot studies in the late 1990s, it was properly launched in 2000 with a first study in Santa Barbara, California (see Pukli 2002). Since then, a number of surveys have been launched in various parts of the world (see e.g. Przewozny this vol., Pukli this vol.) and the project holds regular meetings to improve its goals and its coordination.

Within *PAC*, among other things, we aim at (1) giving a better picture of spoken English in its unity and diversity (geographical, social and stylistic), (2) testing phonological and phonetic models from a synchronic and diachronic point of view, giving pride of place to the systematic study of variation, (3) favouring communication between speech specialists and phonologists, (4) providing data and analyses which will help improve the teaching of English as a foreign language. To achieve these goals our project is involved in the construction of a corpus of spoken English from a wide variety of locations in the anglophone world on the basis of a common protocol. While there are important corpora of spoken English, most of them have been devised on sociolinguistic (rather than explicit phonological) principles and they do not always offer a uniform methodology allowing for a strict comparison of results.

The approach adopted within *PAC* is a well tested one since it has been followed in the international project '*La Phonologie du Français Contemporain*' (*PFC*) coordinated by Jacques Durand (Toulouse II), Bernard Laks (Paris X) and Chantal Lyche (Oslo/Tromsø): see Durand, Laks & Lyche 2002, 2003a,b; Durand & Lyche 2003, and the internet site: <http://infolang.u-paris10.fr/pfc>.

The *PAC* methodology is inspired by the classical work of Labov (e.g. 1966, 1972, 1994, 2001) in that, for each selection of speakers, it involves the reading aloud of a wordlist and a passage as well as formal and informal conversation. But in each area surveyed, the speakers (usually groups between 10 and 20 informants) are selected on a network principle well known in the United Kingdom, particularly from the work of the Milroys and their associates (e.g. Milroy & Milroy 1985, Milroy 1987a,b, Foulkes & Docherty 1999; and see Durand, Slater & Wise 1987, for an application to French).

In the initial phase of the project we favour geographical variation, that is the recording and analysis of cohorts of speakers from as many different locations as possible in the English-speaking world. Within each location, in so far as possible, the groups include an equal number of men and women and well defined age ranges (e.g. 20+, 40+, 70+). Social diversity is less easy to achieve with small groups of speakers and it has been found profitable

to study family networks which allow for better comparison of age-grading especially when the social background of the informants has remained relatively stable.

The purpose of this document is to present the methodological framework adopted in *PAC*, to outline some of the reasons underlying our selection of material to be read aloud and to highlight a few examples from the range of phenomena tested by our protocol (see Durand & Pukli 2004 for further details). Transcriptions in the paper follow the IPA tradition, and are enclosed in /slant brackets/ unless a particular phonetic realisation is commented on, which is in turn enclosed in [square brackets], but our transcriptions are meant merely to guide the reader and not to provide theoretical interpretations. When discussing various phonemes and corresponding sets of words we refer to the lexical sets and keywords first introduced by Wells (1982: 1.xviii) and subsequently taken up in much of descriptive work in accent studies; these are distinguished by the use of SMALL CAPS. Simple citation from *PAC* material (wordlists and passage) is marked by *italics*. Throughout this paper, we have intentionally steered clear of highly technical concepts. The reader is referred to Durand 1990, Carr 1993, 1999, Burton-Roberts, Carr & Docherty 2000, Durand & Lyche 2000, Durand & Laks 2002, as well as to the other articles in this volume, for more advanced discussions.

1. Methodology

As indicated above, the protocol provides for the recording of various styles for each speaker: (1) the reading aloud of two wordlists, (2) the reading aloud of a written passage (hereafter called ‘the text’), (3) a formal interview between the fieldworker and the informant, and (4) an informal conversation. The experience from similar projects is that this approach generally gives rise to a range of registers for each informant. In the ideal case, the recordings show an ascending (or descending!) scale of formality but the interpretation of the data needs a great deal of care. In particular, the use of the terms ‘formal’ and ‘informal’ should not be taken at face value. When referring to formal and informal conversations, we should not make strong assumptions as to the ‘product’, i.e. the actual style of speech recorded, but rather see these terms as referring to the fundamentally different settings of the two conversational contexts (see §1.3)

1.1 The Wordlists

The two wordlists contain 192 items altogether, which is rather long but allows for the examination of a wide sample of segmental phenomena. Although these lists are far from exhaustive, they include elements applicable to a wide range of varieties of English. Whenever necessary, we recommend supplementing our two lists with a further list, if the informants are able and willing to participate in additional recordings. It should, however, be remembered that reading aloud is tedious and that informants may not have sufficient time and patience for this additional exercise. The same point can be made for the possible inclusion of an extra text over and above the *PAC* text.

The first and longer list, mainly conceived to test the speakers’ vowel systems, includes a number of minimal pairs some of which are familiar from well-known studies such as Hughes & Trudgill (1996), Trudgill & Hannah (1985) and Wells (1982). The second list, based on the same principles, is centred on the consonantal system. These lists and the text have been devised to fulfil the three main objectives that we set ourselves within the *PAC* project : (i) establish the phonological inventory of the speaker(s) under study (phonemic oppositions), as well as the main allophonic variants and phonotactic constraints; (ii)

investigate rhoticity in detail; (iii) examine T/D realisations as a way of understanding the complex behaviour of coronal plosives in all varieties of English (aspiration, tapping, glottalisation, etc.). Of course, (ii) and (iii) are subsets of (i), but their inclusion in our goals is intended to give more homogeneity to the descriptions of the PAC corpora. Rhoticity, in any case, is crucial to an understanding of all accents of English. Not only is the quality of the /r/ used in various contexts of great interest, but above all one has to decide whether the accent being studied is rhotic or non-rhotic (or indeed variable). In a nutshell, in a fully rhotic accent, all orthographic ‘r’s are pronounced (e.g. General American, see Carr & Durand, this vol., or Standard Scottish English, see Durand this vol., Ford this vol. Pukli, this vol.). In a non-rhotic accent like British English Received Pronunciation (RP hereafter), London English or Australian English, an ‘r’ which is not in the onset of a syllable is absent: e.g. RP *rat* /ræt/, *carry* /'kæri/ vs. *car* /kɑ:/, *carter* /'kɑ:tə/ (see Carr, this vol., Moore, this vol., Przewozny, this vol.). Note that it is not the quality of the ‘r’ which defines an accent as ‘rhotic’ but the presence of /r/ in syllable-codas. Many of the words in the list and the text allow us to test whether the system under investigation is rhotic, non-rhotic or variable but also explore some of the consequences of the presence or absence of an ‘r’. For instance, numbers 11 to 19 (*stir, steer, stairs, err, far, war, more, purr, moor*) and 90 to 100 (*fierce, bird, scarce, pert, start, horse, hoarse, word, gourd, short, sport*) sample vowel sounds before /r/ (word finally and preceding a consonant, respectively), as vowel distribution in these positions can be radically different for a rhotic and a non-rhotic accent.

Wordlist 1

1. pit	26. fail	51. aunts	76. four
2. pet	27. foal	52. dance	77. fore
3. pat	28. file	53. farther	78. for
4. pot	29. foul	54. father	79. nose
5. put	30. foil	55. row	80. knows
6. putt	31. furl	56. rose	81. cot
7. sea	32. bird	57. rows	82. caught
8. say	33. bard	58. pore	83. meat
9. sigh	34. beard	59. poor	84. meet
10. sue	35. bared	60. pour	85. mate
11. stir	36. board	61. paw	86. naught
12. steer	37. barred	62. paws	87. knot
13. stairs	38. bored	63. pause	88. doll
14. err	39. bode	64. pose	89. dole
15. far	40. bowed	65. wait	90. fierce
16. war	41. bead	66. weight	91. bird
17. more	42. bid	67. side	92. scarce
18. purr	43. bed	68. sighed	93. pert
19. moor	44. bad	69. agreed	94. start
20. feel	45. bard	70. greed	95. horse
21. fill	46. pant	71. brood	96. hoarse
22. fell	47. plant	72. brewed	97. word
23. fall	48. master	73. fir	98. gourd
24. full	49. afterwards	74. fair	99. short
25. fool	50. ants	75. fur	100. sport

101. next
102. vexed
103. leopard
104. shepherd
105. here
106. there
107. weary

108. spirit
109. marry
110. Mary
111. merry
112. sorry
113. story
114. hurry

115. jury
116. bury
117. berry
118. heaven
119. leaven
120. earth
121. berth

122. cook
123. soot
124. look
125. room
126. pearl
127. peril

A complete description of the phenomena covered in this word-list would require far more space than is available here. We shall content ourselves with a few examples. There are many potential ‘minimal pairs’ in our two lists, i.e. words which differ with respect to only one segment. For instance, all the words from (1) to (6) would be distinguished by an RP speaker through the use of a different vowel in the context **p --- t**: e.g. /pit, pet, pæt, pɒt, put, pʌt/ (see Moore this vol.). On the other hand, some of our informants in Lancashire pronounce *put* and *putt* identically (usually with an [ʊ]-like vowel) even in reading aloud. Numbers 24-25, 50-51, 65-66, 81-82 and 86-87, or 95-96 can either exhibit a minimal phonemic contrast, or conversely, be homophonous, depending on the accent under study. *Full* and *fool* sound the same in Scottish Standard English (e.g. [ʌ]) while they have two different vowels, /ʊ/ and /u:/, in other major accents. *Ants* and *aunts* are distinct in RP, /ænts/ and /ɑ:nts/, but both of them are said with an /æ/ in General American, and with an /a/ in most Scottish accents. (It is in fact the realisation of this minimal pair that can help determine whether or not a Scottish speaker has a ‘Modified’ vowel system with a contrast for the lexical sets TRAP BATH and PALM: see Durand this vol.). *Wait* and *weight* are homophonous within most accents of English, while some speakers from Lancashire make a clear distinction between the two, e.g. /weɪt/ vs. /wert/. Another traditional regional trait is the possibility of a contrast between /ʊ/ and /ʌ/ in words like *moan* – *mown* or *sole* – *soul* in East Anglia and this can be tested in our list via numbers 56-57, and 79-80, *rose* – *rows* and *nose* – *knows*. *Cot* – *caught* (81-82) and *knot* – *naught* (86-87) are minimal pairs for RP, /ɒ/ vs. /ɔ:/, but are pronounced identically in Scottish Standard English (SSE hereafter), for instance; on the other hand, *horse* and *hoarse* are homophones for RP speakers (both being /hɔ:s/) while speakers from Scotland have a minimal contrast there (/hɔrs/ and /hors/, respectively). For researchers concerned with SSE, which contrasts *gate* and *get* as /get/ vs. /get/, one should also note the possibility of an extra vowel sound somewhere between /e/ and /ɛ/ but fairly centralized: [ɛ̃] (often called ‘Aitken’s Vowel’). Words which allow us to test the presence of this extra vowel phoneme are *next*, *shepherd*, *heaven*, *earth* and *berth* (101, 104, 118, 120 and 121) as well as the potential opposition between *berry* and *bury* (/ˈbere/ vs. /ˈbɛre/, in the transcription presented here for Scottish English, see Durand this vol.). A final example, interesting from a diachronic point of view, is the testing of the CURE – FORCE merger in RP (Wells, 1982: 237) - i.e. the tendency to realize the centring diphthong /ʊə/ in *cure* with a lowered monophthongal quality so that it becomes [ɔ:] with the same value as the phoneme in *force*. Thus *more* and *moor* (numbers 17 and 19) may or may not be homophonous for RP speakers just like *pore*, *poor* and *pour* (58, 59, 60).

Our second list (see below) provides invaluable information on consonantal phonemes and allophones. Among other things, the list exhibits a great number of voiced-voiceless pairs, in word-initial, -medial and -final positions, the latter being also interesting with regard to vowel quantity. As is well known, in many accents of English a vowel before a voiced consonant is much longer than the same vowel before the same consonant’s voiceless counterpart

and, while this is probably a universal tendency, in English the length is markedly longer than in other languages and arguably phonologised.

Wordlist 2

1. pat	23. wet	45. bedding
2. bat	24. yet	46. written
3. tuck	25. witch	47. ridden
4. duck	26. which	48. singer
5. carter	27. lock	49. stronger
6. garter	28. loch	50. fat
7. fan	29. earthy	51. fad
8. van	30. worthy	52. lap
9. this	31. sinner	53. lab
10. thick	32. simmer	54. sack
11. seal	33. singer	55. sag
12. zeal	34. supper	56. belly
13. bishop	35. rubber	57. berry
14. leisure	36. little	58. bell
15. heart	37. middle	59. bet
16. batch	38. metal	60. chutney
17. badge	39. meddle	61. kidney
18. rum	40. bicker	62. grace
19. run	41. bigger	63. graze
20. rung	42. degree	64. behave
21. lack	43. decree	65. anyhow
22. rack	44. betting	

Several items help explore word- and morpheme-final distribution of nasal sounds. *Rum*, *run*, *rung*, *sinner*, *simmer*, *singer* and *stronger* (numbers 18-20, 31-33 and 49) can tell us about the realisation of the velar nasal for example, which is subject to considerable regional variation in Britain.

‘Special’ phonemes or allophones are also attested in various accents. In Scotland *witch* – *which* and *lock* – *loch* are traditional minimal pairs, and speakers have two additional phonemes in their consonantal inventory: the voiceless labio-velar fricative /ɸ/ (analysed by some as /hw/) and the voiceless velar fricative /x/. Both sounds, however, may be doomed to disappear as younger urban speakers are reported to increasingly replace them by /w/ and /k/ respectively (cf. Chirrey 1999, Stuart-Smith 1999, 2001, Durand this vol.).

In the United States and Canada tapping (or flapping) is a widespread phenomenon and may bring about the neutralisation of the alveolar plosives /t, d/ into a quick tap against the teeth- or the alveolar ridge within feet (see Carr & Durand this vol). Thus *little-middle*, *metal-meddle* and *betting-bedding* (36-39 and 44-45) can all be pronounced with the same consonant sound in the middle of the word, usually designated by the [ɾ] symbol; it also has a retroflex variant [ɽ] which can be heard after /r/ as in *carter*, *garter*, or *thirty*, *forty*, etc. The possible

phonemic status of this neutralising allophone is quite controversial, and, although prevalent in spontaneous casual speech, tapping is often stigmatised. One proof of this is that some of our American informants first realise a tap when reading aloud a word like *little* [lɪɾl] and then correct it to a voiceless plosive [lɪtɫ]. In one case, we even have a hypercorrection with the word *meddle* being pronounced [mɛtɫ] instead of [mɛdɫ] or [mɛɾɫ] (for more discussion on tapping and the PAC corpus from California see Pukli (2002)). While these artificial pronunciations are the product of reading aloud, they do give us information about the speaker's phonological 'conscious competence' - i.e. what range of sounds speakers have control over - and its possible link with orthographic knowledge.

The numerals that precede each item and which must be read aloud are also a very useful way to double-check on actual, non-monitored performance. Speakers are often highly pronunciation-conscious when reading a list, but, as they are actually concentrating on getting the words right, they do not control directly the articulation of the numerals before them. This proves to be helpful exactly when stigmatised features are under study, such as tapping.

Finally, there are also a few items containing the phoneme /h/ (15. *heart*, 64. *behave*, 65. *anyhow*), which, combined with words from the first list (95. *horse*, 96. *hoarse*, 105. *here*, 114. *hurry* and 118. *heaven*), can be used to test H-dropping (or H-deletion), another domain of wide-ranging social and regional variation. Despite the fact that H-dropping in lexical items (nouns, verbs, adjectives, adverbs) is often socially frowned upon, some of our Lancashire informants pronounce many of these words without an 'h' even in the reading aloud of the wordlist (see Noel 2002). This is a good example of what we call the 'asymmetric value of reading aloud': a reading pronunciation which conforms to a given norm (say in preserving a contrast) is no proof that it is part of the spontaneous system of the informant; on the other hand, the presence of stigmatised features in reading aloud is very strong evidence in favour of the hypothesis that these features are part and parcel of the speaker's system.

1.2 The text

The text is a two-page passage originally based on a newspaper article but substantially modified to hide its source and include a number of phonological phenomena worth investigating. First of all, it aims at cross-checking the range of oppositions and realisations tested in the list. For instance, it allows us to establish that some pronunciations are the result of conscious correction or hypercorrection of particular segments. For instance, an informant might well avoid tapping in the word-list (cf. *metal* and *meddle*) but will use it spontaneously in reading the following sentence: "Today, the look is more business-like: metal-rimmed [mɛɾɪrɪmɪd] glasses, a grey suit, etc". Or again, a British speaker who has opposed short 'a' (e.g. *pant*, *ants*) and long 'a' (e.g. *plant*, *master*, *dance*) in the word-list might pronounce words like *glasses* with a short 'a' in the above sentence [glæsɪz] (RP [glɑ:sɪz]), thereby showing that this opposition is not necessarily part of his or her vernacular.

Text

Christmas interview of a television evangelist © PAC Project

If television evangelists are anything like the rest of us, all they really want to do in Christmas week is snap at their families, criticize their friends and make their neighbours' children cry by glaring at them over the garden fence. Yet society expects them to be as jovial and beaming as they are for the other fifty-one weeks of the year. If anything, more so.

Take the Reverend Peter 'Pete' Smith, the 'TV vicar' who sends out press releases in which he describes himself as 'the man who has captured the spirit of the age'. Before our 9 a.m. meeting at his 'media office' on Crawshaw Avenue, South London, he faced, he says, a real dilemma. Should he make an effort 'to behave like a Christian' – throw his door open, offer me a cup of tea – or should he just play it cool, study his fingernails in a manner that showed bored indifference and get rid of me as quickly as possible? In the end, he did neither.

'As a matter of fact, John,' he says in a loud Estuary English twang, 'St Francis said, "At all times preach the gospel and speak whenever you have to." But hey, he didn't mean "Be on your best behaviour and be happy all the time." I could have been extra-polite to you, but the real me would have come out as I was talking. You cannot disguise what you are.'

'And what are you then, Pete?'

'Well, I'm a Christian, John. I've been one since I was 14. And I know for sure that Christianity will be judged more on who you are rather than what you have to say about it. Many church leaders don't appear to understand this. They think we can only be really Christian when we are ramming the doctrine of the Creation down people's throats. But if you try to force-feed people they get sick of it and think you're a pain. It's seen as the job of a Christian leader to wear a dog-collar and dress in purple and always be talking about the real meaning of the New Testament. In reality, that turns people right off!'

In many ways, 'Pete' Smith looks exactly how you'd expect a high-profile, born-again Christian to look: tall, handsome, clean-cut and evenly sun-tanned. He has those scarily white teeth that TV evangelists tend to have, and he doesn't wear a dog-collar. In fact, when doing his various religious programmes on Sunday mornings, he has been known to wear a black leather jacket instead, in casual mode. Today, the look is more business-like: metal-rimmed glasses, a grey suit, a blue open-neck shirt, and fashionable black shoes with large buckles. Smith is 44 but he looks a mere 24.

During the whole interview, there wasn't any talk of the poor or the needy but only of his forthcoming trip to China in February and the masses waiting for his message there. I ventured a few questions relating to the charity trust he founded some ten years ago and which, it is generally agreed, employs eight hundred staff and runs schools, hospitals and hostels around the world. And what about the gambling organization he has been willing to advise? Is that a temporary activity or might it be true that he has accepted to be paid to sit on its Board of Directors? Which side is religion on these days? Does money matter? It was as if I had launched a few missiles in his direction. He just sighed in answer: 'I'm only human, John. God knows I do my best and often fail, But it's no skin off my nose if our enemies sneer at some of the good work we do. Truth will out.'

Naturally, continuous speech opens up a new dimension beyond individual words (what is sometimes called 'utterance phonology' or 'post-lexical phonology'). The text is specifically contrived to examine linking material between words, such as linking-R and intrusive-R (see e.g. Durand 1997), as well as other processes that might operate across word boundaries, like tapping or aspiration, and further phenomena that might be favoured by the connected nature of discourse, such as H-dropping, vowel reduction or palatalization. Once again, we stress that while reading aloud can create artificial pronunciations, it provides access to data which might not crop up in spontaneous discourse (even in an extensive corpus) and it provides an indispensable element of comparability between different speakers. There are categories of speakers who, of course, cannot be studied by this method. On the other hand, literate speakers of English cannot be dealt with as if they were speakers of a language without a written system. There are many demonstrable connections between phonological and orthographic competence. A good way of studying this question empirically is to explore the effect that the graphemic system can have on phonological systems. Our claim is that it can partially be done through the methodology adopted here (see Durand & Eychenne to appear, for a detailed examination of this point in connection with French schwa).

1.3 Formal and informal conversation

Two forms of spontaneous speech set in different contexts are also recorded for each *PAC* corpus. Ideally, we recommend that surveys be done by two investigators: one external to the group being studied, the second who is a member of the group. The formal interview involves the fieldworker and an informant. It has two main objectives: on the one hand, through the interviewer's questions, it provides invaluable background information on the speakers, and, on the other hand, it helps create a speech style which is somewhat more controlled than the informal conversation and which provides a transition between reading aloud and natural conversation between friends. The informal conversation is recorded either with two or three informants without the investigator being present, or with one or two informants and a fieldworker if the latter is a member of the network which is under study. This is the most casual context of the four styles. Readers should be warned that, beyond three speakers, conversations are impossible to unravel for the purpose of phonological analysis: cross-

conversations make it impossible to perform a reasonable acoustic analysis of the data. We therefore recommend that informal conversations should be limited to a maximum of three participants. Of course, if our objective was to study speech interaction (e.g. patterns of dominance within a group), the restriction to two or three informants would not be acceptable.

In terms of linguistic study, the recordings obviously lend themselves to various types of exploitation (including syntax and pragmatics). However, as mentioned earlier, all participants in the project commit themselves to studying phonological inventories, rhoticity and T/D. Beyond these questions, researchers will pursue their own interests relating to specific segmental or phonemic phenomena and/or supra-segmental features such as fast speech processes, stress, rhythm, intonation, etc. It should be stressed that the project remains neutral as to what the best linguistic analysis is for a given set of observations. Our studies are compatible with various theoretical options such as Optimality theory, Lexical Phonology, Dependency Phonology or Laboratory phonology, to name but a few frameworks.

2. Informants and investigators

There are some basic criteria for informants to meet in order to participate in the survey. First of all, informants should clearly belong to the community or area under investigation: this means that they were ideally born and bred there and, if not, that they should have at least come to the area before primary school. Nevertheless, reasonable concessions occasionally have to be made since work within a close network will often favour interviewees who only partially fulfil this geographical criterion. Secondly, a basic level of reading skills is required in order for the speaker to be able to perform the reading task, thus very young and some less educated informants are excluded from the full-scale survey. Thirdly, if possible, distinct social groups should be included within the study. Finally, and most importantly, the target group should be made up of an equal number of male and female speakers falling into at least two (preferably three) well-defined age groups.

As outlined in the introduction, in the *PAC* project we concentrate on small size groups of around twelve speakers, although in specific cases we may want to work with smaller or larger sets of informants. There are manifold reasons for the approach we adopt. First of all, our project is a phonological/phonetic one making room for the study of variation but not devoting itself to the latter exclusively. One difference between our project and a fully sociolinguistic one is that we investigate the whole phonological system characterizing individual speakers rather than a number of variables. Clearly, it is unrealistic to expect to keep track of individual systems once we concentrate on large numbers of speakers (say upwards of one hundred informants). Secondly, as part of our work, we are also attempting to build a reasonably extensive database of orthographic transcriptions and phonological codings for future studies (see Durand & Pukli this vol). Again, this is not realistic if our survey applies to hundreds of speakers. Thirdly, one of our initial aims is geographical coverage with a common methodology and, again, this is unrealistic if each area selected involves large numbers of informants. Finally, if we wish to obtain usable data it is indispensable

to work with small networks of informants. To find informants willing to take part in a recording, talk to a stranger and read lists of words and a text aloud, a situation of mutual trust has to be established. This can only be done if the investigator is the friend of a friend or the friend of a relation. A number of preliminary *PAC* surveys failed because the investigator had not established a sufficiently close rapport with the informants for the recordings to be fully successful.

In a sense, therefore, our approach is nearer to dialectology than to variationist studies. On the other hand, unlike traditional dialectology (see e.g. Chambers & Trudgill 1997), we do not privilege older, male speakers who are supposed to represent the ideal dialect of the area under study. We do build into our methodology a certain amount of variation as enumerated above. Our own approach can therefore be considered as the basis for more extensive sociolinguistic studies for each of the areas we survey.

We note in passing that numbers around thirty have often become standard in recent sociolinguistic surveys. While we cannot but agree that thirty speakers is better than, say, ten, we are not sure that statistical representativeness can truly be achieved with numbers of speakers which are so small and that reliable trends for the speech of a whole community can always be inferred on this basis. When they are, it is probably because the investigator knows the community and its speech so well that the sample has been constructed in the best way to bring out the relevant parameters of the variety under study. It is not an accident if Labov's well-known study of New York speech (Lower East Side), which was based on firm sociological principles, involved 122 informants.

3. Further methodological issues

In the above sections, we have focused on the protocol to be followed in a *PAC* survey. The success of a project such as *PAC* depends not only on the protocol itself, the choice of informants and the ability of the investigator but also on following rigorous procedures in the whole chain from the collection of data to its interpretation. Our own approach is based on methodological principles adopted within the *PFC* project. First of all, the choice of recording equipment and training in recording techniques is a prerequisite for gathering material good enough to be transcribed and analysed auditorily and acoustically (see Tarrier 2003). Secondly, the transcription and coding stages are also crucial. Our own approach requires for there to be an orthographical transcription aligned on the signal devised within a well-known tool used in the computing community, PRAAT, created by David Weenink and Paul Boersma of the University of Amsterdam (see Durand & Pukli this vol., and Delais-Roussarie, Meqqori & Tarrier 2003 for a detailed study). Once orthographical transcriptions aligned with the acoustic signal are available various techniques can be applied to describe and analyse the data. Software tools have been created within the *PFC* project in order to allow us to compare auditorily all the words in the wordlist for either one speaker or a group of speakers (see e.g. Meqqori & Durand 2003a). A more advanced phonetic tool has been devised by Noel Nguyen and Robert Espesser of the 'Laboratoire Parole et Langage' of the University of Aix-en-Provence allowing the user to extract semi-automatically formant charts for the vowels in

our corpus (only word-lists at this stage but extensions are foreseen for the text and spontaneous speech). A widely applied technique used in the *PAC* project is the coding of information within the orthographic tier (which can be duplicated and enriched notationally within PRAAT). Phenomena which have proved amenable to this kind of treatment are schwa and liaison in French but currently a group of researchers is investigating the treatment of prosody with similar techniques (see Bouzon, Auran & Hirst this vol. for further general discussion). Clearly, linking phenomena in English (so-called linking ‘r’ and intrusive ‘r’) can be analysed in the same way. The coding of information has been accompanied by the devising of a set of computational tools which extract and classify the data automatically (see Meqqori & Durand 2003 b,c, as well as Durand & Eychenne, to appear, for recent exemplification).

4. Conclusion

In this article, we have provided an overview of the *PAC* project. In the wake of groundbreaking work in sociolinguistics associated with names like William Labov in the USA or Peter Trudgill, James Milroy and Lesley Milroy in Great-Britain, we have defended an approach where the collection of data and the construction of a database is part and parcel of doing linguistics. While we do not reject the use of ‘intuition’ (grammaticality judgments) as part of the process of obtaining evidence to test hypotheses, it is clear that it needs to be supplemented by other techniques. Through the construction of corpora, we can check and recheck data, analyse it along different dimensions (e.g. statistically, acoustically, etc.) and discover areas where additional testing may be required (e.g. through psycholinguistic experimentation). We believe that, while this is not the only strategy possible, it represents an important way forward for descriptive and theoretical linguistics. While many issues remain open as shown for instance in Carr (1999b, 2000) or Durand (1993), current developments within the *PFC* and the *PAC* projects give us hope that our goals are worthwhile ones.

REFERENCES

Abercrombie, David (1979). The accents of standard English in Scotland. In A.J. Aitken & T. McArthur, Edinburgh: Chambers, 113-123. Reprinted in Abercrombie 1991, 54-70.

Abercrombie, David (1991). *Fifty years in phonetics*. Edinburgh: Edinburgh University Press.

Bouzon, Caroline, Cyril Auran & Daniel Hirst (this volume). *Comparative approaches to prosody across dialects of English*.

Burton-Roberts, Noel, Carr, Philip & Docherty, Gerard (eds.)(2000). *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press.

Calabrese, Laetitia & Marion Coadou (this volume). *Aspects of the rhythm and intonation of Welsh English*.

Carr, Philip (1993). *Phonology*. London: Macmillan.

- Carr, Philip (1999a). *English phonetics and phonology: An introduction*. London: Blackwell.
- Carr, Philip (1999b). Sociophonetic variation and generative phonology. The case of Tyneside English. *Cahiers de grammaire* 24: 7-16.
- Carr, Philip (2000). Scientific realism, sociophonetic variation and innate endowments in phonology. In Noel Burton-Roberts, Philip Carr & Gerard Docherty (eds.). *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press. 67-104.
- Carr, Philip (this volume). *London English*.
- Carr, Philip & Jacques Durand (this volume). *General American and New York City English*.
- Chambers, J. K. & Peter Trudgill (1998). *Dialectology*. (2nd edition.) Cambridge: Cambridge University Press.
- Chirrey, Deborah (1999). Edinburgh: descriptive material. In Paul Foulkes & Gerard Docherty (eds.) *Urban voices*. London: Arnold. 223-229.
- Colman, Fran (this volume). *On diachronic linguistics, variation and English phonology*.
- Cruttenden, Alan (ed.)(2001). *Gimson's pronunciation of English*. London: Arnold.
- Delais-Roussarie, Elisabeth, Abderrahim Meqqori & Jean-Michel Tarrier (2003). Annoter et segmenter des données de parole sous PRAAT. In Elisabeth Delais-Roussarie & Jacques Durand (eds.)(2003) *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse: Presses Universitaires du Mirail. 159-185.
- Durand, Jacques, Catherine Slater & Hilary Wise (1987). Observations on schwa in southern French. *Linguistics* 25: 983-1004.
- Durand, Jacques (1990). *Generative and non-linear phonology*. London: Longman.
- Durand, Jacques (1999). R postvocalique et histoire de l'anglais. A tale of two countries. A tale of two cities. *Sigma (Anglophonia)* : 199-221.
- Durand, Jacques (1993). Sociolinguistic variation and the linguist. In Carol Sanders (ed.) *French Today : Language in its Social Context*. Cambridge : Cambridge University Press. 257-285.
- Durand, Jacques (1997). Linking r in English: constraints, principles and parameters, or rules? " *Histoire, Epistémologie, Langage*. Vol.19/I, pp. 43-72.
- Durand, Jacques (2001). La transcription phonétique de l'anglais : problèmes théoriques et pratiques. In Wilfrid Rotgé (ed.) *Agrégation externe d'anglais : mode d'emploi*. Paris: Ellipses. 181-219.
- Durand, Jacques (this volume). *English in early 21st century Scotland: a phonological perspective*.
- Durand, Jacques & Chantal Lyche (2000). La phonologie : des phonèmes à la théorie de l'optimalité. In Pierre Escudier et alii. (eds) *La parole : des modèles cognitifs aux machines communicantes*. Paris : Hermès.
- Durand, Jacques & Bernard Laks (eds.)(2002) *Phonology, Phonetics and Cognition*. Oxford University Press.
- Durand, Jacques, Bernard Laks & Chantal Lyche (2002). La phonologie du français contemporain: usages, variétés et structure. In: C. Pusch & W. Raible (eds.) *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, 93-106.

Durand, Jacques & Chantal Lyche (2003). Le projet 'Phonologie du français contemporain et sa méthodologie'. In Elisabeth Delais & Jacques Durand (eds.) *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse : Presses Universitaires du Mirail. 213-276.

Durand, Jacques, Bernard Laks & Chantal Lyche (2003a). Linguistique et variation : quelques réflexions sur la variation phonologique. In Elisabeth Delais & Jacques Durand (eds.) *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse : Presses Universitaires du Mirail. 11-88.

Durand, Jacques, Bernard Laks & Chantal Lyche (2003b) Le projet 'Phonologie du français contemporain' (PFC). *La tribune internationale des langues vivantes*. 33 : 3-9.

Durand, Jacques, Bernard Laks & Chantal Lyche (eds.)(2003c) *La prononciation du français dans sa variation*. Vol. 33 of *La tribune internationale des langues vivantes*.

Durand, Jacques & Monika Pukli (2004). PAC project: methods and tools. *Carnets de grammaire*. CNRS et Université de Toulouse-Le Mirail: Rapports internes de l'ERSS, vol. 14.

Durand, Jacques & Julien Eychenne (to appear). Le schwa en français: pourquoi des corpus? *Corpus*.

Durand, Jacques & Monika Pukli (this volume). *How to construct a phonological corpus: PRAAT and the PAC project*

Ford, John (this volume). *Evolution and Devolution: An Examination of the Historical Development of Scottish English*.

Foulkes, Paul & Gerard Docherty (1999). *Urban voices. Variation and change in British accents*. London: Edward Arnold.

Hughes, Arthur & Peter Trudgill (1996). *English accents and dialects. An introduction to social and regional varieties of English in the British Isles*. (Third edition. First edition 1979). London: Arnold.

Labov, William (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1994). *Principles of Linguistic Change. Internal Factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change. Social Factors*. Oxford: Blackwell.

Meqqori, Aberrahim. & Jacques Durand (2003a). Manuel d'utilisation du comparateur. *Bulletin PFC 2* : 33-38 CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Meqqori, Aberrahim. & Jacques Durand (2003b). Manuel d'utilisation du classeur-liaison. *Bulletin PFC 2* : 23-29 CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Meqqori, Aberrahim. & Jacques Durand (2003c). Manuel d'utilisation du classeur-schwa. *Bulletin PFC 2* : 29-32. CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail.

Milroy, James & Lesley Milroy (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339-384.

Milroy, Lesley (1987a). *Language and Social Networks*. 2nd edition. Oxford : Blackwell.

- Milroy, Lesley (1987b). *Observing and Analysing Natural Language*. Oxford : Blackwell.
- Montreuil, Jean-Pierre (this volume). *Fragmenting Weight in Scottish English*.
- Moore, Steven (this volume). *RP – Past , present and future*.
- Noel, Emilie (2003). English phonology in Central Lancashire: a dialectological study. Unpublished MA [Maîtrise] Dissertation. Université de Toulouse Le Mirail : Département des Etudes du Monde Anglophone.
- Pukli, Monika (2002). Le “tapping” en anglais américain: une etude empirique. Unpublished DEA Dissertation. Université de Toulouse Le Mirail : Département des Etudes du Monde Anglophone.
- Pukli, Monika (this volume). *Scottish English and The Scottish Vowel Length Rule – An Empirical Study of Ayrshire Speakers*.
- Przewozny, Anne (this volume). *Variation in Australian English*.
- Stuart-Smith, Jane (1999). Glasgow: accent and voice quality. In Paul Foulkes & Gerard Docherty (eds.) *Urban voices*. London: Arnold. 203-222.
- Stuart-Smith, Jane (2001). Accent change in Glaswegian : a sociophonetic investigation. Final report to the Leverhulme Trust : <http://www2.arts.gla.ac.uk/SESL/EngLang/research/accent/Report.htm>
- Tarrier, Jean-Michel (2003). L’enregistrement et la prise de son. In Delais-Roussarie, Elisabeth & Jacques Durand (eds.)(2003) *Corpus & variation en phonologie du français : méthodes et analyses*. Toulouse: Presses Universitaires du Mirail. 187-212.
- Trudgill, Peter & Hannah, Jean (1985). *International English. A guide to varieties of standard English*. (Second edition. First edition, 1982) London: Edward Arnold.
- Wells, John C. (1982). *Accents of English*. 3 volumes. Cambridge: Cambridge University Press.