Differences in the scope of obstruent voicing assimilation in learners' English as a consequence of regional variation in Polish.

Robert Lew *(rlew@amu.edu.pl)*

1. Introduction

The question of what factors, and to what extent, shape the phonetic performance of second language learners has been the subject of much debate (see for instance Ioup & Weinberger 1987 or James & Leather 1987). One factor whose importance has remained, and is likely to remain, relatively unchallenged is the influence of the sound system of learners' native language. This effect is usually referred to as L1 transfer in the ESL/EFL literature, as parameters (speech habits, features, realisations, rules, processes, etc. – specific formulations vary with different approaches) of native language (L1) pronunciation are *transferred* or carried over to the second (L2, learned) language. In the present study I undertake to investigate if, and how, regional differences in Polish with respect to regressive sandhi voicing assimilation may influence the corresponding aspects of learners' English pronunciation. I will also consider some practical implications of the results for pronunciation teaching.

2. Sandhi voicing in Polish

In Polish, members of a cluster of obstruents almost always agree in voicing, thus resulting in a uniformly voiced or uniformly voiceless cluster. Occasional older forms running counter to this generalisation such as [kv], [sv], [xv] are rapidly disappearing (Madejowa 1990: 85-86; 1993: 25). The constraint disallowing differential-voicing obstruent clusters also generally holds for clusters separated by a word (and some types of morpheme) boundary (Strutyński 1996: 52). Across such a boundary, the voicing feature of the onset obstruent or obstruents of the following word (morpheme) will determine the voicing of the word-final obstruent or obstruents in the preceding word (morpheme). For clarity, I will illustrate this *regressive assimilation of voicing* with some examples: both <kot sam> and <kod sam> will yield a voiceless sandhi cluster, thus [kɔtsam], whereas each of<kot wam> and <kod wam> will be realised as uniformly voiced, thus [kɔdvam] – unless of course a pause is inserted between the two words.

Things get more complicated when the second word begins with a sonorant (I will henceforth use this term in the broader sense, referring to a vowel, nasal, or approximant, i.e. any non-obstruent). In such contexts, the voicing of any preceding word-final obstruent or obstruents varies geographically: they tend to be voiced in Silesia, Wielkopolska, and Małopolska (*voicing* pronunciation): <kot/kod albo> [kɔdalbɔ], but voiceless in Mazowsze and Pomerania (*devoicing* pronunciation): <kot/kod albo> [kɔtalbɔ] (Dejna 1994: Map 6; Nagórko 1996: 55; Strutyński 1996: 52-53).

Two further complications arise. Firstly, prepositions appear to behave in a slightly different way, undergoing regressive voicing for speakers of all varieties, so that <bez awantur> is usually pronounced as voiced [bɛzavantur] irrespective of the accent involved. Secondly, it has been noted (Horwath 1984) that voicing does not always follow the patterns outlined above (that, of course, is the expected state of affairs for any linguist interested in anything beyond the idealised speaker-hearer). There is some evidence (Madelska, personal communication, based on data for Madelska 1987) that devoicing pronunciation may be gaining ground, at least among the younger speakers of Wielkopolska.

3. Possible effect on English pronunciation

If the characterisation of sandhi voicing phenomena given above is correct, and if regressive voicing patterns are subject to L1 → L2 transfer, then it might be predicted that:

- Polish learners of English should generally exhibit a tendency to voice word-final fortis obstruents if the following word begins with a lenis obstruent (as in <less than>, <think that>, <at the>);
- Those Polish learners of English who come from regions of voicing pronunciation should exhibit a tendency to voice word-final fortis obstruents if the following word begins with **any** voiced sound (so apart from the above examples, in contexts of the type <if you>, <think I>, <night long>).

The main aim of the present study is to verify experimentally whether the above predictions are correct.

4. Regressive voicing assimilation in native English

According to authoritative sources, regressive sandhi voicing before consonants is very rare in native English speech but is a noticeable feature of a foreign accent. Peter Roach claims that:

> When C^f [word-final consonant] is fortis ("voiceless") and C^i [word-initial consonant] lenis ("voiced"), a context in which in many languages C^f would become voiced, assimilation of voice *never* takes place [in English]; consider the following example: 'I like that black dog' aɪ laɪk ðæt blæk dɒg. It is typical of many foreign learners of English to allow regressive assimilation of voicing to change the final k of 'like' to g, the final t of 'that' to d and the final k of 'black' to g. This creates a very strong impression of a foreign accent.
>
> (Roach 1991: 125; Roach's emphasis, my editorial clarifications)

Gimson is somewhat more tentative in his estimate of frequency of regressive voicing:

> It is to be noted that word or morpheme final fortis consonants in English rarely show tendencies to assimilate to their lenis counterparts: such pronunciations of *nice boy, black dress, half-done, they both do, wishbone, birthday* as /naɪz bɔɪ, blæg dres, hɑːv dʌn, ðeɪ bəʊð duː, wɪʒbəʊn, bɜːðdeɪ/ are typical of many foreign learners.
>
> (Gimson 1989: 296)

At another point Gimson categorically calls on the foreign learner "*to avoid un-English assimilations such as* /aɪ laɪg ðæt/ *(incorrect voicing)* […]" (Gimson 1989: 307). Peter Trudgill specifically suggests that "*[i]f one does hear it, it is a sure sign of a foreign (and usually Slavic!) accent*" (Trudgill, personal communication). Andrew Butcher (personal communication) points out that regressive voicing is rare before consonants, but happens quite commonly before vowels in Australian English. There is of course the phenomenon of prevocalic t-voicing, also known as tapping or, more loosely, flapping, which happens in some accents of English, primarily North American ones (see, for instance, Bronstein 1960: 74-75), but has not, to my knowledge, been found to be a frequent phenomenon in RP. In view of all of the above, I will assume that regressive voicing in RP (the teaching model for the students serving as subjects in the study) is of negligible frequency.

5. Collection of data

All data was provided in the form of analogue compact tape recordings by Sylwia Scheuer of the School of English, Adam Mickiewicz University.[1] The recordings had been produced as part of the ongoing Phonetic Polish-English Corpus of Learner English project, known for short as the Phonetic PICLE. Details of this project may be found in Scheuer (1996; in press), so only the most important information will be given here.

Seventeen subjects were recorded, all of them first year English majors at the School of English, Adam Mickiewicz University. At no point were the subjects notified that the focus of the study would be on pronunciation. Recordings were completed in three recording sessions. The first of these took place on 11th October 1995, a week after the students serving as subjects were admitted into the University. They were asked to fill in questionnaires specifying, among others, their age, place of birth and details of their previous residence. Each subject was asked to read an unrehearsed passage of text of about 200 words taken from an intermediate-level EFL coursebook. The passages were different for all subjects. Then, in an interview setting, each subject was asked to tell the experimenter how they had spent their summer holidays. Essentially the same procedure was repeated in the two May sessions (15th and 22nd of May 1996), in which each student was given the same reading passage as in October. This time, subjects were asked to share with the interviewer their impressions of their first year at the School of English. The total duration of all recordings is slightly below 2.5 hours. Subject number 10 was discarded for the purposes of this study, as data for this subject was largely incomplete. Of the remaining 16 subjects, most were 19 years old at the time of the first recording session, and all were within the 19 to 22 range. All students were in the same teaching section: they took the same courses with the same British or Polish instructors, and were taught British English with RP as the pronunciation model. They received two

[1] I would like to take this opportunity to express my gratitude to Sylwia Scheuer for making the data available to me.

hours a week of practical training in English phonetics throughout the whole academic year. In an interview with the instructor I have established that the focus of the practical phonetics course was on segmental accuracy, but some attention was given to excessive devoicing and regressive voicing, mainly in an opportunistic, unsystematic manner.

6. Corpus building and tagging

Recordings of the 16 subjects were first transcribed orthographically from tape, yielding a total of approximately 17,000 orthographic words. Then all potentially interesting sandhi environments were tagged as follows:

<ES.> potential voicing environment before a sonorant (fortis obstruent + word boundary + sonorant)
<EO.> potential voicing environment before an obstruent (fortis obstruent + word boundary + lenis obstruent)

The dots within the tag codes above represent an as-yet-unspecified portion of the tags to be filled in in due course (see below). Once the tags were in place, they were extracted together with the immediate context with a GNU awk script written by the present author for this particular purpose, and the resulting KWIC (TIC ?) lists were double-checked for misattributed tags.

Once the <ES.> and <EO.> tags were all in order, recordings were examined auditorily for the presence of voicing or pause in the (previously tagged) contexts of interest. A selection of passages were digitised (at a 16-bit resolution and 22kHz sampling rate) and examined auditorily at half speed and visually for traces of periodicity in the waveform. At this point, the "underspecified" <ES.> and <EO.> tags were supplemented by the third character depending on the actual phonetic realisation:

<..V> voicing of the word-final obstruent,
<..N> no voicing of the word-final obstruent,
<..P> word boundary realised as pause (and, usually, no obstruent voicing),

where the two dots stand for either ES or EO, as appropriate. The <..V> tag was only given if I was satisfied of the presence of uninterrupted voicing extending from the vowel preceding the word-final obstruent(s) to the initial segment of the following word. For those cases for which waveforms were examined, the primary criterion was the presence of a clearly defined periodic component. This method was essentially only applied to those items for which a second auditory examination conducted two weeks after the original one yielded a different determination from the first one. There were about 30 such realisations. The <..P> tag was used when there was a discernible pause between the two words. In all remaining cases, the <..N> tag was transcribed.

Apart from the tags representing phonetic phenomena, the following non-phonetic tags were also used in the mark-up of the corpus. The first three of these served as paragraph headings:

[I=*n*] *n*-th subject as identified by questionnaire number
[P=V] subject from the voicing dialect area of Poland
[P=D] subject from the devoicing dialect area of Poland
[XP] untranscribed fragment by experimenter

The *n* in [I=*n*] is an integer variable with values ranging from 1 to 17 (it will however be recalled that subject number 10 was discarded).
The [XP] tag was used in place of transcription for any fragment spoken by the interviewer, such as questions, explanations, etc.
The values of the [P=.] tags (V for voicing and D for devoicing area) were determined on the basis of Dejna (1994: Map 6) by plotting the locations derived from questionnaires on the map. For the western area of Poland acquired after WW II for which Dejna gives no information I used Strutyński's (1996: 52-53) criteria, treating it as D ('devoicing area'). The narrow area of overlap which Dejna (1994: caption to Map 6) describes as an area of expansion of sandhi devoicing pronunciation with relics of voicing across word-internal morpheme boundaries was classified as D ('devoicing area'). Information as to whether the passages were spoken or read, or came from the October or May recording sessions, was not tagged in paragraph headings. Coding for these two variables is by means of unique strings within filenames and extensions.

It is important to note that the values of the [P=.] tags were determined and entered into the corpus only after the tagging of voicing had been complete. This was a deliberate decision motivated by a desire to avoid one possible source of experimenter bias. While transcribing the phonetic phenomena, I was not aware of what [P=.] tag I would later assign to a particular passage.

7. Analysis of data

7.1 The choice between pause-free and all contexts

One issue in calculating the proportion of voiced realisations is whether the divisor should be the number of all potential contexts, or just of those with no pause. To some extent, the decision rests on what sort of inferences one would like to draw from the data. The former method would provide a measure of the relative frequency of voiced realisations to all contexts. The divisor would then include not only pauses resulting from hesitation- and fluency-related factors, but also those naturally occurring at major syntactic boundaries. Since there were reasons to suspect that the frequencies of fluency-related and hesitation pauses might differ systematically between reading and speaking (due to the need to attend to the informational content in speech and thus different amounts of attention that the reader/speaker is able to give to the monitoring of different aspects of speech), and between October and May (due to the extensive language training received between these points in time and the likely increase in fluency following therefrom)[2], using the first approach to calculating the proportion of voiced realisations might confound the experimental effects with additional undesirable components, and make direct comparisons between certain subsets of the data impossible, or at least questionable. Further, it should be noted that the regressive voicing processes in Polish only apply in pause-free environments. In the presence of a pause, obstruents are generally voiceless. Given the aims of the experiment, it appears that calculating a proportion of voiced realisations with pause-free contexts in the divisor is by far the preferable option, and this has been my choice.

7.2 Quantification of data

Once the tagging was complete, another awk script was written and used to automatically count all tags. With the tag-extraction script, comprehensive lists of all categories of tags were generated and carefully inspected for potential errors and problems. The total number of pre-obstruent contexts (i.e. sequences of fortis obstruent + word boundary + lenis obstruent) was found to be 646, broken down as follows:

<EON>: 208
<EOP>: 215
<EOV>: 223

The total number of pre-sonorant contexts (i.e. sequences of fortis obstruent + word boundary + sonorant) was found to be 2045, broken down as follows:

<ESN>: 1195
<ESP>: 721
<ESV>: 129

Even from these summary figures some useful information can already be extracted. It turns out that the frequency of pre-sonorant contexts exceeds that of the pre-obstruent contexts by a factor of three. Because of the sheer difference in token frequency, there is thus a much greater theoretical potential for pre-sonorant contexts than for pre-obstruent contexts to lead to pervasive pronunciation problems. The proportion of pause realisations is virtually identical for pre-sonorant and pre-obstruent contexts, both figures being close to one third.[3]

Pooling the frequencies for all subjects in those contexts in which pause is not present, roughly every second pause-free pre-obstruent context triggers regressive voicing, as does every tenth pre-sonorant context. Such a generalisation is valid only to the extent that each subject contributed an equal proportion of contexts, which has

[2] As it turned out, none of these suspicions were borne out by the experimental data.
[3] The fact that these two figures turned out to be so similar may provide an indication of the consistency of transcription. An indication, rather than strong evidence, since we have no independent check that pauses in the two contexts are indeed equally frequent. I see no particular reason, however, why this should not generally be the case.

not been the case. Still, these figures do correctly suggest that overall (all subjects, October and May, speaking and reading) a larger proportion of potential pre-obstruent contexts trigger regressive voicing than is the case for pre-sonorant contexts.

Details of experimental data in the form of raw frequencies are given in Table A and Table B. More transparent summaries of the data will be presented in later sections of this paper, and so some readers may at this point wish to proceed directly to section 7.3 or, if they are not interested in methodological details, straight to section 7.6.

Column **Area** of Table A and Table B shows the code representing the value of the geographical grouping variable. The value was assigned on the basis of the questionnaires and reflects the area of Poland where a given subject has spent most of his or her life. The detailed criteria for assigning V or D values have already been given in section 6 above. The **Subj** column identifies the subject by questionnaire number. Note that number 10 is missing: this is the subject who was discarded due to incompleteness of data. Frequencies in the tables are split into October and May, and then, within each of the two, into Reading and Speaking. For each of the four combinations of factors, three types of frequencies are given. **Total** refers to the total number of potential voicing contexts and was calculated as the total number of <E.N>, <E.P> and <E.V> tags, where the dot represents S in Table A and O in Table B. The **NPaus** columns give just the pause-free contexts obtained by adding the <E.N> and <E.V> tags (equivalent to subtracting the <E.P> tags from the **Total**). The **Voice** columns simply report the appropriate number of the <E.V> tags. The **MD** codes appearing for subjects 9, 11, and 16 in the October Speaking part of the tables stand for missing data. October recordings for these subjects did not include oral interviews. The three summary measures given towards the end of both tables represent the ratio of actually voiced contexts to all pause-free contexts expressed as percentages for October data, May data, and all data, respectively. The lowermost rows of the two tables designated as **Totl** give column totals for the raw data part of the table, and percentages based on these pooled (unweighted) subject totals in the three final columns[4].

[4] Please note that the figures in the last three cells of the bottom rows of Table A and Table B are **not** column means; this would obtain only if the contributions of all subjects were exactly equal.

Table A: Frequencies of word-final fortis obstruent realisations before sonorants

		October						May						Oct	May	All
		Reading			Speaking			Reading			Speaking			R&S	R&S	R&S
Area	Subj	Total	NPaus	Voice	Total	NPaus	Voice	Total	NPaus	Voice	Total	NPaus	Voice	%Voi	%Voi	%Voi
V	1	27	19	1	80	52	0	32	24	0	88	53	3	1.41	3.90	2.70
D	2	18	10	0	30	24	0	18	14	0	48	38	1	0.00	1.92	1.16
V	3	49	31	2	15	7	1	50	34	1	43	24	1	7.89	3.45	5.21
V	4	47	27	1	35	24	2	49	25	0	66	33	9	5.88	15.52	11.01
D	5	38	27	1	32	22	0	36	20	0	25	16	1	2.04	2.78	2.35
D	6	26	21	0	16	10	1	24	19	0	53	35	4	3.23	7.41	5.88
D	7	20	14	1	32	19	0	19	15	0	31	17	0	3.03	0.00	1.54
V	8	46	32	0	18	6	0	42	29	7	22	16	4	0.00	24.44	13.25
D	9	18	16	2	MD	MD	MD	31	27	3	31	23	2	12.50	10.00	10.61
D	11	23	13	0	MD	MD	MD	23	13	0	30	17	1	0.00	3.33	2.33
V	12	15	7	1	58	40	7	15	10	1	75	50	7	17.02	13.33	14.95
D	13	18	10	0	64	40	1	17	14	0	40	18	1	2.00	3.13	2.44
V	14	20	13	6	24	9	2	22	14	6	28	17	5	36.36	35.48	35.85
D	15	25	14	1	57	41	1	26	20	1	11	9	0	3.64	3.45	3.57
D	16	24	15	0	MD	MD	MD	23	19	4	33	22	6	0.00	24.39	17.86
V	17	18	10	4	28	20	9	18	12	5	55	34	12	43.33	36.96	39.47
Totl	16	432	279	20	489	314	24	445	309	28	679	422	57	7.42	11.63	9.74

Table B: Frequencies of word-final fortis obstruent realisations before obstruents

		October						May						Oct	May	All
		Reading			Speaking			Reading			Speaking			R&S	R&S	R&S
Area	Subj	Total	NPaus	Voice	Total	NPaus	Voice	Total	NPaus	Voice	Total	NPaus	Voice	%Voi	%Voi	%Voi
V	1	13	9	5	20	17	4	14	11	4	30	18	10	35	48	42
D	2	5	4	3	9	8	7	5	4	3	16	15	13	83	84	84
V	3	12	9	1	3	2	1	12	10	2	8	8	4	18	33	28
V	4	5	1	1	7	4	2	4	2	0	33	22	9	60	38	41
D	5	16	10	2	6	3	3	16	10	5	15	8	6	38	61	52
D	6	11	8	5	7	2	1	10	6	4	23	17	12	60	70	67
D	7	7	3	2	9	6	4	7	5	3	10	6	2	67	45	55
V	8	20	14	11	6	4	2	16	10	7	14	10	4	72	55	63
D	9	5	2	1	MD	MD	MD	9	7	3	11	7	0	50	21	25
D	11	11	8	4	MD	MD	MD	10	5	0	9	4	1	50	11	29
V	12	5	3	1	12	7	5	4	4	2	8	8	4	60	50	55
D	13	4	0	0	20	12	7	4	2	2	30	15	10	58	71	66
V	14	17	13	7	2	0	0	14	11	3	10	4	2	54	33	43
D	15	4	3	2	10	8	5	4	3	2	3	2	1	64	60	63
D	16	6	6	2	MD	MD	MD	6	6	5	7	5	2	33	64	53
V	17	4	2	0	7	4	2	4	3	3	17	11	5	33	57	50
Totl	16	145	95	47	118	77	43	139	99	48	244	160	85	52	51	52

7.3 The choice of method

The primary reason for including Table A and Table B was to provide the raw data for those wishing to use them in ways other than those given further below in the present paper. These tables do not by themselves give a transparent picture of the phenomena being investigated. To verify whether the frequency of voicing is related to the individual experimental factors, a multi-way analysis of variance (ANOVA) will be applied. This procedure makes it possible to compare interaction effects between factors. For a comprehensive overview of ANOVA see Winer (1962) or Kirk (1982).

The design of the present experiment involves measuring one dependent variable – relative frequency of voicing – for the same 16 subjects grouped into two categories (V, 'voicing area'; D, 'devoicing area' - grouping variable) under four different experimental conditions determined by a further two independent variables; for brevity's sake, I will refer to these two variables as Style (with two levels: reading - speaking), and Time (October - May). There are thus three independent variables involved in the experiment. Since the same subjects are measured repeatedly, this is a repeated measures design.

7.4 Assumptions of ANOVA

Before parametric statistical tests involving comparisons of variances can be employed, however, assumptions about the criterion variable have to be met which underlie the theoretical model of analysis of variance. The assumptions are: independence of data, normality of distribution, and homoscedasticity (homogeneity of variance) (Winer 1962, Welkowitz, Ewen, & Cohen 1971: 212, Rietveld & van Hout 1993: 120-123).

In terms of the independence of data, in the present study the data come from different subjects, and in this sense may be considered independent. Of course, repeated measurements are inherently correlated, and this is an aspect that a repeated measures design capitalises on in extracting the within-subjects source of variation, resulting in an increased sensitivity for effects of the two repeated measures factors. However, we should not forget that between the October and May recording sessions all subjects attended the same classes with the same instructors. Those results then that make use of the October vs. May comparison (effect of training), are not independent from the point of view of generalising to students taking other classes, taught by different instructors, or at other institutions. For this reason, one should be very cautious when trying to generalise any effect of training results to students not directly investigated.

As far as normality of distribution is concerned, distribution plots examined by the author show that for all data distribution is clearly unimodal and close to normal. In any case, statisticians agree that the ANOVA procedure is quite robust to minor violations of the normality assumption.

No such agreement exists, however, with reference to the homoscedasticity requirement. Some authors claim that "*ANOVA is robust with regard to these assumptions [...] and will yield accurate results even if population variances are not homogeneous*" (Welkowitz, Ewen, & Cohen 1971: 212). Others, such as Box (1954) or Wilcox (1987), are less optimistic and warn against the risk of positive bias in the α-level values, especially if the differences between variances are very large. Also, the problem may be compounded by a marked difference in sample sizes.

Nonhomogeneity of variance is a potential problem if underlyingly binomial data are presented in the form of proportions or percentages[5] – as is the case in the present study. It can be demonstrated mathematically that variances tend to vary with the squares of proportion means. In addition, proportions are not measured on a true interval scale[6], as pointed out by Cohen & Cohen (1975; cited in Rietveld & van Hout 1993: 126-127; ellipsis by Rietveld & van Hout): "*it is [...] evident [...] that the difference between 0.01 and 0.05 is much more important than the same difference between .48 and .52*". The nonlinearity and nonhomogeneity problems hardly arise as long as the proportion values stay roughly in the 0.5 (50%) area. The rule-of-thumb advice given by Woods, Fletcher & Hughes (1986: 220) is not to worry if "*most of the subjects in the experiment obtained scores in the range 20%-80%*". This is very clearly the case for the pre-obstruent voicing data (see the three final columns of

---

[5] Percentages are simply proportions scaled up by a factor of 100.

[6] This is another (sometimes thought too obvious to mention) assumption of many parametric statistical procedures. An interval measurement scale is a scale on which equal increments correspond to equal changes in the measured phenomenon.

Table B). For pre-sonorant voicing, however (Table A), most of the values are in fact way below the 20% level. There is reason to suspect, then, that analysing raw proportions for pre-sonorant voicing may not be a legitimate procedure. This suspicion may be tested more formally with any of the several homogeneity of variance tests. Levene's test (recommended for use with distributions whose normality cannot be guaranteed – see Rietveld & van Hout 1993: 122) yields an F value of 1.049 (at 14 df) with p=.323 for overall pre-obstruent voicing data (final column of Table B) grouped by the Area variable (V versus D). An analogous procedure for pre-sonorant data results in an F value of 7.436 (at 14 df), p=.016. These results clearly show that there is no reason to worry about the homoscedasticity assumption in the pre-obstruent data (in fact, F is very close to 1, which points to a high degree of homogeneity between the variances of the V subjects and D subjects measured on proportions of overall pre-obstruent voicing). Things look dramatically different for the pre-sonorant context data, where a low p-level indicates a serious violation of the homogeneity of variance assumption. This confirms our initial suspicion that analysing raw proportion figures for pre-sonorant voicing may result in unreliable calculations of α-levels.

To overcome such problems, statisticians generally advocate transforming the data. For data expressed as proportions, some authors recommend the use of the arcsine transformation (e.g. Winer 1962: 221, Woods, Fletcher & Hughes 1986: 220[7]). Others (Rietveld & van Hout 1993: 127) stress the advantages of the logit (or log-odds) transformation. This transformation – usually given as logit(p)=ln(p/(1-p)), where p is the value of the experimental variable expressed as proportion, ln is the natural logarithm, and logit(p) is the transformed value of the variable – efficiently removes the nonlinearity and nonhomoscedasticity of proportions. I have found that the logit transformation stabilises the variances of my data much more satisfactorily than does the arcsine transformation[8]. It was essentially the logit transformation that I applied to pre-sonorant data, with one modification, however. It will be apparent from a careful inspection of Table A that a minority of the proportions will be zero, quite simply because not a single instance of voicing had been found for some combinations of subjects and treatments. A natural logarithm of zero does not exist, however (it tends to $-\infty$). To rectify this problem, a small value (0.2) was added in all cases to both the numerator and denominator thus yielding corrected proportions, which were later subjected to a standard logit transformation. The addition of 0.2 has very little influence on the values of non-offending proportions, but it remedies the division-by-zero problem (see Rietveld & van Hout 1993: 351, who use a value of 0.5 in the same fashion).

As an example of how the transformation was applied, consider pre-sonorant October Reading data. In the first step, a corrected proportion OR1 was calculated as OR1=(ORV+.2)/(ORNP+.2), where ORV (October Reading Voiced) and ORNP (October Reading No Pause) were the frequencies of <ESV> and <ESV> + <ESN> tags respectively, for a given subject, reading style in October. (For comparison, the formula for uncorrected proportion would simply read OR=ORV/ORNP.) Having obtained the OR1 value, the logitOR was calculated as logitOR=ln(OR1/(1-OR1)). This procedure was repeated for all subjects, all combinations of October – May on the one hand and Reading – Speaking on the other. The resulting logits were subjected to statistical testing. The mathematical advantages of logits granted, it must be admitted that plain proportions have a much more intuitive feel for most readers, and so it is (uncorrected) proportions or percentages of voicing that will routinely be given in any *descriptive* statistics below.

7.5 Fixed versus random effects

One more issue that needs to be dealt with before actual results of statistical tests are given is the choice between fixed and random effects. In the present design all effects are treated as fixed, resulting in what is known as a fixed model. In terms of the grouping (Area) variable, the two values are assumed to exhaust the possibilities (either Voicing or Devoicing Area), and thus it is quite natural to consider the variable fixed. The Time variable (October versus May) does not, of course, exhaust all possibilities, but neither is it a result of random sampling of time. The choice of the two points in time is deliberate and claims will be limited to these two points in time. Thus, this variable should also be considered fixed. Finally, in terms of the Style variable (Reading versus Speaking), it is not known whether other settings, such as word-list reading, or completely natural conversation (if there is such a thing for second-language learners at this level of competence) as opposed to an interview would constitute different styles. In any case, the selection of the two styles is not random, and no generalisations will be attempted beyond the two styles present in the design. In consequence, the Style variable is also treated as fixed.

[7] The actual formulae for the arcsine transformation differ slightly between these two sources.
[8] The relevant p-levels for Levine's test calculated as above are .79 for the logit and .16 for the arcsine.

7.6 Results for pre-sonorant contexts

The ANOVA table for pre-sonorant voicing is given in Table C below.

Table C: 3-way ANOVA, repeated measures, fixed model, 1,11 df
1 dependent variable, logit transformed proportions of pre-sonorant voicing
Between: Area (2 levels: Devoicing area, Voicing area)
Within: Time (2 levels: October, May) x Style (2 levels: Reading, Speaking)

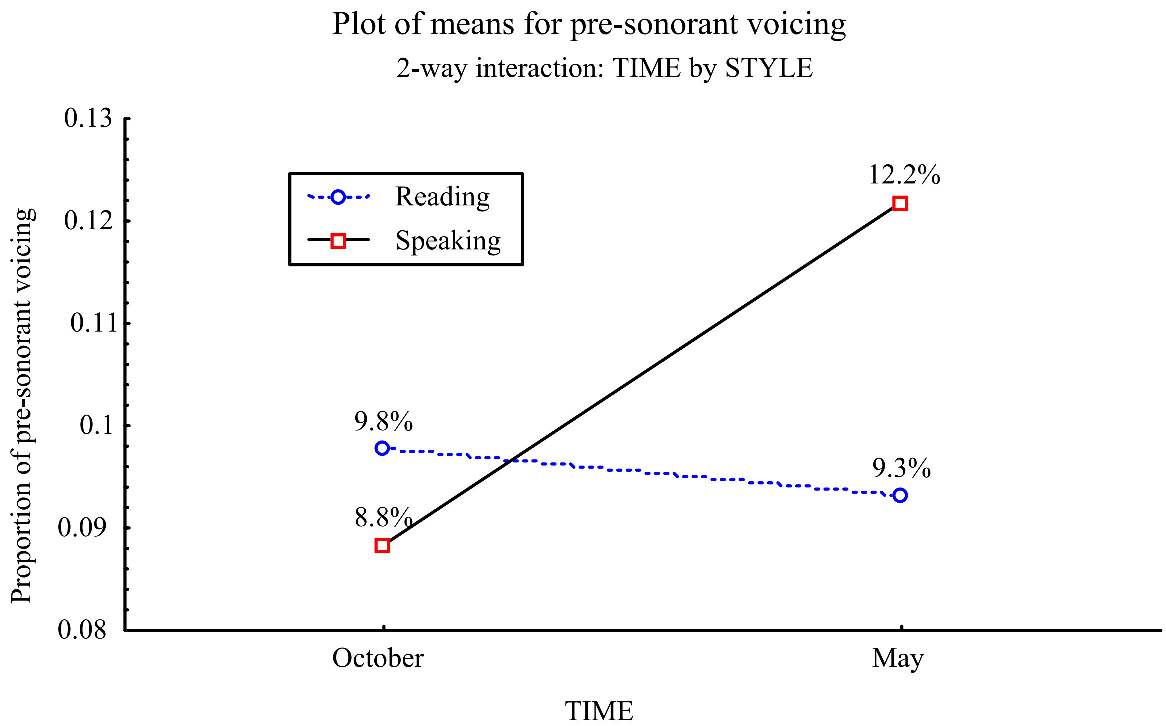
Effect	F	p-level
Area	**7.789**	**.018**
Time	.194	.668
Style	.890	.366
Area by Time	.247	.629
Area by Style	.041	.844
Time by Style	**4.441**	**.059**
Area by Time by Style	.341	.571

Table C above reveals a significant main effect of the Area grouping variable. The group means expressed as percentages across all levels of Style and Time are **17.6%** for the Voicing area group and **5.3%** for the Devoicing area group. From these values and from the significance level given in Table C it can be concluded that subjects from the voicing area of Poland voice fortis obstruents in a significantly greater proportion of pause-free potential contexts before word-initial sonorants than subjects from the devoicing area of the country. In terms of expected values, **V subjects voice over three times as frequently as D subjects**. The probability that there is no real difference here and that the difference found in the analysis is due to mere chance is no more than about 0.02 (in other words, if there is no real difference, then experiments like the present one would be expected to produce differences of at least the magnitude obtained in the present experiment in no more than 2% of the runs).

All other effects, main or interactions, have an F statistic below 1, except for the Time by Style interaction. The significance level of .059 for this interaction falls just short of the customary .05 level commonly accepted as the minimum significance level sufficient for rejecting null hypotheses in behavioural sciences[9]. However, this customary level – to a large extent conditioned by the general availability of statistical tables for this particular α-level in the days when it was not easy to compute the actual p-level but rather values of distribution statistics had to be compared to tabulated values – should not be treated as immutable. It should not be forgotten that a p-level of .059 means that there is at least a 94% chance that the interaction is a genuine one. Neither should it be forgotten that the sample (the number of subjects) is fairly small at quite high inherent variation levels probably largely due to the imperfect correlation between the area of the country a person has been brought up in and the actual Polish accent they end up speaking. These two factors (small sample size and inherent variability) negatively affect the power of a statistical test, understood as its ability to find a relationship where there really is one. At the very least, this result should indicate a need for further study, perhaps on a larger sample. That further study is indeed called for should shortly become evident from the examination of the nature of the Time by Style interaction. A graph of the interaction is given in Figure A below. First, it should be noted that the lines on the graph cross, resulting in what some authors (Rietveld & van Hout 1993: 35) refer to as a disordinal interaction – an interaction that will not be removed by any kind of rescaling. It should be clear from the graph that between October and May the proportion of voicing in reading remained almost unchanged (it fell slightly), while the proportion of voicing in speaking went up from slightly below **9%** in October to more than **12%** in May – an **increase by 38%**. In view of the seriousness of this effect, a p-level of .059 is quite enough to at least sound the warning bell.

[9] The null hypothesis here being, of course, that there is no interaction effect of this kind.

Figure A: Interaction of Time and Style in pre-sonorant voicing.



7.7 Results for pre-obstruents contexts

Turning to pre-obstruent contexts, it would seem on the face of it that apart from the need to transform the pre-sonorant data, the same ANOVA procedure can be used as for pre-sonorants. Actually, there are good reasons speaking against such a solution. If one takes a good look at the data in Table B, it will become evident that if we calculate percentages of voicing for the four possible combinations of treatments for both repeated measures factors (the combinations being: October Reading, October Speaking, May Reading, May Speaking), then quite apart from the three cases of missing data for October Speaking (subjects 9, 11, and 16), we would get two further cases for which the proportions cannot be determined: October Reading for subject 13, and October Speaking for subject 14. The reason the proportions cannot be determined is that all potential contexts for pre-obstruent voicing were resolved through pauses by these two subjects for these particular combinations of treatments, leaving no pause-free contexts. This produces the much-dreaded division-by-zero error in calculating the proportions. Simply, we have no information on the performance of subjects 13 and 14 under these particular combinations of treatments, which is equivalent to missing data. Should we insist on sticking with a three-way ANOVA, we face the difficult choice of case-wise deletion of 5 subjects, leaving only 9 good subjects to work with, few degrees of freedom and a low test power, or we attempt to estimate the missing data from extant data – a controversial procedure which anyhow consumes additional degrees of freedom. Since neither of the two options seems acceptable, the solution that will be adopted here is to run two two-way ANOVA's: one for Time (on the data given in the second last and third last columns of Table B), and another one for Style, by pooling the October and May frequencies for each of the subjects. The respective ANOVA's are given in Table D and Table E below.

Table D: 2-way ANOVA, repeated measures, fixed model, 1,14 df
1 dependent variable, percentages of pre-obstruent voicing
Between: Area (2 levels: Devoicing area, Voicing area)
Within: Time (2 levels: October, May)

Effect	F	p-level
Area	1.723	.210
Time	.072	.793
Area by Time	.055	.818

Table E: 2-way ANOVA, repeated measures, fixed model, 1,14 df
1 dependent variable, percentages of pre-obstruent voicing
Between: Area (2 levels: Devoicing area, Voicing area)
Within: Style (2 levels: Reading, Speaking)

Effect	F	p-level
Area	1.641	.221
Style	.190	.669
Area by Style	.782	.391

A quick inspection of both tables is enough to verify that neither Time, nor Style, nor interactions of any of these with Area – even begin to approach significance. The only effect of any potential interest is the Area factor which may be worthy of further study. In a case like this, one should ask oneself the question of whether the difference would matter practically if found significant (for example, by using a larger sample). To find out, we have to examine the group means. The relevant mean percentages are **55.2%** for the devoicing area group and **45.7%** for the voicing area group, a smallish difference that would have no practical implications in most circumstances. What is striking, is that the D group exhibits *more frequent* pre-obstruent voicing than the V group. This has two important consequences in terms of the experimental design itself.

First, it tentatively points to an absence of experimenter bias: had the decision process discriminating between voicing or lack thereof been biased by the experimenter tending to transcribe <E.V> more frequently for the V subjects[10], then this bias would probably extend to pre-obstruent contexts, as I had generally expected to find more voicing in both contexts in the V subjects. Also, I was listening out for voicing/no voicing/pause before both contexts on the same pass, using a single-press search key, and so was at the time generally unaware of whether a particular context was pre-obstruent or pre-sonorant.

Secondly, finding contradictory tendencies between the two groups relative to the two experimental contexts gives support to the original decision to split the contexts in the manner in which they had been split (i.e. obstruents vs sonorants), and thus also indirectly points to L1 → L2 transfer as the source of the variation in L2 voicing.

8. Implications and conclusions

For the V and D groups considered together, the pre-obstruent environment, though three times less frequent, triggers regressive voicing five times more easily (mean **51.0%**) than the pre-sonorant context (mean **10.7%**). In general then, the frequency of tokens of pre-obstruent voicing in running speech will be higher (by a factor of 2) than will be the frequency of tokens of pre-sonorant voicing. It is pre-obstruent voicing, then, that remains the more troublesome of the two, and remedying this problem alone would result in a greater improvement in pronunciation than would exterminating pre-sonorant voicing alone (assuming for the moment that both types of voicing are pronunciation errors of equal gravity).

However, for subjects from the voicing area the relationship is very different. These subjects voice an average of 45.7% of pause-free pre-obstruent contexts, and 17.6% of pause-free pre-sonorant contexts. Given that the latter are over three times as frequent as the former, tokens of pre-sonorant voicing become at least as frequent as those of pre-obstruent voicing in running speech.

An unexpected and alarming finding of this study is that pre-sonorant voicing probably becomes a more serious problem as students advance through their first year of studies, despite the accompanying intensive phonetic training. Three possible causes come to mind. It might be that students from the devoicing area change their accent through exposure to the local – largely voicing – Poznañ accent[11]. This explanation does not appear to be supported by the data: should this be the case, an Area by Time by Style interaction effect (or Area by Time if

[10] I did not have overt information as to the V - D membership at the time, but just suppose pessimistically that somehow I was able to subconsciously guess this from the pronunciation.
[11] I am indebted to Włodzimierz Sobkowiak for suggesting this possibility, and for his comments on an earlier draft of this paper.

the effect were not limited to Speaking) would show up in the ANOVA Table C; and there is nothing of the sort. In fact, I found that subjects originally from Poznañ considered as a group also display an increase in the frequency of voicing. And yet, through interacting with students from devoicing areas, they were likely to get relatively *less* exposure to voicing accents between October and May than they used to prior to October 1995. A second possible explanation is that the phonetic training the students received somehow contributed to the aggravation of the voicing problem. This could for example come as a result of extensive drills aimed at eradicating excessive devoicing. However, the instructor reported that no such drills were done and obstruent devoicing was attended to opportunistically, as was regressive voicing. Another way phonetic training might have contributed to the problem at hand might have been through exercises aimed at teaching students to link words more smoothly. According to the instructor, some time and attention were given to this particular issue. This possible cause ties up with the third possible explanation for the apparent rise in the frequency of pre-sonorant voicing in speech: a general increase in fluency related to the language-learning process as a whole. In this connection, it should be reiterated that one obvious consequence of an increase in fluency – a decrease in the frequency of pauses – had been factored out of the design by calculating the proportion of voiced realisations in *pause-free* contexts, so the proportion or number of pauses had no consequence whatsoever for the results. In any case, no such decrease in the relative frequency of pauses occurred between October and May, as can be readily verified by referring to Table A and Table B (the proportions of pause realisations are virtually identical, with around one third of all pre-sonorant contexts being resolved through pause). If an increase in fluency is indeed the underlying factor, then the direct cause does not lie in overt phonetic pauses. Perhaps regressive sandhi voicing in Polish is to some extent related to the degree of conscious attention given to one's speech (and there is some indication that it might not be an obligatory process – see e.g. Horwath 1984). If so, then an increase in fluency in English might lead to a lower degree of monitoring, resulting in the Polish voicing process being switched on more often. If this is the underlying cause, then one might expect the trend to continue into further years of study at the School of English. A follow-up study with the same subjects in the higher years might be able to throw some light on this issue and may be undertaken in the future.

Replications with students from other institutions would be necessary before generalisations are made as to the relationship between years of study or amount of phonetic instruction and tendency to voice. In contrast, those results of the present study in which no October versus May difference has been detected, can – with some caution – be extended to students of an equivalent level of competence throughout Poland, as subjects in the study came from a variety of regions (there were four locals though). Those institutions whose enrolment tends to cover candidates from around the country may expect their newly-admitted students to perform similarly to our October results, except insofar as there may be systematic differences in average competence levels between candidates applying to different institutions.

I believe that the present results may have some useful practical consequences as far as the teaching of English pronunciation is concerned. It appears that pre-obstruent regressive voicing is a pervasive problem and a variety of remedial intervention strategies should be systematically tested for efficacy and implemented if found effective. Pre-sonorant voicing may be at least as serious as pre-obstruent voicing for students from the voicing area, and may possibly be getting worse with time. This information should be useful for those institutions, such as some Teacher Training Colleges, whose catchment areas are geographically limited. In such institutions, it might be worth while to custom-cut phonetic course syllabi to accommodate the relative tendency for the two types of voicing in the area. For example, a Teacher Training College in Upper Silesia might place a relatively larger emphasis on eradicating pre-sonorant voicing than would its counterpart in Szczecin.

This seems an appropriate moment to raise the issue of choosing geographical area as a grouping variable. I could have opted to use instead an experimentally derived measure of sandhi voicing in Polish for particular subjects. Perhaps if the aim had been to verify some theoretical models of second language acquisition, this type of design would have given the experimenter greater power, by eliminating part of the variation that follows from an imperfect, no doubt, correlation between where one lives and how one speaks. However, such a design would have had limited practical applications, as it cannot be reasonably expected that candidates for English studies will be pre-tested on their Polish phonetics. And even if they are, what does one do with the results? Average them for each section of students and design a syllabus on that basis? I do not think this is a feasible option. In contrast, if it is known that a given institution mostly admits candidates from the local region, it does not seem at all unreasonable for a school to be able to adopt a syllabus reflecting the general characteristics of a given area, if only in terms of the numbers of hours devoted to particular problems. In view of this, the advantage of the present design lies in its predictive power being based solely on the geographical distribution of candidates, without having to know what Polish accent they actually speak.

At the very least, it is hoped that the present study will help to raise the awareness of pronunciation teachers of the severity of regressive voicing assimilation – be it pre-obstruent or pre-sonorant – in the speech of Polish learners of English. It seems to me that this particular type of pronunciation error has so far been neglected – unlike its perhaps less foreign-sounding phonological *alter ego*, final devoicing.

References:

BOX, G.E.P. (1954), "Some theorems on quadratic forms applied in the study of variance problems, II: Effect of inequality of variance and of correlation between errors in the two-way classification". *Annals of Mathematical Statistics* 25, 484-498.
BRONSTEIN, A.J. (1960), *The Pronunciation of American English*. New York: Appleton-Century-Crofts.
COHEN, J. & COHEN, P. (1975), *Applied Multiple Regression: Correlation Analysis for the Behavioural Sciences*. New York: Wiley & Sons.
DEJNA, K. (1994), *Atlas polskich innowacji dialektalnych*. Warszawa – Łódź: PWN.
GIMSON, A.C. (1989), *An Introduction to the Pronunciation of English*. London: Edward Arnold.
HORWATH, E. (1984), "Udźwięcznienie międzywyrazowe w Krakowie". In: Dunaj B. (ed.), *Studia nad Polszczyzną mówioną Krakowa*, vol. 2, 43-50. Kraków: Uniwersytet Jagielloński.
IOUP, G. & WEINBERGER, S.H. (eds.) (1987), *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge, Ma.: Newbury House.
JAMES, A. & LEATHER, J. (eds.) (1987), *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris.
KIRK, R.E. (1982), *Experimental Design: Procedures for the Behavioral Sciences*. Belmont: Brooks/Cole.
MADEJOWA, M. (1990), "Modern Polish linguistic norm with special reference to the pronunciation of consonants". *Studia Phonetica Posnaniensia* 2, 69-105.
MADEJOWA, M. (1993), "Normative rules of modern Polish pronunciation". *Studia Phonetica Posnaniensia* 4, 19-30.
MADELSKA, L. (1987), "Mowa spontaniczna. Analiza wariantywności fonetycznej w mowie studentów Uniwersytetu im. Adama Mickiewicza". [Unpublished PhD dissertation, Adam Mickiewicz University, Poznañ]
NAGÓRKO, A. (1996), *Zarys gramatyki polskiej*. Warszawa: PWN.
RIETVELD, T. & VAN HOUT, R. (1993), *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin - New York: Mouton de Gruyter.
ROACH, P. (1991), *English Phonetics and Phonology: A Practical Course*. Cambridge: Cambridge University Press.
SCHEUER, S. (1996), "The application of a spoken corpus of learner English to teaching phonetics to Polish students". In: Waniek-Klimczak E., *Teaching English Phonetics And Phonology in Poland*, 27-29. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
SCHEUER, S. (in press), "The application of a spoken corpus of learner English to teaching phonetics to Polish students". In: Lewandowska-Tomaszczyk B. & Melia P.J., *Proceedings from the International Conference on Practical Applications in Language Corpora, £ódŸ, 10-14 April 1997*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
STRUTYŃSKI, J. (1996), *Gramatyka polska. Wprowadzenie. Fonetyka. Fonologia.* Kraków: Księgarnia Akademicka.
WELKOWITZ, J., EWEN, R.B. & COHEN, J. (1971), *Introductory Statistics for the Behavioural Sciences*. New York - London: Academic Press.
WILCOX, R.R. (1987), "New designs in analysis of variance", *Annual Review of Psychology* 38, 29-60.
WINER, B.J. (1962), *Statistical Principles in Experimental Design*. New York: McGraw-Hill.
WOODS, A., FLETCHER, P. & HUGHES, A. (1986), *Statistics in Language Studies*. Cambridge: CUP.