

## Aspects of identity in a second language: ING variation in the speech of Polish migrants living in Manchester, UK

ROB DRUMMOND

*Manchester Metropolitan University*

### ABSTRACT

This study uses variationist methods in an L2 context to investigate ING variation in the English speech of UK-based Polish migrants. The results suggest that such variation is as consistent in an L2 context as it is in an L1 context, with several of the expected L1 factors emerging as statistically significant constraints. However, two social factors emerge as particularly worthy of further discussion: the reversal of the typical gender pattern (in this case women are more likely to use the alveolar variant than men), and the question of whether the speaker intends to return to Poland in the future. Gender is discussed from the perspective of contexts of language use (occupation), and future plans is discussed in terms of its role in the construction of identity. There is a strong indication that the intention to return to Poland encourages the use of an ING variant, [ɪŋk], that signals this allegiance to the L1 identity.

The recent wave of increased migration from Eastern Europe to the UK offers an ideal context within which to investigate questions of sociolinguistic variation in a second language. The idea that L2 speech is an appropriate arena for the application of variationist methods of analysis is one that has gained acceptance over the last fifteen to twenty years. Time and again, learner speech (or interlanguage) has been shown to be highly systematic and just as open to influence from linguistic and social factors as any other natural language variety (see Tarone, 2007; Young, 1999, for a detailed overview of the various studies). Research in this area of interlanguage variation can be split into two types: that which investigates elements of linguistic competence – the acquisition of “obligatory” native speaker (NS) target forms, and that which investigates sociolinguistic competence – the acquisition of NS patterns of variability. Linguistic competence has been referred to as “the vertical continuum” (Adamson & Regan, 1991; Corder, 1981; Young, 1988) or ‘Type 1 variation’ (Mougeon et al., 2004), and sociolinguistic competence as ‘the horizontal continuum’ or “Type 2 variation.” However, the two types are clearly not entirely separate, as movement along the horizontal continuum is not possible without a certain degree of movement along the vertical continuum first.

I would like to thank Maciej Baranowski for his valuable advice on earlier versions of this text. I would also like to thank the three anonymous reviewers who each offered very useful comments.

This paper investigates the acquisition of ING variation in the speech of Polish adult migrants living in Manchester, UK. As such, it is primarily a study into the acquisition of Type 2 variation, as ING is a well-known stable linguistic variable in NS English, but one that does not exist in a similar form in Polish. In the course of the study, the following research questions are addressed:

1. To what extent are the patterns of ING variation being acquired by non-native speakers (NNSs) similar to those patterns exhibited by NSs?
2. What social factors influence the variation?

#### THE POLISH COMMUNITY IN MANCHESTER<sup>1</sup>

Manchester is a traditionally industrial city in the North West of England with a population of 483,800 (Office for National Statistics, 2010). While the city has been home to a (now very established) Polish community since the Second World War, the recent expansion of the EU in 2004 brought with it a significant wave of migration from Eastern European countries, with Poles outnumbering the other nationalities arriving in the UK. In contrast to the more traditional route of staying in and around London, these migrants moved to all parts of the UK, with the North West of England proving to be a popular destination.

Various migratory patterns can be identified in the movement of Polish migrants. Some move regularly between Poland and the UK, some come just once to make some money before returning to Poland, some younger individuals follow opportunities wherever they are in the UK or elsewhere, and others come with the intention of staying in the UK permanently (Eade et al., 2006:10–12). These different patterns help to maintain a cycle of migration, such that each group type either relies on, or provides opportunities for, another.

The nature of the relationship between Polish migrants and the local (non-Polish) community is difficult to determine. On the one hand, one hears of a certain degree of resentment toward the Polish migrants from a section of the UK population who feels that the newcomers are taking their jobs and benefits (Garner et al., 2009). On the other hand, however, in the conversations carried out with Polish migrants in the course of the current study there were very few negative stories on the subject of interaction with the local community. That is not to say that all the participants showed a similar degree of integration within the community. There were in fact a wide variety of individual situations, ranging from people whose lives appeared indistinguishable (in terms of social networks) from those of their native friends, to those who barely had any voluntary contact with NS members of the local community.

#### ING VARIATION

‘A staple of sociolinguistics’ (Hazen, 2006:581), the variable ING has been studied in a wide variety of contexts since the 1950s. As a sociolinguistic variable, the

focus has generally been on its variable realization as [ɪn] and [ɪŋ] in unstressed syllables in multisyllabic words.

Labov (2001:86) claimed ING to be ‘the first sociolinguistic variable to be studied quantitatively, [having] the widest range and most uniform pattern of all variables in English’. Central to this uniformity is the constraint of grammatical category, which has been shown to remain consistent across studies. The underlying nature of this constraint is described as some kind of nominal-verbal continuum (e.g. Abramowicz, 2007; Adamson & Regan, 1991; Houston, 1985; Labov, 2001) with the more verbal structures showing a greater occurrence of [ɪn], and the more nominal structures favoring [ɪŋ]. Labov (2001) made the point that it is difficult to determine the level of detail along the continuum, due to the large number of possible syntactic categories, some with very low frequency. This is compounded by the difficulty in first determining the boundaries of these categories. While some categories sit neatly at the two ends of the continuum, for example, progressive verbs such as *he is running* and simple nouns such as *ceiling*, others, such as the status of so-called gerunds, are more problematic. This has led to a variety of solutions and categorizations, with different studies opting for more or less detailed categories.<sup>2</sup> However, few would argue that their own system of categorization is perfect, instead perhaps accepting that the precise details are not the most important factor, rather it is the observation that ‘there are two distinct groups: a verbal and a nominal use of /ɪŋ/, which cluster at radically different levels’ (Labov, 2001:88).

While grammatical category is seen as one of the most consistent constraints at work on the ING variable, other constraints are equally well-researched, often with consistent results. From a linguistic point of view, three important constraints are those of priming (the idea that the realization of one ING can then affect the realization of a subsequent ING, favoring similarity), regressive homorganic assimilation (a following velar encourages the use of [ɪŋ] and a following alveolar encourages the use of [ɪn]) and progressive homorganic dissimilation (a preceding velar discourages the use of [ɪŋ], and a preceding alveolar discourages the use of [ɪn]) (Houston, 1985). These tendencies, however, are not as consistent as the grammatical category constraint appears to be, with Labov (2001:87) finding ‘no strong phonological conditioning before following velars or apicals’.

There are also social constraints at work in the realization of ING that also appear to be consistent across studies, such as socioeconomic class, style, and gender (see Labov [2001] or Hazen [2006] for an overview). Generally speaking, on the basis of what has previously been found, one would expect a higher rate of [ɪn] lower down the socioeconomic scale, in more informal speech, and in the speech of men. Although social class is not a consideration in the present study due to the complex nature of class in a migrant community, gender is central to this research. Style will play a peripheral role.

### *ING variation in the UK*

Those studies which look specifically at ING variation in British English tend to reflect the findings of other research. In terms of linguistic factors, the

nominal-verbal ordering remains constant, with verbs favoring [ɪn] and nouns favoring [ɪŋ] (Houston, 1985; Tagliamonte, 2004; Watts, 2005). In terms of social factors, gender, social stratification, and style have all been shown to follow the now familiar patterns (e.g., Mathisen, 1999; Trudgill, 1974). However, there is an additional element to ING variation in certain areas of the UK, including Manchester, often referred to as ‘velar nasal plus’ (Wells, 1982). This describes the variant [ɪŋg] whereby the words *finger* and *singer* use the same velar stop, which is then also used word-finally in words such as *during* and *watching*. This is a common variant in this and other (mainly northern) areas of the country and has been identified in some areas as a local prestige form (Mathisen, 1999).<sup>3</sup>

One particularly relevant study is by Schlee, Meyerhoff, and Clark (2011). The study compared ING variance in the speech of local and Polish-born adolescents in Edinburgh and London and found that it was the Polish-born groups in both cities who replicated the classic nominal-verbal pattern rather than the locally born groups. Schlee et al. explain this with the suggestion that the speech of the Polish-born adolescents is being influenced by supra local norms or constraints rather than by the local constraints exhibited by their locally born peers. Other differences between the local and Polish groups in terms of the significance, strength, and ordering of constraints are explained by the idea of imperfect learning. Polish-born adolescents, by definition, will not have been exposed to the same depth and variety of sociolinguistic information as their locally-born peers and thus will not have had the opportunity to refine their own production. They argue that the complexity of the task of replicating these established patterns renders the ING variable a very different type for L2 speakers. What is seen very much as a stable variable for NSs, might in fact not be so stable for NNSs who are, after all, language learners. The imperfect learning therefore shows itself in the ‘re-ordering or non-replication of variable constraints’ (Schlee et al., 2011: 226).

## IDENTITY

There is a clear connection between issues of identity and the context of L2 speech, particularly in relation to pronunciation accuracy. In the case of advanced speakers, the issue of ‘passing’ becomes relevant; that is, the extent to which an individual is able to pass as a NS, or, more importantly, the extent to which an individual *wants* to pass as a NS (cf. Piller, 2002). Marx (2002:273) described a phase of her own personal experience of living in an L2 context (Germany) and then returning to the L1 context (Canada), the ‘construction of an L2 identity and attrition of the L1’, during which she appropriated the L2 accent and ‘deemed it a great success when [she] could ‘fool’ someone into believing [she] was indeed German’.

In contrast, there is research to show that individuals might consciously avoid acquiring native like pronunciation so as to reinforce their L1 identity.

Gatbonton et al., (2005) studied the relationship between ethnic group affiliation and L2 pronunciation accuracy, drawing on data from two separate studies. The general findings were that 'the more learners sound like the speakers of their target language, the less they are perceived by their peers to be loyal to their own group' (Gatbonton et al., 2005:504). This was found to be true both in a situation where the L1 and the L2 were in conflict (French and English in 1970s Quebec) and in a situation where there was no conflict (Chinese and English in Montréal). It should be borne in mind, however, that the status of English in both Montréal and Quebec is different from the status of English in Manchester.

Lybeck (2002) used elements of Schumann's acculturation model (Schumann, 1978) along with elements of social network theory (Milroy, 1987; Milroy & Milroy, 1992) in her study of the L2 pronunciation accuracy of Americans living in Norway. Those speakers with the lowest level of cultural distance (developed through 'supportive engagement in exchange networks' [p. 179]) were the ones who had the highest level of pronunciation accuracy. They were also the ones who felt they had accepted a new identity. This contrasts with the group with the highest level of cultural distance, who had the lowest level of pronunciation accuracy, and who felt that to lose one's foreign accent was to risk losing one's American identity (p. 181).

The relationship between a person's L1 and their identity is a fundamental one that has always been recognized (see Tabouret-Keller, 1997, for a historical overview), yet it is a relationship that is exceedingly complex. Much of this complexity stems from the multivalent nature of identity itself, in contrast to the essentialist nature of much of the (early) research involving identity in both sociolinguistics and anthropology (Bucholtz & Hall, 2004:374; Mendoza-Denton, 2002). Mendoza-Denton (2002) describes three broad types of studies which, though not entirely separate, exist on a continuum from analysts' categories to participants' categories. The three types represent studies based on: (1) sociodemographic category-based identity, for example, Labov's work in New York City; (2) practice-based identity, e.g. Eckert and McConnell-Ginet's (1992) interest in how identities are constructed by individuals' participation in various communities of practice; (3) practice-based variation, in which identity is seen as shifting during interaction e.g. Johnstone and Bean (1999). Similarly, Eckert (2010) identified three waves of variationist studies, the third of which puts stylistic practice in the center of the process of constructing and negotiating identity, rather than seeing identity as being reflected by the variables people use.

The present study adopts a social network rather than practice-based approach, albeit with a certain degree of assumption as to the precise makeup of the participants' networks. The problem is, the participants in the study, while arguably members of the same (large) community of practice with regard to being Polish migrants in Manchester, do not use English in that community. Instead, we are observing the result of their use of English in their own, smaller communities (or networks) that make up their contact with the local community (to varying degrees). It is the perceived quality as well as the quantity of linguistic interaction with the local variety that is being used as a possible

explanatory factor in the degree of acquisition of local forms. The quantity/quality distinction is one that Holmes and Meyerhoff (1999) identified as one that distinguishes a social network and a community of practice, while acknowledging that there is a substantial degree of overlap between the two. Although future work into the Polish community in Manchester would benefit from taking a more robust practice-based approach using ethnographic observation, this preliminary study is best positioned more tentatively in the overlap just mentioned.

The acquisition of local speech features could be viewed, in a limited way, as indicative of a growing sense of local identity, particularly when the feature in question does not show anything like the same variable nature in the L1. As the local variants themselves are fairly salient, there is the possibility that their acquisition represents a conscious construction of a local (L2) identity. By the same token, lack of acquisition may signal resistance to the local culture and a determination to maintain one's L1 identity. Even within the methodological limitations just described, the findings presented here lend weight to that possibility.

## METHODOLOGY

### *Participants*

The participants for the study consisted of Polish adults who had grown up in Poland, but who were now living in Manchester. As individuals, they all fulfilled the following criteria:

1. They grew up in Poland and came to England as adults.
2. They were aged between 18 and 40.
3. They had at least a basic proficiency in English before coming to England.<sup>4</sup>
4. The vast majority (37 out of 40) had lived nowhere else in the UK apart from in the Manchester area.

The final sample consisted of 40 individuals (see Table 1).

### *Gathering data*

Meetings were arranged with individuals throughout 2009 (Drummond, 2010). Although there were other elements to the meetings (a picture description task and a word list), all speech data presented here come from an informal conversation with each participant.<sup>5</sup> The term 'conversation' is used intentionally here, as the idea was to replicate an informal chat. Every effort was made to elicit as much speech as possible from the participant, resulting in the conversations being desirably one-sided, but they remained conversations rather than interviews. The reason for this approach was an awareness that the participants were not using their first language, which for many would be a challenging task. It was therefore important to ensure that the meeting in no way resembled any

TABLE 1. *Participants*

Length of Residence (LOR) (years)	Male	Female	Total
0–2	5	6	11
2–4	9	9	18
4–6	6	5	11
	20	20	40

kind of language test, where an interlocutor would ask a series of questions and offer little in return.

The purpose of the conversation was to elicit speech that was as natural as possible by accessing information, explanations, and most importantly stories, that might usually be shared between friends. Certain core topics such as the participant's life in Poland, life in Manchester, problems faced when living in a different country and future plans were covered with each participant through leading questions. Other topics developed naturally depending on the individual. The length of the conversations varied with each speaker, with the shortest being 18 minutes and the longest 1 hour and 10 minutes (average 34 minutes). The most important factor determining length was level of English [LoE], with some speakers with lower LoE finding it understandably challenging to maintain a conversation in a second language for an extended period. While these conversations could be considered short in the context of wider sociolinguistic research, the evidence suggests they are long enough to illustrate possible patterns of acquisition.

The recorded conversation was also used to assess the participants' level of spoken English. This was an impressionistic score made by the researcher and a colleague (both experienced English teachers) on overall fluency, accuracy, and use of vocabulary. A numerical scale from 1 (low) to 10 (high) was used.

### *Coding ING*

As described earlier, the interest of the ING variable usually lies in the alternation between [ɪŋ] and [ɪn] in unstressed syllables. However, the present study includes two other variants, [ɪŋg] and [ɪŋk]. The first of these was included on the basis that it is a common variant among the NSs of the local area; it was initially felt that the use of [ɪŋg] by the Polish speakers might possibly indicate an acquisition of a local form. However, in a pilot study it soon became clear that a fourth variant, [ɪŋk], was common amongst the Polish speakers.<sup>6</sup> This is perhaps not surprising, given that in Polish, the velar nasal only occurs before a velar plosive (Gussman, 2007). The nature of this velar plosive (voiced or voiceless) generally depends on the following sound. In coda position, obstruent voicing is not always contrastive in Polish; the stop will assimilate to what follows. This would suggest that when a stop is present, [ɪŋk] is to be expected before a voiceless obstruent or a pause, and [ɪŋg] is to be expected before a voiced obstruent.<sup>7</sup> Before nasals, approximants, and vowels, the situation is more complex as it is

dialect-dependent to an extent.<sup>8</sup> This made it impossible to determine if any realizations of [ɪŋ] were as a result of a move away from standard [ɪŋ] towards the local variant or as a result of L1 interference. Nevertheless, the four variants were coded separately with a view towards looking briefly at the [ɪŋk] / [ɪŋg] alternation in addition to the main focus of the use of [ɪŋ]. All examples were categorized according to five linguistic features:

1. Variant – [ɪŋ], [ɪn], [ɪŋg], [ɪŋk]
2. Preceding consonant – alveolar, velar, other
3. Following segment – alveolar, velar, other, pause
4. Grammatical category (see Table 2)
5. Previous variant [ɪŋ], [ɪn], [ɪŋg], [ɪŋk].

The difficulty in categorising certain ING forms was mentioned earlier, particularly with regard to the gerund, which has been identified in different ways by different people for many years (if it has been identified as a separate form at all). When it is identified as a separate category, it is usually described as simultaneously exhibiting properties of a verb and properties of a noun. The approach taken in the present study is to use two categories for gerund: ‘gerund (nominal)’ and ‘gerund (verbal)’ to indicate this separation. It should also be noted that unlike much previous research, all *-thing* words (*something, nothing, anything, everything*) have been included. This decision was based on the understanding that it is perhaps unwise to automatically assume similarities with previous findings, especially when the data come from such a different group of speakers (NNSs rather than NSs). However, these words were categorized separately as pronouns, so as to separate them from simple nouns.

Initially, 30 tokens were identified for each speaker where possible and were categorized into the four variant types auditorily. This was felt to be a satisfactory process, despite a certain degree of subjectivity on a few occasions in distinguishing between [ɪŋg] and [ɪŋk]. If those 30 tokens showed no variation from [ɪŋ], then no further tokens were sought. However, if there was evidence of any variation from [ɪŋ], then a further 20 tokens were identified where possible.

TABLE 2. *Grammatical categories for ING*

Category	Example
Pronoun	<i>I know everything about it.</i>
Progressive verb	<i>He is watching TV.</i>
Present participle	<i>Considering he is not English ...</i>
Gerund (verbal)	<i>He's started working there.</i>
Gerund (nominal)	<i>A big swimming pool.</i>
Noun	<i>I liked the beginning ...</i>
Adjective	<i>It is more exciting than ...</i>
Preposition	<i>I went during the holidays.</i>
Discourse marker	<i>... for three years or something.</i>



In the case of a few speakers, the number of tokens available fell below the numbers just mentioned. In total, 1677 tokens of ING were analyzed, an average of just under 42 tokens per participant.

### *Coding social factors*

In addition to the speech data, social and attitudinal data were gathered by means of a questionnaire which was completed after the conversation, but within the meeting. It was decided to have the entire questionnaire translated into Polish so as to avoid both possible misunderstanding and fatigue on the part of the speakers. The first section of the questionnaire targeted information such as self-assessed English proficiency, amount of English instruction, amount of use of L1/L2, and future plans. The second section focussed on attitudinal factors and used multi-item seven-point Likert scales to investigate various aspects of individuals' attitudes towards living in Manchester and their spoken English. The internal consistency of the questions was measured using Cronbach's alpha (a measurement of correlation between items intending to measure the same aspect in a questionnaire of this kind), and the existence of correlations amongst the factors was checked by calculating a Pearson correlation coefficient for each combination, with any problematic factors being discarded.<sup>9</sup> As a result, the following aspects were retained:

- Attitude towards Manchester, its people, and living there (ATT)
- Awareness of a Manchester accent (AW)
- Desire to lose one's Polish accent and sound like NS (not specifically Manchester English) (CHA)
- Motivation to improve pronunciation. (MOT)

### *Lexical frequency*

To test for the effects of lexical frequency on ING, it was first necessary to identify a suitable account of lexical frequency in spoken English. Leech, Rayson, and Wilson (2001) provided word frequency lists for the spoken section of the British National Corpus for all words with a frequency of 10 or more per million words.<sup>10</sup> Of the words making up the 1677 ING tokens in the present Polish data, those which did not appear on the spoken frequency lists (i.e., those with a frequency of less than 10 in 1 million words) were excluded from the analysis. The resulting list was then checked against individual speaker, and any word that was not used by three or more individual speakers was also excluded. The BNC frequency lists use only five grammatical categories: NOUN, VERB, ADJECTIVE, PREPOSITION, PRONOUN, necessitating the conflation of some of the Polish data categories for comparison purposes (noun and gerund [nominal] became NOUN and, progressive, participle, gerund [verbal] became VERB).

The result was a list of 75 words, each with a corresponding BNC frequency value (frequency per million words), its frequency within the Polish dataset (total count), and the number of individual speakers who used that particular word. In addition to this, each word's proportion of [ɪn] was calculated. For

example, the verb *coming* had a BNC frequency value of 522 (per million words), a Polish dataset frequency of 35, was used by 22 different speakers and had an [ɪn] proportion of .11.

The frequency data were normalized using the log10 transformation<sup>11</sup> and a Pearson correlation coefficient was calculated to assess the relationship between the BNC frequency values and the Polish dataset values; there was found to be a modest/strong correlation between the two ( $r = .674, p < .01$ ).

## RESULTS

Figure 1 shows the overall proportions for each variant in the conversation element of the interview for all 40 speakers (for individual results see Table A1 in the appendix).

The standard (in terms of a pedagogical model) variant of [ɪn] was by far the most common in the group as a whole, accounting for 70.3% of the total number of ING tokens. Seven of the 40 speakers showed no variation from this standard form, leaving a majority showing some degree of variation. Nine of these speakers exhibited the use of all four variants. In terms of traditional ING research, arguably the most important variant is the alveolar [ɪn], and this will be the focus of the first part of the analysis here. This variant accounted for only 6% of the total, yet was found in the speech of 16 of the speakers. Figure 2 shows the proportion of each variant for each speaker.

### *Regression analysis*

Multiple logistic regression analyses were carried out using Rbrul (Johnson, 2008), including individual speaker as a random effect. Rbrul is a variable rule program in the mold of Goldvarb (Sankoff et al., 2005), yet which incorporates mixed-effects modelling. The result is a model which “can still capture external effects, but only when they are strong enough to rise above the inter-speaker variation” (Johnson, 2009:365). Rbrul expresses coefficients in log-odds rather than factor weights, although both are given in the analysis presented here.

Due to the fact that 24 of the speakers did not produce any tokens of the variant that is of particular interest ([ɪn]), it was decided to carry out two regression analyses in the first instance. The initial analysis included all the speakers and aimed to explore patterns behind which speakers are more likely to produce [ɪn] and under which linguistic conditions, while the second analysis included only the subset of speakers who exhibited [ɪn], aiming to explore in more detail the variables which encourage or inhibit its use. In both cases, the dependent variable was the ING variable, with the application value as [ɪn] and the non-application values as the other three possible variants. However, in these and any subsequent analyses it must be remembered that the overall rate of [ɪn] was very low. Two changes were made to the data for the regression analysis, and that was the exclusion of one grammatical category (‘preposition’) and the recoding of another (‘gerund [nominal]’ was recoded to be part of ‘noun’). This

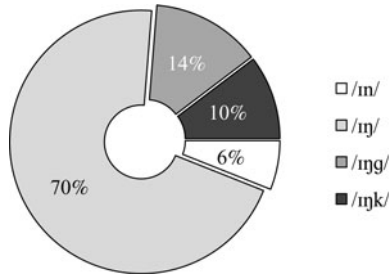


FIGURE 1. Total proportion of each variant of ING, all 40 speakers.

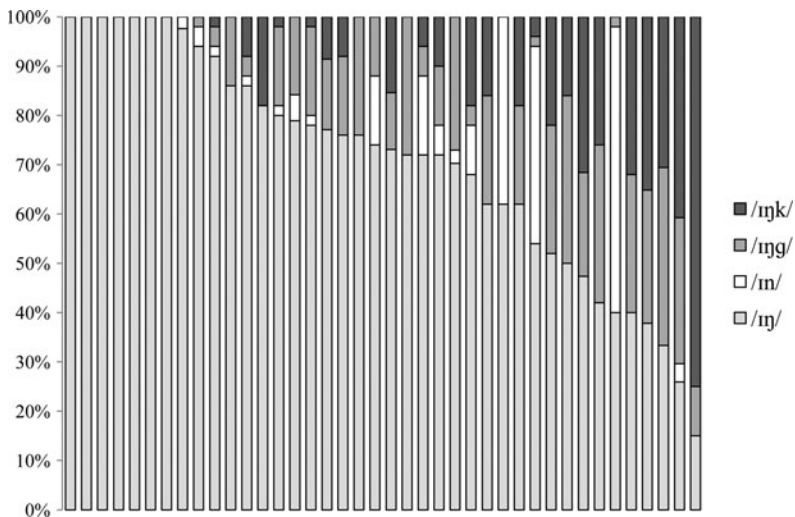


FIGURE 2. Chart showing the proportions for each ING variant, all speakers, ordered by proportion of standard /ɪŋ/.

was done because neither category showed any examples of [ɪn], thus creating so-called knockout categories, a situation which makes any results unreliable at best.<sup>12</sup> An alternative solution was considered which involved simply excluding the gerund (nominal) category rather than conflating it with the noun category; however, the deviance measures showed that this solution did not provide a better fitting model. The results of the analysis can be seen in Table 3.

The four statistically significant factors can be divided into two types: linguistic and social. The linguistic constraints grammatical category and preceding consonant are both highly significant; furthermore, they both largely reflect the patterns found in previous research. Recall that Houston (1985) found a pattern of progressive dissimilation whereby a preceding velar consonant disfavored the

TABLE 3. Regression analysis (*Rbrul*) of the effect of linguistic and social factors on (*ING*), all speakers

Application value: [m]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Preceding consonant $p < .01$	velar	1.285	268	.164	.783
	other	-.383	939	.045	.405
	alveolar	-.902	448	.033	.289
Grammatical category $p < .01$	progressive	1.015	634	.112	.734
	participle	.980	220	.068	.727
	gerund (verbal)	-.040	177	.023	.49
	adjective	-.468	63	.016	.385
	noun	-.688	181	.011	.334
	pronoun	-.798	380	.021	.31
Gender $p < .05$	female	1.127	867	.099	.755
	male	-1.127	788	.019	.245
Future plans $p < .05$	stay in UK	1.286	532	.081	.783
	no plans	1.205	728	.078	.769
	return to Poland	-2.491	395	.003	.076
Not significant:	LOR, LoE, ATT, AW, CHA, MOT, NS partner, Use of L1/L2, Following sound, Previous variant, Formal English instruction, No. of syllables				
Model	deviance	df 12	intercept	-6.274	mean .002
	422.298				
	Speaker ID random standard deviation: 2.177				

use of [m] and a preceding alveolar disfavored the use of [m]. This is clearly the case in the current data, in which a preceding velar strongly favors the application value of the dependent variable ([m]) and a preceding alveolar strongly favors one of the three velar variants. It should be noted, however, that the effect of the following sound (and therefore the process of regressive assimilation) is not statistically significant. The ordering of the grammatical category constraints clearly follows the established nominal-verbal continuum, with progressive verbs and participles quite strongly favoring the alveolar variant; simple nouns, pronouns, and adjectives quite strongly disfavoring the alveolar variant; and verbal gerunds in the middle (Figure 3).

With regard to the social factors, the gender difference is quite striking. Of the 16 speakers who exhibited [m], 11 were female and 5 were male. Furthermore, the mean proportion of [m] produced by those speakers was .16 for females and .07 for males. The statistical significance of future plans is also worthy of further comment. It is perhaps best interpreted as the intention to return to Poland acting as a strong inhibitor of the use of the alveolar variant as it is not clear that an intention to stay in the UK (as opposed to having no clear plans) is enough to encourage its use. These four factors will be explored further in the discussion section.

A second regression analysis was carried out including only those speakers who showed some use of [m]. The results can be seen in Table 4.

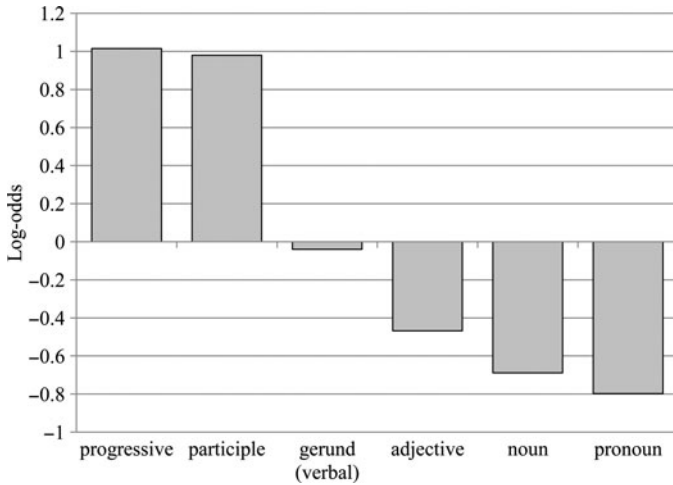


FIGURE 3. Chart showing log-odds for the ING variant. Application value [m].

TABLE 4. Regression analysis (Rbrul) of the effect of linguistic and social factors on (ING), 16 speakers who produced [m]

Application value: [m]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Preceding consonant $p < .01$	velar	1.278	115	.383	.782
	other	-.405	387	.109	.4
	alveolar	-.873	217	.069	.295
Grammatical category $p < .01$	participle	1.062	81	.185	.743
	progressive	1.028	291	.244	.736
	gerund (verbal)	-.194	65	.062	.452
	adjective	-.484	38	.026	.381
	pronoun	-.665	157	.051	.34
	noun	-.746	87	.023	.322
LOR $p < .05$	continuous scale 1-72 months	+1	.04	719	
Level of English $p < .05$	continuous scale 1-10	+1	.826	719	
ATT $p < .01$	continuous scale 1-7	+1	-2.048	719	
Not significant:	Gender, Future plans, AW, CHA, MOT, NS Partner, Use of L1/L2, Following sound, Previous variant, Formal English instruction, No. of syllables				
Model	deviance	df 12	intercept	-.139	mean
		375.364			.14
	Speaker ID random standard deviation: .588				

The first thing to note from these results is the continued statistical significance of the two linguistic constraints. The first, preceding consonant, is almost identical in its strength and pattern as in the initial analysis, and the second, grammatical category, is fundamentally the same despite a slight reordering. That they both

appear largely unchanged in both analyses strengthens the explanatory power of these constraints. The two social constraints from the first analysis are no longer statistically significant in this smaller dataset, and three different ones have taken their place. LOR has emerged as significant, with a greater LOR encouraging a higher rate of the alveolar variant. Although it is not particularly surprising that it should appear, if we are to consider it as a real explanatory factor, its absence in the initial analysis is perhaps a little unexpected. Level of English is working in the expected direction, with higher proficiency equating to an increased likelihood of [m]. However, the final significant social constraint, attitude towards Manchester, is unexpected and not immediately easy to interpret. The difficulty lies in the fact that there appears to be a negative correlation between attitude and the use of [m], which does not make intuitive sense. It should be borne in mind that the ING variation being considered is by no means a feature that is specific to Manchester, so the inclusion in the regression analysis of 'attitude towards Manchester' is perhaps not justified in the first place.<sup>13</sup> There is a possibility that individuals' response to the attitude questions could be interpreted as measuring a more general attitude towards living in the UK, but the specificity of the questions does highlight the local rather than the general.

### *Lexical frequency*

In order to explore a potential relationship between lexical frequency and use of [m], Pearson correlation coefficients were calculated for both frequency measures (BNC and Polish dataset) and the proportion of [m] for each word.

Measures	<i>r</i> value	<i>p</i> value	<i>N</i>
BNC freq and proportion of [m]	.279	.016	75
Polish dataset freq proportion of [m]	.115	.327	75

The results suggest a statistically significant weak correlation between the BNC frequency and the use of [m], but no statistically significant relationship between the Polish frequency measure and the use of [m].

The BNC frequency data were then added to the regression analysis as an independent variable. Note that due to the exclusions described above, the number of tokens was reduced from 1655 to 1029. In addition, the 'grammatical category' variable was recoded to reflect those used by the BNC frequency lists. However, of the five categories, two were excluded on the basis of there being no examples of [m] in either. The first was 'preposition', the second was 'adjective'. In addition, 'return to Poland' was excluded from the future plans variable for the same reason. The results of the regression analysis can be seen in Table 5.

Despite the (slight) correlation when isolated, the effect of lexical frequency did not reach statistical significance when assessed along with the other variables.

TABLE 5. Regression analysis (Rbrul) of the effect of linguistic and social factors on (ING), all speakers, with the addition of lexical frequency

Application value: [m]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Preceding consonant $p < .01$	velar	1.248	193	.223	.777
	other	-.343	629	.064	.415
	alveolar	-.906	207	.058	.288
Grammatical category $p < .01$	verb	1.318	648	.131	.789
	noun	-.034	75	.027	.491
	pronoun	-1.284	306	.026	.217
Syllables $p < .05$	3	.707	175	.051	.67
	2	-.707	854	.101	.33
Gender $p < .05$	female	1.118	550	.145	.754
	male	-1.118	479	.031	.246
Not significant:	Lexical frequency, LOR, LoE, ATT, AW, CHA, MOT, NS Partner, Use of L1/L2, Following sound, Previous variant, Formal English instruction, Future plans				
Model	deviance	df 8	intercept -4.72	mean .009	
	371.299				
	Speaker ID random standard deviation: 2.259				

When the steps of the analyses were consulted it was noted that frequency was working in the expected direction, with +1 adding .473 to the log-odds coefficient, but the  $p$  value of this addition to the model was .152. The continued statistical significance of both the preceding consonant and grammatical category, albeit in a simplified form, is additional confirmation of the strength of these two constraints, which have remained the same in each analysis. The number of syllables in the ING words was statistically significant in this model, with words of 3 syllables favoring the alveolar variant, although this category only accounts for 6 of the 75 words (*anything, beginning, everything, happening, studying, traveling*). Gender retains statistical significance, with females more likely to use the alveolar variant.

### Velar variants

Having investigated patterns behind the use of [m], another set of analyses was carried out in order to explore any patterns behind the distribution of the three velar variants, [ŋ], [ŋg] and [ŋk]. This involved excluding [m] tokens completely from the analysis so as to focus on the three velar variants. In addition to the exclusion of [m], grammatical category was recoded in a similar way as the first analysis, with gerund (nominal) becoming part of noun. Initially, [ŋ] was chosen as the application value, with the other two variants together as the non-application value. There was no need to exclude any other factors on the basis of knockouts. The results of the regression can be seen in Table 6.

Once again, two linguistic constraints and two social constraints reached statistical significance, although only one, grammatical category, remained from

TABLE 6. *Regression analysis (Rbrul) of the effect of linguistic and social factors on (ING), focusing on [ɪŋ], [ɪŋg], and [ɪŋk]. All speakers*

Application value: [ɪŋ]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Following sound $p < .01$	velar	1.903	25	.960	.87
	alveolar	.251	310	.861	.563
	other	-.570	1034	.752	.361
	pause	-1.584	207	.527	.17
Grammatical category $p < .05$	progressive	.289	563	.766	.572
	participle	.164	205	.785	.541
	adjective	.130	63	.774	.532
	pronoun	.111	372	.755	.528
	preposition	.048	22	.773	.512
	gerund (verbal)	-.175	173	.694	.456
	noun	-.566	179	.670	.362
LOR $p < .05$	continuous scale 0-72 months	+1	.028	1576	
Level of Eng $p < .05$	Continuous scale (1-10)	+1	.417	1576	
Not significant:	Gender, Preceding consonant, ATT, AW, CHA, MOT, NS Partner, Use of L1/L2, Formal English instruction, Future plans, No. of syllables.				
Model	deviance	df 13	intercept	-1.696	mean
	1397.907				.747
	Speaker ID random standard deviation: 1.511				

the previous analyses. It is perhaps unwise to speculate too much on the basis of the ordering of grammatical category, as the differences between the categories is so small, and the overall effect is quite weak; however, there is a suggestion of quite an intriguing pattern. Although there is a mixture in the middle, the two extremes suggest the continued existence of a nominal-verbal continuum, with verbs favoring [ɪŋ], and nouns favoring [ɪŋg] or [ɪŋk]. Recall that there was a strong version of the continuum at work in the initial ING analysis, illustrating that verbal forms favor [ɪŋ] and disfavor the three velar variants. Yet here, in the absence of [ɪŋ] variants, the verbal forms favor one of these velar variants over two others.

The fact that a following velar strongly favors [ɪŋ] is only to be expected. Although there were a few examples of the velar variant being released before the following velar, thus allowing for one of the other variants to be distinguished, the vast majority simply assimilated and were heard as [ɪŋ]. The finding that a following pause strongly disfavors [ɪŋ] might be as a result of the suggestion made earlier that a following pause encourages the use of [ɪŋk] due to the influence of the L1. This will be explored in more detail in the next analysis.

The two social constraints are of great interest. Both suggest a move towards a standard variant from a variant influenced by the L1. Interestingly, LoE and LOR are themselves not correlated ( $r = .021$   $p = .899^{14}$ ), so each represents a different process. The effect of LoE is independent of location, so it exists whether a speaker has spent time in the UK or not. However, looking at the relationship



TABLE 7. Regression analysis (Rbrul) of the effect of linguistic and social factors on (ING), focusing on [ɲg] and [ɲk]. All speakers

Application value: [ɲk]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Following voice $p < .01$	pause	1.909	98	.765	.871
	nasal	.862	18	.611	.703
	voiceless obstruent	-.399	69	.406	.401
	voiced obstruent	-.693	18	.278	.333
	approximant	-.698	42	.262	.332
	vowel	-.981	153	.281	.273
	return to poland	1.152	137	.555	.76
Future plans $p < .05$	stay in uk	-.312	139	.417	.423
	no plans	-.840	122	.320	.302
	Not significant:	Grammatical category, Preceding consonant, Gender, LOR, ATT, AW, CHA, MOT, NS Partner, Use of L1/L2, Formal English instruction, No of syllables			
Model	deviance	df 9	intercept	-.536	mean .369
	423.021				
	Speaker ID random standard deviation: 1.413				

between LoE and the use of a standard variant is a somewhat circular argument. An increased frequency of the standard variant in someone's speech might just as easily be playing a part in the evaluation of that person's speech as proficient, as it is a result of increased proficiency. Yet the separate effect of LOR suggests that spending time in the UK and being exposed to more examples of [ɲ] does play a small part in its increased use.

The second regression analysis looked for patterns in the use of the two velar + plosive variants [ɲg] and [ɲk]. All [ɲ] and [ɲ] were excluded and the independent variable 'following sound' was replaced with 'following voice'<sup>15</sup> in order to provide more insight into the distribution of the two variants (in reference to the role of voicing mentioned earlier in Polish ING). Previously, 'following segment' was a variable with three options: *velar*, *alveolar*, or *other*, which was included to test for regressive assimilation with regard to [ɲ]; 'following voice' is a variable with six options: *voiced obstruent*, *voiceless obstruent*, *nasal*, *approximant*, *vowel*, *pause*, which aims to explore any patterns of voicing assimilation. The results of this analysis can be seen in Table 7.

The earlier assertion that a following pause would favor [ɲk] rather than [ɲg] is supported, although it was also predicted that a voiceless obstruent would favor this addition of a voiceless plosive, which it does not appear to do. Instead, a nasal is the only other sound showing the same tendency. The reappearance of future plans as statistically significant is of interest, with those speakers who intend to return to Poland favoring [ɲk] and those who intend to stay in the UK or with no plans favouring [ɲg]. However, these results should all be treated with caution due to the low number of tokens (398).

TABLE 8. *Regression analysis (Rbrul) of the effect of linguistic and social factors on (ING), focusing on [ɪŋ]. All speakers*

Application value: [ɪŋ]	Factor	Log-odds	Tokens	Response proportion	Factor weight
Following voice <i>p</i> < .01	vowel	1.003	497	.221	.732
	approximant	.295	234	.132	.573
	voiceless obstruent	-.086	399	.103	.479
	pause	-.156	218	.106	.461
	nasal	-.382	83	.084	.406
	voiced obstruent	-.674	246	.053	.338
	Not significant:	rammatical category, Preceding consonant, Gender, ATT, AW, CHA, MOT, NS Partner, Use of L1/L2, Formal English instruction, No of syllables			
Model	deviance	df 7	intercept -1.906		mean .134
	1154.044 Speaker ID random standard deviation: 1.302				

TABLE 9. *Summary of regression analyses of ING variation*

Application value	[ɪn]	[ɪn]	[ɪn]	[ɪŋ]	[ɪŋk]
Nonapplication value	[ɪŋ] [ɪŋg] [ɪŋk]	[ɪŋ] [ɪŋg] [ɪŋk]	[ɪŋ] [ɪŋg] [ɪŋk]	[ɪŋg] [ɪŋk]	[ɪŋg]
Tokens	1655	719	1029	1576	398
Notes	40 speakers	16 speakers who exhibited [ɪn]	40 speakers.	40 speakers	40 speakers
Statistically significant constraints	<ul style="list-style-type: none"> <li>• Preceding consonant</li> <li>• Grammatical category</li> <li>• Gender</li> <li>• Future plans</li> </ul>	<ul style="list-style-type: none"> <li>• Preceding consonant</li> <li>• Grammatical category</li> <li>• LOR</li> <li>• LoE</li> <li>• Attitude</li> </ul>	<ul style="list-style-type: none"> <li>• Lexical frequency</li> <li>• Preceding consonant</li> <li>• Grammatical category</li> <li>• No. of syllables</li> <li>• Gender</li> </ul>	<ul style="list-style-type: none"> <li>• Following sound</li> <li>• Grammatical Category</li> <li>• LOR</li> <li>• LoE</li> </ul>	<ul style="list-style-type: none"> <li>• Following voice</li> <li>• Future plans</li> </ul>

One final analysis was carried out in an attempt to deal with the issue of [ɪŋg] being a local variant. The point was made that it was impossible to determine whether use of [ɪŋg] in the speech of the participants was a result of local acquisition or L1 interference. While this remains the case, a little insight can be gained by running an analysis of all tokens with [ɪŋg] as the application value. Following segment was once again replaced with ‘following voice’, but all other variables remained the same as in the very first analyses. The results can be seen in Table 8.

The fact that following voice emerged as statistically significant above all other factors suggests that it is more likely to exist as a result of L1 interference than of

local acquisition. The nature of the L1 interference is not clear, as the ordering of the categories does not correspond with what we know about Polish velar nasals. In order to be sure that ‘following voice’ was not simply an overriding powerful variable, the analysis was repeated with [m] as the application value (as in the very first analysis) but the variable was not statistically significant. On the basis of these results, in the following discussion [ɱŋ] will be treated as an L1-influenced variant rather than a local variant, as this appears to be the more likely scenario.

## DISCUSSION

From the five analyses carried out, a variety of constraints emerged as statistically significant (see Table 9). Some of the more predictable ones (e.g. LOR and LoE in the second analysis) have already been commented on briefly in the results section. The following discussion will therefore explore the implications of some of the more noteworthy findings, or those which potentially provide a greater insight into the possible processes at work.

### *Linguistic constraints*

The two linguistic constraints of preceding consonant and grammatical category clearly have a consistent influence on the distribution of ING. This consistency lends further strength to the argument for employing variationist methods in a second language setting, illustrating the fact that L2 speech can and does exhibit systematic variation. Moreover, the fact that both constraints reflect the patterns identified in previous research (e.g. Houston, 1985; Labov, 2001) suggests that these speakers are acquiring NS patterns of variation. However, not all the expected constraints proved to be statistically significant. For example, although there was evidence of progressive dissimilation, there was no sign of any regressive assimilation. In fact, even when an additional Rbrul analysis was carried out without individual speaker as a random effect, thus giving a much less conservative output, following sound still failed to reach statistical significance. This mirrors Labov’s (2001:87) findings when he reported no evidence of this type of phonological conditioning. Alternatively, this might be a case of L1 interference, with the Polish rules for the regressive assimilation of the -ing coda overriding, or at least affecting, the patterns generally seen in English.

### *Gender*

There appears to be a clear gender effect at work in the distribution of [ɱn], with women more likely than men to use this alveolar form. This effect represents a deviation from what is usually expected in L1 speech, where numerous studies have shown the reverse to be the case. Even in L2 studies, this traditional gender pattern has held or even been exaggerated. In Adamson and Regan’s (1991)

study into Cambodian speakers' use of ING, male speakers not only showed a higher rate of [ɪn], but this rate was higher still when more attention was paid to speech, a finding that the authors explained in terms of covert prestige. However, the recent findings of Schlee et al. (2011) reflected those presented here. In their London data, they found that the Polish females were more likely than the Polish males to use [ɪn] and explained the pattern in terms of ING not being a stable sociolinguistic variable for L2 speakers. Schlee et al. go on to interpret the differences found between the constraint hierarchies and rankings of the locally born speakers and those of the Polish-born speakers as examples of a reinterpretation or transformation of the constraints by the L2 speakers. This is a useful interpretation, and one that reflects the findings of recent research into long-term language and dialect contact (e.g. Buchstaller & D'Arcy, 2009; Meyerhoff, 2009) in which different strengths of transfer between the model and replica varieties are discussed. While it would be relatively simple to apply a similar interpretation to the present data, the lack of comparative, current data from local NSs weakens the hypothesis slightly. The studies mentioned all have a relevant, local comparison to explore, whereas the data being discussed here are relying on more general comparisons with a wider range of previous research. Further research into local NS patterns is required in order to determine whether this gender pattern can be attributed to constraint reinterpretation.

An alternative (and perhaps complementary) interpretation of the gender difference is one which takes a 'gender-as-practice' type approach as espoused by, for example, Eckert and McConnell-Ginet (1992). Indeed, this is effective in accounting for variation in another linguistic feature from the same study, glottal variation in /t/ (Drummond, 2011). As mentioned earlier, this can only be a tentative interpretation given the lack of ethnographic, practice-observing data. However, within these limitations, there is value in taking this kind of approach. The point is, isolating gender from other social factors is neither possible nor desirable, especially in this group of speakers where there is the added complication of potentially different Polish and British gender norms and identities. In the t-glottaling study, the suggestion was made that the real source of (gender) difference might lie in the contexts in which English is used, and that it is these contexts which differ with respect to gender. Clearly attempting to ascertain an individual's context of L2 use is challenging and complex, but a person's occupation offers some insight, particularly as this is the situation in which most contact with NSs is likely to occur. If we look at a list of identifiable occupations of the participants divided by sex, an interesting picture emerges (Table 10). The highlighted occupations are those of the 16 speakers<sup>16</sup> who exhibited [ɪn]. It could be argued that the use of the variant in question is influenced by context of L2 use rather than by gender; it just happens to be the case that those contexts of use are divided along gender lines. What is striking about the female side of the list is that with one exception (bookmakers), the occupations which do not coincide with the use of [ɪn] are those which one would expect to involve the least contact with NSs, and the occupations which do coincide with the use of [ɪn] are all potentially high contact. The male

TABLE 10. *Identifiable occupations of the participants, categorized by sex. Use of [m] is highlighted*

Male	Female
Factory	Café
Warehouse	Shop manager – department store
Bus driver	Bar manager
University canteen	Office – insurance
Office – small software company	Bookmakers
Hospital – mental health nurse	Waitress
Mechanic	Office – hotel admin
Welder	University researcher
Warehouse	Shop assistant – department store
Security guard – industrial estate	Housewife
Student	Polish office
	Housewife / Classroom assistant
	Student

occupations are mostly the kind in which minimal contact with NSs would be expected. Two of those which do suggest more NS contact are highlighted as coinciding with the use of [m]. Admittedly, the pattern is not so clear cut for the males as it is for the females, as there are several jobs on the male side which do suggest a higher level of contact than others (e.g. bus driver and nurse); however, there is clearly an underlying trend.

### *Aspects of identity*

The statistical significance of future plans in two of the regression analyses paints an interesting picture, perhaps on the importance of identity in the acquisition of local features. It should be borne in mind that future plans is but one aspect of an individual's identity and can only be seen as playing a small part in terms of overall identity construction. However, it can still offer insights into what might be happening in this particular community. Those speakers who were planning on returning to Poland were found to be less likely to produce [m] in the analysis which included all four variants and more likely to produce [ɲk] in the analysis which looked only at the two nonstandard velar nasal + plosive variants. It is perhaps possible to view the four variants as existing on a continuum, with the most L1-influenced variant at one extreme, and the most L2-influenced variant at the other. It should be noted that this interpretation entails the acceptance of the argument made earlier about [ɲg] more likely being an L1-influenced variant than a local variant. While none of the speakers is (or is likely to be) categorically at one end or the other, the results of the analyses suggest that those speakers who intend to return to Poland are towards one end, and those speakers who intend to stay in the UK or who have no plans are towards the other. Figure 4 provides a visual representation of this idea. It shows that while those who plan to stay in the UK or who have no plans exhibit

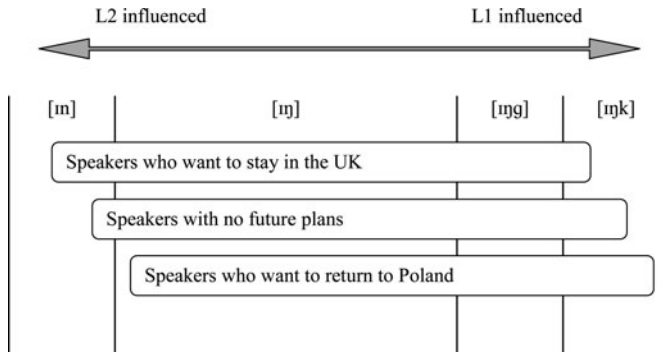


FIGURE 4. A visual representation of a possible [ɪn] - [ɪŋk] continuum.

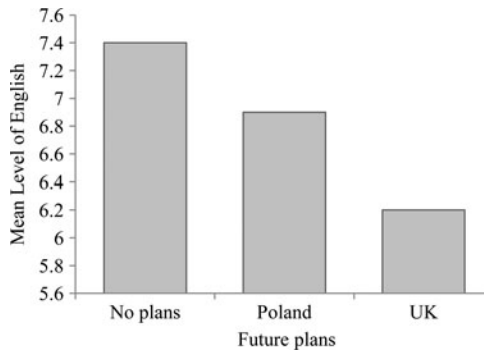


FIGURE 5. Chart showing the mean level of English categorized by future plans.

all four variants (but to slightly different degrees), those speakers who plan to return to Poland exhibit no [ɪn] tokens yet more [ɪŋk] tokens than the other two groups. The [ɪŋ] category is given the largest area in the diagram to reflect its status as the most common form.

Viewing the variants as lying on a continuum is especially plausible due to the fact that it is very unlikely that a speaker will produce [ɪŋk] without also producing [ɪŋg]. In fact, of the 23 speakers who produce [ɪŋk], only one shows no tokens of [ɪŋg].

The three categories of future plans do not correlate with any other factors, suggesting that this is a real constraint on the variation of ING. Most notably, there is no relationship between future plans and level of English, a factor which one intuitively feels might affect the distribution of [ɪŋk]. This lack of relationship is made clear in Figure 5, which shows that the mean LoE of those speakers who plan to return to Poland is actually higher than that of those speakers who plan to stay in the UK. Therefore, the increased use of [ɪŋk] in the Poland group cannot be put down to a lower level of spoken English.

Instead, the results could be interpreted (in an admittedly limited way) as a measure of identity towards the L2 or the L1 culture. Those speakers who intend to return to Poland arguably feel a stronger sense of identity and allegiance towards their native country and culture, and this is reflected in their use of a variant which signals that connection. On the other hand, it is likely that those speakers who intend to settle in the UK, while still identifying themselves as Polish, will also identify to a certain extent with the target culture. This reconstruction of their Polish identity could be reflected in their reduced use of [ɪŋk] and increased use of [ɪn].

## CONCLUSION

This article began by asking two questions:

1. To what extent are the patterns of ING variation being acquired by NNSs similar to those patterns exhibited by NSs?
2. What social factors influence the nature of variation?

Clearly, the data presented here have shown that there is considerable variation in the speech of NNSs. Although the majority of tokens collected from the 40 speakers were of a single variant ([ɪŋ]), three other variants ([ɪn], [ɪŋg], [ɪŋk]) are clearly possible. More importantly, however, this variation shows a considerable degree of systematicity, thus further reinforcing the idea that L2 speech is a valid arena for variationist analysis. In answer to question 1, it would appear that certain constraints do indeed exert similar patterns of influence over ING variation in both L1 and L2 speech. The most obvious comparisons can be drawn between the linguistic constraints of grammatical category and preceding consonant, where both showed consistently similar patterns to results of comparable L1 studies. It would be interesting in future work to see if these patterns are replicated in studies involving L1s other than Polish.

Interestingly, not all the statistically significant constraints exhibited the same direction as the existing L1 data, and the reversed patterning of gender is of particular interest here. In the discussion above, the suggestion was made that this result was attributable to the context of L2 use, particularly in relation to occupation, and it is these occupations/contexts of use which are divided along gender lines. This partly answers question 2, along with the other social factor of identity. Of course identity can be interpreted and constructed in many different ways, and the idea of using future plans as signaling aspects of identity is just one possibility. However, given that a decision to stay in a country automatically gives an individual the label of 'immigrant', it is perhaps a valid interpretation. With this in mind, the suggestion that the use of a certain variant, whether that use be conscious or unconscious (or more likely a combination of the two), can signal some kind of allegiance to one or other culture, is quite a justified one to make.

## NOTES

1. This is a brief overview. For more details please see Drummond (2011, 2010).
2. For example, Adamson & Regan (1991) used six categories, Abramowicz (2007) used four, and the Philadelphia study described in Labov (2001) used eight.
3. Mathisen's (1999) 'local' refers to Sandwell in the West Midlands.
4. By its very nature this had to be self-assessed proficiency.
5. The conversation was recorded using a Zoom H2 Handy Recorder placed unobtrusively on a surface near the participant. Recordings were made as .wav files using a 44.1 kHz sampling rate with 16-bit precision, saved onto an SD memory card then transferred onto a PC.
6. Although [ɪŋk] is actually a possible variant in British English, it is confined to a small subset of (ing) words (*anything, something, everything, and nothing*). It is also unlikely to be used in local Manchester speech (Watts, 2005).
7. There are difficulties, however, in distinguishing between different velar variants when there is a following velar. This will be addressed briefly in the results section.
8. Thanks to Jarosław Weckwerth of Adam Mickiewicz University in Poznań (personal communication) for clarifying these points.
9. 'The anxiety measure, for example, was discarded due to its correlation with several other factors.
10. The study Leech, Rayson, and Wilson (2001) is based on the 100,000,000 words of the British National Corpus (including 10,000,000 words transcribed from spoken English, with 4,000,000 from spontaneous conversation). Further details can be found at <http://www.natcorp.ox.ac.uk/docs/URG/BNCdes.html#spodes> [Accessed March, 2011]
11. A standard procedure for normalizing skewed data, especially when dealing with lexical frequency. See e.g. Clarke and Trousdale (2009).
12. In theory, Rbrul is slightly more forgiving than Goldvarb when it comes to knockouts in that results are still generated (Johnson, 2009). However, in reality, the extreme log-odds and factor weights that are created tend to skew the other results, making the whole output rather unreliable. For a discussion of knockouts in Goldvarb, see Tagliamonte (2006) or Paolillo (2002).
13. This study is part of a larger project which included some locally specific features, hence the inclusion of questions relating to Manchester.
14. This lack of correlation between LOR and LoE is somewhat surprising, suggesting that an individual's proficiency in English does not improve after living in the UK. However, it is in fact highly likely that this improvement takes place, but the effect is masked by recently arrived high-level speakers (i.e. high LoE but low LOR).
15. The two variables could not be used in the same regression analysis for reasons of collinearity.
16. There are not 16 occupations highlighted, as there is some degree of repetition. For example, four of the males who exhibited [ɪn] were students.

## REFERENCES

- Abramowicz, Łukasz. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics* 13(2):27–37.
- Adamson, H. D., & Vera M. Regan. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: a preliminary study. *Studies in Second Language Acquisition* 13(1):1–22.
- Bucholtz, Mary, & Kira Hall. (2004). Language and identity. In A. Duranti (ed.), *A companion to linguistic anthropology*. Malden, MA: Blackwell. 369–394.
- Buchstaller, Isabelle, & Alexandra D'Arcy. (2009). Localized globalization: A multi-local, multivariate investigation of quotative be like. *Journal of Sociolinguistics* 13(3):291–331.
- Clark, Lynn, & Graeme Trousdale. (2009). Exploring the role of token frequency in phonological change: Evidence from TH-fronting in east-central Scotland. *English Language and Linguistics* 13(1):33–55.
- Corder, S. Pit. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Drummond, Rob. (2010). *Sociolinguistic variation in a second language: The influence of local accent on the pronunciation of non-native English speakers living in Manchester*. Unpublished PhD thesis, University of Manchester.
- . (2011). Glottal variation in /t/ in non-native English speech: Patterns of acquisition. *English World-wide* 32(3):280–308.
- Eade, John, Stephen Drinkwater, & Michal Garapich. (2006). Class and ethnicity—Polish migrants in London. *CRONEM, University of Surrey, December*.



- Eckert, Penelope. (2010). Three waves of variation study: The emergence of meaning in the study of variation. (forthcoming).
- Eckert, Penelope, & Sally McConnell-Ginet. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21(1):461–488.
- Garner, Steve, James Cowles, Barbara Lung, & Marina Stott. (2009). *Sources of resentment, and perceptions of ethnic minorities among poor white people in England*. D. f. C. a. L. Government. London: Department for Communities and Local Government.
- Gatbonton, Elizabeth, Pavel Trofimovich, & Michael Magid. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly* 39: 489–512.
- Gussmann, Edmund. (2007). *The phonology of Polish*. New York: Oxford University Press.
- Hazen, Kirk. (2006). The /in/ing variable. *Encyclopedia of Language and Linguistics* 5:581–584.
- Holmes, Janet, & Miriam Meyerhoff. (1999). The community of practice: Theories and methodologies in language and gender research. *Language in Society* 28(2):173–183.
- Houston, Ann Celeste. (1985). *Continuity and change in English morphology: The variable (ING)*. Unpublished Ph.D. thesis. University of Pennsylvania.
- Johnson, Daniel Ezra. (2008). *Rbrul, Version 1.9*. [www.ling.upenn.edu/~johnson4/Rbrul.R](http://www.ling.upenn.edu/~johnson4/Rbrul.R).
- . (2009). Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3(1):359–383.
- Johnstone, Barbara, & Judith M. Bean. (1999). Self-expression and linguistic variation. *Language in Society* 26(2):221–246.
- Labov, William. (2001). *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- Leech, Geoffrey, Paul Rayson, & Andrew Wilson. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow: Longman.
- Lybeck, Karen. (2002). Cultural identification and second language pronunciation of Americans in Norway. *Modern Language Journal* 86(2):174–191.
- Marx, Nicole. (2002). Never quite a 'native speaker': Accent and identity in the L2—and the L1. *Canadian Modern Language Review/La Revue canadienne des langues vivantes* 59(2):264–281.
- Mathisen, Anne. (1999). Sandwell, West Midlands: Ambiguous perspectives on gender patterns and models of change. In P. Foulkes & G. Docherty (eds.), *Urban voices: Accent studies in the British Isles*. London: Arnold:107–123.
- Mendoza-Denton, Norma. (2002). Language and identity. In J. Chambers, P. Trudgill & N. Schilling-Estes (ed.) *The handbook of language variation and change*. Malden, MA: Blackwell. 475–499.
- Meyerhoff, Miriam. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21(3):297–317.
- Milroy, Leslie. (1987). *Language and social networks*. Oxford: Wiley-Blackwell.
- Milroy, Leslie & James Milroy. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21(1):1–26.
- Mougeon, Raymond, Katherine Rehner, & Terry Nadasdi. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics* 8(3):408–432.
- Office for National Statistics. (2010). *Mid Year Population Estimates 2009*.
- Paoilillo, John C. (2002). *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: CSLI Publications.
- Piller, Ingrid. (2002). Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics* 6(2):179–208.
- Sankoff, David, Sali Tagliamonte & Eric Smith. (2005). *Goldvarb X: A variable rule application for Macintosh and Windows*. Toronto: Department of Linguistics, University of Toronto.
- Schleef, Erik, Miriam Meyerhoff, & Lynn Clark. (2011). Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-wide* 32(2):206–236.
- Schumann, John. (1978). The acculturation model for second language acquisition. In R. C. Gingras (ed.), *Second language acquisition and foreign language teaching*. Arlington, VA, Center for Applied Linguistics. 27–50.
- Tabouret-Keller, Andree. (1997). Language and identity. In F. Coulmas (ed.), *The handbook of sociolinguistics*. Malden, MA: Blackwell. 315–326.
- Tagliamonte, Sali. (2004). Something's going on! In L. Gunnarsson, G. Bergstrom, S. Eklund, et al. (eds.), *Language variation in Europe*. Uppsala: Uppsala University. 390–403.
- . (2006). *Analyzing sociolinguistic variation*. Cambridge: Cambridge University Press.

- Tarone, Elaine. (2007). Sociolinguistic approaches to second language acquisition research 1997–2007. *The Modern Language Journal* 91(s1):837–848.
- Trudgill, Peter. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Watts, Emma L. (2005). *Mobility-induced dialect contact: A sociolinguistic investigation of speech variation in Wilmslow, Cheshire*. Unpublished PhD thesis. Essex.
- Wells, John. (1982). *Accents of English Vol 2*. Cambridge: Cambridge University Press.
- Young, Richard. (1988). Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition* 10(3):281–302.
- \_\_\_\_\_. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics* 19(1):105–132.

## APPENDIX

TABLE A1. Total count for ING for all speakers.

Speaker	[ɪŋ]	[ɪn]	[ɪŋg]	[ɪŋk]	Total
1	43 86%	0 0%	7 14%	0 0%	50
2	41 97.6%	1 2.4%	0 0%	0 0%	42
3	31 62%	0 0%	11 22%	8 16%	50
4	36 72%	0 0%	14 28%	0 0%	50
5	39 78%	1 2%	9 18%	1 2%	50
6	41 82%	0 0%	0 0%	9 18%	50
7	46 92%	1 2%	2 4%	1 2%	50
8	30 100%	0 0%	0 0%	0 0%	30
9	27 77.1%	0 0%	5 14.3%	3 8.6%	35
10	43 86%	1 2%	2 4%	4 8%	50
11	18 46.4%	0 0%	8 21.1%	12 31.5%	38
12	15 79%	1 5.3%	3 15.2%	0 0%	19
13	14 37.8%	0 0%	10 27%	13 35.2%	37
14	3 47.4%	0 0%	2 10%	15 75%	20
15	20 48%	29 50%	1 2%	0 0%	50
16	26 70.3%	1 2.7%	10 27%	0 0%	37
17	40 80%	1 2%	8 16%	1 2%	50
18	37 74%	7 14%	6 12%	0 0%	50

Continued

TABLE A1. *Continued*

Speaker	[ɪŋ]	[ɪn]	[ɪŋg]	[ɪŋk]	Total
19	27 54%	20 40%	1 2%	2 4%	50
20	19 73.1%	0 0%	3 11.5%	4 15.4%	26
21	21 42%	0 0%	16 32%	13 26%	50
22	30 100%	0 0%	0 0%	0 0%	30
23	20 40%	0 0%	14 28%	16 32%	50
24	7 25.9%	1 3.7%	8 29.6%	11 40.7%	27
25	30 100%	0 0%	0 0%	0 0%	30
26	38 76%	0 0%	8 16%	4 8%	50
27	47 94%	2 4%	1 2%	0 0%	50
28	26 52%	0 0%	13 26%	11 22%	50
29	30 100%	0 0%	0 0%	0 0%	30
30	31 62%	19 38%	0 0%	0 0%	50
31	30 100%	0 0%	0 0%	0 0%	30
32	31 62%	0 0%	10 20%	9 18%	50
33	30 100%	0 0%	0 0%	0 0%	30
34	36 72%	8 16%	3 6%	3 6%	50
35	34 68%	5 10%	2 4%	9 18%	50
36	38 76%	0 0%	12 24%	0 0%	50
37	25 50%	0 0%	17 34%	8 16%	50
38	36 72%	3 6%	6 12%	5 10%	50
39	30 100%	0 0%	0 0%	0 0%	30
40	12 33.3%	0 0%	13 36.1%	11 30.6%	36
<b>Total</b>	<b>1178</b> <b>70.3%</b>	<b>101</b> <b>6%</b>	<b>225</b> <b>13.5%</b>	<b>173</b> <b>10.3%</b>	<b>1677</b>