The permeability of dialect boundaries:

A case study of the region surrounding Erie, Pennsylvania

a *dissertation proposal* by

Keelan Evanini, University of Pennsylvania

manuscript, 5/14/08

1 Introduction

The Atlas of North American English (Labov et al. 2006), henceforth ANAE, represents the first comprehensive phonological study of the entire North American continent. It thus provides a detailed overview of the various dialect regions and the sound changes that are taking place in each one. ANAE, however, is a study of urban speech—the survey's methodology sampled two speakers from every Metropolitan Statistical Area with 50,000 or more inhabitants. Thus, while ANAE is able to precisely define the characterists of the dialect regions of North America, it is not able to describe how the areas at the dialect boundaries look, since these boundaries normally lie in less populated geographic regions between urban areas. The aim of this dissertation is to address this lack of coverage by studying the dialect boundary areas around a city of theoretical interest: Erie, Pennsylvania.

Erie holds a unique place in the dialectology of North America, since it is the only city to have switched from the North to the Midland. The earliest dialectological records of the region (Kurath 1949, Kurath and McDavid 1961) show Erie to pattern together with the North with respect to nearly all lexical and phonological isoglosses. However, ANAE shows that the phonology of Erie is clearly no longer Northern, and shares two crucial phonological traits with the Midland: the merger of /o/ (as in *cot*) and /oh/ (as in *caught*) and a pattern of raising /æ/ before nasals (Labov et al. 2006:205). This shift of allegiance from the North to the Midland is surprising, since the North is perhaps the most cohesive dialect region in North America—as evidenced by the high rate of homogeneity and consistency of its defining isoglosses (Labov et al. 2006:151)—and the boundary between the North and the Midland is one of the sharpest boundaries in North America.

The goals of this dissertation are both empirical and theoretical. On the one hand, I will build a

corpus of speech samples collected from Erie and the surrounding boundary regions. This research will supplement the ANAE by describing the speech of less populated areas that have never been studied before and by determining the exact locations of the dialect boundaries on all sides of Erie. On the other hand, after the data has been collected, I will seek to address several theoretical questions in dialect geography through an analysis of the corpus. Section 1.1 summarizes three of the most important theoretical questions to be addressed.

1.1 Theoretical Questions

1.1.1 What Causes the Diffusion of Linguistic Change across Dialect Boundaries?

One of the main theoretical questions facing dialect geographers is to explain the mechanism by which linguistic changes diffuse across dialect boundaries. The types of theories that have been put forth to explain the spread (or lack thereof) of a linguistic change from one dialect region to a neighboring one often make reference to the demographic situation that obtains in the two regions. For example, the Gravity Model (Trudgill 1974) proposed that the spread of sound change from one area to another is proportional to the population of the two areas, and inversely proportional to the square of their distances. In a similar approach, the Cascade Model (Labov 2003) proposed that changes spread from large cities to smaller ones, skipping over the sparsely populated areas in between. Both of these models are based on the idea that linguistic change is brought about through increased communication with speakers from another dialect region (Labov 1974). All of these types of explanations are ultimately based on the Principle of Density (Bloomfield 1933).

On the other hand, structural factors can also play a role in promoting or inhibiting the spread of a linguistic change across a dialect boundary. For example, the merger of /o/ and /oh/ would be expected to spread into all dialect regions in North America according to Herzog's Principle that "mergers expand at the expense of distinctions" (Labov 1994:313). However, other structural factors have inhibited its spread across certain boundaries, such as the fronting of /o/ as part of the Northern Cities Shift in the North. An example of a structural factor influencing the spread of a change is the correlated behavior of /ow/ and /ʌ/: the backing of /ʌ/ in the North appears to be influenced by the extreme back position of /ow/ in that region, whereas the fronting of /ʌ/ in the

Midland and South appears to be influenced by the fronting of /ow/ (Labov et al. 2006:143).

The most important linguistic change to spread to Erie was the merger of /o/ and /oh/. Section 2.2 shows how these two phonemes were originally distinct in Erie, and Section 4 presents apparent time and real time data to demonstrate that the merger must have taken place in Erie over 100 years ago. Other studies that have examined areas of recent merger in depth have proposed explanations based on demographic shift to explain the merger, either through foreign immigration (Herold 1990) or through an influx of residents from a neighboring merged region (Johnson 2007). Using these earlier studies as a guide, this dissertation will attempt to determine the relative influence of social and structural factors on the spread of the merger of /o/ and /oh/ to Erie and its continuing diffusion to neighboring regions.

1.1.2 Are All Dialect Boundaries Alike?

In conducting fieldwork around Erie, Pennsylvania, I will be examining three disjoint dialect boundary regions consisting of two separate dialect contact situations:

1. Erie vs. the North: This boundary exists in the two geographically distinct regions between Erie and Buffalo to the northeast of Erie and between Erie and Cleveland to the west. On the Erie side of this boundary, /o/ and /oh/ are merged, and there is no evidence of the Northern Cities Shift. On the North side of the boundary, /o/ and /oh/ remain distinct, and the NCS is represented in varying degrees. Figure 1 shows how Erie is located outside of most of the 8 isoglosses for the North in ANAE. Furthermore, the two that do include Erie are tenuous there, at best.[1]

2. Erie vs. the Midland / Western Pennsylvania: This boundary exists between Erie and the Midland region centered around Pittsburgh to the south of it. On the Erie side of this boundary,

[1]The barred blue isogloss showing the outer limits of the North is defined by the three characteristics that are thought to be necessary conditions for the Northern Cities Shift: a short-*a* system that is not split into tense and lax class, no fronting of /ow/, and the lack of a merger between /o/ and /oh/. While the first two characteristics clearly hold for the two Erie ANAE speakers, one speaker provides very slight evidence that the merger is not complete for her: two out of the five minimal pairs were judged close in production, although the rest were judged to be exactly the same. The other speaker is clearly merged across the board in production and perception. The second Northern isogloss that is shown to include Erie Figure 1 is the ED line, inside of which F2 of /e/ is less than 375 Hz higher than F2 of /o/. However, the ANAE speaker from Erie who has these two vowels the closest actually has them separated by 388 Hz. So, Erie should not actually have been included inside of the ED isogloss.

/ow/ and /aw/ are not being fronted as they are in much of the Midland. Furthermore, near Pittsburgh, on the southern side of the boundary, the Pittsburgh Shift is apparent in many speakers who show monophthongization of /aw/.

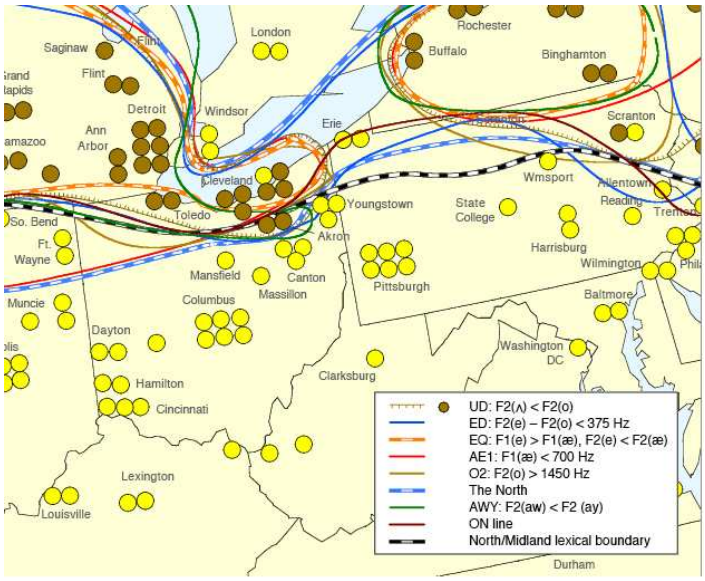

Figure 1: Erie and the surrounding region (Map 14.11 from ANAE)

One of the main goals of the fieldwork for this dissertation will be to determine the nature of these dialect boundaries in the areas of transition between the two regions. Several theoretical possibilities exist, based on the amount of overlap between the features of the two regions in the boundary area. For example, Chambers and Trudgill (1999:104) distinguish between *abrupt* and *gradual* transition areas. A slightly more refined taxonomy is presented by Dinkin (2007:1–2): *sharp*, *fading*, *overlapping*, and *null* boundaries. Research into the two boundary regions around Erie will determine what type of boundary exists in each area. It is hypothesized that the boundary between Erie and the North will be a sharper boundary, since the merger of /o/ and /oh/ prevents the other stages of the Northern Cities Shift from taking place. ANAE has already shown that the boundary between the North and the Midland consists of a bundle of several closely related isoglosses, and that this is boundary is one of the sharpest in North America (p. 205). On the other hand, it is hypothesized that the boundary between Erie and the area to the south with strong fronting of /ow/ will be more gradual, since there are no structural barriers to the fronting of /ow/ in

Erie.

Sections 2 and 5 go into detail about the linguistic features that define the regions, and provide evidence for Erie's current relationship with them. Section 4.6 provides evidence for the current location of the boundary between the area with the merger of /o/ and /oh/ around Erie and the area of the North with the distinction.

1.1.3 What is the Relationship between Different Types of Isoglosses?

Chambers and Trudgill (1999) define six different types of isoglosses, based on the level of linguistic structure involved: *lexical*, *pronunciation*, *phonetic*, *phonemic*, *morphological*, and *syntactic*. Most studies that have attempted to define dialect regions to date have focused only on one or two of these types of isoglosses. For example, Carver (1987), Cassidy and Hall (1985–2002), and Kurath (1949) are all concerned with lexical isoglosses. On the other hand, (Kurath and McDavid 1961) examines only pronunciation and phonetic isoglosses. ANAE is the most wide-ranging single study to date: its main focus is on phonetic and phonemic isoglosses, but it also collected some data on lexical, pronunciation, and syntactic isoglosses.

There has not been any systematic research into how the different types of isoglosses pattern differently, especially around dialect boundary areas. ANAE has shown how phonetic and phonemic isoglosses bundle together along dialect boundaries in cases of large-scale sound shifts with many related components, such as the Northern Cities Shift and the Southern Shift. However, there has been less research into the correlation of isoglosses from levels of linguistic structure that are not related. One such case that has been investigated is the close correlation between the lexical boundary dividing the North and the Midland in Kurath (1949) and the phonological boundary separating the two regions in ANAE (Labov et al. 2006:205). ANAE showed that Erie was the only city to switch from being on the Northern side of the lexical boundary to the Midland side of the phonological boundary. However, it is not known whether this shift in phonological status also coincided with a shift to Midland features in other levels of linguistic structure. Since many of the lexical items surveyed in (Kurath 1949) are now obsolete, I will investigate the relationship between the phonological isoglosses around Erie and a few syntactic isoglosses that are characteristic of the

Midland areas with heavy original Scots-Irish settlement, specifically positive *anymore* and *need* + Past Participle. Section 5.2 discusses these features and the methodology that will be used in investigating them.

1.2 Methodology

This dissertation will also contribute to the field by employing a methodology for data analysis that has not been used before in dialect geography. After the interviews have been collected, they will also be transcribed in their entirety. Then, automatic speech recognition will be done on the interview recordings, using the transcriptions as a reference, in order to determine the locations of the phoneme and word boundaries. This annotated corpus of the field recordings will then enable the extraction of linguistic information—such as pitch, formant trajectories, and duration—that has been neglected by earlier studies that only took measurements by hand at a single point in time. Another benefit of such a corpus is that it will enable quick searches for all tokens of a given phoneme or word for any given speaker or across all speakers. Section 6.2 describes this methodology in more detail.

1.3 Outline

This document is structured as follows: Section 2 presents the evidence for Erie's original status as part of the Northern dialect region. Section 3 presents data on the early settlement history of the area in an attempt to explain the original Northern influence. Section 4 examines both apparent time and real time data in an attempt to establish the chronology of the merger of /o/ and /oh/ in Erie. Section 5 discusses Erie's current relation to the Midland region and, specifically, Western Pennsylvania. Finally, Section 6 describes the methodology that will be used for further data collection and for analysis.

2 Erie's original status as a Northern city

All of the earliest dialectological studies indicate that Erie was aligned with the North for at least the early part of the 20th century. The evidence for this based on lexical items is quite strong, and comes primarily from *A Word Geography of the Eastern United States* (Kurath 1949), henceforth *WG*, as well as the data collected for the *Dictionary of American Regional English*, henceforth DARE, as presented in Carver (1987). The sources for the phonological evidence are *The Pronunciation of English in the Atlantic States*, henceforth *PEAS*, (Kurath and McDavid 1961) and Wetmore (1959), both based on the fieldwork done for the *Linguistic Atlas of the Middle and South Atlantic States*, henceforth *LAMSAS*.

2.1 Lexical

WG provides two types of evidence for Erie's position within the North. First of all, Erie is located within 10 of the 11 defining isoglosses of the North (Maps 5–8 in *WG*) and all 6 isoglosses that are characteristic of both the North and the Midland (Maps 39 and 40). Table 1 lists these words along with their non-Northern counterparts, showing the Northern version used in Erie in italics.

Northern form	non-Northern form	map in WG
whiffletree, whippletree	swingletree	5a
pail	bucket	5a
darning needle	dragonfly	5a
teeter, teeterboard	seesaw	5b
stone boat	vehicle for dragging field stones	5b
spider	frying pan	6
skaffle	scaffold (in a barn)	6
buttry	pantry	6
stoop	porch	7
dutch cheese	cottage cheese	8
stone wall	fence built of loose stone	39
hay mow	hay loft	39
grist of corn	turn of corn	39
whinny	nicker / whicker	40
corn husks	corn shucks	40
string beans	snap beans	40

Table 1: Northern isoglosses in *WG* that contain Erie (italicized variant used in Erie)

Table 2 shows that *belly-gut* is the only one of the 11 Northern lexical items that does not contain Erie.

Northern form	non-Northern form	map in WG
belly-gut	*face-down on a sled*	7

Table 2: Only Northern isogloss in *WG* that does not contain Erie

Secondly, evidence for Erie's original status as a Northern city is provided by the Midland isoglosses in *WG*. Erie falls outside of 8 isoglosses that define the Midland (Maps 15-18) and 5 isoglosses that are characteristic of both the South and the Midland, all of which reach northward past Pittsburgh (Maps 41 and 42) . In Table 3 I refer to these two types of isoglosses as the *non-Northern isoglosses*, in contrast to the isoglosses in Table 1 that were either distinctly Northern or shared by the North and Midland.

non-Northern form	Northern form	map in WG
I want off	*I want to get off*	15
Sook!	*call to cows*	15
snake feeder	*dragonfly*	15
blinds	*roller shades*	16
bawl	*noise a calf makes*	16
poke	*paper bag*	17
sugar tree	*maple tree*	17
worm fence	*a rail fence laid zigzag*	18
corn pone, pone	*cornbread*	41
paling fence, pale fence	*picket fence*	41
roasting ears	*sweet-corn*	41
pole cat	*skunk*	42
Christmas gift!	*Merry Christmas!*	42

Table 3: Non-Northern isoglosses in *WG* that do not contain Erie (italicized variant used in Erie)

Conversely, Erie shares only two of the lexical items characteristic of the Midland: *run* and *smear case* (used to define the North Midland in Map 18), and one of the items common to both the Midland and South: *spicket*. These three lexical items that Erie shares with the Midland are shown in Table 4.

Thus, Erie behaves like a Northern city for 88% (29 out of 33) of the relevant lexical items from *WG*. Finally, Erie is also situated outside of the three isoglosses that Kurath provides for Western

non-Northern form	Northern form	map in *WG*
run	a small stream	18
smear case, smear cheese	cottage cheese	18
spicket	faucet	42

Table 4: Only non-Northern isoglosses in *WG* that contain Erie

Pennsylvania (Map 25 in *WG*). These three extend northward from Pittsburgh into northwestern PA, but none of them quite reach Erie. This suggests that Pittsburgh's influence in Erie was not yet strong at that time. These three lexical items are show in Table 5.

Western PA form	non-Western Pa form	map in WG
hap	*quilt*	25
doodle, hay doodle	*haycock*	25
drooth	*drought*	25

Table 5: Three Western Pennsylvania isoglosses that do not contain Erie

The evidence from DARE is available only indirectly through Carver (1987), who trolled through the DARE fieldwork data to compile maps that capture the regional patterning of some of the lexical items used in the survey. Instead of the more traditional concept of dialect region, Carver prefers to use dialect *layers* as his descriptive apparatus. He defines a dialect layer as "the composite of a unique set of areal isoglosses, the geographical spread of its lexicon" Carver (1987:16). This concept of dialect layer is useful for the lexicon (which, in contrast to the phonology, does not have structural relationships between its elements), because it does not force the researcher to posit discrete boundaries, and thus enables the landscape to be viewed more as a continuum. Any given community can belong to several different layers, each having a different strength at that location, based on the number of items from that layer used in the community. The different layer strengths thus provide information about how strongly that community is affiliated with each dialect region.

The evidence from the maps relevant to Erie is presented in Table 6. The first column in the table represents the name of the dialect layer, as defined by Carver. The next two columns represent the number of DARE isogloss terms for the layer that occur in Erie, and the total number of DARE

isoglosses used to define the layer, respectively. Unfortunately, even though Carver does provide lists of all of the isogloss terms he used to define the layers, there is no way to know, without consulting the original fieldwork data, which of them occur in any given geographical point—due to space limitations on the maps, Carver only depicts the number of terms, not the specific terms themselves. So, a direct comparison with the distribution of the words from WG in Tables 1–5 is not possible. The fourth column in Table 6 shows the strength of the boundary within which Erie is situated for each of the dialect layers. The possible types of boundaries are *primary*, *secondary*, *tertiary*, and *quaternary*, with a primary boundary containing the area where the highest percentage of terms for the layer are found. Finally, the fifth column shows the label of the corresponding map from Carver (1987).

Dialect Layer	# of terms in Erie	total # of terms for layer	boundary strength	map from Carver
North	33	82	primary	3.3
Upper North	20	62	secondary	3.7
Inland North	18	51	primary	3.9
Midland	1	40	N / A	6.5
Lower North	4	53	N / A	6.15

Table 6: Erie's position with regard to dialect layers in Carver (1987)

The three boundaries for the layers of the North provide good evidence for Erie's status as a Northern city at the time of the DARE fieldwork. Erie is located within primary boundaries for the North and Inland North layers; furthermore, Erie falls just outside of the primary boundary for the Upper North, which extends westward to Lake Erie, stopping just at the Pennsylvania-New York state line. On the other hand, the data provide very little evidence for associating Erie with the Midland. Map 6.15 situates Erie outside of the tertiary boundary for the Midland layer (which extends northward to around Youngstown, OH). Similarly, Erie falls outside of the two boundaries provided for the Lower North[2] layer, although the secondary boundary does stretch northward almost to Erie.

[2]Carver's Lower North layer corresponds closely to what is traditionally referred to as the North Midland. Carver's terminology reflects his disbelief in the existence of a separate Midland dialect region.

2.2 Phonological

Table 7 presents the features mapped in PEAS that are evidence for Erie's original affiliation with the North. For all of these features, the isogloss falls just south of Erie, indicating that Erie was always just on the edge of the boundary between the North and the Midland. The first column in Table 7 describes the feature that Erie shares with the North, and lists the lexical items that PEAS uses to illustrate this.[3] The second column describes the contrasting feature that is found just south of Erie. In some cases, this feature is widely distributed throughout the Midland region (e.g., /uw/ in *due*, Map 163), whereas in others the contrasting feature is more characteristic of Western Pennsylvania in particular (e.g., [druθ] for *draught*, Map 142). In either case, there is a clear boundary between Erie to the north and the area of Western Pennsylvania surrounding Pittsburgh to the south. The speakers from two counties immediately south of Erie appear to be transitional for many of these features, with Crawford County aligning more frequently with Western PA, and Warren County aligning more frequently with the North.

The large number of features that Erie shares with the North can be contrasted with the features in Table 8. These are the only three listed in PEAS that have Erie aligned with the Midland or Western Pennsylvania in opposition to the North.

Even more probative of Erie's original phonological alignment with the North is the status of the low-back vowels. This (along with the status of /aɛ/) is one of the two main structural features that determine the status of a dialect of present-day North American English (Labov 1991). The fact that /o/ and /oh/ are kept distinct by the two PEAS speakers from Erie proves the regions's original alignment with the North as opposed to Western Pennsylvania, where the two phonemes are merged as a low, back rounded vowel ([ɒ] in the PEAS notation).

Table 9 presents the /o/ and /oh/ words that are mapped in PEAS. First of all, the merger in vowel quality is shown clearly for the Western PA speakers: they have a rounded low-back vowel for all tokens. Interestingly, however, the atlas does mark a distinction in length: the three /oh/ words (*law*, *salt*, and *dog*) have a half-length mark, and are also characterized by the possible

[3]The isoglosses providing evidence for the low-back vowels are omitted from this table, since they are discussed separately below.

Erie feature (shared with the North)	contrasting Midland / Western PA feature	PEAS Map #
/i/ monophthongal in *crib*	ingliding diphthong [ɪ^{ə}]	4
/e/ monophthongal in *bed*	ingliding diphthong [ɛ^{ə}]	4
/ey/ more close (i.e. [eɪ]) in *day, bracelet*	[ɛɪ]	18, 19
non-fronted /ow/ in *ago, coat*	fronted to [ɜʊ]	20, 21
raised nucleus for /ay/ in *nine*	not raised	26
fronted [ɛʊ] in *mountain, (worn) out*	[aʊ ∼ ɑʊ]	28, 29
unrounded [a] in *father*	rounded [ɒ]	32
/iw/ in *dues* and *tube*	/uw/	33
vowel in *four* and *forty* distinct	merged	44
[e] before /r/ in *married*	[æ]	51
[ʊ] in *root*	[u]	113
[ʌ ∼ ɵ] in won't	[o]	125
[drauθ] for *draught*	[druθ]	142
/iw/ in *blue, chew,* and *suit*	/uw/	147
[ɪ] in final unstressed syllable of *careless, houses, haunted,* and *bucket*	[ə]	148
/iw/ in *due, new,* and *Tuesday*	/uw/	163, 164, 165
yeast pronounced as [jist]	[ist]	166
/ð/ in *without*	/θ/	170
/s/ in *greasy*	/z/	171

Table 7: Northern phonological isoglosses that contain Erie in *PEAS*

presence of a central offglide. The Northern speakers, on the other hand maintain a clear distinction in quality between the two types. The /o/ words have an unrounded low-central vowel, possibly slightly fronted in *oxen*, whereas the /oh/ words all have a consistently low-back rounded vowel.

Wetmore (1959) reaches the same conclusion based on a larger body of evidence from the LAMSAS fieldnotes. In addition to the data presented in PEAS, he examined the lexical items *pot, fought, shock, god, off, cloth, sauce, costs, frost, all, John, gone, launch, strong, saw, swamp,* and *wasp* (Wetmore 1959:109) for speakers from Western Pennsylvania (although it is unclear exactly which of these had data from the two Erie speakers). Based on this evidence he lists both Erie informants as having a distinction between /ɑ/ and /ɔ/ (Wetmore 1959:113).

Thus, it is clear from the earliest survey data available that Erie's linguistic original linguistic affiliation was with the North. Both the lexical data from WG and DARE as well as the phonological

Erie feature (shared with the Midland / Western PA)	contrasting Northern feature	PEAS Map #
/i/ is ingliding [rᵊ] in *whip*	[ɪ]	5
/e/ is ingliding [ɛᵊ] in *fence*	[ɛ]	9
trisyllabic pronunciation of *mushroom* ending in /n/	disyllabic ending in /n/	177

Table 8: Midland / Western PA phonological isoglosses that contain Erie in *PEAS*

Lexical Item	Erie	North	Western PA	Map #
oxen	[ɑ ∼ ɑ̟ ∼ a̟]	[ɑ ∼ ɑ̟ ∼ a̟]	[ɒ ∼ ɔ]	15
wash	[ɑ]	[ɑ]	[ɒ ∼ ɔ]	135
fog	[ɑ]	[ɑ]	[ɒ ∼ ɔ]	136
on	[ɑ]	[ɑ]	[ɒ ∼ ɔ]	138
law	[ɒˑ ∼ ɒˑᵊ]	[ɔˑ ∼ ɔˑᵊ]	[ɒˑ ∼ ɒˑᵊ]	22
salt	[ɒˑ ∼ ɒˑᵊ]	[ɔˑ ∼ ɔˑᵊ]	[ɒˑ ∼ ɒˑᵊ]	23
dog	[ɒˑ ∼ ɒˑᵊ]	[ɔˑ ∼ ɔˑᵊ]	[ɒˑ ∼ ɒˑᵊ]	24

Table 9: /o/ and /oh/ words in PEAS in Erie, the North, and Western PA

data from PEAS and further LAMSAS field records show that Erie was located inside of most of the Northern isoglosses and outside of most of the Midland / Western PA isoglosses. The following chapter will attempt to document the early settlement history of Erie and show how its original linguistic status as a Northern city is a consequence of the geographic origins of the early settlers to the area.

3 Settlement history of Erie

The first recorded inhabitants of the coastal area along Lake Erie were the Erie Indians. Their tribe was defeated in battle in 1654 by the Senecas, who killed many of the Eries and scattered those remaining alive among neighboring tribes. The Senecas remained the main inhabitants of the area up until the turn of the 19th century. The British and French each did establish forts in the area, and were vying for control of this strategically important bridge between the eastern and western settlements. However, neither the French or the British established non-military settlements.

American control over the area started in 1784 when Pennsylvania acquired the rights to the land through a treaty with the Six Nations. The British did not leave their military forts immediately, but were out-maneuvered by the American military who were able to establish alliances with the Senecas. The first American settlers arrived in 1795, aided by inexpensive land grants through the Pennsylvania Population Company. However, relations with the Senecas cooled drastically as American settlers began to move in, and the frequent violent raids discouraged settlement in the first years. However, military reinforcements were quickly sent to the area, and they dealt harshly with the hostile Senecas. This paved the way for a rapid increase in the population in the early 19th century.

Erie was incorporated as a borough in 1805, and experienced a sizeable growth in population throughout the 19th century. Table 10 shows the population growth in the city of Erie and Erie County for this time period (Sanford 1894, FWP 1938).

Year	City Pop.	County Pop.
1800	81	1,468
1810	394	1,358
1820	635	8,553
1830	1,329	17,041
1840	3,412	31,344
1850	5,858	38,742
1860	11,113	49,697
1870	15,516	
1880	27,737	
1890	40,634	

Table 10: Population growth in the city of Erie and Erie County, 1800–1890

In considering the effect of settlement patterns on the subsequent linguistic system in Erie, it is necessary to consider the geographical origins of the early settlers. It has been known for some time among cultural geographers that a small group of the earliest settlers in a region can have a profound and lasting impact on the culture of the region. This was formulated clearly by Zelinsky (1973:13–14) as the Doctrine of First Effective Settlement:

> Whenever an empty territory undergoes settlement or an earlier population is dislodged by invaders, the specific characteristics of the first group able to effect a viable, self-perpetuating society are of crucial significance for the later social and cultural geography of the area, no matter how tiny the initial band of settlers may have been...Thus, in terms of lasting impact, the activities of a few hundred, or even a few score, initial colonizers can mean much more for the cultural geography of a place than the contributions of tens of thousands of new immigrants a few generations later.

Recent studies have also shown that this is true for linguistic structure: Mufwene (1996) formulated a similar idea which he called the Founder Principle, and demonstrated its applicability to creole genesis; Labov (2007) provides an explanation of the diffusion of the New York City short-/æ/ system to Cincinnati based on early settlement data; and Dinkin (2008) correlates the early preponderance of Dutch settlers in Amsterdam and Oneonta in eastern upstate New York to the fact that these two towns are not participating in the Northern Cities Shift.

As Section 2 has shown, the earliest linguistic evidence indicates that Erie originally patterned with the North. Thus, we would expect a large portion of the earliest settlers to have arrived from sources similar to the ones that settled nearby cities in the North such as Buffalo and Rochester. Published sources indicate that there were two main sources of early settlers to Erie County: New England and Southeastern Pennsylvania. The New Englanders who arrived in Erie were for the most part of British origin, and came to Erie through New York state primarily from Massachusetts and Connecticut. On the other hand, the settlers arriving from Southeastern Pennsylvania were of either Scots-Irish or German descent.

However, published sources do not provide the information necessary to account for the early linguistic patterns, namely, the proportion of the two groups among the early settlers. FWP (1938:23) does state that most of the settlers prior to 1800 came from New England and New York, and that subsequent migrations were also from the same sections. However, the authors provide no specific

data to support this claim. On the other hand, Sanford (1894) claims that the early settlers were a mix of New Englanders and Scots-Irish: "the first settlers in Erie County were mostly...from moral, thrifty, intelligent New England; or...perhaps a more numerous class, of the illustrious, historic race of Scotch-Irish." However, she also does not provide any specific numbers to support this claim.

Two primary sources were consulted in an attempt to document the early settlement history of the region more accurately: a publication documenting the location of the burial sites for all of the "Revolutionary Patriots" interred in Erie County, and an early history of the county with biographical information for the prominent early settler.

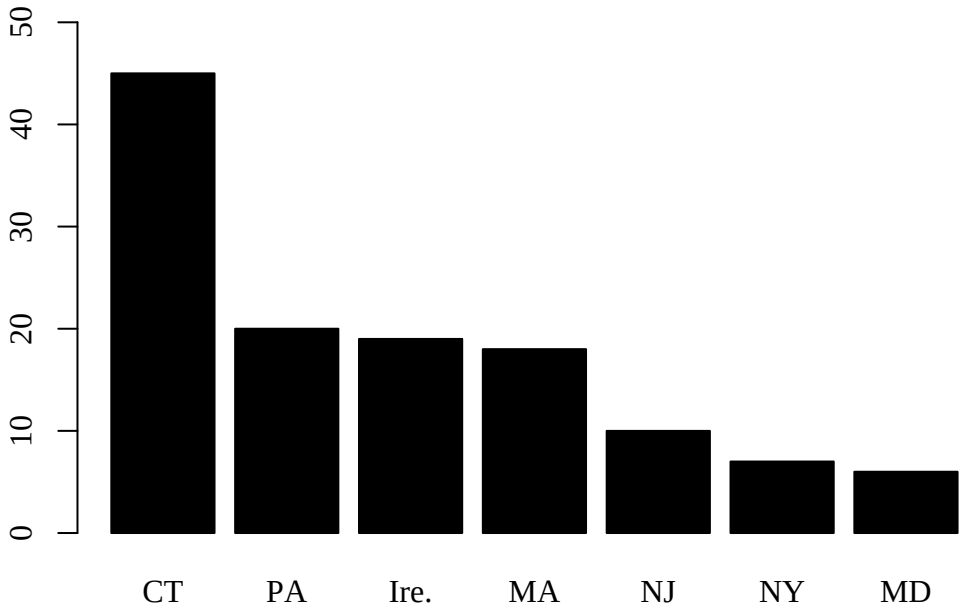

Figure 2: Place of birth of Revolutionary Patriots from Erie County

The first source is entitled *Revolutionary Patriots in Erie County, PA*, and was published by the Sons of the American Revolution as an aid in genealogy. The National Association of the Sons of the American Revolution defines a Revolutionary Patriot as an individual who has given "acceptable service to this nation". Examples of such service include (but are not limited to) such actions as:

signing the Declaration of Independence, serving in the Revolutionary army between April 19, 1775 and Nov 26, 1783, serving in the Continental Congress, etc. The Erie Chapter of the Sons of the American Revolution compiled a list of all known Revolutionary Patriots who resided in and were buried in Erie County. Out of the 212 such Patriots, the place of birth is known for 143. Figure 2 shows that almost one third of these early settlers came from the state of Connecticut.

When the data in Figure 2 are grouped into three representative regions (North, Mid-Atlantic, and Europe), it is clear that settlers from the North far outnumbered those from other areas. For these purposes, the North is defined as New England plus New York state, and the Mid-Atlantic region includes Pennsylvania, New Jersey, and Maryland. Grouped together this way, the majority of the Revolutionary Patriots came from the Northern states: 57.3% from the North vs. 25.2% from the Mid-Atlantic states and 17.5% from Europe. The average year of birth for all of these Revolutionary Patriots is 1754, and the average year of death is 1833. Thus, this group must have been among the very first settlers in the region, since Erie's first non-indigenous settler arrived in 1795.

A further source on the settlement of Erie County comes from a history of the county written in 1884, whose second volume contains biographies of 1,077 important residents of the county at the time of writing. Of these, 615 were born outside of Erie County, and their places of birth are presented in Figure 3 (excluding those locations that are represented by 10 or fewer settlers).

The data from this group of prominent residents represents a later stage of immigration than the Revolutionary Patriots data—the average year of birth for the residents considered in the 1884 book is 1825, a good three generations later than the earlier group. However, the proportion of Northern settlers to Mid-Atlantic and European ones remains quite similar: 60.1% to 19.9% to 13.0%, respectively. Again, the North contributes the majority of settlers, but a sizable portion also comes from Pennsylvania. The Northern settlers in this later group are dominated by those born in New York (as opposed to Connecticut for the Revolutionary Patriots). This reflects the massive westward migration from New England into New York state in the early / mid-nineteenth century.

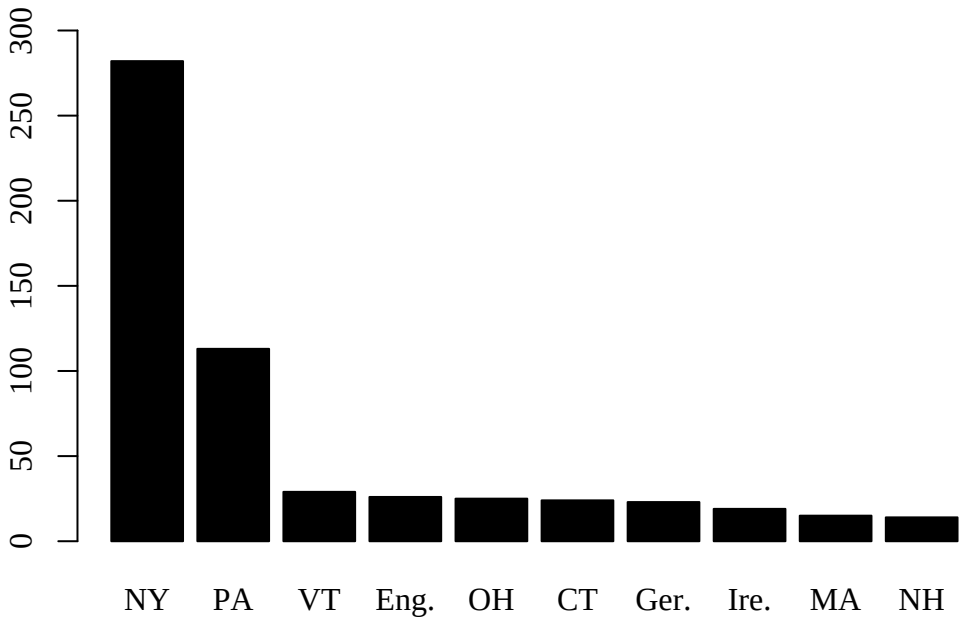

Figure 3: Place of birth of prominent 19th century Erie County residents

4 The merger of /o/ and /oh/ in Erie

As Section 2 showed, Erie patterned like a Northern city with respect to most lexical and phonological isoglosses investigated by the early linguistic atlas projects. However, more recent data show that Erie has lost its original Northern affiliation. The clearest diagnostic of this shift is the merger of /o/ and /oh/. The earliest study to document the merger of /o/ and /oh/ in Erie is Herold (1990). She conducted a telephone survey of all of the counties in Pennsylvania that were reported as distinct in PEAS and Wetmore (1959) in order to track the progress of the merger in the state. She interviewed one speaker from Erie County: a 63-year-old female from the city of Erie. Through a series of elicitations and minimal pair tests she concluded that this speaker had the merger of /o/ and /oh/. Subsequently, two female Erieites were interviewed in 1995 for the ANAE survey. At

the time, they were 31 and 39 years old, and both had a solid merger of /o/ and /oh/ in production. Finally, my own research so far includes 20 residents of Erie County, all who have a solid merger of /o/ and /oh/ in production and perception.

4.1 The unmerged LAMSAS informants

Based on these reports it seems clear that the merger of /o/ and /oh/ must have taken place in Erie sometime between the time when the youngest LAMSAS subject was born and sometime before the time when the oldest merged informant was born. In order to examine the chronology more closely, Table 11 presents the demographic information for the two LAMSAS speakers from Erie County.

Informant #	Township	Type	Sex	Age at Interview	Year of Birth	Year Interviewed
PA67a	Venango	IA	M	76	1864	1940
PA67b	Amity	IIB	M	37	1903	1940

Table 11: Demographic information for the two LAMSAS informants from Erie County

Also, it should be noted that both LAMSAS informants come from very rural communities, not from the city of Erie itself. This information will become relevant later in this section, after the apparent time data for the merger is presented. However, based on this data alone, the fact that the younger of the two LAMSAS informants from Erie County was born in 1903 would seem to indicate a time around 1910 as the *terminus post quem* for the merger of /o/ and /oh/ in Erie.

4.2 Sun Valley residents

In order to test the hypothesis that the merger of /o/ and /oh/ occurred in Erie sometime after 1910, an effort was made to record elderly Erieites with the hope of finding some who were born before the merger took place. If present-day speakers with the /o/ $\sim$ /oh/ distinction could be found, it would enable us to pinpoint the date of the merger with a high degree of certainty.

I established contact with the director of an upscale retirement community in Erie, which I will call Sun Valley. She invited me into the community, and arranged one-on-one interviews with several of the residents who ranged in age from 66 to 95. The oldest was born in 1912, and would thus be a good candidate for maintaining the distinction, assuming the chronology in Section 4.1 is

correct.

However, it is clear from minimal pair tests and word list data that all of the native Erieites interviewed at Sun Valley have a complete merger between /o/ and /oh/. None of them had a difference in production of perception for any of the minimal pairs, and the word list data show almost total overlap between the two classes. Figure 4 shows /o/ and /oh/ for the oldest Sun Valley informant, Dan R., who was born in 1912.

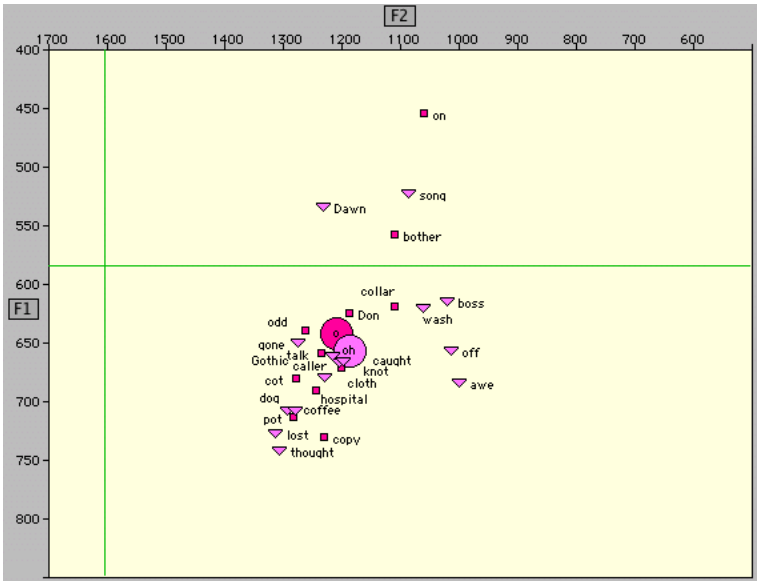

Figure 4: /o/ and /oh/ from Dan R., born 1912 in Erie

Table 12 displays the F1 and F2 differences for /o/ and /oh/, as well as the Euclidean distances between the two vowels, for all of the Sun Valley residents from Erie. The means are all quite close—no speaker has a difference greater than 50 Hz in both F1 and F2, and no speaker has a difference greater than 100 Hz in either domain. Furthermore, the vowel plots for all speakers are similar to Dan R.'s in Figure 4, and show that the clouds for the two classes overlap.

As a comparison point, Table 13 shows the same values for the one non-merged Sun Valley resident, Walter K., who is from Buffalo. The Euclidean distance between his /o/ and /oh/ means is more than 6 times larger than the average distance for the 9 merged Sun Valley residents, and more than 3 times larger than the individual merged speaker with the largest distance. The individual tokens from Walter's reading of the word list are plotted in Figure 5. This plot shows complete

Name	Year of Birth	Age	F1(/o/) – F1(/oh/)	F2(/o/) – F2(/oh/)	Dist(o, oh)
Dana W.	1941	66	9	17	19.2
Sally W.	1928	79	13	54	55.5
Dottie A.	1926	81	35	72	80.1
Eloise B.	1925	82	25	71	75.3
Charles B.	1925	82	98	-4	98.1
Mary D.	1919	88	40	21	45.2
Flora R.	1919	88	49	39	62.6
Robert E.	1916	91	-7	39	39.6
Dan R.	1912	95	-14	21	25.2
		Averages	28	36	55.6

Table 12: /o/ and /oh/ from word lists for 9 Sun Valley residents from Erie

separation between the two classes (except for the word *on*, which is closer to the /oh/ cloud—this likely indicates that *on* actually has the phoneme /oh/ for Walter K., not /o/).

Name	Year of Birth	Age	F1(/o/) – F1(/oh/)	F2(/o/) – F2(/oh/)	Dist(o, oh)
Walkter K.	1927	80	187	326	375.8

Table 13: /o/ and /oh/ from the word list for a Sun Valley resident from Buffalo

The clear evidence for the merger of /o/ and /oh/ among several Sun Valley residents aged 80 and above indicates a time around 1920 as the *terminus ante quem* for the merger of these two vowels in Erie. This evidence, along with the LAMSAS data presented in Section 2.2, would seem to indicate a short window in the second decade of the 20th century for its occurrence. In order to shed more light on this chronology, I attempted to find older recordings of Erieites born before the Sun Valley residents. These finding will be presented in the following two sections.

4.3 The Seasonal Workers in Viticulture Corpus

The Seasonal Workers in Viticulture (SWV) corpus was compiled as part of an oral history project conducted in 1988 to document the local grape growing and picking industry around North East, PA. The town of North East is located in the northeastern corner of Erie County, about 20 miles from downtown Erie, and directly across the state line from New York. The town has always been

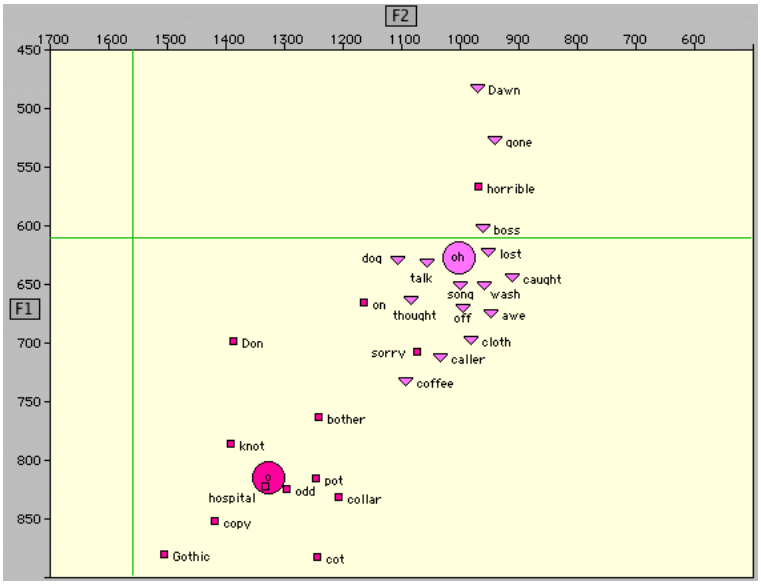

Figure 5: /o/ and /oh/ from Walter K., born 1927 in Buffalo

a center for grape production, focusing mainly on Concords for juice. A few other varieties are also grown, and several wineries exist along both sides of Route 20. Viticulture has always been the main source of jobs for North East residents.

The SWV project attempted to interview older native residents of North East who had owned vineyards or who had worked as grape pickers, although a few younger people and a few in-migrants were also interviewed. In all, 50 recordings were made, each about one hour in length. The interviews are available to the public as cassette tapes at the Erie County Historical Society.

The two oldest, native North East residents from the corpus were selected for analysis, since they were considered to be the most likely ones to still maintain a distinction between /o/ and /oh/. Their demographic data is provided in Table 14. Both of these informants grew up on farms in North East and lived in the town their entire lives. Their families owned grape farms, and they sold their grapes at local markets, and, later, to nearby plants for the production of grape juice.

Figures 6 and 7 show plots of the vowels /o/ and /oh/ for Richard O. and Benjamin S., respectively, and Table 15 displays the differences between the vowels in the F1 and F2 dimensions for each speaker. Richard O. has a small amount of overlap between the two classes, with *a lot*,

	Place of Birth	Year of Birth	Age at interview
Benjamin S.	North East, PA	1907	81
Richard O.	North East, PA	1906	82

Table 14: Demographic data for the two SWV speakers analyzed

operator, and *Concord* clearly falling in the /oh/ cloud and *thought*, *Catawbas*, and *crossroads* approaching the /o/ cloud. This distribution suggests that Richard O. maintained a distinction between /o/ and /oh/; however, the two phonemes are quite close, and are close to merging. The F2 difference between /o/ and /oh/ for Richard O. is only 154 Hz, compared to 326 Hz for the clearly unmerged Walter K. from Buffalo. This is the type of distribution that exists for many of the Midland ANAE speakers who are labeled as "transitional" with regard to the /o/ $\sim$ /oh/ merger.

Benjamin S. shows a much greater degree of overlap between the two classes, with several tokens from each class falling clearly within the cloud of the other class. Furthermore, the F2 distance between the means of the two classes is only 40 Hz. All of this evidence suggests that the merger of /o/ and /oh/ is quite advanced for Benjamin S., and has probably already reached completion.

Name	Year of Birth	Age	F1(/o/) – F1(/oh/)	F2(/o/) – F2(/oh/)	Dist(o, oh)
Richard O.	1906	82	56	154	163.9
Benjamin S.	1907	81	36	40	53.8

Table 15: /o/ and /oh/ from interviews for two SWV informants from North East, PA

4.4 H. O. Hirt

Another bit of early evidence comes from an interview conducted with H.O. Hirt, the founder of Erie Insurance Exchange. He was born in 1887, founded the company in 1925, and served as its CEO until 1976. An interview was conducted with him in 1977 and a 20-minute VHS version of this interview was released as a publicity tape by the Erie Insurance Group. This interview thus represents the oldest recorded Erieite that I have so far been able to discover. Table 16 summarizes his demographic information.

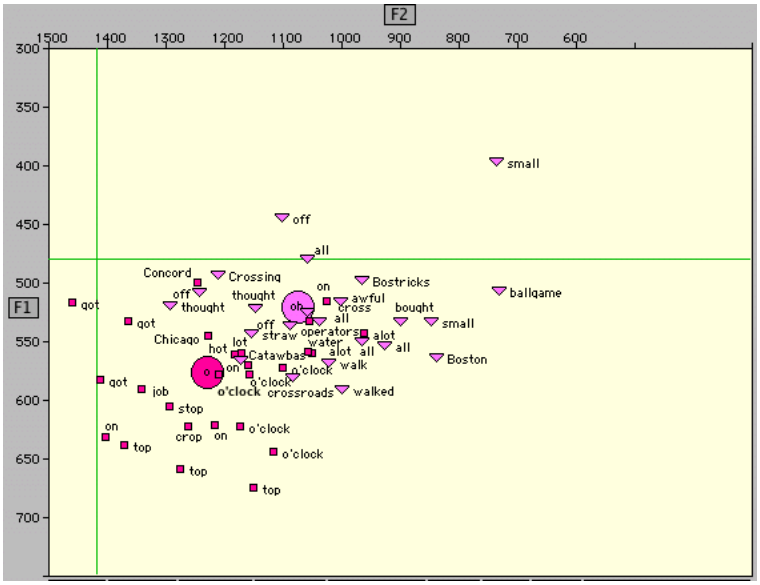

Figure 6: /o/ and /oh/ from Richard O., born 1906 in North East, from the SWV corpus

Name	Place of Birth	Year of Birth	Age at interview
H. O. Hirt	Erie, PA	1887	90

Table 16: Demographic data for H. O. Hirt

Figure 8 shows a plot of all tokens with /o/ and /oh/. As the figure shows, there is a large amount of overlap between /o/ and /oh/. Two tokens of *thought* and one of *office* are clearly within the /o/ distribution, and several /o/ words fall within the /oh/ cloud: *policy*, *dollars*, *modest*, *profit*, *pocket*, *top*, and *not*. Furthermore, several tokens from the /o/ class sound quite rounded, especially *top*, *policy*, *pocket*, and *dollars*.

Table 17 shows the differences between the means for the two vowel classes. Again, the difference in F1 is quite small (30 Hz), and the difference in F2 is only slightly larger than 100 Hz, much smaller than the normal distance for unmerged speakers.

Name	Year of Birth	Age	F1(/o/) – F1(/oh/)	F2(/o/) – F2(/oh/)	Dist(o, oh)
H. O. Hirt	1887	90	30	114	117.9

Table 17: /o/ and /oh/ from interview with H.O. Hirt, born 1887 in Erie

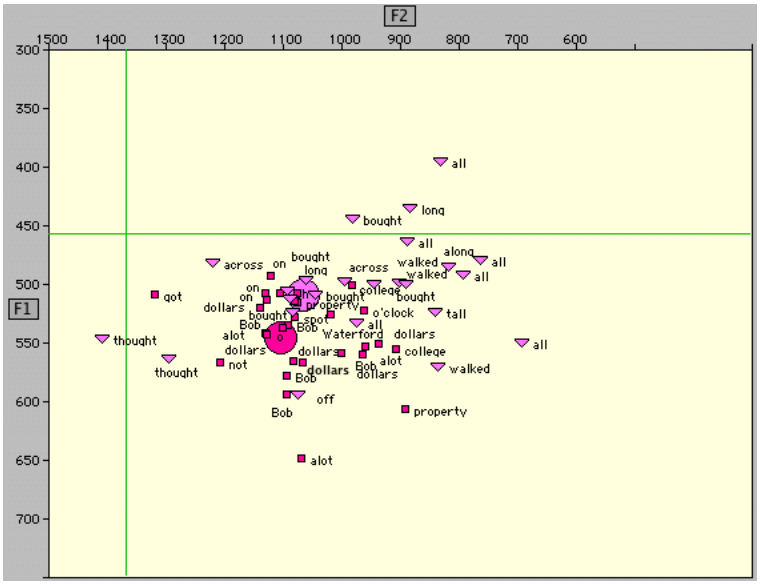

Figure 7: /o/ and /oh/ from Benjamin S., born 1907 in North East, from the SWV corpus

Finally, Table 18 compares /o/ and /oh/ for H. O. Hirt in four allophonic environments, as defined by the type of segment that follows it. The Table shows that the difference between /o/ and /oh/ when a stop follows (e.g. *not* vs. *thought*) and when a fricative follows (e.g. *profit* vs. *office*) is minimal, and indicates that the merger is near completion for those environments. The only allophonic environments that still maintain a difference between /o/ and /oh/ are before nasals (e.g. *honorable* vs. *long*) and laterals (e.g. *policy* vs. *called*), although the number of tokens before nasals is quite small, so it is hard to make and evaluation on those tokens alone. Furthermore, the difference between /ol/ and /ohl/ is not by itself probative of a distinction between the two classes, since the tokens that appear for each class are further differentiated by the fact that most of the /ohl/ tokens have /l/ in the coda position at the end of the word after /oh/ (such as *all*, *small*, and *Hall*, whereas the /o/ tokens before /l/ are mostly in open syllables (e.g. *policy*, *dollars*). Tokens like *all* would be expected to have a lower F2 even for completely merged speakers, due to the effect of the coda /l/ on the vocalic nucleus.

All of the evidence from H. O. Hirt indicates that the distinction between /o/ and /oh/ was tenuous even for him, and that the merger was already in progress in Erie when he was acquiring

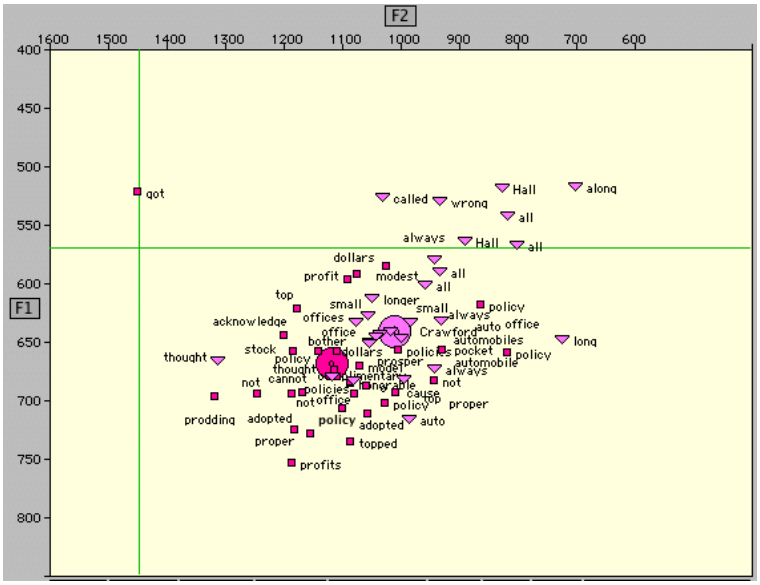

Figure 8: /o/ and /oh/ from H. O. Hirt, born 1887 in Erie

Following Segment	N(/o/)	N(/oh/)	F1(/o/) – F1(/oh/)	F2(/o/) – F2(/oh/)	Dist(o, oh)
stop	17	6	7	52	52.1
fricative	4	6	15	85	85.8
nasal	2	4	112	259	281.7
lateral	11	12	78	127	148.7

Table 18: /o/ and /oh/ for H. O. Hirt by allophonic environment

the language in the early 1890's. Again, this seems to contradict the LAMSAS evidence, but could be consistent with the evidence presented in Section 4.2 for the 9 elderly Erieites interviewed in 2007 at the retirement community and in Section 4.3 for the two elderly participants in the SWV project interviewed in 1988.

4.5 Explaining the LAMSAS data

All of the pieces of evidence presented above do not seem to present a coherent picture, especially the LAMSAS data, which seems to suggest that the merger must have taken place later in Erie County than it clearly did based on the evidence from the older recordings. However, so far the discussion has focused primarily on the temporal dimension of the merger, but has disregarded the

geographic dimension. If we consider how the merger must have progressed through both time *and* space, the picture becomes clearer.

The younger non-merged LAMSAS speaker who seems to provide evidence for a later date for the merger, PA 67b, was born in 1903. However, he was not from the city of Erie itself; rather he was born and raised in Amity township, a small farming community in the southeastern part of Erie County. On the other hand, the speakers who provide evidence for an earlier date for the merger are much more connected to the city than PA 67b: H.O. Hirt and the Sun Valley residents are all from the city of Erie itself, and the two SWV speakers are from North East. North East is only slightly closer to Erie than Amity in terms of distance, but is much more closely connected with Erie, since it is a larger community and a major road passes between Erie and North East.

So, if all of the temporal and geographic evidence is taken at face value, then it indicates that the merger first occurred in the city of Erie, and then spread gradually to the nearby townships in Erie County, first to the more populous ones, then, finally, to the smaller, more isolated ones, in accordance with the Cascade Model of Labov (2003). H.O. Hirt's data indicates that the merger probably took place in Erie already before the turn of the 20th century. The two SWV speakers indicate that it had spread to North East by around 1910. Finally LAMSAS speaker PA67b indicates that the merger had not yet reached Amity township by 1910.

If this chronology is correct, then it might still be possible to find non-merged speakers in the small, rural areas of Erie County that the merger spread to last. In order to investigate this, I plan to interview elderly residents of farming townships in Erie County such as Amity. Furthermore, I hope to obtain archival audio recordings from other speakers who were born around the time of H. O. Hirt or earlier, in order to provide evidence for the origins of the merger in the city of Erie.

4.6 The Current Geographic Extent of the Merger around Erie

Figures 9 and 10 show the geographic extent of the merger of /o/ and /oh/ for 17 speakers from Erie County and one from Warren County in Pennsylvania and 9 speakers from Chautauqua County in New York. The maps show that the isogloss seems to coincide pretty well with the state boundary between Pennsylvania and New York. The town of Ripley, NY, appears to be a transition area which

warrants more detailed study, particularly of younger speakers.

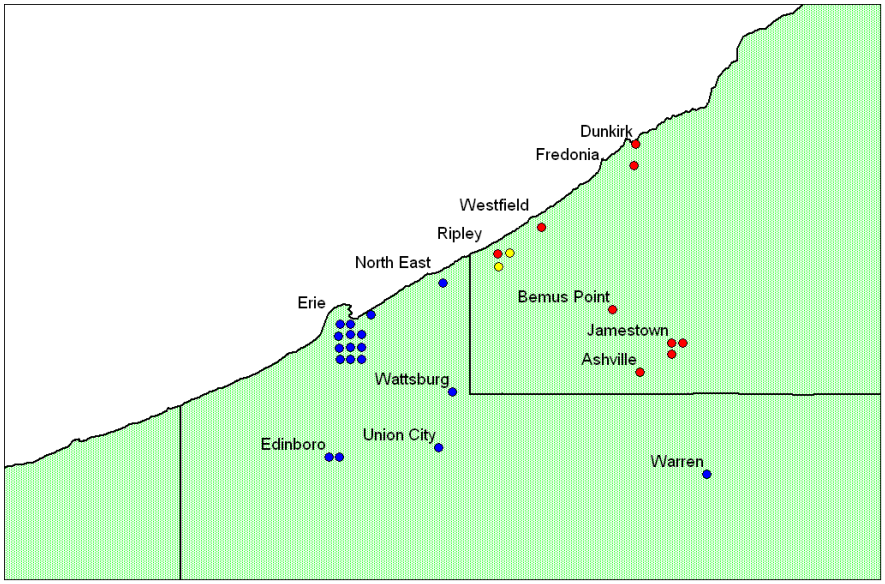

Figure 9: Geographic extent of the merger of *cot* and *caught* around Erie; blue = merged, yellow = close, red = distinct

A 55-year-old male speaker from Ripley produced tokens of *cot* and *caught* in a minimal pair test that were only slightly different, and judged them to be identical. This discrepancy is not surprising, since categorization normally precedes pronunciation in cases of merger in progress (Labov 1994:319). Furthermore, the same speaker was merged for *Don* and *Dawn* in both production and perception. Additionally, a 44-year-old female produced nearly indistinguishable tokens of *cot* and *caught*, although she still judged them to be different. I also ran the minimal pair test with three waitresses from Ripley who didn't have time to be recorded. The youngest one was 36 years old, and she produced both pairs identically, and also judged them to be the same. The two older waitresses (mid-40's and mid-50's), however, produced them distinctly. Finally, several high school seniors that I met at Ripley Central School produced identical tokens of *cot* and *caught*.

If these findings are correct, they would indicate that the merger was in progress in Ripley 30 to 40 years ago, and has already gone to completion for the younger two generations. It will be necessary to interview more teenagers and younger adults, as well as middle-aged adults from Ripley in order to provide apparent time evidence for the spread of the merger. Furthermore, I will

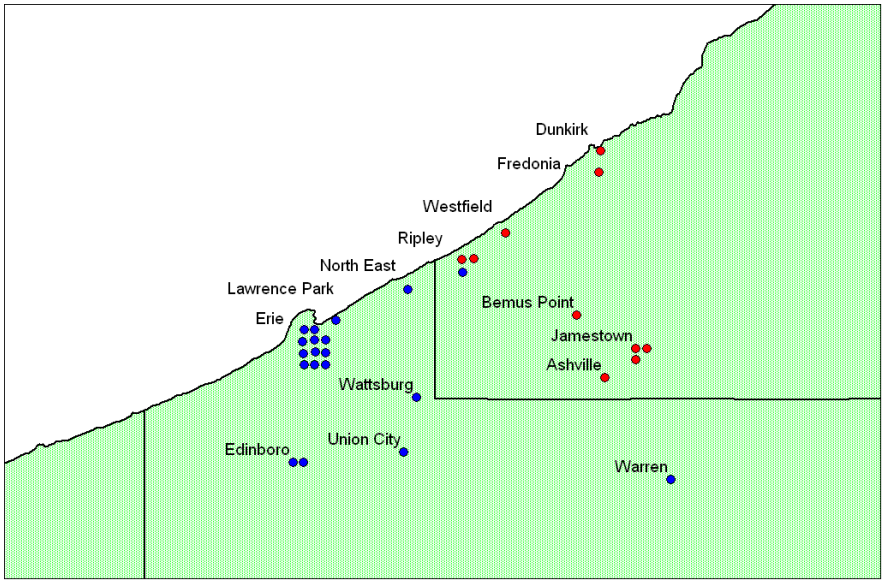

Figure 10: Geographic extent of the merger of *Don* and *Dawn* around Erie; blue = merged, red = distinct

attempt to interview younger speakers in Westfield, NY, the next town towards Buffalo after Ripley. The three speakers that I have interviewed so far from Westfield are all clearly distinct; however, the youngest one is only 48. If the merger is indeed spreading away from Erie, it is possible that younger speakers in Westfield will show signs of it.

5 Erie and the Midland

5.1 Phonology

5.1.1 /æ/ and /o/ ∼ /oh/

As Section 4 showed, the merger of /o/ and /oh/ has been complete in the city of Erie for at least a century. This fact alone is sufficient evidence to demonstrate that Erie is no longer part of the North, since the distinction is maintained throughout that entire region. It also provides good evidence of Erie's phonology now being aligned with the Midland, since the merger of /o/ and /oh/ is in progress or has gone to completion throughout most of the Midland (Labov et al. 2006:264).

A further Midland characteristic that Erie shares is the distribution of /æ/. Most Midland speakers have words with /æ/ before nasals much higher and further front than before other coda consonants. ANAE shows that the F1 distance between /æ/ before nasals and before /t/ is at least between 100 and 200 Hz for nearly all Midland speakers, and is greater than 200 Hz for many (Labov et al. 2006:176). Figure 11 shows a typical distribution for the Midland from a Pittsburgh speaker from ANAE.

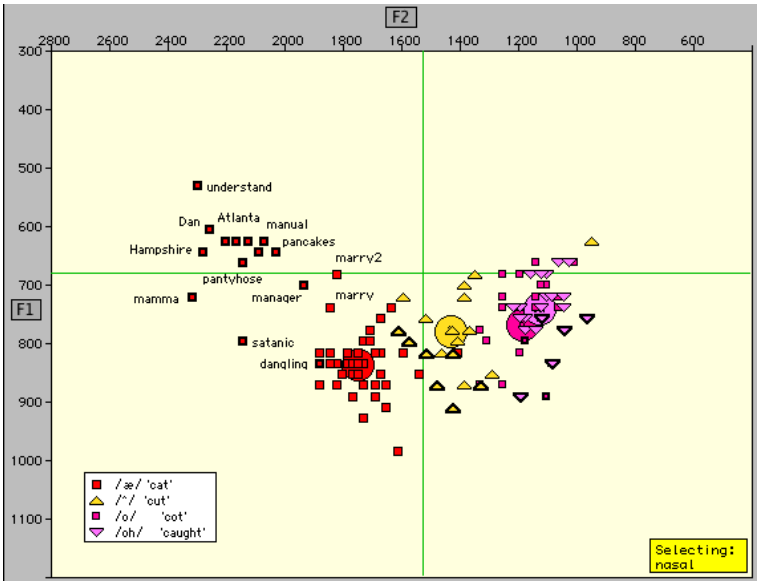

Figure 11: Nasal /æ/ system for Telsur speaker 545, Keith Koch, 35, from Pittsburgh, PA

In Figure 11, all tokens of /æ/ before nasals are raised and fronted, in both closed and open syllables. No other tokens with /æ/ approach the cloud before nasals, except for one instance of the word *jacket*.

Figures 12 and 13 show the nasal /æ/ system for two Erie residents, Sally L. and Tom L. (who is Sally's father). Sally has an extreme version of the nasal system, with /æ/ over 400 Hz higher before nasals and over 600 Hz fronter (excluding the two tokens before /ŋ/, which are intermediate). Her father, Tom, also has a clear nasal pattern, with almost no overlap between the two distributions. Sally's distribution of /o/ and /oh/ is interesting in that several of the tokens before nasals (*Dawn*, *gone*, *strong*, and *on*) are separated from the cloud of merged tokens by over 100 Hz in the F1 dimension.

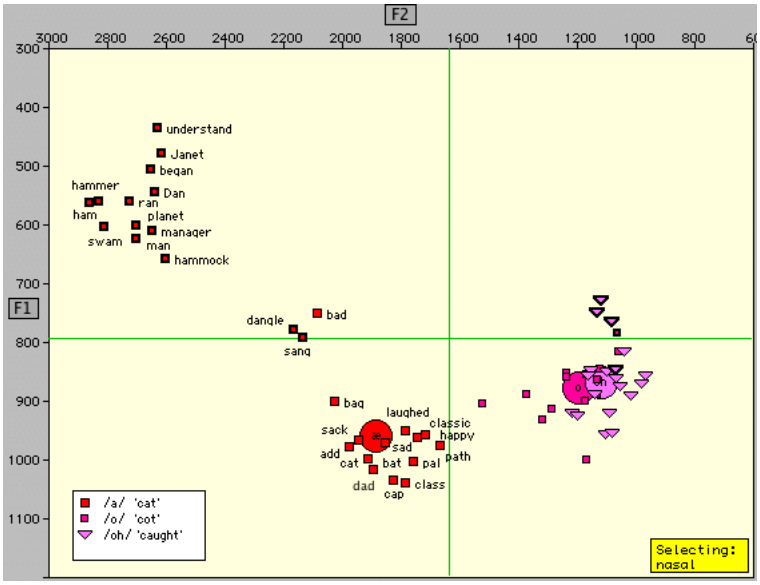

Figure 12: Nasal /æ/ system for Sally L., 25 [2006], from Erie, PA; word list tokens only

All Erie residents interviewed to date have shown a nasal /æ/ system with raised tokens before nasals in both closed and open syllables, and low front tokens before non-nasals. Some variations exist with respect to the treatment of /æ/ before /d/ and /ŋ/, but the basic pattern is clearly Midland, as opposed to Northern.

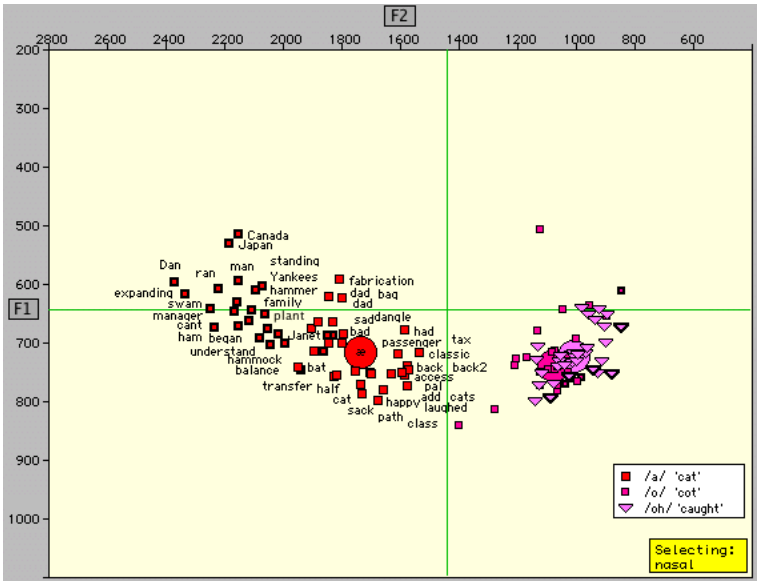

Figure 13: Nasal /æ/ system for Tom L., 53 [2006], from Erie, PA; word list and interview tokens

5.1.2 /ow/

While Erie's vowel system does share the two defining phonological characterists of the Midland (the merger of /o/ and /oh/ and a nasal /æ/ system), it is clearly not a prototypical Midland city. This can be seen in the behavior of the back upgliding vowels: /aw/ and /ow/. The Erie residents that I have interviewed so far have shown little, if any, fronting of the nuclei of these two vowels, and thus pattern with speakers from the North in this case.

The difference between Erie and the Midland in the vowel /ow/ is especially marked. The fronting of /ow/ is quite advanced in the Midland, a feature it shares with the Southeastern super-region, setting it apart from the neighboring West and North regions. Nearly all of the ANAE speakers from the Midland have a mean F2 value greater than 1200 Hz, and several have extreme fronting of /ow/ past 1550 Hz, the midpoint of the normalized F2 distribution (Labov et al. 2006:265). Figure 14 displays the mean F2 values for /ow/ from word list tokens for 22 speakers from around Erie.

The speaker labels contain the Federal Information Processing Standard (FIPS) state code followed by an underscore, followed by the FIPS county code, followed by an underscore, followed by

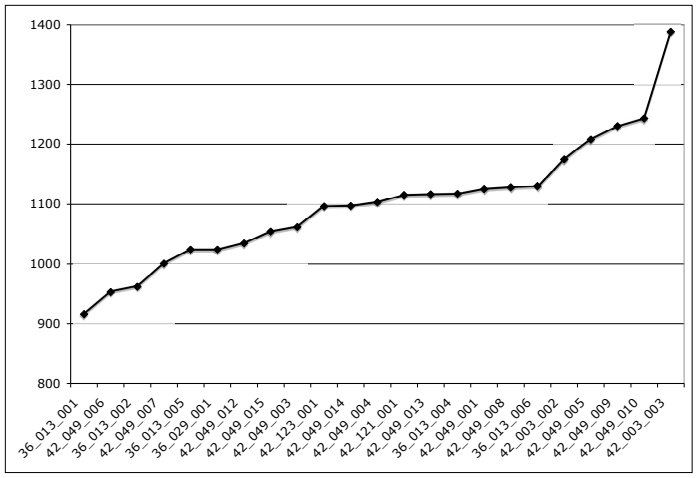

Figure 14: F2 means for /ow/ from 13 word list tokens for 22 speakers

a number that represents the order of recording. Thus, the first speaker in Figure 14, 36_013_001, was the first speaker recorded from Chautauqua County (FIPS county code 13) in New York State (FIPS state code 36). Most of the speakers in Figure 14 come from either Chautauqua County, NY, (36_013) or Erie County, PA, (42_049). In general, these speakers have very little, if any, fronting of /ow/. There are only three speakers from Erie County who have an F2 of /ow/ greater than 1200—two of these speakers, 42_049_009 and 42_049_010, come from Edinboro, a city near the southern edge of Erie County. The speaker with the most fronting of /ow/, 42_003_003, who has a mean F2 of 1388, is an 88-year-old woman from Pittsburgh. The only other Pittsburgh speaker in this dataset, 42_003_002, has more fronting of /ow/ than most of the Erie speakers, with an F2 mean of 1175 Hz.

There is thus a dialect boundary between Erie and the Midland area to the south of it that has strong fronting of /ow/, although this boundary is not as sharp as the one between Erie and the North. A boundary must also exist between the area of Northwestern Pennsylvania around Erie and the area of Western Pennsylvania around Pittsburgh which is characterized by the Pittsburgh Shift and the monophthongization of /aw/. One of the goals of this dissertation will be to determine the location of this boundary between Erie and the areas to the south of it, and to compare it qualitatively to the

boundary between Erie and the North, e.g., according to the taxonomy of dialect boundary areas set forth by Dinkin (2007).

5.2 Morphosyntax

Section 5.1 argued that Erie is a Midland city on the basis of the two most important features of its vowel system: the merger of /o/ and /oh/, and the nasal short-a system. In addition, Erie is also aligned with the Midland with respect to two morphosyntactic variables: positive *anymore* and *need* + Past Participle.

5.2.1 Positive *anymore*

The use of *anymore* without an accompanying negative or question marker is widespread throughout the United States, and is a feature of most of the Midland region. Precise isoglosses for its geographic distribution throughout the country have not been determined, however, despite numerous studies. Two factors contribute to the lack of precision in our knowledge of the extent of its use. First of all, positive *anymore* is relatively infrequent in normal speech. Thus, targeted formal methods are required to elicit judgments about its use. However, introspective judgements about positive *anymore* have been consistently shown to disagree with actual usage (Labov 1973). Thus, any data obtained about positive *anymore* must be treated with caution.

Individual studies have shown positive *anymore* to be in common use in specific areas of the Midland, such as Missouri (Youmans 1986), Southeastern Pennsylvania (Shields 1997), etc. Furthermore, two large-scale surveys have investigated positive *anymore*, and their results show that its area of acceptance overlaps considerably with the Midland. The isogloss in Labov et al. (2006:294) extends as far east as Philadelphia and as far west as Idaho, but stays south of the North / Midland boundary. The evidence from DARE (Cassidy and Hall 1985–2002) also shows that positive *anymore* is concentrated most heavily in the Midland, showing especially high rates of use in Kentucky, Indiana, and West Virginia.

The fact that the use of positive *anymore* is strongest in the Midland corresponds well with the theory that it originated in the speech of Scots-Irish immigrants (Crozier 1984, Montgomery 2004),

since the Midland was the main area of Scots-Irish settlement. A Scots-Irish origin would also help to explain why positive *anymore* appears to be stronger in the South Midland, and does not extend as far north as the boundary between the North and the Midland in ANAE.

Preliminary data collected in Erie indicate that positive *anymore* is a feature of the Erie dialect. The two ANAE speakers from Erie both indicate that they thought they had heard the construction in the area, but that they didn't use it themselves. During my initial fieldwork in the area, I have observed six instances of positive *anymore* in natural conversation from five different speakers (examples 1 and 2 were uttered by the same speaker):

(1) That's the world's excuse to do anything anymore.

(2) A: It's amazing how much equipment kids need.
B: Oh, *anymore*.

(3) The way they strap these toys in anymore.

(4) Any little town I go to anymore has a local espresso place.

(5) I'm sure you've probably gone to a GNC or a health foods store or even a grocery store anymore and I mean it's crazy looking at all that stuff.

(6) It's so hard the way we build things anymore.

In order to investigate the status of positive *anymore* in Erie and the neighboring region that forms the boundary with the North, I propose to incorporate a survey that elicits speakers' judgments about the construction into my interviews. The survey will use the methodology employed by ANAE (Labov et al. 2006:293) for this construction, namely, a three-point forced choice scale. The speakers will be presented with the following three sentences:

(7) Ticket prices are so high anymore, I never go to the movie theater.

(8) Anymore, there's too much crime in this neighborhood.

(9) John eats fast food so much anymore, it's no wonder that he's becoming overweight.

They will be asked to rate them as either 1) "I could say a sentence like this." 2) "I wouldn't say this, but I've heard people around here say something like it." or 3) "I've never heard anything like this before—it sounds like bad English." Examples 7 and 8 express complaints, which (Labov et al. 2006:293) have argued to be the most natural pragmatic context for positive *anymore*. Example 8

has preposed positive *anymore*, which is generally judged to be less acceptable. The survey will also include Example 10 as a control sentence with *anymore* in a negative context, which should be judged to be perfectly natural by all speakers. Refer to Appendix C for a list of all of the sentences used in the survey.

(10) I was a pitcher when I was young, but now I don't play baseball anymore.

5.2.2 *need* + Past Participle

A second grammatical feature that appears to have spread from the Midland up to Erie is the use of *need* + Past Participle (V-*en*), as in *The car needs washed*. This use contrasts with the use of *need* + Present Participle (V-*ing*) in other dialect regions, as in *The car needs washing*. Regional studies have demonstrated that the use of *need* + V-*en* is widespread throughout the Midland and limited in other regions (Murray et al. 1996). ANAE shows that the geographical range of *need* + V-*en* is smaller than that of positive *anymore*, and the isogloss of the former is almost wholly surrounded by the isogloss of the latter (Labov et al. 2006:295).

As is the case for positive *anymore* (Labov 1972:309), the alternation between *need* + V-*en* and *need* + V-*ing* operates below the level of consciousness for most speakers (Murray et al. 1996), and can thus be appropriately studied with a written questionnaire. I propose to track the northern extent of the use of *need* + V-*en* by including a forced-choice sentence completion task between *need* + V-*en* and *need* + V-*ing* in my survey, as in Examples 11 and 12:

(11) I drove through a big, muddy puddle yesterday. Now my car needs
washed.
washing.

(12) I haven't cleaned my kitchen in weeks. The floor really needs
mopped.
mopping.

Of the two ANAE speakers from Erie, one reported that she herself uses *need* + V-*en*, and the other reported that people in the area use it. My own initial studies of the feature in Erie show that Erieites prefer *need* + V-*en* to *need* + V-*ing* slightly more than residents of Chautauqua County, NY, as shown in Table 5.2.2.

	Now my car needs...		The floor really needs...	
	washed	*washing*	*mopped*	*mopping*
Erie Co. and nearby counties in Western PA	9	9	9	8
Chautauqua Co., NY	2	8	4	6

Table 19: Preliminary results for *need* + V-*en* vs. *need* + V-*ing* around Erie

In addition to the use of *need* + V-*en*, Murray and Simon (1999) and Murray and Simon (2002) have shown that a similar use of the Past Participle exists with the verbs *want* and *like*. These uses are also confined to the Midland region, and are both substantially more restricted than the use of *need* + V-*en*. Their research into the three constructions shows an implicational scale of acceptability such that if a speaker accepts *like* + V-*en* she will also accept *want* + V-*en*; similarly, if she accepts *want* + V-*en*, she will also accept *need* + V-*en*. Their maps for want + V-*en* and like + V-*en* show a heavy concentration of positive attestations in Western Pennsylvania around the Pittsburgh area. However, few speakers from Erie so far have accepted these constructions (even though most of these speakers use *need* + V-*en*), which means that the northern boundary of the *want* + V-*en* and *like* + V-*en* isoglosses must lie between Erie and Pittsburgh. In order to determine more precisely to location of this boundary, my survey will also elicit acceptability judgments about Examples 13 and 14:

(13) My cat looks really hungry. I think he wants fed.

(14) Every newborn baby likes cuddled.

For a complete version of the survey with the sentences eliciting judgments on *need*, *want*, and *like* + V-*en*, refer to Appendix C.

6 Methodology

6.1 Data Collection

The fieldwork for this dissertation will examine the three dialect boundaries surrounding Erie, Pennsylvania. On the one hand, the boundary between Erie and the North will be examined between Erie and Buffalo and between Erie and Cleveland. On the other hand, the boundary between Erie and the Midland / Western Pennsylvania will be examined between Erie and Pittsburgh. In order to determine the location of the boundaries, I will start out in Erie and head to the northeast towards Buffalo, to the west towards Cleveland, and to the south towards Pittsburgh. The region, along with the towns of interest for this dissertation are displayed in Figure 15.

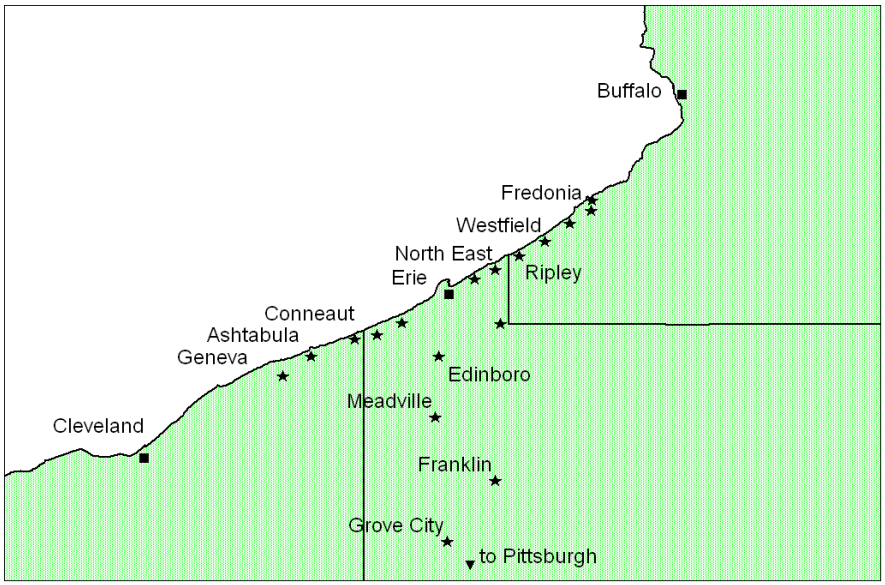

Figure 15: Towns in which exploratory interviews will be conducted

Pilot recordings will be made in each of the towns in Figure 15, modeled on the "Short Sociolinguistic Encounters" described in Ash (2002); thus, a targeted effort will be made to record the greatest amount of relevant linguistic and demographic information in the shortest time possible. The purpose of these interviews will be to determine the dialectological affiliation of the town, in order to pin down the location of the boundaries between the regions.

These interviews will also be designed to elicit speech samples from a wide variety of stylistic

contexts, including minimal pair tests (see Appendix A), a word list (Appendix B), and a short interview. Speakers will also be asked to read the sentences in the written survey in Appendix C out loud, which will provide samples of read connected speech.

The collection of data from several different stylistic contexts is crucial in determining the full extent of the spread of linguistic change to a new area. In particular, in the case of the merger of /o/ and /oh/ several studies have demonstrated that some speakers consistently produce words from the two classes distinctly in unmonitored conversational speech, but produce them quite close or identically in minimal pair tests. These speakers also judge the minimal pairs to be identical (Labov et al. 1972:235), (Herold 1990:182–186), (DiPaolo 1992). Such speakers appear to be sensitive to the incoming merged norm in the community, and are thus able to reproduce it in their most conscious style of speech. These speakers are usually older speakers living in communities in which the merger has already gone to completion for the youngest generations. Thus, another way to provide a more complete characterization of each town from the initial short interviews will be to record speakers from different generations—ideally, I will record at least one young speaker and at least one old speaker in each town.

Once the towns at or near the dialect boundaries in each direction are reached, several additional speakers will be interviewed. In these towns, speakers from all age groups will be sought out, so that a complete apparent time picture will show whether the boundary is stable or in flux.

6.2 Data Analysis

Once the interviews have been collected, they will be transcribed. These transcriptions will then be used as a reference during *forced alignment*, a process in which the combination of the audio file and the transcription are compared to a pronunciation dictionary and acoustic models trained on a separate corpus of American English. The forced alignment procedure uses Viterbi decoding to determine the most likely sequence of phonemes given the acoustic data (Jelinek 1976), implemented using the Hidden Markov Model Toolkit (Young et al. 2006). The output of the procedure is, for each interview, a set of starting and ending times for all of the words in the interview, and a set for all of the phonemes in each word. While the procedure is not foolproof—errors can be caused by

background noise, disfluencies, etc.—the results for all of the interviews I have transcribed so far have very few gross errors.

Once a corpus is created with time-stamps for all of the words and phonemes in each interview, a wider variety of phonetic data can be extracted than is normally done with sociolinguistic interviews, such as pitch, duration, formant contours, etc. Although the automatic processes of extracting this information will also introduce sources of error into the dataset, the benefit of the richer set of linguistic data will hopefully outweigh the fact that some tokens will need to be excluded due to erroneous measurements.

As an example of the type of results I plan to obtain through the process of forced alignment and automatic phonetic data extraction, I will present a case study of automatic F1 and F2 extraction. For this study, the word list recordings from 26 speakers were coded by hand according to the methodology in ANAE (p. 37). The word list contained 160 words in random order. F1 and F2 were extracted at the vowel's midpoint for all 11 words containing the phoneme /o/. Table 20 compares the manual formant measurements to the automatic formant measurements for each word. The fourth and seventh columns in the table display the absolute value of the differences in the measurements.

word	manual F1	auto F1	diff	manual F2	auto F2	diff
tock	756	772	16	1492	1464	28
Don	720	697	23	1487	1459	28
cot	788	783	5	1550	1526	24
on	705	712	7	1341	1345	4
pot	744	744	0	1386	1418	32
hospital	772	772	0	1392	1401	9
collar	747	746	1	1323	1327	4
Gothic	720	736	16	1574	1597	23
copy	724	735	11	1541	1531	10
odd	745	752	7	1385	1409	24
bother	750	764	14	1377	1395	18
	743	747	4	1441	1443	2

Table 20: Manually vs. automatically extracted F1 and F2 measurements for all words with /o/ for speaker 36_013_001, a 77-year-old male from Dunkirk, NY

The automatic formant extraction procedure worked quite well for this speaker—the largest

difference between the manual and the automatic measurements is 32 Hz. These values are all within the range of error that would be expected between two sets of formant values manually extracted by different people. The final row in Table 20 shows that the means produced by the two processes are nearly identical, and a *t*-test confirms that the differences are not statistically significant.

The formant extraction procedure does not perform as well for all speakers, though. A frequent source of error comes from missing one of the first two formants. For example, one speaker (42_049_015, an 82-year-old male from Erie) had an average F2 for /o/ by manual extraction of 1010 Hz, whereas the average by automatic extraction was 1496. These distributions are statistically different (p = 0.034). This difference in the means is caused by tokens such as 'Don', which had manual F1 and F2 values of 798 and 955 Hz, respectively, but automatically extracted values of 991 and 2891 Hz. Cases like this occur because F1 and F2 for the vowel /o/ are so close to each other for this speaker that the normal settings (10 poles) do not separate them. In order to measure such vowels accurately, it is necessary to manually increase the number of poles. However, such cases of extreme mis-measurement can be detected easily automatically, since they will fall well outside of the expected distribution for that formant. For this study, since there are hand measurements for comparison, I simply excluded all tokens for which the automatic measurement differed by more than 300 Hz. Table 21 shows the mean values from the two methods for all 26 speakers, excluding the tokens with extreme mis-measurements. Again, most of the differences between the mean values are quite small (less than 50 Hz), and there is only a single case of a difference larger than 100 Hz. Also, *t*-tests show that only one of the differences between the manual and the automatic measurements is different at the 0.05 level: the F1 means for speaker 42_003_003 (p = 0.049).

Table 21 demonstrates that the procedure of forced alignment combined with automatic formant extraction can be comparable to manual formant extraction. However, since the vowel /o/ is normally a steady-state vowel (except for the onset and coda consonant transitions), this task was much simpler than it would be for vowels that are characterized by extreme formant transitions, such as a heavily fronted Midland /ow/, or a raised, in-gliding /æ/ in the North. For these cases, I will also attempt to implement automatic formant extraction algorithms that produce measurements that are

speaker	avg. manual F1	avg. auto F1	diff	avg. manual F2	avg. auto F2	diff
36_013_001	743	747	4	1441	1443	2
36_013_002	861	837	24	1590	1546	43
36_013_004	984	897	87	1376	1319	58
36_013_005	809	801	8	1252	1253	1
36_013_006	703	701	2	1203	1217	14
36_013_007	626	664	38	1166	1219	54
36_013_008	700	714	14	1178	1212	34
36_029_001	732	788	56	1159	1190	31
36_059_001	870	856	13	1312	1303	9
42_003_002	829	829	0	1208	1264	56
42_003_003	706	749	43	1191	1250	58
42_049_001	657	686	28	1060	1096	35
42_049_003	812	838	26	1191	1217	26
42_049_004	698	727	30	1215	1186	29
42_049_005	884	840	44	1304	1343	39
42_049_006	816	832	16	1198	1243	45
42_049_007	766	790	24	1099	1117	18
42_049_008	594	603	8	1143	1157	14
42_049_009	774	785	11	1210	1244	33
42_049_010	704	757	52	1197	1268	71
42_049_012	755	752	3	1300	1275	25
42_049_013	552	552	1	1021	1046	25
42_049_014	776	792	16	1130	1151	21
42_049_015	706	815	109	1046	1117	71
42_121_001	722	752	30	1141	1156	15
42_123_001	754	768	15	1105	1137	32

Table 21: Manually vs. automatically extracted F1 and F2 means for 26 speakers for the vowel /o/, 11 word list tokens

not significantly different from the manual ones, as in the case of /o/ above. In the event that any of these fail, I will simply revert to measuring those vowels by hand.

References

Ash, Sharon. 2002. The distribution of a phonemic split in the Mid-Atlantic region: Yet more on short A. *Penn Working Papers in Linguistics* 8:1–15.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Carver, Craig M. 1987. *American regional dialects: A word geography*. Ann Arbor: University of Michigan Press.

Cassidy, Frederic G., and Joan Houston Hall. 1985–2002. *Dictionary of American Regional English*. Cambridge, MA: Harvard University Press, Belknap.

Chambers, J.K., and Peter Trudgill. 1999. *Dialectology (2nd Edition)*. Cambridge University Press.

Crozier, Alan. 1984. The Scotch-Irish influence on American English. *American Speech* 59:310–331.

Dinkin, Aaron. 2007. Boundary communities in the dialectology of New York state. Unpublished manuscript, University of Pennsylvania.

Dinkin, Aaron. 2008. Settlement patterns and the Eastern boundary of the Northern Cities Shift. Paper presented at the American Dialect Society annual meeting, Chicago, January 2008.

DiPaolo, Marianna. 1992. Hypercorrection in response to the apparent merger of (ɔ) and (a) in Utah English. *Language and Communication* 12:267–292.

FWP, Federal Writers' Project. 1938. *Erie: A guide to the city and county*. Philadelphia, William Penn Association: WPA.

Herold, Ruth. 1990. Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania. Doctoral Dissertation, University of Pennsylvania.

Jelinek, Frederick. 1976. Continuous speech recognition by statistical methods. *IEEE* 64:532–556.

Johnson, Daniel Ezra. 2007. Stability and change along a dialect boundary: The low vowels of Southeastern New England. Doctoral Dissertation, University of Pennsylvania.

Kurath, Hans. 1949. *A word geography of the Eastern United States*. Ann Arbor: University of Michigan Press.

Kurath, Hans, and Raven I. McDavid. 1961. *The pronunciation of English in the Atlantic states*. Ann Arbor: University of Michigan Press.

Labov, William. 1972. *Sociolingusitic Patterns*. Phialdelphia: University of Pennsylvania Press.

Labov, William. 1973. Where do grammars stop? In *Report on the twenty-third annual round table meeting on linguistics and language studies*, ed. Roger Shuy, 43–88. Georgetown University Press.

Labov, William. 1974. Linguistic change as a form of communication. In *Human communication: Theoretical explorations*, ed. A. Silverstein, 221–256. Hillsdale, N.J.: Erlbaum.

Labov, William. 1991. The three dialects of English. In *New ways of analyzing sound change*, ed. Penelope Eckert, 1–44. New York: Academic Press.

Labov, William. 1994. *Principles of linguistic change, vol. I: Internal factors*. Oxford: Basil Blackwell.

Labov, William. 2003. Pursuing the cascade model. In *Social dialectology*, ed. David Britain and Jenny Cheshire, 9–22. John Benjamins.

Labov, William. 2007. Transmission and diffusion. *Language* 83:344–487.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English*. Mouton de Gruyter.

Labov, William, Malcah Yaeger, and Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Montgomery, Michael. 2004. Solving Kurath's puzzle: Establishing the antecedents of the American Midland dialect region. In *Legacies of colonial English: Studies in transported dialects*, ed. Raymond Hickey, 311–325. Cambridge University Press.

Mufwene, Salikoko S. 1996. The Founder Principle in creole genesis. *Diachronica* 13:83–134.

Murray, Thomas E., Timothy C. Frazer, and Beth Lee Simon. 1996. Need + past participle in American English. *American Speech* 71:255–271.

Murray, Thomas E., and Beth Lee Simon. 1999. Want + past participle in American English. *American Speech* 74:140–164.

Murray, Thomas E., and Beth Lee Simon. 2002. At the intersection of regional and social dialects: The case of like + past participle in American English. *American Speech* 77:32–69.

Sanford, Laura G. 1894. *The History of Erie County, Pennsylvania, from its first settlement, revised and enlarged edition*. Erie, PA: published by the author.

Shields, Kenneth. 1997. Positive *anymore* in Southeastern Pennsylvania. *American Speech* 72:217–220.

Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3:215–246.

Wetmore, Thomas H. 1959. *The low-central and low-back vowels of the English of the Eastern United States*. Tuscaloosa, AL: American Dialect Society.

Youmans, Gilbert. 1986. Any more on anymore?: Evidence from a Missouri dialect survey. *American Speech* 61:61–75.

Young, Steve, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, and Phil Woodland. 2006. *The HTK Book*. Cambridge University Engineering Department. URL `http://htk.eng.cam.ac.uk/docs/docs.shtml`.

Zelinsky, Wilbur. 1973. *The cultural geography of the United States*. Prentice Hall.

A List of Minimal Pairs Tested

Minimal Pairs

pin vs. pen

hoarse vs. horse

cot vs. caught

Mary vs. merry

merry vs. marry

fool vs. full

whale vs. wail

poor vs. pour

collar vs. caller

pool vs. pull

ferry vs. furry

don vs. dawn

which vs. witch

barn vs. born

stock vs. stalk

tour vs. tore

berry vs. bury

near Minimal Pairs

father vs. bother

nearer vs. mirror

spa vs. paw

on vs. Don

B Word List

hood
bag
here
news
today
dangle
toe
creek
duck
awe
knot
pen
Janet
goal
poor
huge
merry
mole
food
core
den
better
high
Oklahoma
lift
cot
toy
Ed
bitter
left
pot
cable
party
writer
witch
lost
both
bird
manager
sorry
pin

coffee
hammock
boat
man
height
Moe
go
found
ferry
hide
fool
soon
Sunday
bus
Dan
pal
bat
open
house
forty
tock
orange
hug
out
cloth
laughed
hay
spider
beer
Gothic
marry
don’t
had
boy
odd
downtown
gone
hurt
mother
Tuesday
spa

caught
Dawn
sad
Mark
send
talk
full
boss
on
Don
butter
now
planet
down
bother
sew
horrible
bee
understand
deck
sack
bad
farm
made
cut
wash
head
agree
bike
hoe
dog
bit
Mary
who’d
elementary
tire
up
collar
heed
ham
good

route
off
boot
how
put
four
sang
dad
began
hid
know
caller
class
hospital
pour
path
sin
fairy
ran
hope
which
month
hammer
dude
happy
roof
name
cap
bet
thought
classic
copy
home
tiger
father
song
rider

C Sentences for judgement elicitation task

Circle the word that sounds most natural in each sentence:

1) I drove through a big, muddy puddle yesterday. Now my car needs washed. / washing.

2) I haven't cleaned my kitchen in weeks. The floor really needs mopped. / mopping.

Rate the following sentences as 1 ("I could say a sentence like this."), 2 ("I wouldn't say this, but I've heard people around here say something like it.") or 3 ("I've never heard anything like this before—it sounds like bad English."):

1)	I was a pitcher when I was young, but now I don't play baseball anymore.	1	2	3
2)	My cat looks really hungry. I think he wants fed.	1	2	3
3)	Ticket prices are so high anymore, I never go to the movie theater.	1	2	3
4)	I really should redd up the living room before the guests come over.	1	2	3
5)	Anymore, there's too much crime in this neighborhood.	1	2	3
6)	Every newborn baby likes cuddled.	1	2	3
7)	I got into an accident last week, and now my front bumper needs repaired.	1	2	3
8)	John eats fast food so much anymore, it's no wonder that he's becoming overweight.	1	2	3