MORPHOPHONOLOGICAL VARIATION IN URBAN ASTURIAN SPANISH: LANGUAGE CONTACT AND REGIONAL IDENTITY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of The Ohio State University

By

Sonia Barnes

Graduate Program in Spanish and Portuguese

The Ohio State University

2013

Dissertation Committee:

Professor Scott A. Schwenter, Advisor
Professor Rebeka Campos-Astorkiza
Professor Terrell A. Morgan
Professor Donald Winford
ABSTRACT

Asturian, the regional language of Asturias (Spain), and Spanish have been in contact since the 14th Century. Some linguistic features of Asturian are frequently used in the Spanish of Asturian speakers (Sánchez Álvarez 1979, D’Andres 2001). Among these features are the raised vowels that appear in the masculine singular morpheme /-u/, as opposed to Spanish /-o/; and the feminine plural morpheme /-es/, as opposed to Spanish /-as/. The appearance of the Asturian variant in the urban varieties of the language is a consequence of the borrowing of the Asturian morpheme; however, speakers show variation in its use, alternating between the two options.

The first part of this dissertation quantitatively explores the linguistic and extra-linguistic factors that govern this alternation analyzing production data from 24 speakers from Gijón. Mixed linear and logistic regression models were developed in R to test the effects of the independent variables on the choice between Asturian and Spanish features. Vowel height was treated as both a categorical variable and a continuous one, measuring the frequencies of the first two formants. The results show that the selection of one vowel over the other is constrained by the adjacent phonological context and the variant used in the previous form. In the case of /-as/ and /-es/, it is also constrained by the lexical category of the token. The use of Asturian variants is more likely in the speech of speakers who work in commerce, in food and restaurant industries and in manufacturing. Speakers who have a lower educational level and male speakers are also more likely to
use /-u/ and /-es/. In the case of /-o/ vs. /-u/, the social tie between the speaker and their interlocutor also has a significant effect on the alternation, with Asturian /-u/ being more likely when there is a first order relationship between them.

The second part of this dissertation examines the role that Asturian vowels have in the construction of urban identity and test what social meaning is associated with each linguistic variant. Using a Matched Guise Technique (Lambert et al. 1960), 192 participants in a perception experiment rated speakers' utterances containing one of the variants (/-u/ or /-o/ and /-es/ or /-as/ respectively) with regard to various social characteristics. The results show that participants rate speakers who use Asturian variants as significantly less urban, less educated, less feminine and as having a lower economic status than those who use the Spanish one. These results confirm that the social correlations that were found to be significant in the quantitative analysis of production data are meaningful and used by listeners to evaluate the identity of other speakers. The analysis also revealed that /-u/ is a stronger marker of the social categories explored in this experiment than /-es/.

The results reported in this dissertation indicate that speakers strategically increase or decrease the amount of Asturian features according to how Asturian or “Gijonés” they want to sound, contributing to the idea that speakers actively use variation to position themselves in the social world.
To Drew, Eva and Elsie
ACKNOWLEDGMENTS

En primer lugar, me gustaría agradecer a mi tutor y mentor, el Dr. Scott Schwenter, su constante apoyo y sus enseñanzas. Gracias por despertar en mí el interés por la sociolingüística variacionista y por darme las herramientas necesarias no solo para completar esta tesis, sino para el desarrollo de mi carrera investigadora. Le agradezco inmensamente todos los comentarios y críticas que me ha proporcionado durante el proceso de escribir esta tesis y la confianza que siempre ha depositado en mi trabajo.

También estoy eternamente agradecida a la Dra. Rebeka Campos-Astorkiza, por su dedicación, sus conocimientos y su excepcional ojo para los detalles. Sus comentarios e ideas han sido absolutamente esenciales en el desarrollo de esta tesis y le agradezco todo el tiempo que le ha dedicado a este proyecto. Gracias al Dr. Terrell Morgan por sus consejos en lo personal y lo profesional, por su sabiduría y por todas las observaciones que sin duda han mejorado la calidad de este trabajo.

I would also like to thank Dr. Don Winford, who, from the very beginning of my degree, helped me develop the ideas that would eventually result in this dissertation. I am grateful to Dr. Winford for sharing his knowledge of sociolinguistics and language contact with me, for his valuable advice and for always encouraging me to go a step further in my research.
I also thank the faculty in the Linguistics Department at OSU, who have, in one way or another, helped me in the development and completion of this project: Beth Hume, Cynthia Clopper and Kathryn Campbell-Kibler. I would also like to thank the graduate students in the Linguistics Department for their assistance during different stages of this dissertation, especially Katie Carmichael, Marivic Lesho and Rory Turnbull.

The collection of the data used in this dissertation took place during two trips to Asturias, which were made possible by different sources of funding. The first trip was funded by a Small Research Grant from The Ohio State University Graduate School and the Mario Iglesias Award from the OSU Department of Spanish and Portuguese. The second trip to Asturias was funded by an International Affairs Grant from the OSU Office of International Affairs.

También me gustaría darles las gracias a todos los familiares, amigos y conocidos de Gijón que participaron y/o me ayudaron a reclutar más participantes para los dos estudios. Sin su ayuda jamás habría podido conseguir todos los datos en los que se basa esta tesis, así que les estoy eternamente agradecida.

A continuación, me gustaría reconocer el apoyo de todos mis compañeros de Hagerty Hall. Gracias a Mary Johnson por su amistad durante los años del doctorado. Su compañía durante la escritura de esta tesis hizo el proceso mucho más ameno y pasar juntas por cada etapa me ayudó a mantener la cordura. Gracias a Jenny Barajas por su apoyo y su amistad, y por siempre estar dispuesta a escuchar. También le agradezco a Christy García todas las discusiones y comentarios que me han ayudado a ir progresando.
en este proyecto. Finalmente, le agradezco a Meghan Armstrong todos los consejos que me ha dado durante este último año, sin los cuales me habría sentido bastante perdida.

I am forever grateful to Ed and Jeanie for always being there for us and for making babysitting their second job. Their unconditional love and support have allowed me to get through every milestone of my doctoral degree and, for that, I am very thankful. I also thank Thom and Connie for their encouragement and constant support in this endeavor. Their confidence in me has been a tremendous help throughout my graduate studies.

También me gustaría agradecerles a mis padres, Marisol y Jose, y a mi hermano, Diego, todo el apoyo que siempre me han dado, sin el cual jamás habría llegado a donde estoy hoy. A pesar de la distancia, su cariño y sus palabras de ánimo siempre me llegan con fuerza. Por encima de todo, le agradezco a mi madre los dos fríos meses que pasó en Columbus para que yo pudiera completar esta tesis.

Por último, quiero darles las gracias a las tres personas más importantes de mi vida. A mis hijas, Eva y Elsie, les agradezco su amor incondicional. Gracias por obligarme a cerrar el portátil de vez en cuando e ir a jugar. Sois el motor de mi vida y os quiero con locura. And last but not least, I want to thank my husband, Drew, for always supporting me and for being by my side during the toughest moments. Thank you for your patience and your encouragement. I love you.
VITA

2005.............................................................. B.A. Spanish Philology, Universidad de
Oviedo
2007.............................................................. M.A. Spanish, Miami University

Publications

Johnson, Mary and Sonia Barnes. 2013. Haya vs. Haiga: An analysis of the variation
observed in Mexican Spanish using a mixed effects model. Selected Proceedings of the
6th Workshop on Spanish Sociolinguistics, to appear.

Barnes, Sonia. 2012. ¿Qué dijistes?: A Variationist Reanalysis of Non-standard -s on
Second Person Singular Preterit Verb Forms in Spanish. Selected Proceedings of the 14th
Somerville, MA: Cascadilla Proceedings Project.

Fields of Study

Major Field: Spanish and Portuguese
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ ii
ACKNOWLEDGMENTS .................................................................................................................. v
VITA .................................................................................................................................................. viii
LIST OF TABLES .............................................................................................................................. xv
LIST OF FIGURES .......................................................................................................................... xviii

CHAPTER 1: INTRODUCTION ........................................................................................................ 1
  1.1 Description of the phenomenon .................................................. 1
  1.2 Language contact, language shift and variation ..................... 2
  1.3 Variationist sociolinguistics ...................................................... 7
  1.4 Sociophonetics .......................................................................... 8
  1.5 Dissertation goals and overview ............................................. 9

CHAPTER 2: SOCIOHISTORICAL BACKGROUND ON ASTURIAN/SPANISH CONTACT .................................................................................................................. 12
  2.1 Historical perspective on Asturian/Spanish contact in Asturias ........................................ 12
    2.1.1 Origins of Asturian .......................................................... 13
    2.1.2 The expansion of the Kingdom of Castile ..................... 14
    2.1.3 Industrialization and the beginning of normalization efforts ........................................ 16
  2.2 Current sociolinguistic situation in Asturias ........................................ 17
    2.2.1 Unbalanced bilingualism .................................................. 18
    2.2.2 Current revitalization efforts ......................................... 21
    2.2.3 Language attitudes towards Asturian and Spanish ............ 22
<table>
<thead>
<tr>
<th>2.3</th>
<th>Interference and linguistic variation in Asturian Spanish</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1</td>
<td>Asturian Spanish as a continuum</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Differentiating features and implicational scales</td>
<td>29</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Linguistic variation in Asturian Spanish</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Asturian and regional identity</td>
<td>36</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Linguistic interference and regional identity</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusions</td>
<td>41</td>
</tr>
</tbody>
</table>

CHAPTER 3: THE LINGUISTIC VARIABLES | 43 |
| 3.1 | Introduction | 43 |
| 3.2 | The linguistic variables | 45 |
| 3.2.1 | Unstressed vowels in Asturian | 46 |
| 3.2.2 | /-o/ vs. /-u/ | 47 |
| 3.2.2.1 | Distribution of /-o/ and /-u/ in Asturian | 49 |
| 3.2.2.2 | Envelope of variation | 52 |
| 3.2.3 | /-as/ vs. /-es/ | 58 |
| 3.2.3.1 | Distribution of /-as/ and /-es/ in Asturian | 58 |
| 3.2.3.2 | Envelope of variation | 59 |
| 3.2.4 | Acoustic qualities of Asturian and Asturian Spanish vowels | 63 |
| 3.3 | Previous research on vocalic variation in Asturian Spanish | 68 |
| 3.4 | Conclusions | 71 |

CHAPTER 4: ANALYSIS OF PRODUCTION DATA | 73 |
| 4.1 | Introduction | 73 |
| 4.2 | Methodology | 74 |
| 4.2.1 | Data collection | 74 |
| 4.2.2 | Participants | 75 |
| 4.2.3 | Envelope of variation and exclusions | 79 |
| 4.2.4 | Dependent variables | 83 |
| 4.2.4.1 | Acoustic analysis | 84 |
| 4.2.4.2 | Principal Components Analysis and clustering | 85 |
| 4.2.5 | Independent variables | 89 |
# 4.2.5.1 Internal/Linguistic variables

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the stressed vowel</td>
<td>89</td>
</tr>
<tr>
<td>Stress pattern</td>
<td>91</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>91</td>
</tr>
<tr>
<td>Preceding segment</td>
<td>92</td>
</tr>
<tr>
<td>Lexeme language</td>
<td>93</td>
</tr>
<tr>
<td>Frequency</td>
<td>94</td>
</tr>
<tr>
<td>Lexical category</td>
<td>97</td>
</tr>
<tr>
<td>Prior form</td>
<td>100</td>
</tr>
<tr>
<td>Mass/count distinction</td>
<td>101</td>
</tr>
<tr>
<td>Following segment</td>
<td>101</td>
</tr>
<tr>
<td>Lexical gender</td>
<td>102</td>
</tr>
</tbody>
</table>

# 4.2.5.2 External/social variables

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>103</td>
</tr>
<tr>
<td>Education and occupation</td>
<td>104</td>
</tr>
<tr>
<td>Age</td>
<td>105</td>
</tr>
<tr>
<td>Social tie</td>
<td>106</td>
</tr>
</tbody>
</table>

# 4.2.5.3 Summary

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>108</td>
</tr>
</tbody>
</table>

# 4.2.6 Statistical analysis

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
</tr>
</tbody>
</table>

# 4.3 Results

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>/-o/ vs. /-u/</td>
<td>111</td>
</tr>
<tr>
<td>Categorical analysis</td>
<td>114</td>
</tr>
<tr>
<td>Continuous analyses</td>
<td>119</td>
</tr>
<tr>
<td>First formant (F1)</td>
<td>119</td>
</tr>
<tr>
<td>Second formant (F2)</td>
<td>123</td>
</tr>
<tr>
<td>Effect of stressed vowel and stress pattern</td>
<td>127</td>
</tr>
<tr>
<td>Mass/count distinction in nouns</td>
<td>129</td>
</tr>
<tr>
<td>/-as/ vs. /-es/</td>
<td>130</td>
</tr>
<tr>
<td>Categorical analysis</td>
<td>132</td>
</tr>
<tr>
<td>Continuous analyses</td>
<td>137</td>
</tr>
</tbody>
</table>

xi
4.3.2.2.1 First formant (F1) .......................................................................................... 137
4.3.2.2.2 Second formant (F2) ....................................................................................... 141
4.3.2.3 Effect of stressed vowel and stress pattern ......................................................... 143
4.3.2.4 Effect of lexical gender ......................................................................................... 144
4.3.3 Summary .................................................................................................................. 144

4.4 Discussion ................................................................................................................... 145
4.4.1 Internal/linguistic variables ..................................................................................... 146
  4.4.1.1 Preceding segment ............................................................................................... 146
  4.4.1.2 Following segment .............................................................................................. 150
  4.4.1.3 Vowel duration .................................................................................................. 151
  4.4.1.4 Structural priming and spreading activation ....................................................... 152
  4.4.1.5 Quality of the stressed vowel ............................................................................. 153
  4.4.1.6 Lexical category .................................................................................................. 155
  4.4.1.7 Non-significant factors ....................................................................................... 159
    4.4.1.7.1 Mass/count distinction, lexical gender and morpho-phonological variation .................................................. 160
4.4.2 Social variables ......................................................................................................... 170
  4.4.2.1 Occupation and Education ............................................................................... 174
  4.4.2.2 Gender .............................................................................................................. 179
  4.4.2.3 Social tie ............................................................................................................ 180
  4.4.2.4 Non-significant factors ....................................................................................... 181

4.5 Conclusions .................................................................................................................. 184

CHAPTER 5: SOCIAL PERCEPTIONS OF THE LINGUISTIC VARIABLES .................................................. 188

5.1 Sociolinguistic perception ............................................................................................. 188
  5.1.1 Social meaning ........................................................................................................ 189
  5.1.2 Social perceptions of Asturian ................................................................................ 191

5.2 Methodology ................................................................................................................ 192
  5.2.1 Matched Guise Technique ...................................................................................... 193
  5.2.2 Stimuli .................................................................................................................... 193
  5.2.3 Questionnaire ......................................................................................................... 195
APPENDIX A: STIMULI USED IN THE PERCEPTION EXPERIMENT .......... 265
APPENDIX B: QUESTIONNAIRE USED IN THE PERCEPTION EXPERIMENT ........................................................................... 266
LIST OF TABLES

Table 1: Comparison of formant frequencies in Hz between Asturian Spanish final vowels and Spanish unstressed vowels (Quilis and Esgueva 1983) ........ 65
Table 2. Summary of independent linguistic variables and factor levels ..................... 108
Table 3. Summary of independent social variables and factor levels ......................... 109
Table 4. Overall distribution of /-o/ and /-u/ as determined by the DAPC ............... 111
Table 5. Distribution of speakers by Education and Occupation ............................. 115
Table 6. Linguistic and social factors contributing to the use of /-u/ in the complete data set .............................................................................................................. 116
Table 7. Effect of Education on the use of /-u/ in the complete data set ..................... 118
Table 8. Linguistic and social factors contributing to higher F1 frequency of /-o/ and /-u/ in the complete data set .............................................................................................................. 121
Table 9. Effect of Education on higher F1 frequency of /-o/ and /-u/ in the complete data set .............................................................................................................. 123
Table 10. Linguistic and social factors contributing to higher F2 frequency of /-o/ and /-u/ in the complete data set .............................................................................................................. 125
Table 11. Effect of Stressed vowel on the use of /-u/ ................................................. 127
Table 12. Effect of Stressed vowel on higher F1 frequency of /-o/ and /-u/ .............. 128
Table 13. Effect of Stressed vowel on higher F2 frequency of /-o/ and /-u/ .............. 129
Table 14. Overall distribution of /-as/ and /-es/ as determined by the DAPC .......... 130
Table 15. Comparison of the ranking of predictors in the two linguistic variables ...... 133
Table 16. Linguistic and social factors contributing to the use of /-es/ in the complete data set .............................................................................................................. 134
Table 17. Effect of Education on the use of /-es/ in the complete data set ............... 135
Table 18. Effect of Occupation on the use of /-es/ in the complete data set .......... 135
Table 19. Comparison of the effects of social predictors on the categorical choice between linguistic variants. .......................................................... 136
Table 20. Linguistic and social factors contributing to higher F1 frequency of /-as/ and /-es/ in the complete data set .......................................................... 139
Table 21. Effect of Occupation on F1 frequency of /-as/ and /-es/ in the complete data set .................................................................................. 140
Table 22. Linguistic and social factors contributing to higher F2 frequency of /-as/ and /-es/ in the complete data set .......................................................... 142
Table 23. Significant linguist predictors across the six dependent variables ............ 144
Table 24. Significant social predictors across the six dependent variables .......... 145
Table 25. Distribution of tokens of /-o/ and /-u/ across Stressed vowel and Preceding segment ............................................................................. 154
Table 26. Distribution of tokens of /-as/ and /-es/ across Lexical category and Preceding segment .................................................................. 155
Table 27. Effect of Lexical category on the use of /-es/ in tokens that have an alveolar or a dental consonant as the preceding segment. ......................... 156
Table 28. Use of /-es/ by lexical category .............................................................................................................................. 159
Table 29. Effect of social predictors on Pillai score for the /-o/, /-u/ variable ............. 164
Table 30. Effect of Education on Pillai score for the /-o/, /-u/ variable ......................... 164
Table 31. t=15.90, df=2127, p<0.01 ......................................................................... 168
Table 32. Effect of social predictors on Pillai scores for the /-as/, /-es/ variable ........ 169
Table 33. Distribution of participants by gender and education level ..................... 198
Table 34. Distribution of participants according to whether they speak Asturian and whether they believe that Asturian should be an official language in the region .... 198
Table 35. Effect of feature language on intelligence rating of utterances that contain the /-o/, /-u/ variable ............................................................................... 204
Table 36. Factors contributing to masculinity/femininity rating of utterances that contain the /-o/, /-u/ variable ............................................................................... 205
Table 37. Effect of speaker gender on masculinity/femininity rating of utterances that contain the /-as/, /-es/ variable ............................................................................... 207
Table 38. Factors contributing to formality rating of utterances that contain
the /-o/, /-u/ variable................................................................. 208
Table 39. Factors contributing to formality rating of utterances that contain
the /-as/, /-es/ variable................................................................. 210
Table 40. Factors contributing to rural/urban rating of utterances that
contain the /-o/, /-u/ variable................................................................. 211
Table 41. Factors contributing to rural/urban rating of utterances that
contain the /-as/, /-es/ variable................................................................. 211
Table 42. Factors contributing to rural/urban rating of utterances that
contain the /-o/, /-u/ variable................................................................. 213
Table 43. Effect of speaker gender on economic status rating of utterances
that contain the /-as/, /-es/ variable................................................................. 215
Table 44. Effect of speaker gender on familiarity rating of utterances that
contain the /-o/, /-u/ variable................................................................. 216
Table 45. Effect of speaker gender on familiarity rating of utterances that
contain the /-as/, /-es/ variable................................................................. 216
Table 46. Effect of feature language on the selection of educational level
in utterances that contain the /-o/, /-u/ variable................................................................. 219
Table 47. Effect of feature language on the selection of age range in
utterances that contain the /-as/, /-es/ variable................................................................. 221
Table 48. Effect of feature language on the selection of occupation in
utterances that contain the /-o/, /-u/ variable................................................................. 223
Table 49. Effect of feature language on the selection of occupation in
utterances that contain the /-as/, /-es/ variable................................................................. 225
Table 50. Effect of feature language on the selection of place of origin in
utterances that contain the /-o/, /-u/ variable................................................................. 227
Table 51. Significant predictors across the six variable ratings........................................ 229
LIST OF FIGURES

Figure 1. Geographic extension of Asturian in the 13th Century (Academia de la Llingua Asturiana 2002)................................................................. 14
Figure 2. Vowel plot comparing the realization of Spanish unstressed vowel (Quilis and Esgueva 1983) to that of Asturian Spanish final vowels by male speakers.. 66
Figure 3. Vowel plot comparing the realization of Spanish unstressed vowels (Quilis and Esgueva 1983) to that of Asturian Spanish final vowels by female speakers66
Figure 4. Location of the districts of El Llano, Pumarín and Contrueces .............. 75
Figure 5. Social network of informants ............................................................ 77
Figure 6. Spectrogram of the word “tiempu” [tjempu]......................................... 82
Figure 7. Spectrogram of the word "esas" [esas]................................................... 83
Figure 8. Distribution of /o/ and /u/ in a vowel space as coded auditorily ............ 87
Figure 9. Distribution of /o/ and /u/ in a vowel space as grouped by the DAPC........ 87
Figure 10. Distribution of /a/ and /e/ in a vowel space as coded auditorily............. 88
Figure 11. Distribution of /a/ and /e/ in a vowel space as grouped by the DAPC....... 88
Figure 12. Frequency distribution of tokens that show /-o/, /-u/ variation ............. 96
Figure 13. Frequency distribution of tokens that show /-as/, /-es/ variation............ 96
Figure 14. Distribution of F1 (ERB) frequency by vowel (t=22.3, df=1318.33, p<0.01)............................................................................................................. 112
Figure 15. Distribution of F2 (ERB) frequency by vowel (t=27.5, df=1312.11, p<0.01)............................................................................................................. 113
Figure 16. Random forest with all predictors of /-o/ and /-u/ alternation.............. 114
Figure 17. Random forest with all predictors of higher F1 frequency of /-o/ and /-u/ .. 120
Figure 18. Random forest with all predictors of higher F2 frequency of /-o/ and /-u/ .. 124
Figure 19. Distribution of F1 (ERB) frequency by vowel (t=20.84, df=760.96, .......
Figure 20. Distribution of F2 (ERB) frequency by vowel ($t=-21.4$, $df=634.087$, $p<0.01$) ................................................................. 131

Figure 21. Random forest with all predictors of /-as/ and /-es/ alternation ................. 132

Figure 22. Random forest with all predictors of higher F1 frequency of /-as/ and /-es/. 138

Figure 23. Random forest with all predictors of higher F2 frequency of /-as/ and /-es/. 141

Figure 24. Effect of preceding segment on the F1 and F2 of [o] and [u] ..................... 147

Figure 25. Effect of preceding segment on the F1 and F2 of [a] and [e] ..................... 149

Figure 26. Effect of following segment on the F1 and F2 of [o] and [u] ..................... 151

Figure 27. Distribution of vowel duration across lexical categories ....................... 158

Figure 28. Correlation between Pillai score and usage of /-u/. $R=0.1442$, $p<0.01$ .... 163

Figure 29. Correlation between Pillai score and age. $R=0.2011$, $p<0.01$ ............... 166

Figure 30. Correlation between Pillai scores and usage of /-es/. $R=0.07533$, $p<0.01$ .... 169

Figure 31. Conditional inference tree of social factors that significantly predict
the use of /-u/ ........................................................................................................ 171

Figure 32. Conditional inference tree of social factors that significantly predict
the use of /-es/ ........................................................................................................ 173

Figure 33. Correlation between Age and usage of /-u/. $R=0.04087$, $p<0.01$ .......... 183

Figure 34. Correlation between Age and usage of /-es/. $R=0.04076$, $p<0.01$ .......... 183

Figure 35. Conditional inference tree of factors that have a significant effect on
masculinity/femininity rating of utterances that contain the /-o/, /-u/ variable ............ 206

Figure 36. Conditional inference tree of factors that have a significant effect on
formality rating of utterances that contain the /-o/, /-u/ variable ................................ 209

Figure 37. Conditional inference tree of factors that have a significant effect on
urban rating of utterances that contain the /-o/, /-u/ variable .................................. 212

Figure 38. Conditional inference tree of factors that have a significant effect on
the rating of economic status of utterances that contain the /-o/, /-u/ variable .......... 214

Figure 39. Bar plots illustrating what proportion of the utterances that were
assigned to each education level contained Asturian or Spanish features ............... 218

Figure 40. Bar plots illustrating what proportion of the utterances that were
assigned to each age range contained Asturian or Spanish features. .......................... 220

Figure 41. Bar plots illustrating what proportion of the utterances that were
assigned to each occupation contained Asturian or Spanish features. ......................... 222

Figure 42. Bar plots illustrating what proportion of the utterances that were
assigned to each place of origin contained Asturian or Spanish features. ..................... 226

Figure 43. Mean rating of the utterances used in the perception experiment,
classified by category, variable and feature language ................................................. 231
CHAPTER 1 : INTRODUCTION

1.1 Description of the phenomenon

Situations of bilingualism in which the two languages in contact are maintained are common in Spain. Languages like Basque, Catalan and Galician enjoy co-official status together with Spanish in their respective regions. Despite the existence of these examples, we do not find this type of situation for all the ancestral languages spoken in the Iberian Peninsula. In Asturias, a region situated in the Northwest of Spain, the autochthonous language of the community, Asturian or Bable, does not have official status. Most scholars agree that language shift is gradually taking place in Asturias: due to socio-historical reasons, Spanish is being extended to most communicative contexts thereby usurping the use of Asturian. The death of Asturian is so imminent that some researchers have claimed a total lack of monolingual speakers of this language (Konvalinka 1985, Amado Rodríguez et al. 1993). Konvalinka (1985) suggests that speakers can be situated in a continuum that reflects different degrees of interference between the two languages, locating speakers that are dominant in Asturian to one of the ends of this continuum at which Asturian shows minimal influence from Spanish. The idea of a linguistic continuum between Asturian and Spanish will be further explored in section 2.3.1. Due to the different degrees of linguistic interference between Spanish and
Asturian (and facilitated by the typological similarities between the two languages) the majority of today’s population speaks a mixed variety of language to which speakers and scholars usually refer as *Amestāu* (Asturian for ‘mixed’). Research on Asturian has traditionally focused on the language maintained in the small rural areas of the region and has tried to identify the features that distinguish different varieties of Asturian. Little attention has been paid to the linguistic contact between Asturian and Spanish. Even though the population in the urban centers of Asturias normally identify themselves as native speakers of Spanish (or Castilian1), many linguistic features from Asturian are found in their speech. Among these features are the presence of raised vowels in the masculine singular morpheme /-u/, as opposed to Spanish /-o/ as in “el perru” (‘the dog’) vs. Spanish “el perro”, and the feminine plural morpheme /-es/ as opposed to Spanish /-as/, as in “les cases” (‘the houses’) vs. Spanish “las casas” (D’Andrés 2002). The appearance of the raised variants in the urban varieties of the language (which I will refer to as Asturian Spanish) is a consequence of the borrowing of morphemes from Asturian; however, it does not occur categorically. Instead, speakers alternate between the Asturian and the Spanish features.

### 1.2 Language contact, language shift and variation

One of the possible outcomes of situations in which two or more languages are in contact is language shift. Winford (2003:15) describes language shift as “the partial or

---

1 Even though the term “Castilian” is normally used in the literature to refer to the dialect of Latin that originated in the region of Cantabria, today it is commonly used in Spain as a synonym of “Spanish”.
total abandonment of a group’s native language in favor of another”. The linguistic result of language shift can vary depending on the contact situation. For example, in some cases the languages involved are reserved for different social functions while in other situations the shift results in the complete replacement of the first or native language (L1), with more or less influence of its linguistic system on the target language (TL). This type of influence has often been referred to as “substratum” influence and it is normally characterized by being more structural than lexical. According to Winford, the formation of these contact varieties can be explained from a Second Language Acquisition (SLA) perspective, since processes involved in group SLA are comparable to those in the individual’s acquisition of a second language. Winford proposes the use of the SLA framework to classify the outcomes of language shift, placing them on a continuous scale depending on how successful the acquisition of the second language is. As he explains, “the chief criterion for placement would be the degree of L1 influence and perhaps simplification in the contact variety” (Winford, 2003:255). The variety of Asturian Spanish under study could be placed towards the side of the scale where there is the least influence of the L1, since, as we will see in chapter 2, the retention of the L1 in this contact variety is limited to a set of phonological and morphosyntactic features. The type of linguistic contact observed in Asturias is not uncommon. Winford provides the example of Irish English. Many of the features present in this variety can be attributed to the influence of Irish, which is observable at all structural levels.

The socio-historical events that led to the contact between Spanish and Asturian are explained in chapter 2; in the remainder of this section I review the diachronic and synchronic linguistic processes that result in the language varieties spoken today in
Asturias. In order to do so, I adopt Winford’s (2005) classification of contact-induced changes, which is based on Van Coetsem’s (1988) framework. Both authors differentiate between two types of transfer: borrowing and imposition. Whether one or the other takes place is determined by the agentivity in the crosslinguistic influence or transfer; i.e. whether the agent of the transfer is the speaker of the source language (SL) or that of the recipient language (RL). As Winford (2005:376) explains, “The direction of the transfer of material is always from the SL to the RL and the agent of the transfer is either the RL speaker (RL agentivity) or the SL speaker (SL agentivity)”. Whenever the speaker of the RL is the agent of the transfer we have borrowing, while whenever there is SL agentivity, imposition occurs. Van Coetsem and Winford explain the types of transfer as a matter of dominance. RL agentivity and borrowing occur in a situation in which speakers borrow materials from their non-dominant language into the language in which they are most dominant. This type of transfer involves mostly, but not exclusively, the borrowing of vocabulary. SL agentivity and imposition take place when the speaker is most dominant in the SL and less proficient in the RL. This type of phenomenon involves the transfer of mostly structural features.

Borrowing and imposition are not mutually exclusive. Instead, in many situations of language contact they can operate complementarily. In fact, in a particular contact situation, the same speakers can show both types of transfer. Van Coetsem (1988:35) mentions the example of Norwegian immigrants in America, explaining that “when speaking English, they imposed Norwegian articulatory habits (sl) upon English (rl), and when speaking Norwegian, they borrowed vocabulary items from English (sl) and incorporated them into their Norwegian (rl)".
In the contact situation between Asturian and Spanish we find cases of borrowing and imposition. As I explain in the following chapter, Asturian was the autochthonous language that developed from Latin in the region. During the fourteenth century, the expansion of the Kingdom of Castile led to the imposition of Castilian as the written, administrative and, in general, prestige language. As a result, the two languages were reserved for different social functions. However, the migration from rural areas to the cities that took place during the industrialization of the nineteenth century initiated the language shift. The rural population was dominant in Asturian while the majority of the population in urban areas was dominant in Spanish. As Asturian speakers learned Spanish, they imposed structural and articulatory features of their native language (SL) upon Spanish (RL). The use of Spanish as the language of the administration, education and, later on, the media, resulted in the expansion of Spanish in both urban and rural areas. Thus, speakers that were proficient in Asturian all over the region borrowed vocabulary items from Spanish when they used their native language. As González Quevedo (2001:175) states, “we see Spanish vocabulary appearing as neologisms referring to realia that were not originally part of traditional country life”. However, Fernández Álvarez (1989) points out that together with these types of ‘new’ lexical items, a substitution of already existing words in Asturian has also taken place. It is interesting to note that, on some occasions, a term adopted from Spanish coexists with the autochthonous equivalent. In these cases, the words in the two languages might develop a situation of complementary distribution, according to which the Asturian item is specialized for one of the original meanings of the word and the item in Spanish for the
other. D’Andrés (1997) provides the example Asturian “fueya” used for ‘leaf’, and the Spanish equivalent “hoja”, reserved for ‘sheet (of paper)’.

The situation that we find today in the urban areas of Asturias is different from that observed immediately after the migration period in the nineteenth century. The use of Spanish was extended to more communicative contexts and today the great majority of speakers in these areas are dominant in this language. In their speech, however, they borrow not only vocabulary, but also morphosyntactic and phonological features from Asturian. This type of structural borrowing is facilitated by the typological similarity between the two languages involved. According to Gonzalez Quevedo (2001), the presence of Asturian structural features in this variety has become stabilized.

It is possible that the origin of structural borrowing is the borrowing of particular lexical items. As Winford (2005:386) explains, “certain structural innovations in an RL appear to be mediated by lexical borrowing”. As an example, he mentions the introduction of certain phonological distinctions and new derivational suffixes into Middle English via the extensive lexical borrowing from French that took place during the fourteenth and fifteenth centuries. In Asturian Spanish the incorporation of, for example, Asturian /-u/ could have started by the borrowing of particular lexical items that contained that morpheme; i.e. initially, speakers did not isolate the morpheme /-u/ and incorporate it into Spanish stems. The resulting linguistic system, however, does not show the use of Asturian features in a categorical way. Variation between Asturian and Spanish morphemes occurs within the speech of the same speaker and, on some occasions, in the use of the same lexical item. This observed behavior is what motivates the adoption of a variationist approach to the study of urban Asturian Spanish.
1.3 Variationist sociolinguistics

Variationist sociolinguistics emerged as a departure from dialectology and historical linguistics. The first variationist studies were concerned with how language change arises and how it spreads. Labov (1966) suggested that by looking at how language is variable in a given moment and how different linguistic variants are correlated to social categories, such as socioeconomic status and gender, one could attempt to answer questions regarding the process of language change. Labov’s incorporation of social and stylistic factors into a more extensive dialectological approach constitutes the roots of variationist sociolinguistics. Within this framework, alternations between linguistic forms, such as those observed in Asturian Spanish, are not random. Instead, variation, which constitutes the inherent state of language, occurs in an orderly and systematic manner (Labov 1966, 1969). Because of this systematicity, it is possible to model the behavior of linguistic variants quantitatively, allowing us to discover the way in which the use of a particular linguistic variant is linguistically and socially constrained. In order to do so, variationist sociolinguistics analyzes the recurrent choices that speakers make in their linguistic production. The use of statistical analyses reveals meaningful patterns of use of the linguistic variables and allows us to make generalizations from a sample of speakers to the speech community they represent. As Tagliamonte (2011:9) explains, “The choices are taken to represent the (underlying) variable grammar of the speaker as well as the grammar of the speech community to which she belongs.”

Recent research on sociolinguistic variation has been increasingly concerned with how this linguistic variation relates to social meaning, with the idea that speakers use variation to construct their identity and position themselves in the social world. In order
to do so some scholars have focused on the variation that takes place in smaller communities and using locally-defined social categories (Eckert 2001, Mendoza-Denton 2008).

In this dissertation I explore the distribution of the Spanish and Asturian morphemes under study from a variationist perspective, exploring the linguistic and social distribution of the variants and also their social meaning, in order to provide a description of the variable grammar and discover how speakers use the variation to construct a particular identity.

1.4 Sociophonetics

The nature of the features that are the focus of this dissertation calls for a phonetic analysis of the vowels that are involved in the variation. Thus, an approach that combines the principles of sociolinguistic variation but that also includes sophisticated phonetic analysis is necessary to fully account for the social and linguistic distribution of the variables under study.

Hay and Drager (2007: 90) define sociophonetics as “The study of socially-conditioned phonetic variation in speech”. While earlier work on sociophonetic variation relied heavily on the auditory coding of linguistic variants and usually treated them as a binary choice, more current research emphasizes the need to perform sophisticated phonetic analysis in order to account for the gradient nature of phonetic variables. This type of analysis allows the researcher to capture the fine phonetic detail that can potentially have social meaning. Researchers in this area have found that speakers are
aware of these subtle differences in speech and that they use them to construct social meaning (Docherty and Foulkes 2005, Podesva 2006, among others). They also maintain that listeners use this phonetic detail to evaluate the social characteristics of other speakers.

Performing an acoustic analysis of the vowels that are the object of my study will allow me to capture the gradient nature of the variation. The morphological nature of the variation also calls for a categorical approach, since, as I explained above, the presence of /-u/ and /-es/ in Asturian Spanish is a result of the borrowing of these morphemes from Asturian. However, basing this categorization on the acoustic properties of the vowels instead of the auditory coding will increase the reliability of the classification.

1.5 Dissertation goals and overview

The first objective of this dissertation is to investigate the linguistic and social factors that govern speakers’ choice between the Asturian and the Spanish features in the language variety spoken in Gijón, the largest city of Asturias. Previous research has emphasized the need to describe the linguistic variety spoken in the urban areas of Asturias (Sánchez Álvarez 1979, D’Andrés 2002). This need stems from the primary focus that has traditionally been given to the description of rural Asturian within the field of dialectology. Thus, interference that takes place between the two linguistic systems has not received much attention. The few sociolinguistic studies that explore the use of Asturian and Spanish features emphasize the importance of social factors in the variation but fail to provide quantitative evidence that supports the conclusions reached regarding
the effect of external variables. In addition, the possible internal factors that might govern the variation have been underexplored and no studies have taken into consideration the gradient nature of the differentiation between Asturian and Spanish vowels.

The second objective of this dissertation is to discover whether the speakers are aware of the correlations observed in the quantitative analysis and whether they use the linguistic variables to interpret social identity. If there is social meaning associated with the variation in the use of Asturian or Spanish final vowels, the ultimate goal is to discover how this social meaning is used by speakers. Prieto (1991) reports that the use of Spanish is stigmatized in Gijón and monolingual speakers of Spanish (or speakers who try to erase any interference from Asturian) are referred to as *falantes finos* (‘fancy speakers’). There seems to be a sense of regional identity linked to the use of Asturian features. Dyzmann (2000) claims that cases of linguistic interference between Spanish and Asturian have a double function: on the one hand, they reflect the speakers’ desire to communicate in the regional language; on the other hand, interference can also indicate the social necessity of showing competence in Spanish, motivated by a desire on the part of the speakers to distance themselves from the rural world and its characteristic speech, which is also stigmatized by urban speakers.

I hypothesize that the social meaning associated with the raised vowels that are the object of my study is utilized by speakers to construct this urban identity. Previous studies that address the possible connection between Asturian and regional identity have done so in a general and impressionistic way, without examining the role that particular linguistic features play in the process. This study is the only one to date to explicitly address the connection between language use and social meaning in Asturian Spanish.
This is important because, as has been demonstrated for different varieties of English (Eckert 2000, Mendoza-Denton 2008, Campbell-Kibler 2009), speakers actively use linguistic variation to construct their own identities. My research findings will therefore contribute to the general understanding of how identity is constructed through language, utilizing data from a heretofore-ignored language variety.

The dissertation is organized into six chapters divided as follows. In chapter 2, I review the socio-historical background on Asturian/Spanish contact, I explore the linguistic result of such contact and the studies that have addressed the variation between Asturian and Spanish features. I also discuss the relation between the use of Asturian and regional identity. In chapter 3, I provide a description of the linguistic variables under study and their envelope of variation, and I review previous studies that have explored vocalic variation in Asturian Spanish. In chapter 4, I present the quantitative analysis of production data, describing the methodology used in the collection and analysis of the data, the results obtained and discussing the effect that linguistic and social variables have on the alternation between Asturian and Spanish vowels. In chapter 5, I explore the social perceptions of the linguistic variables. I explain the methodology used in the design and implementation of the perception experiment, and describe and discuss the results obtained in the analysis of the data. Finally, in chapter 6, I provide conclusions and contributions of this dissertation and outline future possibilities for research.
CHAPTER 2: SOCIOHISTORICAL BACKGROUND ON ASTURIAN/SPANISH CONTACT

2.1 Historical perspective on Asturian/Spain contact in Asturias

Asturian is a Romance language that developed in the kingdom of Asturias and León as a result of the spread of Latin in the Iberian Peninsula. Even though it does not have the co-official status that other regional languages in Spain have, such as Basque, Catalan or Galician, Asturian is considered a language distinct from Spanish and not a regional variety of the official language. Metzeltin (2001) indicated that documents written as early as the 13th century show clear differences between Asturian and Castilian in all levels of the language. Despite diachronic and synchronic evidence in favor of an independent development of Asturian on the Iberian Peninsula, a Castilian-centered tradition within Hispanic Philology imposed early on a classification of Asturian as a dialect of Spanish. This view was adopted by the educational system in Asturias and quickly became part of popular opinion, with speakers assuming that the use of Asturian constituted a deviation from the “correct” variety of the Spanish language (D’Andrés 1993, 1997). Nowadays this idea has been abandoned in academic circles and in popular opinion in favor of the characterization of Asturian as an autonomous language whose
development parallels that of Galician and Spanish. This section tracks the historical development of Asturian on the Iberian Peninsula and its process of stigmatization.

2.1.1 Origins of Asturian

The region that is now Asturias was populated by Astur groups, one of the ethnicities that inhabited the northwest of what is Spain today (together with Galaicos and Cantabrians) prior to the Roman conquest. The arrival of the Roman Empire resulted in the imposition of Latin in all these territories and in the development of different linguistic varieties of Latin in each region due to the influence of the substratum languages. During the Muslim domination of the Peninsula, the linguistic differences that characterized the speech of different territories were partially erased. The north of Spain, however, maintained linguistic differentiation to a greater degree (Konvalinka 1985). In this period, Asturias became the first Christian kingdom of the Peninsula (Kingdom of Asturias) and had Oviedo as its capital. In 924 Asturias became the Kingdom of León, and the Capital shifted from Oviedo to León. The Kingdom of León expanded to the south and east to include the regions of Leon, Zamora, Salamanca, Extremadura and Cantabria. In 1230 the Kingdom of León was incorporated into the Kingdom of Castile and the Kingdom of Castile and León was formed. The map in Figure 1 shows the geographic domain of Astur-Leonés, the Romance language spoken in the Kingdom of Asturias and later León (Academia de la Llinguà Asturiana 2002). The diagonal shading indicates the part of this kingdom into which Castilian, the Romance language that
developed in the region of Castile parallel to Astur-Leonês, would later expand and completely substitute Astur-Leonês.

Figure 1. Geographic extension of Asturian in the 13th Century (Academia de la Llingua Asturiana 2002)

2.1.2 The expansion of the Kingdom of Castile

According to Konvalinka (1985), the shift of the capital of the Kingdom to León, and with it the change of what was the cultural center, was decisive in the development of Asturian regional identity. The author explained that, until that moment, historically Asturias had followed a similar path to that of regions like Galicia and Catalonia and until after the Reconquest all the regional languages of Spain were prestige varieties. However, other regions maintained their political and cultural centers, which contributed to promoting cultural and linguistic regional awareness among speakers. In contrast to this situation, the few years during which Asturias was the center of the Kingdom were not sufficient to achieve the linguistic normalization that took place in Galicia and
Catalonia. Instead, Asturias became an extension of Castile, erasing any trace of the regional differentiation that had characterized the previous political and administrative situation.

The political expansion of the Kingdom of Castile in the fourteenth century led to the imposition of Castilian as the language of the administration and the higher social strata, also becoming the official written language and giving rise to a classic situation of diglossia (Ferguson 1964). This is registered in the written documentation of the period. As González Quevedo (2001) pointed out, we can find documents written in Asturian until the sixteenth century. After that date, Spanish substituted Asturian as the written language, becoming the high variety that was employed in public domains while Asturian, the low variety, was reserved for informal domains, like the family and the neighborhood. García Arias (2002) attributed this shift to the presence in Asturias of civil and ecclesiastic officials who were originally from Castile.

Despite the expansion of Castilian as the public language, Asturian was maintained in spoken registers, especially in the most rural and remote areas of the region. After the seventeenth century we start finding examples of literary work in Asturian and, even though Spanish had become the dominant language, based on the opinions of authorities and writers on literary and linguistic matters, Asturian was considered to be just another of the languages native to the Peninsula until the twentieth century (González Quevedo 2001:167).
2.1.3 Industrialization and the beginning of normalization efforts

According to several authors the language shift that led to the current sociolinguistic situation started in the twentieth century as one of the immediate consequences of the process of industrialization experienced in the community. Viejo Fernández (2004) signaled two main factors as the sources of the linguistic change in Asturias: first, the migration of peoples from rural areas to the cities promoted linguistic and cultural uprooting; second, the Asturian bourgeoisie favored the use of Spanish as the prestige language, contributing to the stigmatization of Asturian as a rustic language that impeded social ascent. Even though there were initiatives throughout the twentieth century to recover the social status enjoyed by Asturian in previous centuries, it is only at the end of Franco’s dictatorship in the 70’s that we find attempts to normalize Asturian and reverse the language shift.

As a first step in the process of normalization, in 1969 the association “Amigos del Bable” (‘Friends of Asturian’) was created. Even though the organization did not play an active role in the normalization efforts, it reflected a desire to cultivate the Asturian linguistic tradition (García Arias 2002:18). In the late seventies, at a time when nationalist movements and efforts of linguistic revitalization were taking place in Galicia, Catalonia and the Basque Country, the “Conceyu Bable” (‘Asturian Council’), a linguistic and political organization, was created by a group of young Asturian speakers. This organization had a crucial role in shaping the current social attitudes towards the regional language. Its function was to defend the linguistic rights of Asturian speakers. The Conceyu was very successful in getting a large portion of the population to be concerned with the status of the regional language and it brought about significant
momentum in the revitalization efforts. Their activity culminated in 1980 with the creation of the “Academia de la Llingua Asturiana” (‘Academy of the Asturian Language’). The establishment of the Academy resulted in the official recognition of Asturian as one of the languages of Spain (even though it has not been granted co-official status) and the promotion of linguistic research on Asturian together with the blossoming of literary works written in the regional language. The publication of the *Gramática de la Llingua Asturiana* (Asturian Language Grammar) in 1998 and the *Diccionariu de la Llingua Asturiana* (Asturian Language Dictionary) in 2000 by the Academy were crucial steps towards the standarization of the language.

### 2.2 Current sociolinguistic situation in Asturias

Even though new positive attitudes towards language shift reversal have emerged during the last decades, Asturian is still in an asymmetrical and inferior position to Spanish. In the last few years the demand for linguistic normalization has increased among the Asturian population, together with an increasing awareness of the situation of language endangerment. The favorable attitude towards language shift reversal contrasts with the opposition of an influential minority to the process of normalization. This minority is formed by some academics, linguists among them, and highly educated individuals who hold positions of intellectual power over the rest of the population. This sector of the population considers Asturian a rural language that is inferior to Spanish and, consequently, only suited for the home and the rural world. One of the arguments often given against the normalization of the regional language is the dialectal
fragmentation of the language. In adopting this perspective this influential minority maintains that the Asturian population cannot identify with a language that is not homogeneous. They characterize standardized Asturian as an artificial variety that does not correspond to the linguistic reality. Despite the change in the linguistic attitudes towards Asturian, as Viejo Fernández (2004:174) pointed out, “the situation in general is more of passive acceptance than of active commitment”.

2.2.1 Unbalanced bilingualism

D’Andrés (1996) characterizes the Asturian sociolinguistic situation as one of unbalanced or diglossic bilingualism. While the use of Spanish is accepted in all communicative situations, the use of Asturian is restricted to informal situations and avoided in public and official contexts. D’Andrés (1993) also claims that this type of unbalanced bilingualism rarely constitutes a stable situation. Instead, he notes that there is gradual regression of Asturian in favor of Spanish (D’Andrés 1996). Evidence of the regression of regional language comes from sociolinguistic studies and sociolinguistic surveys that note a decline in the use of Asturian or Asturian linguistic features in the younger generations.

Half of the population of Asturias lives in one of the three largest urban areas of the region. Of the one million Asturians that were reported in the census of 2012, 25.68% live in Gijón, 20.89% in Oviedo, and 7.68% in Avilés. The rest of the population (45.75%) lives in smaller towns (of 50,000 inhabitants or less) and rural areas. Even though there are no data regarding the exact number of monolingual and bilingual
speakers of the two languages in Asturias, D’Andrés (1997:77) proposes the following classification of the Asturian population, based on their linguistic competence in one or both languages. First, there is a monolingual group of Spanish speakers, constituted by people with higher socioeconomic status that live in urban areas, speakers that have higher education levels and immigrants (from other areas of Spain or from outside of the country). However, as González Quevedo (2001:176) noted, even outsiders show incorporation of Asturian features in their speech if they have been living in the region for several years. The degree to which immigrants assimilate to Asturian depends on the area in which they live. Thus, for example, it has been found that native Spanish speakers that live in coal-mining areas show a higher degree of linguistic assimilation, while those that immigrate to urban centers tend to resist Asturian assimilation. The second group is that of bilingual speakers of Spanish and Asturian. There are two types of bilingual speakers: those who have Asturian as their native language and acquired Spanish in school, and those who have Spanish as their native language and learned Asturian as adults in a formal setting. The existence of the latter is a result of the process of normalization and regularization of the language. González Quevedo (2001:176) explained that these speakers constitute a minority of the population and describes them as “speakers with a high degree of self-awareness who are fully aware of their linguistic usage. In many instances – though not all – these are younger people who move in university circles, in education or left wing and union politics.” Their speech is also different from that of bilinguales that have acquired Asturian in the home and community as their native language, since they carefully use Asturian without interference from Spanish and not according to a particular variety of the regional language but to a
standarized variety of the language. This normalized system includes features mainly from central Asturian but it also incorporates some features that characterize the speech of Western and Eastern Asturian. As González Quevedo explains, this is a process of “Re-Asturianization” of the language originating from the educational change that is beginning to reshape the attitudes of speakers towards the language. The last group are the monolingual speakers of Asturian. González Quevedo claims that this group is clearly limited to rural areas and the aging population. Despite having had contact with Spanish through education and the media, the influence of the language has not been strong enough in these speakers to completely replace Asturian. Instead, they show some degree of linguistic interference (mostly in their lexicon) but much less noticeable than that of bilingual speakers. Despite these claims, the existence of monolingual speakers of Asturian has been questioned by some scholars. For example, Konvalinka (1985) and Amado Rodríguez et al. (1993) stated that the only monolingual speakers in Asturias are those of Spanish. The rest of the population can situate their speech at some point along the continuum between Asturian and Spanish, including more or fewer features from the regional language.

As part of a sociolinguistic survey of Asturias, Llera Ramo and San Martín Antuña (2003) obtained information regarding the native language of a group of 1,300 speakers from all areas of Asturias. They found that 20.7% of the speakers in the sample claim to be native speakers of Asturian, 53.6% identify themselves as native speakers of Spanish and 22.5% claim to have both as their native language. Even though we do not have information regarding the degree of linguistic interference present in the speech of
these informants, the results of the survey provide us with an idea of how the languages are distributed in Asturian society.

2.2.2 Current revitalization efforts

The revitalization efforts initiated in the last part of the twentieth century continue at present. In the last decades various political and cultural groups, together with the Academia de la Llingua Asturiana, have worked towards reversing language shift. These efforts are directed, on the one hand, at the Asturian population and, in this respect, they have resulted in a progressive change in attitudes towards the language. On the other hand, part of the Asturianist movement seeks a change in the political status of the language that would grant it official status in the region. While Asturian still lacks official status, it now has a greater presence in the public realm. The Statute of Autonomy of Asturias (1982) declares the need to protect Asturian and to promote its use in the media and include it, in an optional way, in the education system. It also states that speakers of Asturian will not be discriminated against because of their use of the regional language. Currently, Asturian is used by the administration in forms, public documentation and signage. Asturian has also been introduced in the educational curriculum as an optional subject in primary and secondary schools. The language is also present in the media, performing arts and literature. However, while more positive attitudes towards the regional language are developing among Asturians, scholars believe that this is not sufficient for reversing language shift and they see granting Asturian official status as the only possible way of truly increasing the vitality of the language.
2.2.3 Language attitudes towards Asturian and Spanish

Sociolinguistic surveys have been used to explore the current social distribution of the two languages and the attitudes that speakers have towards them. While our understanding of the sociolinguistic situation cannot be solely based on this type of information, the results of these surveys provide useful information that can be used to further evaluate the relationship between the use of Asturian and certain social categories.

Llera Ramo (1994) carried out the most comprehensive sociolinguistic survey to date in Asturias. He divided the region into three sections and recruited 400 participants from each section, 26% of whom were from urban areas. He found that the majority of the speakers (35.9%) consider Asturian to be their native language, while 23.7% claim to speak a mixture of Asturian and Spanish. Only 18.9% of the participants believe that Spanish is the language of the region, with Asturian (43.9%) and both (34.6%) being the most common answers. Despite identifying Asturian as the language of Asturias, 46.7% of the participants claim to be able to communicate more easily in Spanish. The majority of the participants (61.2%) have a positive attitude towards granting Asturian co-official status and towards teaching Asturian in schools (88.4%). Only 10.5% of the participants believe that speaking Asturian is speaking incorrectly, and 64.6% would like to formally learn Asturian. Llera Ramo also researched the distribution of both languages in different domains. Participants claimed to speak more Asturian than Spanish with family and friends while the use of Spanish predominated at work and at school. It should be noted that the percentage of speakers that claim to use both languages indifferently is higher in the family and friendship contexts.
Similar results were obtained in the survey carried out by Armesto Fernández (2003). The author surveyed 200 Pre-Education students at the University of Oviedo regarding their attitudes regarding the use and normalization of Asturian and its inclusion in the academic curriculum. The results show favorable attitudes towards granting Asturian co-official status and towards including it in the educational curriculum. Despite the favorable attitudes, the majority of the participants claim not to be proficient in the regional language and to reserve its use to the family and friendship contexts. Participants believe that Asturian should have a larger presence in the media and are also aware of the endangered status of the language.

D’Andrés (1996) used the results from previous sociolinguistic surveys to examine the relationship between linguistic attitudes and the social variables of age and gender. He found that, in the surveys that have been carried out in Asturias to date, men normally claim to use Asturian more than women and also to have better knowledge of it. He pointed out that men also have a more favorable attitude towards Asturian and that women more often identify Asturian speech as different from that of other regions in the country. With respect to age, he stated that older generations use Asturian more often and claim to have better knowledge of it but that younger generations have more positive attitudes towards the language. The author pointed out that the most favorable attitudes are maintained by speakers that range from 18 to 25 years old.

Dyzmann (2000) surveyed fifty students from a high-school in Gijón. She found that 62% of the students disagree with the opinion that speaking Asturian is the same as speaking Spanish incorrectly. Only 20% of the students considered Asturian to be the language of Asturias, while 70% assigned this role to Spanish. When asked about
language use in several domains, Spanish was the most common answer for all contexts except for the family and in conversations with classmates, where 68% of the participants claimed to use a mixture of the two systems.

Rose Alcorta (2000) used a different technique to explore sociolinguistic attitudes towards Asturian. The author carried out a perception experiment in which she used a Matched Guised Technique. To create the stimuli she asked the same speaker to read different paragraphs in 4 different varieties: Asturian, Spanish, a mixed variety and Catalan. Rose Alcorta explained that Catalan was included in the study because it is a minority language that has more prestige than others spoken in the Peninsula (such as Galician) and that can be understood by Asturian speakers. She found that higher ratings are given to Spanish and Catalan in status categories, such as intelligence, efficiency, etc. In contrast to this result, Asturian and Asturian Spanish are rated higher than Spanish and Catalan in solidarity dimensions such as trustworthiness, kindness, friendliness, etc. The different evaluations provided by the participants support the results obtained in the sociolinguistic surveys that I reviewed in this section regarding the domains of use of each language and the social status of each variety.

2.3 Interference and linguistic variation in Asturian Spanish

The prolonged contact between Spanish and Asturian has led to several degrees of linguistic interference, facilitated by the typological similarity between the local and the national language. A manifestation of such interference is the coexistence of Asturian features with the corresponding Spanish ones in the language variety spoken in the urban
areas of Asturias. The hybrid nature of this variety (which I will refer to as Asturian Spanish from here on), has led some scholars to focus on the linguistic effects that the contact situation between Spanish and Asturian has had.

2.3.1 Asturian Spanish as a continuum

Part of the research on Asturian Spanish has focused on describing in general terms the nature of the mixture between the two linguistic systems. Many researchers have defended the existence of a continuum of mixed language varieties between Asturian and Spanish (Konvalinka 1985, Llera Ramo 1994, Antón 1995, D’Andrés 1997, Dyzmann 2000, Arnold 2001, González Quevedo 2001). In previous research the denomination “Amestau” (Asturian for “mixed”) has been used to refer either to the whole spectrum of varieties or to a particular point on it. Dyzmann (2000), for example, stated that “Amestau” denominates the variety spoken in the cities and that it is sometimes understood as an intermediate point between Asturian and Spanish. She then explained that, instead of a three-way distinction between Asturian, Amestau and Spanish, the different linguistic varieties present in Asturias should be located at different points of the Asturian/Spanish spectrum. Thus, even though “Amestau” is used in sociolinguistic studies to refer to different manifestations of interference between the two languages, because it is possible to interpret it only as a particular point in the continuum and not as the continuum itself, I have opted for the expression “Asturian Spanish” in my research.
Speakers of the mixed system are normally aware of the fact that their linguistic variety is neither Asturian nor Spanish (D’Andrés 1993, 2002; González Quevedo 2001). D’Andrés (2002) mentioned the use of different names for Asturian and for the mixed system respectively. For the mixture he included terms like “mezcla” (‘mixture’), “asturiano” (‘Asturian’), “asturiano corriente” (‘normal Asturian’), “asturiano de aquí” (‘Asturian from here’) and “amestáu” (‘mixed’). In my own research I have also registered names like “español asturianao” (‘asturianized Spanish’) or “español malo” (‘bad Spanish’). In contrast with this, speakers use terms like “Bable”, “asturiano cerrao” (‘strong Asturian’), “muy asturiano” (‘very Asturian’) and “asturiano de verdad” (‘real Asturian’) to refer to Asturian. The idea of a continuum between the two linguistic systems is thus reflected in some of the labels used by the speakers, such as “muy asturiano” or “asturiano cerrao”. In a sociolinguistic survey performed throughout the entire region of Asturias, Llera Ramo (1994:63) concluded that only half of Asturians consider “asturiano” and “bable” to be the same thing, reserving the term “asturiano” for “what people speak in their normal life” and “bable” for Asturian. When asked about their own linguistic variety, Llera Ramo reported the following responses from speakers:

(1) “hablo mezclado el castellano y el asturiano”
(‘I speak a mixture of Spanish and Asturian’)

(2) “hablo así, una mezcla de asturiano, alguna palabra en castellano, otra en asturiano”
(‘I speak like that, a mixture of Asturian, a Spanish word here, an Asturian one there’)

26
“Es un asturiano castellanizado, con palabras, acentos. No es asturiano lengua, es esta mezcla de los dos”
(‘It’s a castilianized Asturian, with words, accents, it’s not Asturian the language, it’s this mixture between the two’)

According to Antón (1995), who also mentioned the distinction made by the speakers between “asturiano” and “bable”, many Asturians believe that they speak “incorrect” Spanish but at the same time they know that their speech is not too far from standard Spanish and that “bable” is in some way present in their variety.

D’Andrés (1997, 2002) establishes two stages between the Asturian and Spanish poles of the continuum: Minimal Asturian, where the matrix language is Asturian and has a maximum degree of influence from Spanish, and Minimal Spanish, where the matrix language is Spanish. In order to determine the matrix language, the author uses the relative “weight” of the features from each language but he does not explain how this weight can be measured. He also recognizes the possibility of additional “contact varieties” that could be situated at other points in the continuum. D’Andrés’ characterization raises questions regarding the matrix language of the mixed systems. First, the differentiation between Minimal Spanish and Spanish seems empirically unmotivated, since the characteristic features of Minimal Spanish coincide with alternations also observed in other geographic varieties of Spanish and cannot be attributed to Asturian influence (such as deletion of intervocalic /d/ in past participles and the absence of leísmo). Second, claiming Asturian as the matrix language of Minimal Asturian – the system that actually reflects interference from the autochthonous language – is arbitrary. The author justified an Asturian matrix language by alluding to the heavier
weight that Asturian features have in the mixed variety and to the diachronic
development by which Spanish was imposed in the area on an existing Asturian system.
However, it is unclear how the relative weight of features from each language could be
measured. Other authors have also assumed Asturian as the matrix language in their
descriptions. For example, Arnold (2001) defined “Mixed Asturian” (“asturiano
amestao”) as an Asturian system that shows interference from Spanish.

Dyzmann (2001) maintains a different position regarding the matrix language. In
her study, she evaluated cases of interference from Asturian based on the awareness
speakers have of such interference. Her participants were a group of 50 students in a high
school in Gijón. She concluded that, because of the students’ high degree of awareness of
grammatical interference from Asturian and their higher linguistic competence in
Spanish, their variety should be characterized as Asturianized Spanish. Even though the
connection between the awareness of linguistic interference and the existence of a matrix
language is not clear in her explanation, the statement serves as evidence that different
interpretations are possible regarding the nature of the matrix language.

It is also possible that a different matrix language can be determined for different
demographic groups. For example, Amado Rodríguez et al. (1993) examined the
presence of Asturian features in the spontaneous speech of speakers from Oviedo and
observe that as age decreases, the use of Asturian features also decreases. They
concluded that in speakers older than 27 the matrix language is Asturian, whereas in
younger generations the matrix language is Spanish.

The denomination of “Asturian Spanish” does not attempt to make any claims
regarding the existence of a matrix language. Instead, it will cover any variety in the
continuum between Asturian and Spanish found in the region, regardless of whether it
could be considered mostly Spanish or mostly Asturian, as long as it shows some degree
of interference between the two linguistic systems.

2.3.2 Differentiating features and implicational scales

D’Andrés (2002) provided a detailed characterization of the continuum that was
described in the previous section. In his study he considered the variety spoken in the
urban areas (which he called Minimal Asturian) as an originally Asturian system that has
experienced massive influence from Spanish. According to the author, this system is
characterized by the presence of a set of Asturian features, which he referred to as
“differentiating features”. These features are phonological and morphosyntactic and
include the following:

- Use of the morpheme /-u/ to mark masculine singular in nouns, adjectives and
  pronouns, where Spanish uses /-o/ (“el perru” vs. “el perro” – ‘the dog’).

- Use of the morpheme /-es/ to mark plural in nouns, adjectives, determiners and
  pronouns where Spanish uses /-as/ (“las cases” vs. “las casas” – ‘the houses’).

- Use of /-i/ instead of Spanish /-e/ in the demonstratives “esti” and “esi”.

- Use of /-es/ and /-en/ in the second person singular and third person plural
  respectively in verbal endings where Spanish uses /-as/ and /-an/ (“ellos bailen”
  vs. “ellos bailan” – ‘they dance’).
o Use of definite articles with prenominal possessives (“el mi coche” vs. “mi coche” – ‘my car’).

o Enclisis of unstressed pronouns in cases where the prescriptive position in Spanish is proclitic (“hizote gracia” vs. “te hizo gracia” – ‘you thought it was funny’).

o Use of the direct object and indirect object pronoun “vos” for second person plural where Spanish uses “os” (“si vos llamen” vs. “si os llaman” – ‘if they call you (plural)’)

o Use of the indirect object pronouns “y” and “yos” for the third person singular and plural respectively, where Spanish uses “le” and “les” (“dije-y la verdad” vs. “le dije la verdad” – ‘I told him/her the truth’). The Asturian forms are derived from the Latin ILLI, with “yos” being a result of the extension of the pattern found in the other plural forms, “nos” and “vos”.

o Deletion of /-r/ in the infinitive forms when an enclitic pronoun follows (“tengo que decítelo” vs. “tengo que decírtelo” – ‘I have to tell you’)

o Absence of present perfect tense and exclusive use of preterit (“canté” vs. “he cantado” – ‘I sang’ vs. ‘I have sung’).

o Use of diphthongized forms in the second and third person singular of the present indicative of “ser” (“tú yes” vs. “tú eres”, “él ye” vs. “él es” – ‘you are’, ‘he is’)

o Deletion of final /-e/ in the forms of the third person singular in the present indicative of verbs where the stem ends in a vowel preceded by /n/ or /θ/ (“él diz” vs. “él dice”, “él tien” vs. “él tiene” – ‘he says’, ‘he has’).
D’Andrés stated that speakers recognize this set of features as belonging to the Asturian linguistic system. He also claimed that the features appear as a cohesive block in the speech of the urban population, explaining that speakers of Asturian Spanish choose a particular variety of the language and not single features. However, Barnes and Schwenter (2013) found that this is not true for all the differentiating features and that only some of them imply the presence of the rest. In the analysis of the speech of 16 speakers of Asturian Spanish we found that there exists a hierarchical implicational scale of features (Barnes and Schwenter 2013, Rickford 2002). As Rickford (2002:143) explains, “Implicational scales depict hierarchical co-occurrence patterns in the acquisition or use of linguistic variables by individuals or groups, such that $x$ implies $y$ but not the reverse”. In our study we found that the use of /-u/ and the diphthongized forms of the verb “ser” do not imply the presence of any other Asturian feature. The use of raised vowels in the verb paradigm implies the presence of raised vowels in the nominal categories and the presence of Asturian clitics implies the placement of unstressed pronouns in enclitic position. Finally, we observed that the mean number of Asturian features used by speakers that are older than 35 years old is higher than that of speakers younger than 35. We concluded that, while features that are reductive in nature, such as the deletion of /-r/ in infinitives or the deletion of /-e/ in third person singular forms of present indicative, are persistent across all speakers, substitutive or additive features, such as the use of the definite article before possessives or the use of Asturian
object pronouns, are being lost. This hierarchical organization supports a strategic use of Asturian features by the speakers in order to index more or less “Asturianness”.

Whether a speaker incorporates one, some, or all of the differentiating features in their speech, the use of each one of them is usually not categorical; instead, they normally alternate with the Spanish variants. Thus, speakers do not only show variation in how many or what differentiating features they use but also in how often and when they use the Asturian variants as opposed to the Spanish ones.

2.3.3 Linguistic variation in Asturian Spanish

The rural varieties of Asturian have received most of the attention in the field of Asturian linguistics, in an attempt to describe the most archaic varieties of the language before language shift takes place. However, some scholars have seen a need to describe the urban varieties of the language, paying special attention to interference between Spanish and Asturian.

The first study to point to the necessity of exploring the linguistic diversity found in urban areas of the region of Asturias is that of Sánchez Álvarez (1979). The author explored the variety of the language spoken in Oviedo, the second largest urban area in the region and capital of the region. She noted that the linguistic varieties of the urban areas are not homogeneous and that speakers alternate between the use of Spanish and Asturian features. Her data came from recorded conversations between family members and groups of friends. All of her informants were native to the city of Oviedo and resided in an area that has not received much immigration. She excluded informants that had a
college education and those that worked in administrative positions, claiming that the attitudes towards the regional language in these groups varied significantly from speaker to speaker.

In her qualitative analysis of linguistic data Sánchez Álvarez noted that features from Asturian are present in the lexicon, the phonology and the morphosyntax of the language, but, according to her, it is at the morphosyntactic level where their presence is more stable and less noticed by the speakers. She explored the use of unstressed pronouns, final vowels, the verbal paradigm and certain lexical items. The author noted that Asturian variants are reserved for informal conversations that take place in a relaxed and familiar environment. She also found that the Asturian lexical items that are used in place of their Spanish equivalents, have an affective or expressive meaning that the Spanish variants lack. As an example she provided the utterance reproduced in (4):

(4) “¡Ay Dios! ¿vístelu? Ta gordu como un xatu.”
(‘Oh God! Did you see him? He’s fat as a calf’)

In (4) the use of Asturian “xatu” instead of Spanish “ternero” makes the expression more exaggerated and adds humor to the characterization.

Prieto (1991) analyzed the speech of two women from Gijón, who were characterized as speakers of a hybrid variety. Both informants were customers in a market that is located in a central part of the city. The author performed a qualitative analysis of two hours of conversation, evaluating the presence of phonological and morphosyntactic features, as well as the use of the lexicon, focusing most of the analysis
in the use of final vowels and the form and placement of unstressed pronouns. In her conclusions she noted a “general confusion” of the two linguistic systems, with the Asturian variants being prevalent in the morphology.

Prieto did not address the influence of linguistic factors on the choice between the Asturian and Spanish variants. She did point to some possible external factors that might favor the use of Asturian features in the conversations analyzed. According to the author, it is not the social class to which the speaker belongs that constrains the amount of Asturian features that are included in their speech. Instead, she observed an effect of the familiarity between speakers and their interlocutors, finding more Spanish features if conversations take place in an official setting, where the speakers are less likely to know each other. Related to this is the observation that the topic of conversation seems to also have an effect on the choice of Asturian or Spanish features, with the latter being reserved to talk about “serious” and “prestigious” topics of conversation.

Amado Rodríguez et al. (1993) made a similar observation regarding topic of conversation. The authors mentioned the example of a speaker that uses more Spanish features when reporting what a doctor said in a conversation than when reproducing her own words. In their study they examined the linguistic results of the contact between Spanish and Asturian in the urban area of Oviedo, particularly in the parroquias of Trubia and Godos. Their data came from interviews of 250 speakers that belong to the middle social class. The data was collected using linguistic and lexical surveys and sociolinguistic interviews. They analyzed the use of phonological, syntactic and lexical variables. They concluded that the most important factors constraining the variation between Spanish and Asturian forms were age and context: older speakers used more
Asturian features in informal contexts than younger speakers (under 27) and the percentage of Asturian features decreases in formal contexts for both groups. Women also seemed to favor the use of Asturian features.

D’Andrés (1993) analyzed the maintenance of Asturian features in Deva, a community close to the urban center of Gijón. He used a questionnaire in which he asked each of the 32 participants to provide the lexical item that referred to a particular concept. Even though he did not analyze the distribution of particular linguistic variables, he classified the answers provided by the speakers according to three different categories, based on the phonemes present in the lexeme and morphemes of each lexical item: predominantly Spanish, a mixture between Spanish and Asturian and predominantly Asturian. He concluded that the answers that were predominantly Asturian were the most common among the informants (57%), with the older speakers favoring this type of answer. This contrasts with the preference for Spanish answers observed in the younger speakers. He also pointed out that Asturian was preferred when the lexical items referred to objects associated with rural life.

As we can see, all the studies that have addressed the general linguistic variation present in urban areas of Asturias provide some external causes, such as the age and gender of the speaker, that might constrain the choices made by the speakers. However, most of the studies ignore the possible linguistic factors that may also have an effect on the appearance of Spanish or Asturian features. The analysis of linguistic factors is crucial due to the existence of intra-speaker variation, which has been pointed out in previous research. This variation could be governed by constraints that go beyond the style shifts mentioned by some authors; i.e. even within the same topic of conversation
and in interactions with the same addressee, speakers might still alternate between Asturian and Spanish features. Thus, in order to fully describe the distribution of the linguistic variables under study it is necessary to include internal factors. Analyzing the role of linguistic factors further allows us to discover patterns of behavior that might also be observed in other language varieties, which enable the identification and understanding of general linguistic mechanisms that guide language variation and change (Tagliamonte, 2011). These linguistic factors could also be intricately linked to social variables. In this dissertation I explore both social and linguistic factors, analyzing the way in which they constrain the variation between Asturian and Spanish features.

2.4 Asturian and regional identity

Speakers use language to relate to a particular geographic location. This relationship is not a straightforward one; instead, it is a complex relationship that interacts with other social categories, such as gender, age, social class and education. Recent work on sociolinguistics has paid an increasing attention to the relationship between language, regional identity and ethnicity (Woolard 1989, Eckert 2004, Johnstone, Andrews and Danielson 2006, Myers 2006, Johnstone and Kiesling 2008, Johnstone 2010).

For instance, Johnstone et al. (2006) explored the process of “enregisterment” in Pittsburghese. According to the authors, certain linguistic forms in the dialect under study, which were once unnoticed by the speakers, became forms that were used to mark social class first and then regional identity. This process was enabled by economic and
geographic changes in the society under study. The first shift was triggered by the increasing social mobility of working-class Pittsburghers, which made them aware of the existence of linguistic variants different from their own. Once the forms became variable, the choice of one variant over the other was associated with a particular social identity (in this case class and correctness). In the enregisterment process, the non-standard forms, while they continued to index social class, were linked to a particular local identity and became markers of place. Speakers who use these linguistic features variably are capable of controlling how much they use each of the variants involved depending on how much they want to sound like they belong to a particular regional identity.

I hypothesize that a similar process, which resulted in the use of Asturian features to index certain social categories but also a particular regional identity, has taken place in Asturian Spanish. The main external change that triggered this phenomenon in Asturias was the process of industrialization and the consequent migration of population from rural to mining and industrial areas. As I mentioned in section 2.1.3, this migration resulted in the contact between Spanish and Asturian speakers and an increased awareness of the differences between the linguistic systems. Asturian was associated with rural and uneducated speech and its features began to index low socioeconomic status, incorrectness and low education but also locality and regional identity. In the following section I explore the connection between the general use of Asturian and regional and local identities.
2.4.1 Linguistic interference and regional identity

The link between regional identity and the use of the regional language in Asturias does not seem to be as strong as the one observed in other bilingual communities in the Peninsula, such as Catalonia and Galicia. In the case of Catalan, this could be attributed to the fact that the use of Catalan is not stigmatized. Woolard (1989) examined the relationship between ethnicity and language use in Catalonia. She tested the listeners’ perceptions of Catalan and Spanish and found that Catalan is given higher scores in solidarity but also status scores. This shows that speakers value in-group use of Catalan but it also points to the prestige of the regional language. In this case, it seems that the official status of Spanish has not secured the linguistic prestige that one might expect. Woolard attributes this to the greater economic power of Catalans, as compared to that of other regions in Spain.

Another possible reason is that the relationship to the outside, which is crucial in determining the social identity of a group (Eckert 2004), is slightly different in the case of Asturias. Bucholtz and Hall (2004) talk about distinction, the opposite of linguistic adequation, as a tactic to construct identity and mention how it is used by speakers of minority languages in some cases to differentiate themselves from the state. As the authors describe, “differentiation of identity is a way of resisting the relentless march of the assimilating forces of modernity and the nation-state” (Bucholtz and Hall, 2004, p.384).

Konvalinka (1985) explained that the historical lack of tension between Asturias and Castile (and the centralist government) has led to a lack of strong regionalism. This tension is present in Galicia, Catalonia and the Basque Country, leading to a completely
different linguistic outcome, in which the regional language is still maintained and used by speakers for in-group communication. In their study of the use of Catalan by high school students in Cataluña, Giles and Viladot (1994) concluded that the strongest predictor of self-reported linguistic differentiation with out-group speakers was disassociation with the Spanish identity. In contrast with this, Rose Alcorta (2000) qualified Asturian identity as “passive” or “latent” and manifested in an absence of the strong animosity towards Castilian that we find in the other bilingual regions. This is also reflected in the social information collected by Llera Ramo (1994) according to which the majority of Asturian speakers identify themselves as both Asturian and Spanish, contrary to what is observed in the Basque Country and Cataluña, where most speakers identify with the regional identity. Similar results were reported by Rose Alcorta, with almost 80% of the speakers claiming to be equally Asturian and Spanish.

Despite the weaker nature of regionalism found in Asturias, the link between the use of Asturian and regional identity is still present. Furthermore, Prieto (1991) claimed that there is a strong connection between Asturian Spanish and the city of Gijón. The city of Gijón is different from other cities in Asturias in that it is an industrial city that attracted much more population from the rural areas. As a consequence the presence of Asturian has been more prominent in this city. The connection between Asturian Spanish and the city of Gijón seems to have a double function: linking Gijón to the regional identity and distancing the urban dialect from the rural world. The desire to disassociate from the non-urban identity can also be perceived in some sociolinguistic studies. For example, Rose Alcorta noted that when asked what the language of the community is, the response of most speakers in the small city of Candás was Spanish (followed closely by
“both languages”); however, when they were questioned about the language of Asturias, the majority of the speakers responded with Asturian (with “both languages” again following at a short distance). Thus, even though speakers from urban areas recognize that Asturian is the language of Asturias, they perceive the variety spoken in the city differently from that of the region.

Sánchez Álvarez (1979) considers the city as a linguistic unit opposed to the village and claims that the speakers are very aware of this opposition. Even though speakers recognize that the speech in the urban areas is different from that of rural areas, they might not be able to identify all the features that come from Asturian. Sánchez Álvarez included the following quotes from the participants from her study:

(5) ¿Aquí?, ¿en las ciudades?; el bable no se habla. Eso ye en les aldees.
   ‘Here? in the cities?; Asturian is not spoken here. That’s in the villages’

   ‘Those things are not from here. They speak very rural. Of course, silly, they grew up in the village’

In these examples the speakers from an urban area use some Asturian features (in bold) while they state that Asturian is not spoken there but only in the villages. The author also pointed out that even though the city constitutes a linguistic unit, the system is not homogeneous due to the heterogeneity of its inhabitants. The observations made by Sánchez Álvarez illustrate the contrast that exists between urban and rural identities and
support the idea that the alternative use of Asturian and Spanish variants contributes to the construction of both local (urban in this case) and regional identities.

2.5 Conclusions

In this chapter I have outlined the historical context that gave rise to the contact situation that is observed today in Asturias and the social and political events that shaped the language attitudes that Asturians have toward the regional and national languages. I showed that the development of Asturian is diachronically parallel to that of Spanish; i.e. that Asturian is not a dialect of Spanish. The characterization of Asturian as a dialect of Spanish is only a product of negative attitudes towards the language and not a linguistic reality. Despite a change towards more positive attitudes and the efforts made to normalize and revitalize the use of the regional language, Asturian and Spanish are in a situation of unbalanced bilingualism that reserves Asturian for informal contexts and trivial topics of conversation.

In this section I also reviewed the nature of the mixed system and the studies that have addressed linguistic interference between Spanish and Asturian. These studies focus on the variety spoken in the urban areas of the region and base their conclusions on the qualitative analysis of spontaneous speech and the information obtained in sociolinguistic surveys. Most of these studies draw conclusions regarding the social and stylistic factors that govern the choice between Asturian or Spanish variants. The traditional use of a dialectological approach together with the lack of attention to linguistic factors in the few sociolinguistic studies that address variation in Asturian Spanish call for a new analysis
that can fully account for the distribution of Asturian and Spanish variants. This type of analysis is a necessary first step to elucidate the role that variation has in the construction of regional and local identities, whose relationship with Asturian and Asturian Spanish was outlined in the last section of this chapter.
CHAPTER 3: THE LINGUISTIC VARIABLES

3.1 Introduction

The use of Asturian /-u/ as the masculine singular morpheme, marked in Spanish with /-o/, and /-es/ as the plural morpheme, where Spanish uses /-as/, are two of the differentiating features outlined by D’Andrés (2002). The presence of these features in speech characterizes the linguistic variety spoken in the urban areas of Asturias. Both features involve the use of a higher vowel than the one used in the corresponding Spanish morphemes, thus I will refer to the Asturian variants as raised vowels from here on. The use of raised vowels in Asturian Spanish is not categorical. Instead, speakers alternate between the Asturian and the Spanish variants. Examples (7) and (8) illustrate how the same speaker alternates between the two variants:

(7) “Ta’ to’l dia tocándote les narices” (GI11-04)
    ‘She’s bothering you all day’

(8) “bueno, me toca un poco las narices” (GI11-04)
    ‘She bothers me a little’
The variation observed is a consequence of the borrowing of morphemes from Asturian; i.e. the presence of /-u/ and /-es/ in the variety under study is constrained morphologically so that the raised variants only appear as a marker of particular morphological categories. Even though the existence of phonetic constraints on the variation is possible, the envelope of variation is limited to certain morphological contexts. Variation between /-o/ and /-u/ is only observed in masculine singular nouns (mass and count), adjectives that qualify masculine singular count nouns, direct object pronouns that have masculine singular count nouns as referents, possessive and indefinite adjectives and ordinal numbers. The alternation between /-as/ and /-es/ is restricted to feminine plural nouns (or masculine plural nouns for which the corresponding singular form has /-a/), adjectives that qualify feminine plural nouns, demonstratives, possessive and indefinite adjectives, subject pronouns and direct object pronouns that have feminine plural nouns as referents.

In the remainder of this chapter I describe the pertinent elements of the Asturian vocalic system and outline the envelope of variation of the two linguistic variables, indicating what the possible variants are and in what contexts the alternation between them is observed. I also review the contributions made by previous research to the understanding of the distribution of the linguistic variables, focusing on the internal and external factors that have been found to govern the alternation between the Asturian and the Spanish variants.
3.2 The linguistic variables

There are several reasons that justify the choice of the variables that are the object of this dissertation. First, raised vowels in final position have been described as one of the most salient features that are associated with Asturian identity. For example, Antón (1995) stated that final /-u/ is the feature that is most often associated with Asturian pronunciation. Furthermore, the use of raised vowels in final position is still a vital phenomenon in Gijonés, likely to appear even in the speech of speakers that show minimal interference from Asturian.

Raised vowels can also appear in Asturian Spanish in unstressed word-internal positions. In this dissertation I decided only to include the variation in vowels in final position (whether they are in an open or in a closed syllable) for two reasons. First, the analysis of the data showed that raising in other unstressed positions is very rare in the variety under study, registered sporadically in only three lexical items: “piquiñin” (as opposed to “pequeñín” – ‘little’), “puquitín” (as opposed to “poquitín” – ‘a little’) and “prubitín” (as opposed to “probitín” – ‘poor thing’). Second, the nature of the two processes is different: as I mentioned above, even though it is possible that both phenomena are at least partially phonetically motivated, the appearance of raised vowels in final position is constrained morphologically, while the raising observed in other positions is not.

\[2\] In the Asturian lexical item used for Spanish “pobre” (‘poor’), /r/ is in the onset of the first syllable instead of in the second syllable, resulting in “probe”.

45
3.2.1 Unstressed vowels in Asturian

The vocalic system of Asturian followed a similar path to that of Spanish in its development from Latin. Today, Central Asturian and Spanish share the same set of vowels in stressed and unstressed positions: /i/, /e/, /a/, /o/ and /u/. However, there are some differences between both systems in unstressed positions. Even though in most varieties of Asturian, including the area where the variety of Asturian Spanish under study is spoken (Central Asturian), a differentiation between the five vowels in unstressed position (word-final and internal) is still maintained, there is a general tendency to raise these vowels. This tendency is also registered in Asturian Spanish, for which acoustic evidence will be provided in section 3.2.4. In general we observe that the difference in formant values between [o] and [u], and [a] and [e], respectively, is greater in Spanish than in Asturian Spanish; i.e. each pair of vowels is closer to merging in Asturian Spanish than in Spanish.

In some varieties of Western Asturian, such as the variety spoken in Palacios del Sil, vowel raising has resulted in a reduction of the vocalic system to three units in unstressed positions: /I/, /a/ and /U/ (García Arias, 2003). This three-unit system is general in word-final position in all varieties of Western Asturian and limited to specific geographic areas in word-internal position (Dyck 1995).

García Arias claimed that while the vocalic system of Central Asturian is a direct outcome of the Latin system, several internal readjustments have taken place that differentiate the Asturian system from the Spanish one. The development of the pertinent sounds is reviewed in sections 3.2.2 and 3.2.3 respectively.
Raising of final vowels is not an uncommon phenomenon in other languages, including some varieties of Spanish. For example, Holmquist (1998) and Oliver Rajan (2007) analyzed the raising of /-e/ and /-o/ in word-final position in the coffee zone in Puerto Rico. Parodi and Santa Ana (1998) also registered raised final vowels in the Spanish spoken in Michoacán, Mexico, a phenomenon that is further explored by Barajas (2013) in the Spanish of Colongo, Michoacán. Vowel raising was also found as a result of language contact, such as in the Quechua-Spanish contact in Perú (O’Rourke 2010) and Ecuador (Guion 2003), and as a mechanism of hiatus resolution (Garrido 2007, Hernández 2009). The fundamental difference between these cases of vowel raising and the phenomenon observed in Asturian Spanish is that the use of raised variants in Asturian Spanish is restricted to particular morphological environments.

3.2.2 /-o/ vs. /-u/

In Central Asturian /-o/ and /-u/ are contrastive morphemically. While /-u/ is used in masculine singular nouns, and adjectives and pronouns that agree with masculine singular count nouns, /-o/ is used to mark “mass” in adjectives and pronouns (Penny 1994, D’Andrés 1994, 1998; Camblor Portilla and Wood Bowden, 2005; Fernández Ordóñez, 2006). The examples in (9) show the contrast between the two morphemes:

(9) “el vinu frío sí lo tomo” (mass) ‘cold wine, I do drink’
    “esi llibru gordu no lu llei” (count) ‘that big book, I didn’t read’
Several theories have been proposed to explain the use of /-o/ and /-u/ to express the mass/count distinction.

García Arias (2003) claimed that the potential merger between /o/ and /u/ in Asturian created a desire among the speakers of Central Asturian to phonologize the existing semantic distinction between mass and count. Thus, he postulated a pre-literary stage in which only one phonological unit was maintained for the back vowels, /U/, which had different possible realizations that varied in vowel height, a situation that we currently find in Western Asturian (Dyck 1995). In an attempt to further differentiate mass from count units, the higher variant [-u] was phonologized to refer to count and the lower variant [-o] to refer to mass. Western Asturian followed the same path but the distinction between /-o/ and /-u/ was only maintained in third person singular unstressed pronouns.

The question that arises is why the realization [-o] was associated with mass and [-u] with count. García Arias explained that, the fact that in Latin [-o] was found in word-final position in many adverbs and verbal endings, where an association with any gender was not possible, can explain the current distribution of final back vowels. Penny (1970) claimed that Asturian [-o] was adopted from dative and ablative /-o/ when the case system was still partially productive. Finally, other authors, such as Lüdtke (2003), maintain that the distinction arose from the existing opposition between ILLUM and ILLUD, which resulted in “lu” and “lo” respectively in Asturian. The opposition present in the unstressed pronouns would then be extended to adjectives.

Whether it was due to the influence of adverbs and verbal endings or to an expansion of the contrast between “lu”/“lo”, Central Asturian maintained the opposition
in back vowels in word-final position, which was lost in Spanish in favor of /-o/ and in Western Asturian in favor of /-u/, to represent the semantic distinction between mass and count.

3.2.2.1 Distribution of /-o/ and /-u/ in Asturian

As we saw in the previous section, the two back vowels are possible in final position in some lexical categories in Asturian. In addition to descriptive adjectives, the morphemic contrast between /-o/ and /-u/ can be observed in the third person singular direct object pronouns (“lu” vs. “lo”), ordinal numerals (“primero” vs. “primeru”) and in quantifier (“muncho/u”, “demasiao/u”), possessive (“nuestro/u”, “mío/u”) and indefinite adjectives (“mesmo/u”, “otro/u”) when they appear postponed to the noun (D’Andrés 1994, Academia de la Llingua Asturiana 1998).

It is important to remember that the morphemic contrast between the two back vowels in final position does not occur in other varieties of Asturian, such as Eastern Asturian, where mass is also marked with /-u/ and the distinction can only be observed in the agreement between feminine mass nouns and their accompanying adjectives or pronouns that have them as referents. In Western Asturian the opposition between mass and count nouns does not exist. However, since the variety spoken in Gijon is normally classified as belonging to Central Asturian, the mass/count distinction is of most relevance.

---

3 Mass/count agreement in prenominal position has been occasionally observed in the case of quantifiers (Academia de la Llingua Asturiana 1998).
As Dyck (1995) explained, the mass category “denotes mass objects, abstract qualities, and substances in general, as opposed to count objects, concrete qualities, and specific items” (p. 31). It should be noted that what Asturian displays is not a three-way distinction between masculine, feminine and mass; instead, mass marking is superimposed on that of gender so that singular nouns are assigned a masculine or feminine gender and then categorized as either mass or count. D’Andrés (1994) provided the following examples to illustrate the combination of gender and mass marking:

(10) el carbón prieto (masc. sg. mass) ‘black coal’
la fariña blanco (fem. sg. mass) ‘white flour’

The agreement between nouns and adjectives and referents is illustrated in (11) and (12):

(11) el vinu malo (masc. sg. mass) ‘bad wine’
la leche malo (fem. sg. mass) ‘bad milk’
la neña mala (fem. sg. count) ‘bad girl’
el neñu malu (masc. sg. count) ‘bad boy’

(12) el vinu, nun lo quiero ‘wine, I do not want it’
la leche, nun lo quiero ‘milk, I do not want it’
a la neña, nun la quiero ‘the girl, I do not love her’
al neñu, nun lu quiero ‘the boy, I do not love him’
Even though nouns can be semantically classified as mass or count, prescriptively they do not show any morphological marking of this classification in Asturian and have \(-u\) categorically (D’Andrés 1994, Fernández Ordóñez 2006). The mass/count distinction in nouns is only reflected in the agreement of adjectives that accompany them or in the pronouns that have them as referents (as shown in (4) and (5)). While in prescriptive Asturian masculine singular nouns appear invariable with \(-u\), alternations between back vowels in nouns have been reported in spoken Asturian in a limited number of cases. D’Andrés (1994, 1998) referred to the few examples in which there exists an alternation between the two back vowels in Asturian nouns. According to the author, in some nouns, the innovative use of \(-o\) might actually respond to a desire on the part of the speaker to mark mass nouns. That is the case for nouns that can be interpreted as both count and mass, such as “fieru/fierro”, in which the former refers to a single piece of iron while the latter is used to denominate the material. D’Andrés also included in this category nouns that are normally not used in their plural form, such as “mundo” (‘world’) and “cielo” (‘sky’). García Arias (2003) also noted the presence of an alternation between \(-o\) and \(-u\) in nouns in spoken Central Asturian and agrees with D’Andrés in that the presence of \(-o\) in nouns responds to an attempt to expand the mass/count distinction found in adjectives and pronouns to nouns, and not to the influence of Spanish.

A similar explanation has been provided by some authors to account for the variation between \(-o\) and \(-u\) in Asturian Spanish. In her study, Prieto (1991) analyzed the variation found in the speech of two women in a market in Gijón. Even though she did not provide any quantitative analysis of the variation, she registered examples like “pescao” (‘fish’), “tabaco” (‘tobacco’), and “hígado” (‘liver’). It is unclear whether the
presence of /-o/ in these lexical items in Asturian Spanish could be attributed to the Spanish masculine singular morpheme (due to language contact) or to the mass/count distinction present in Asturian (or to both). Prieto opted for the latter interpretation, claiming that, because such nouns could be semantically interpreted as mass nouns, the presence of /-o/ is a reflection of the mass/count distinction described for Asturian. The rest of the variation in final back vowels is explained by alluding to the Spanish origin of some words that show /-o/, in cases when an Asturian equivalent exists (e.g. “el marido” vs. “l’home” (‘husband’)).

Because of the existence of the mass/count distinction in Asturian, one of the objectives in the analysis of this variable will be to determine whether the appearance of the mid back vowel is constrained in nouns by the possible mass categorization of the noun as mass.

3.2.2.2 Envelope of variation

Variation between final /-o/ and /-u/ in Asturian Spanish is observed in the following lexical categories and contexts:

a) All masculine singular nouns that follow the /-u/ (sg.) and /-os/ (pl.) paradigm in Asturian (/-o/ and /-os/ respectively in Spanish). This category includes both mass and count nouns, as the following examples illustrate:
(13)  a. “el picar el escombru pa sacar toda la tubería y demás” (masc. sg. mass)
    ‘smashing the debris to get the piping and the rest out’ (GI11-11)
    b. “él puso les perres y yo puse el trabajo” (masc. sg. mass)
    ‘he put the money and I put the labor’ (GI11-04)

(14)  a. “nunca jamás-y conocí un noviu a mi madre” (masc. sg. count)
    ‘I never met any of my mom’s boyfriends’ (GI11-04)
    b. “yo no salí de esti barrio” (masc. sg. count)
    ‘I never left this neighborhood’ (GI11-04)

b) Descriptive adjectives that have three endings (/-u/, /-o/, /-a/) in the singular paradigm in Asturian (including past participles used as adjectives) and that accompany or refer to masculine singular nouns:

(15)  a. “tuve un problema muy gordu” (GI11-15)
    ‘I had a really big problem’
    b. “es guapo esi salón” (GI11-04)
    ‘That living room is pretty’

(16)  a. “tuvo que comprar un baju pegáu” (GI11-11)
    ‘He had to buy an adjacent local’
    b. “yo sabía que estaba aprobáo (el examen)” (GI11-25)
    ‘I knew that it was passed (the exam)’
During the analysis of the data I did not record any adjectives that show agreement in /-u/ with masculine singular mass nouns, indicating a categorical use of /-o/ in adjectives that qualify masculine singular mass nouns; i.e. variation between /-o/ and /-u/ in adjectives only takes place in adjectives that qualify masculine singular count nouns\(^4\). Consequently, only adjectives that accompany or refer to masculine singular count nouns were included in the analysis, excluding examples like (17), where the adjective refers to a masculine mass noun:

\[(17) \quad \text{“ta el tiempu muy malo” (GI11-18)}
\]

The following possessive adjectives when they agree with masculine singular count nouns:

i. First and second person singular forms and third person form (singular and plural) when they are not in a prenominal position: “míu”/“mío”, “tuyu”/“tuyo” and “suyu/suyo”. Note that these forms differ from the Asturian equivalents, where the second and third person singular possessives are “to” and “so” (equal for masculine and feminine).

ii. First and second person plural forms in any position: “nuestru/nuestro” and “vuestru/vuestro”.

\(^4\) Even though in Asturian adjectives that qualify feminine mass nouns also appear with /-o/, this use has been lost in the variety of Asturian Spanish under study, such that adjectives that qualify any feminine noun adopt a feminine morpheme. Thus, in Asturian Spanish we find “la leche mala” and not “la lleche malo” (‘bad milk’), which is the prescriptive form of the adjective in Asturian.
d) Indefinite adjectives that accompany or refer to masculine singular count nouns, whether they appear in prenominal position or not.

(18)  a. “tuve que ir a otro maestro de otro pueblu” (GI11-10)
     ‘I had to go to another teacher in another village’
   b. “marcho pa los toros con otro amigu” (GI11-10)
     ‘I leave to watch the bullfight with another friend’

e) Third person singular direct object pronouns when they have masculine singular count referents:

(19)  a. “Écho lu mucho de menos (al gato)” (GI11-04)
     ‘I miss it a lot (the cat)’
   b. “Tápeslo con un trapu (el chuletero)” (GI11-10)
     ‘You cover it with a sheet (the rack of chops)’

(20)  a. “Fui de madrina y ahí lu casamos (a mi padre)” (GI11-04)
     ‘I was the godmother and we married him (my dad)’
   b. “Lo limpias un día y al día siguiente está blancu otra vez (el mueble)” (GI11-04)
     ‘You clean it one day and the next day is white again (the piece of furniture)’
As we can observe, the variation is observed in enclitic (19) and proclitic (20) positions and when the referent is animate\(^5\), as in (20)a., or inanimate, as in (20)b. In the case of direct object pronouns, there is also a lack of variation when the referent of the pronoun can be semantically interpreted as mass or is a propositional referent, as in (21):

\[(21)\]
\[\begin{array}{ll}
    \text{a. “mi hermano lógicamente lo hizo (trabajar en la mina)” (GI11-11)} \\
    \quad \text{‘logically, my brother did it (working in the mine)’} \\
    \text{b. “hablen cada cosa que hay que pensalo pa entendelo” (GI11-10)} \\
    \quad \text{‘they say such things that you have to think about it to understand it’}
\end{array}\]

Thus, in the case of unstressed pronouns, examples like the ones in (21) were excluded from the analysis, considering only pronouns that refer to masculine singular count nouns.

f) Ordinal numerals when they accompany or refer to masculine singular count nouns.

\[(22)\]
\[\begin{array}{ll}
    \text{a. “El primeru tuvo cuatro moces (el hijo)” (GI11-10)} \\
    \quad \text{‘The first one had four girlfriends (the son)’} \\
    \text{b. “Subía a un segundo piso” (GI11-11)} \\
    \quad \text{‘He went up to a second floor’}
\end{array}\]

g) Adverbial expressions formed using a masculine singular noun, such as “al cabu” (“al cabo”), “en mediu” (“en medio”), “al lau” (“al lado”), etc.

---

\(^5\) Despite being registered in Asturian Spanish, leismo is not as frequent as in other varieties of peninsular Spanish. A possible reason could be the difference between the dative pronouns in Asturian, “-y” and “-yos” and in Spanish “le” and “les”.

56
(23)  a. “Vivieron un tiempo ahí al lao de mi casa” (GI11-04)
     ‘They lived there for a while, next to my house’
  b. “Trabaja al lau mío además, lo veo bastante” (GI11-02)
     ‘He also works next to me, I see him pretty often’

In addition to the exceptions noted for the categories above, we do not expect to
find variation between /-o/ and /-u/ in Asturian Spanish in the following cases, where /-o/
is the final vowel in both Asturian and Spanish:

• Feminine nouns that end in /-o/, such as “mano” (‘hand’).
• Nouns that end in /-o/ as a result of shortening, as in “foto” (from
  “fotografía” – ‘photograph’) and “micro” (from “micrófono” –
  ‘microphone’).
• Neuter demonstratives, such as “esto”, “eso” and “aquello”.
• Numerals like “cero”, “cuatro” and “cinco”.
• First person singular forms of present indicative: “yo canto”, “yo bebo”,
  etc.
• Gerund forms of the verb, such as “cantando” and “bebiendo”.
• The conjunction “pero”.
• Adverbs, such as “dentro” and “tampoco”.
• Adverbial expressions that are formed using an adjective, as in “poco a
  poco” and “en serio”.

57
In addition to these contexts, variation was not found in the names of days of the week, months and seasons, even if they show final /-u/ in Asturian. Thus, all the tokens that fall in these categories were excluded from the analysis.

3.2.3 /-as/ vs. /-es/

Latin Ā and Ā merged in Asturian, as well as in Spanish, into a single vocalic phoneme /a/ in unstressed final position. However, in many varieties of Asturian, especially in Central and Eastern Asturian, unstressed /-a/ was raised and fronted, resulting in /e/. García Arias finds realizations like the following in nouns in some Central and Eastern varieties of Asturian: “apueste” (instead of “apuesta” – ‘bet’), ruede (instead of “rueda” – ‘wheel’) and “lleñe” (instead of “leña” – ‘firewood’). He also postulates a possible influence of a palatal in the sequence of /-a/.

The process of raising and palatalization resulted in the appearance of the realizations [-es] to mark plurality in Asturian. The realization [e] is also found in the verbal endings that developed from Latin /-AS/ and /-ANT/ and was first documented in the 10th Century (García Arias, 2003: 119). In these cases, the [e] variant was phonologized and is now categorically used in the plural morpheme /-es/ and the verbal morphemes /-es/ and /-en/.

3.2.3.1 Distribution of /-as/ and /-es/ in Asturian

Asturian /-es/ and Spanish /-as/ alternate in the plural morpheme of nouns, adjectives and pronouns that have /-a/ as the morpheme for the singular form. A crucial
difference between this alternation and the one explained for back vowels is that /-as/ is not a morpheme in the Central variety of Asturian; i.e. /-es/ and /-as/ do not contrast in the regional language. Thus, /-es/ appears in Asturian in all the cases in which Spanish marks plurality with /-as/. Because /-es/ is not registered in the Western or Eastern varieties of the language, where /-as/ is the corresponding morpheme, it is possible that the presence of /-as/ in the speech of a speaker could be attributed to a connection with Western or Eastern Asturian (through family origin, for example), instead of to the influence of Spanish.

3.2.3.2 Envelope of variation

Variation in Asturian Spanish between /-as/ and /-es/ is found in the following categories:

a) Plural forms of feminine or masculine nouns that end in /-a/ in their singular form, such as “cosa/coses” and “problema/problemes”:

(24)  a. “Él puso les perres y yo puse el trabajo” (GI11-04)
     ‘He put the money and I put the labor’
   b. “Quedó con sus taras, como queden todos” (GI11-04)
     ‘He ended up with some scars, like they all do’

b) Feminine plural forms of adjectives that have gender distinction (including adjectives derived from past participles), as in “buena/buenes”, and plural forms of adjectives that
do not have a gender distinction and whose singular form ends in /-a/,
like “egoista/egoistes”:

(25)  a. “son hores concretaes\(^6\) con el cliente” (GI11-11)
‘they are hours that you specify with the customer’
b. “les cases eran todas pequeñas” (GI11-11)
‘the houses were all small’

c) Definite articles in their feminine plural form:

(26)  a. “así pasábamos allí les hores” (GI11-10)
‘that’s how we spent the hours’
b. “gusten-y mucho las rancheres” (GI11-04)
‘he really likes rancheras’

d) Indefinite articles in their feminine plural form:

(27)  a. “va a unas naves a Juan Alvargonzález” (GI11-04)
‘she goes to some industrial buildings in Juan Alvargonzález’
b. “va de unas niñas que mataron en un camping” (GI11-04)
‘it’s about some girls that were killed at campsite’

e) Demonstrative adjectives and pronouns in their feminine plural form:

\(^6\) “Concretaes” also shows intervocalic deletion of /d/. The Spanish equivalent is “concretadas”. 60
(28)  a. “**estes** hueveres no puen tar aquí” (GI11-10)  
‘these egg cases can’t be here’

b. “pa la boda, pondremos rancheras y **esas** cosas” (GI11-04)  
‘for the wedding, we’ll play rancheras and things like that’

f) The following possessive adjectives when they agree with feminine plural nouns:

i. First and second person singular forms and third person form (singular and plural) when they are not in a prenominal position: “**mías**”/“**míes**”, “**tuyas**”/“**tuyes**” and “**suyas**/**suyes**”.

ii. First and second person plural forms in any position:

“**nuestras**/**nuestres**” and “**vuestras**/**vuestres**”.

g) Quantifiers in their feminine plural form:

(29)  a. “Trabajé **muches** hores yo” (GI11-04)  
‘I worked many hours’

b. “Salimos a cenar casi **todas** las semanas” (GI11-04)  
‘We go out to dinner almost every week’

h) Feminine plural forms of indefinite adjectives:
(30) a. “Llego y hago les mismes coses” (GI11-04)
   ‘I get here and I do the same things’

   b. “No como… hay otras novias que por ir detrás de ellos van hasta el fin del mundo” (GI11-04)
   ‘Not like… there are other girlfriends that because they want to follow them everywhere, they go to the end of the world’

   i) Subject pronouns “nosotres”, “vosotres” and “elles”:

   (31) a. “No sé si son de elles ya, o pertenecen a otra persona” (GI11-11)
   ‘I don’t know if it’s theirs anymore, or if they belong to someone else’

   b. “Ellas cedieron una parte que entraba al Ayuntamiento” (GI11-11)
   ‘They gave a part that went in to the city hall’

   j) Direct object pronouns when the referent is a feminine plural noun, whether it appears in enclitic or proclitic position:

   (32) a. “Amarreles pol rabu de dos en dos (a les vaques)” (GI11-10)
   ‘I tied them at their tails, in pairs (the cows)’

   b. “Hubo que mandarlas pintar (las furgonetas)” (GI11-11)
   ‘We had to have them painted (the vans)’

   k) Certain adverbial expressions such as “gracies” and “apenes”: 
(33)  a. “así empezó, gracias al MB” (GI11-04)
‘that’s how it started, thanks to the MB’
b. “Gracias a Dios apareció Pedrín” (GI11-04)
‘Thank God, Pedrín showed up’

h) The interrogative “cuántes” when it refers to a feminine plural noun.

(34)  a. “pienso, cuántes armé yo” (GI11-10)
‘I think, how much trouble did I get into?’
b. “¿cuántas horas de vuelo tienes? (GI11-20)
‘How many hours is your flight?’

In contrast with this, variation between /-as/ and /-es/ is not observed in the following categories:

- Plural forms of nouns whose singular form ends in a consonant and that have /-es/ as the plural morpheme also in Spanish, such as “vacación/vacaciones” (‘vacation’), “nariz/narices” (‘nose/s’) and “maldad/maldades” (‘evilness’).
- Plural forms of nouns whose singular form ends in /-e/ and add /-s/ to form the plural, resulting in a final /-es/ sequence, as in “nave/naves”.

3.2.4 Acoustic qualities of Asturian and Asturian Spanish vowels

Even though Spanish and Asturian share the same set of vowels (/i/, /e/, /a/, /o/ and /u/), the realization of the back vowels in Asturian Spanish differs from that of the
Spanish spoken in other areas of the Peninsula. Konvalika (1985) mentioned how unstressed /o/ and /u/ are not as clearly distinguished in Asturian and in Asturian Spanish. In her phonological description of the Spanish spoken in Asturias, Muñiz Chacón (2002) pointed out that [o] is realized higher than the Spanish equivalent, while [u] is lower than the Spanish [u]. She also signaled that in unstressed positions, particularly in word-final position, the realization of /o/ shows great variation, making it possible for the mid vowel to be merged with /u/ in this position. In the data analyzed by the author, the F1 of unstressed [o] occurs normally at a lower frequency than what has been described for other varieties of Peninsular Spanish. In contrast with this, the F1 of unstressed [u] is normally higher than the values reported for the Spanish equivalent. In the production data from Asturian Spanish analyzed for this dissertation I found that, while the F1 of final [u] is higher than the values reported by Quilis and Esgueva (1983) for the unstressed position, agreeing with the observations made by Muñiz Chacón, this is also the case for the F1 of final [o], going against what has been described for Asturian. Nevertheless, the difference in frequencies between the F1 of the two back vowels is still greater in Spanish (as reported by Quilis and Esgueva) than in Asturian Spanish, indicating that final [o] and [u] are closer in vowel height than unstressed [o] and [u] in Spanish. This raises the question of whether listeners can perceive a difference between both back vowels and whether we can maintain that there is a phonetic distinction between /o/ and /u/ in word-final position in Asturian Spanish. These questions will be addressed in chapters 4 and 5.

Muñiz Chacón did not report formant frequencies for [a] and [e]. In the Asturian Spanish production data that I will analyze in chapter 4 I found a similar tendency for [e]
and [a] in final position to that observed in the back vowels. The F1 of [a] is lower than that of Spanish unstressed [a] and the F1 of final [e] in Asturian Spanish is higher than that reported for unstressed [e] in Spanish. The opposite tendency is observed for the F2, so that the F2 of final [a] in Asturian Spanish is higher and the F2 of [e] lower than the corresponding Spanish equivalents. We see that the difference in frequencies between the F1 and the F2 of both vowels are also greater in Spanish than in Asturian Spanish. Table 1 shows the comparison of average formant frequencies in Hz in Asturian Spanish, using the production data analyzed in this dissertation, and Spanish, utilizing the data reported by Quilis and Esgueva for the information about Spanish. The Spanish formant values represent vocalic production in Peninsular and in different varieties of Latin American Spanish.

<table>
<thead>
<tr>
<th></th>
<th>Male speakers Asturian Spanish</th>
<th>Female speakers Asturian Spanish</th>
<th>Male speakers Spanish</th>
<th>Female speakers Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1 539</td>
<td>F2 479</td>
<td>F1 1426</td>
<td>F2 1118</td>
</tr>
<tr>
<td></td>
<td>F2 479</td>
<td>F1 283</td>
<td>F2 895</td>
<td>F1 653</td>
</tr>
<tr>
<td>[a]</td>
<td>F1 578</td>
<td>F2 1445</td>
<td>F1 648</td>
<td>F2 1220</td>
</tr>
<tr>
<td></td>
<td>F2 1445</td>
<td>F1 1698</td>
<td>F2 1935</td>
<td>F1 1985</td>
</tr>
<tr>
<td>[e]</td>
<td>F1 497</td>
<td>F2 1698</td>
<td>F1 545</td>
<td>F2 481</td>
</tr>
<tr>
<td></td>
<td>F2 1698</td>
<td>F1 1935</td>
<td>F2 1985</td>
<td>F1 2214</td>
</tr>
</tbody>
</table>

Table 1: Comparison of formant frequencies in Hz between Asturian Spanish final vowels and Spanish unstressed vowels (Quilis and Esgueva 1983)

The vowel plots in Figure 2 and Figure 3 illustrate the differences reported in Table 1:
Figure 2. Vowel plot comparing the realization of Spanish unstressed vowels (Quilis and Esgueva 1983) to that of Asturian Spanish final vowels by male speakers.

Figure 3. Vowel plot comparing the realization of Spanish unstressed vowels (Quilis and Esgueva 1983) to that of Asturian Spanish final vowels by female speakers.

The plots show that, in general, the Asturian Spanish final vowels are more centralized than the Spanish unstressed vowels reported by Quilis and Esgueva. We also
see that [o] and [u], especially the latter, are articulated lower in Asturian Spanish than in Spanish. Especially in the case of male speakers, both back vowels are more differentiated with respect to backness than in vowel height.

We should note that some of the differences observed in Table 1 could be due to the type of data that is analyzed in each case. While the Asturian Spanish data comes from spontaneous speech, collected using sociolinguistic interviews, the Spanish data reported by Quilis and Esgueva comes from a read list of words. The difference in speech mode could translate in differences in the production of vowels. Another aspect that should be noted is that the unstressed vowels analyzed by Quilis and Esgueva come from both word final and word medial position, while the Asturian Spanish vowels were always in word final position (in an open syllable in the case of [o] and [u] and in a closed syllable in the case of [a] and [e]). Furthermore, the authors analyzed the vocalic production in different varieties of Spanish, also including Peninsular Spanish. However, they did not find significant differences in formant values from one variety to another.

Even with the limitations outlined above, the differences between Spanish and Asturian Spanish in the articulation of vowels should be considered in any analysis of the variation in final vowels. First, we cannot rely on the formant frequencies described for other varieties of Spanish. Second, the close articulation between vowel pairs observed in Asturian Spanish makes it necessary to determine acoustically what constitutes /o/ and /u/ and what /e/ and /a/, in order to avoid contradictions between auditory coding and the acoustic properties of the vowel.
3.3 Previous research on vocalic variation in Asturian Spanish

In chapter 2 I reviewed the research that has explored variation in Asturian Spanish. Most of these studies included observations on the presence of /-u/ and /-es/ in the contact varieties. For example, in her analysis of the speech of Oviedo, Sánchez Álvarez (1979) provided evidence that the mass/count distinction in singular nouns and adjectives characteristic of Central Asturian is still maintained in the speech of Oviedo, using /-u/ to mark count and /-o/ to mark mass. She also mentioned the frequent appearance of /-es/ as the morpheme marking feminine plural in nouns, adjectives, articles, and unstressed pronouns.

Prieto (1991) analyzed the speech of two women from Gijón and part of her analysis focused on the realization of final vowels that is the object of my research. With respect to the alternation between /-o/ and /-u/, she recorded a similar number of occurrences for each ending and, as Sánchez Álvarez also suggested, she claimed that /-o/ is often used to mark certain nouns as mass. With respect to the plural markers /-as/ and /-es/, she noted that the Asturian morpheme, which is recorded around 80% of the time, is more frequent than the Spanish one. However, the confusion between the two variants was noticeable in the data analyzed, as the example in (35) illustrates (Prieto 1991: 17):

(35) Necesito ochocienta y dos mil pesetes y ni aspiraciones a tenelas

‘I need eight hundred and two thousand pesetas and I don’t even aspire to having them’
In (35) we can see that the Asturian variant is used in the noun but the Spanish one is used in the unstressed pronoun that refers to “pesetes”.

The only linguistic factor that Prieto considered is the mass/count distinction in the alternation between final back vowels in nouns, so that the presence of /-o/ instead of /-u/ is sometimes associated with the mass meaning that nouns like “pescao” (‘fish’) and “dinero” (‘money’) have, and not to the possible influence of the /-o/ that is used as the masculine singular morpheme in Spanish.

The only study that provides a quantitative analysis and that includes linguistic and extra-linguistic variables is that of Antón (1995). Her research focused on the pronunciation of unstressed vowels in the urban mining area of Langreo. The data for her study were collected using sociolinguistic interviews of 60 informants, ranging from 14 to 60 years old. She divided the speakers into three different socioeconomic levels, using information about education level, occupation, type of residence, area of residence and income. The variable under study was the appearance or not of raised vowels in unstressed positions, including word-final and internal positions. The author did not differentiate between raising that occurs because of the borrowing of an Asturian morpheme, which constitutes most of the cases of raising in word-final position, and raising of unstressed word-medial vowels that is phonetically conditioned, as in the following examples:

(36) “mulín” vs. “molín”  “molino”  (‘mill’)
(37) “riñir” vs. “reñir”  “reñir”  (‘to argue’)

69
In examples (36) and (37), as was the case for tokens like “piquiñín”, “puquitín” and “prubitín”, the raising of the unstressed vowels is triggered by the presence of a stressed final vowel that is also high (normally /i/). As I explained above, the raising that we find in the variables that are object of this dissertation is morphologically constrained to certain lexical categories.

Antón analyzed the effect of four social factors in the appearance of Asturian raised vowels: gender, age, socioeconomic level and the origin of the speakers. Among the linguistic factors, she included the lexical category of the word, the position of the raised vowel and the height of the stressed vowel. She divided her results into two categories, those where the outcome of the raising is a high vowel (/i/ and /u/) and those where the outcome is the mid front vowel /e/. She pointed out that the raising of the mid back vowel is more frequent than that of the front vowel. She also concluded that final /i/ and /u/ were more frequent in lower socioeconomic levels and that high vowels were more frequent among men and older speakers, with age being the most important social factor. She attributed the effect observed in age to the increase in education level in the younger generations. The author also emphasized the fact that there was an abrupt change in the use of Asturian features between older and younger speakers instead of a progressive decline. Speakers younger than 35 years old used fewer raised vowels than older speakers. However, the author also noted that the appearance of the Asturian vowels in the younger generations does not continue to decline but is stabilized instead. Similar results were observed in the raising of the low vowel, which, according to the author, is more frequent than that of mid vowels.
Antón claimed that the linguistic factors examined were not as important as the social factors in constraining the variation between Asturian and Spanish vowels. However she did point out that raising was more common in final position than in word-internal position. The variant /-u/ was also more frequent in nouns and adjectives than in verbs. This contrasts with the observations made for the Asturian mid vowel /e/, which appeared more frequently in verbal morphemes than in nouns and adjectives. Finally, she noted that the height of the stressed vowel did not seem to have any effect on whether the unstressed vowels were raised or not.

3.4 Conclusions

This chapter describes in detail the linguistic variables that are analyzed in this dissertation. I have reviewed the diachronic development of the Asturian vowels and explained how they contrast with the vocalic system in Spanish. I have explained how the variation between Asturian and Spanish vowels is a result of the borrowing of Asturian morphemes and how this type of variation is different from vowel raising that is phonetically conditioned. I have also defined the envelope of variation, describing the contexts in which variation between Spanish and Asturian is possible; i.e. the lexical categories in which speakers can alternate between the two variants and the contexts in which variation never occurs. Finally, a review of the observations that previous research has made on the variation between Spanish and Asturian vowels reveals that further study is necessary to fully account for the distribution of the two variants, since most of the studies that mention the alternation approach it in an impressionistic manner.
As we saw, Antón’s study provides more information about the possible linguistic and social factors that might govern the variation between Spanish and Asturian vowels. However, her analysis only includes information about percentages and the frequency distribution of each of the variants in each of the social and linguistic categories described above. Thus we cannot assume that the differences observed between different groups are significant. A more detailed statistical analysis of production data is necessary to conclude whether the variables mentioned in this study have an effect on the variation in the realization of unstressed vowels. This dissertation fills the gaps left by these studies and incorporates the latest advances in Sociolinguistics for the description of Asturian Spanish as spoken in Gijón.

In the following chapter I describe the quantitative analysis of production data that was performed to achieve the objectives outlined above. First, I describe the methodology used in the collection and analysis of the data to later report the results obtained for each of the linguistic variables under study. Finally, I discuss the results explaining the outcome of the statistical analysis and addressing its implications.
4.1 Introduction

The objective of the quantitative analysis of production data is to explore the distribution of Spanish and Asturian vowels in the morphemes described in Chapter 3, investigating the correlation between the use of particular linguistic variants and internal and external factors that might affect the variation observed. Previous research has emphasized the need to describe the linguistic variety spoken in the urban areas of Asturias (Sánchez Álvarez 1979, D’Andrés 2002). This need stems from the primary focus that has traditionally been placed on the description of rural Asturian within the field of dialectology. Thus, interference that takes place between the two linguistic systems has not received much attention. As we have seen, the few sociolinguistic studies that explore the use of Asturian and Spanish features emphasize the importance of social factors in the variation but fail to provide quantitative evidence that supports the conclusions reached regarding the effect of external variables. In addition, the possible internal factors that might govern the variation have been underexplored.

In section 4.2 of this chapter I explain the methodology used in the collection and analysis of the data. I describe the dependent variables and explain how they were determined, the statistical analysis performed and the linguistic and social factors
included in that analysis. In section 4.3 I present the results obtained in the statistical analyses, including the results for the categorical distinction between the two vowels in each pair and the continuous analyses of the first two formants. In section 4.4 I discuss the results obtained for the linguistic and social factors and I explore the possible implications of these results.

4.2 Methodology

4.2.1 Data collection

The data used in this part of the analysis was collected in a community in Gijón, the largest urban area in Asturias, using sociolinguistic interviews, the primary method for data collection in first wave variationist studies (Labov 1966, 1984). This one-time conversation between the researcher and the informant is designed to encourage story telling and to elicit as much natural and spontaneous speech as possible (Labov, 1984). A flexible interview protocol was used to elicit the data. The conversations were organized around topics like biographical information, education, family, employment, hobbies, food, life in the city, etc. All the interviews began by eliciting biographical information but the organization of the rest of the modules was adapted to each conversation, altering the order and presence of some topics based on the interests of the informant. Thus, not all the topics were addressed in all the interviews. At the end of each interview each participant was asked to identify their native language and to describe the way people from Gijón in particular and from Asturias in general speak. Each interview lasted between 40 minutes and one hour.
All the recordings were made using a Sony IC digital recording (SX712) with a sampling rate of 44.1kHz. Most of the interviews were recorded in the speakers’ homes and some of them were recorded in public spaces, such as coffee shops and terraces.

4.2.2 Participants

All the participants in the study live, grew up or work in the districts of El Llano, Pumarín and Contrueces, three contiguous middle-class neighborhoods located in the south of Gijón. The location of the three districts is highlighted in the map in Figure 4.

Figure 4. Location of the districts of El Llano, Pumarín and Contrueces

Even though not all the informants were born in Gijón (some were born in rural areas and a few outside of Asturias), they have all spent the majority of their lives in
Gijón, having moved to the city during childhood or early adolescence. González Quevedo (2001) notes that even more recent immigrants will use Asturian features in their Spanish if they have been living in Asturias for several years.

The three districts received a large amount of immigration from rural areas in the second half of the 20th Century, when they were relatively rural and marginal to the center of the city. El Llano is currently the most populated neighborhood in the city and, because of the abundance of retail and grocery stores and the continuous growth of the urban area, it is increasingly becoming more central to the city. In the last decade it has also experienced an increase in population from Romania, North Africa and South America. The population in the three districts live in single-family apartments that normally have two or three bedrooms and that are located in buildings ranging from 4 to 12 stories.

Social relations within the neighborhood are established through different contexts: associations of neighbors within a building (“la comunidad de vecinos”), bars and cafeterias, local stores and markets and sharing of common public space, such as parks and municipal buildings (libraries, swimming pools, etc.).

Because I am a native to the neighborhoods under study, I did not enter the community as a stranger. In order to select the participants I followed a social network or “friend of a friend” approach, by which the researcher uses initial contacts to establish an indirect connection with other participants. As Milroy (2002:550) explains, “A social network may be seen as a boundless web of ties which reaches out through a whole society, linking people to one another, however remotely”. Different types of social ties have been identified: first order ties are direct relationships maintained with family,
friends (strong) and acquaintances (weak), while second order ties are indirect relationships made through a first-order tie. In my case, I had a direct relationship (or first order tie) with some of the speakers interviewed. Second order or indirect connections were made through three members of the community with whom I had a first order tie, and who were not interviewed. As a result, I had first or second order ties with all of the informants. The resulting network is illustrated in Figure 5, where 1 represents the researcher and 2, 3 and 4 represent the three main connections to the rest of the informants, represented by the respective numbers:

Figure 5. Social network of informants
In Figure 5 the thicker links represent strong ties, such as family relations and close friendship, while the thinner links represent weak ties or acquaintances (coworkers, store owners and customers, neighbors, etc). This type of approach explores the effect of interpersonal relations on language variation. Because previous research has emphasized the role of interpersonal relations on the use of Asturian or Asturian features (Llera Ramo 1994, D’Andrés 1997), it was important to record communicative situations in which I had different types of social ties with the interlocutors so that social tie could be included in the quantitative analysis.

Being an insider in the community provided me with some of the advantages found in ethnographic approaches, such as deeper knowledge about the organization of the community, and it also helped to reduce the effect of the observer on the data.

The informants were selected by trying to compile a representative sample that included speakers in different age ranges, of both genders and different levels of education. A total of 30 speakers were interviewed, 14 females and 16 males, with ages ranging from 25 to 65 years old.

Six speakers were excluded from the sample due to problems with the quality of the recording, which did not allow for clear formant identification. In most of the cases that were excluded the quality of the sound was compromised due to background noise. The final sample includes 24 speakers, 12 males and 12 females, with ages ranging from 25 to 60 years old.
4.2.3  Envelope of variation and exclusions

For the quantitative analysis only 15 minutes of each interview were analyzed.
The first ten minutes of each interview were excluded, assuming that during the
beginning of an interview the participant is likely to be adopting a more formal style. As the interview progresses, speakers tend to relax and more natural speech is expected.

All the tokens that fell into the lexical categories described in Chapter 3 were extracted for the analysis. Contexts in which variation was not observed were excluded from the analysis. As explained in section 3.2.2.2, for the /-o/ vs. /-u/ variable, the only adjectives and unstressed pronouns that were included in the analysis were those that qualify or refer to masculine count nouns, since variation was not observed in cases where adjectives and pronouns refer to masculine mass nouns and only /-o/ was used in those tokens, even if the corresponding noun showed /-u/. The following exclusions were also made:

a) The expressions “por ejemplo” (‘for example’) and “sin embargo” (‘however’) were not included in the analysis, because they are normally used as formulaic utterances or pause fillers and categorically appeared with final /-o/.

b) The names of days of the week, months and seasons were also excluded from the analysis. These items, when used in Asturian can be considered lexical borrowings. In

(38) "bueno, ahora, por ejemplo, digamos que nos ocupa más el tiempo… a mí personalmente ya el tema de los hijos, de los nietos…” (GI11-19) ‘well, now, for example, our time is filled more… me personally, with our children, our grandchildren…)

b) The names of days of the week, months and seasons were also excluded from the analysis. These items, when used in Asturian can be considered lexical borrowings. In
some cases, the lexical items used in each language are entirely different, as in the item for ‘autumn’ which is “otoño” in Spanish but “seronda” in Asturian, or the word used for ‘december’, which is “diciembre” in Spanish but “avientu” in Asturian. As we saw for the tokens that fall into category a), variation was not observed in this category.

c) A lack of variation was also observed in toponyms, such as “Oviedo” (“Uviéu” in Asturian). Thus, this class of nouns was also excluded from the analysis.

d) All the tokens in which the identification of the variant could be compromised were also excluded from the analysis. This category includes tokens in which the target vowel is preceded or followed by another vowel of the same quality as the one present in the variants under study. Thus, all the tokens of the variable /-o/, /-u/ that were preceded or followed by either /o/ or /u/ were not included in the analysis. Example (39) illustrates this context:

(39) “entos, hace un año o así que fuimos” (GI11-25)
‘so, it’s been about a year since we went’

In example (39) it would be impossible to determine if the vowel uttered is the final /-o/ of ‘año’ or the following conjunction “o”. In the case of /-as/ and /-es/, only the tokens in which the target vowel was preceded by /a/ or /e/ were excluded from the analysis, since the following context was invariably /s/. Example (40) illustrate such examples:
(40) “taben amarr... agarrás po’l rabu” (GI11-10)
    ‘the were attached by their tail’

In example (40), the deletion of intervocalic /d/ in “agarradas” results in the creation of vocalic sequence /aa/ or /ae/. In this particular case, it is impossible to determine what part of the resulting vowel is the stressed /a/ in “agarradas” and what part is the morphemic vowel.

e) In the /-as/, /-es/ variable, unstressed pronouns in which the appearance of “les” could be interpreted as a case of “leismo” were also excluded from the analysis. Leísmo is a common phenomenon in the north of Spain, by which the pronoun “le” or “les”, prescriptively reserved for the indirect object, is used to refer to masculine and feminine direct objects. In this geographic area leísmo is restricted to cases in which the referent of the direct object is animate. Thus, tokens of “las” or “les” in which the pronoun has an animate referent were not included in the analysis. Example (41) illustrates this use:

(41) “el día seis fuimos a Santander a buscarlas” (a su sobrina y hermana) (GI11-02)
    ‘we went to Santander to get them on the sixth’ – her niece and sister’

(42) “la puta que les parió (a las mujeres)” (GI11-10)
    ‘the whore that gave birth to them (women)’

In (41) “las” is referring to the speaker’s niece and sister. Even though “las” is used, the potential appearance of “les” could be attributed to leísmo and not to Asturian influence. In (42) “les” is referring to women and, even though it could be interpreted as
an example of vowel raising in the direct object pronoun, the presence of /-es/ could also be attributed to leismo. Thus, examples like (41) and (42) were excluded, whether they contained “las” or “les”.

f) In some occasions the vowels under study were devoiced and pronounced with a creaky voice quality that did not allow for formant identification and measurement. In most cases devoicing occurred before a pause. Figure 6 contains a spectrogram of the word “tiempu” (‘time’). Highlighted is the target vowel and we can see how, even though a first formant (F1) is visible, a second formant (F2) cannot be identified. This contrasts with the other vowels in the word (the diphthong [je]), where the first three formants are clearly visible.

Figure 6. Spectrogram of the word “tiempu” [tjempu].
In some occasions devoicing was not a result of the voice quality employed at the end of an utterance but of the phonological context surrounding the target vowel. Particularly, when the vowel appeared between two voiceless fricative sounds, it was often devoiced. Figure 7 illustrates this process in the word “esas” (‘these’):

Figure 7. Spectrogram of the word "esas" [esas].

As we can see, the segment between the two fricatives is almost completely devoiced, making it impossible to identify and measure the first two formants.

4.2.4 Dependent variables

A total of 1321 tokens that contained the /-o/, /-u/ variable and 813 tokens that showed /-as/, /-es/ variation were included in the statistical analysis. Three different
dependent variables, associated with the acoustic measurement of the vowels, were used in the analysis.

4.2.4.1 Acoustic analysis

The first two variables were the frequencies of the F1 and F2 of each vowel which were determined using Praat (Boersma and Weening 2012). These measurements were treated as continuous variables. The formants were determined taking a single measuring point at the midpoint of the vowel. Because all the target vowels appear in unstressed final syllables, on some occasions their duration was too short to consider taking multiple measuring points.

The anatomical differences between speakers in vocal tract configurations result in acoustic differences in the production of vowels. In order to be able to compare the vowel productions of different speakers, it is necessary to normalize the data. As Watt et al. (2011) explain, it is necessary to convert mathematically Hz values into psychoperceptual equivalents, which will give us a sense of how the listeners perceive differences in vowel quality. Because I was not interested in describing the entire vocalic system but only in the realization of some vowels in specific contexts, a vowel-intrinsic method (which uses information within a single vowel token) was the most appropriate. Also, due to the spontaneous nature of the data, the quality of the recordings did not allow me, on most of the occasions, to measure a third formant reliably. As a consequence, a formant intrinsic method, where the measurement of the third formant or the fundamental frequency is not necessary, was appropriate for this study. The only
normalization method that is vowel intrinsic and formant intrinsic is the ERB scale (Traunmuller 1997). Using the ERB values, the statistical analysis allows me to determine whether any of the independent variables have an effect on the frequency of both formants, i.e. if they are correlated with a change in the quality of the vowel.

4.2.4.2 Principal Components Analysis and clustering

The third dependent variable is a categorical grouping of each token based on formant values. Even though auditory coding was also performed for each of the linguistic variables, the plotting of each token in the vowel space revealed overlapping results in the auditory coding with respect to the formant frequencies. Thus, I employed Discriminant Analysis of Principal Components (DAPC) (Jombart, Devillard and Balloux 2010, Jombart 2008), a statistical method that allows for calculating the probabilities of individual tokens of belonging to a particular group using exclusively intrinsic information (in this case the frequencies of the F1 and the F2). This analysis was performed using the adegenet package (Jombart 2008) in the R software (R Development Core Team 2011). The formant values for each token, which constituted the input for the analysis, were orthogonally transformed using Principal Components Analysis, a step that ensures that the variables used in the DAPC are perfectly uncorrelated (Jombart et al. 2010). Using the Principal Components and a predetermined number of clusters – two in this case, since we want to distinguish between two vocalic realizations, DAPC maximizes between-group variation and finds groups that minimize within-group variation. Finally, the probabilities of an individual token belonging to one of the clusters
were calculated. Using this information I assigned each token to the category to which they were more likely to belong, based on the probabilities determined by DAPC.

In Figure 8 and Figure 9 we can see the comparison of the distribution of the two back vowels as determined in the auditory coding to the distribution that resulted from the DAPC. The general location of the two vowels in the vowel space is similar in both plots, pointing to the reliability of the DAPC. There are several advantages to using DAPC. First, the distinction between the two groups is made based on the combination of both formants. Second, using intrinsic information and probabilities, it reduces the arbitrariness of the discrimination. Finally, it eliminates the contradictions and subjectivity associated with the auditory coding, while still allowing us to perform a categorical analysis.
Figure 8. Distribution of /o/ and /u/ in a vowel space as coded auditorily

Figure 9. Distribution of /o/ and /u/ in a vowel space as grouped by the DAPC

The comparison between the auditory coding and the DAPC for /a/ and /e/ is illustrated in Figure 10 and Figure 11:
Figure 10. Distribution of /a/ and /e/ in a vowel space as coded auditorily

Figure 11. Distribution of /a/ and /e/ in a vowel space as grouped by the DAPC
4.2.5 Independent variables

In this section I describe the internal/linguistic and external/social independent variables that were included in the study. The regression analysis that will be outlined in 4.2.6 tests the correlation of these independent variables with each of the dependent variables that were described in section 4.2.4.

4.2.5.1 Internal/Linguistic variables

The following sections describe the linguistic variables that were considered in the quantitative analysis.

4.2.5.1.1 Quality of the stressed vowel

The first linguistic factor that was explored is the quality of the preceding stressed vowel. In her study Antón (1995) considered this factor as part of her linguistic variables and concluded that the effect was not significant. However, there has been work in vowel raising in other varieties of Spanish where the height of the stressed vowel has had an effect on the outcome of the variation. Holmquist (1998), for example, explores the phenomenon of word-final vowel raising in Puerto Rican Spanish, specifically in the community of Castañer. In this variety of Spanish vowel raising is not subject to a morphological constraint. Even though the nature of this phenomenon is different from the alternation between raised and non-raised variants that we observe in Asturian Spanish, it is possible that the variation observed in both cases is motivated by similar linguistic factors. In his study he found that preceding stressed high vowels favor the
appearance of high final vowels in the Castañer variety. Despite the relation between the height of the stressed vowel and that of the final vowel, according to Oliver Rajan (2007) the phenomenon observed in Puerto Rican Spanish is not a manifestation of vowel harmony, since vowel harmony normally applies categorically given an appropriate phonetic environment. In contrast to this, in the data reported by Holmquist we find that there exists variation even when the stressed vowel is high. Nevertheless, a potential relation between the appearance of raised vowels in Asturian Spanish and vowel harmony processes should not yet be rejected.

Vowel harmony processes triggered by the stressed vowel are also found in some varieties of Spanish. As an example, raising of pretonic vowels triggered by the stressed vowel is found in the Pasiego dialect, in the region of Cantabria, Spain (Hualde 1989, McCarthy 1984, Penny 1969, Vago 1988). The difference between this phenomenon and a possible influence on vowel height in the Asturian Spanish linguistic variables that are the object of this dissertation is that, while the harmony effect of the stressed vowel operates to the left in Pasiego Spanish, any effect found in Asturian Spanish would operate to the right of the stressed vowel. Walker (2005) explains that processes of vowel harmony triggered by the stressed vowel, such as the one observed in Pasiego and potentially in Asturian Spanish if an effect of the stressed vowel is found, are driven by vowel-to-vowel coarticulation and motivated by the prosodic strength of the stressed position.

Each vowel was considered as a separate factor level, since different effects have been found for example for /i/ in certain languages (Walker 2005). As a result, five factor levels were initially coded for quality of the stressed vowel: /a/, /e/, /i/, /o/ and /u/. The
analysis that included stressed vowel as an independent variable excluded unstressed pronouns that were in proclitic position, as in (43), since it would be impossible to determine the stressed vowel in those cases.

(43) “tú lo ves y piensas que es un hotel (el hospital)” (GI11-20)
‘you see it and you think it’s a hotel (the hospital)’

4.2.5.1.2 Stress pattern

Related to the previous linguistic variable is the stress pattern of the word. All the lexical items were coded according to where the stress was located within the word. Two categories were determined: paroxytone, which included words with stress on the penultimate syllable (exemplified in (44)), and proparoxytone, including words with stress on the antepenultimate syllable or before (as in (45)).

(44) /tra.'ba.xo/ (‘work’)

(45) /'mu.si.ko/ (‘musician’)

4.2.5.1.3 Vowel duration

Vowel duration was also included as a continuous variable in the statistical analysis. Previous research has found that lower vowels are longer in duration than higher vowels (Chladkova, Escudero and Boersma 2011). Thus, we expect to find longer duration in the Spanish variants than in the Asturian ones.
4.2.5.1.4 Preceding segment

The next internal factor included in the regression analysis is the segment that precedes the final vowel. Holmquist found that the preceding consonant had an effect on vowel raising in Castañer Spanish. According to the author, a preceding palatal consonant favored the appearance of a raised vowel in final position. In her study of pretonic vowel raising in Brazilian Portuguese (a phenomenon similar to that observed in Pasiego), Bisol (1989) found that a preceding velar favors raising and so does a preceding labial when the target vowel is /o/.

Based on previous research, the preceding consonants were grouped by place of articulation, resulting in five consonant factor levels for each linguistic variable: bilabial, dental and interdental, alveolar, palatal, velar/uvular.

It should be noted that in the data from Asturian Spanish the raised vowel can also be preceded by a vowel if an intervocalic /d/ is deleted, a phenomenon common in the variety under study, resulting in tokens like “prau” instead of “prado” (‘field’), “mojaes” instead of “mojadas” (‘wet’), “metiú” instead of “metido” (‘introduced’), etc. Thus, vowels were considered as separate factor levels. Since contexts in which it was possible to have two identical consecutive vowels were excluded from the analysis, different vocalic factor levels were determined for each linguistic variable. For the /-o/, /-u/ variable, /a/, /e/ and /i/ constituted the three possible vocalic contexts, whereas for the /-as/, /-es/ variable, only /i/ occurred before the target vowel. For both linguistic variables /i/, /e/ and palatal consonant were later combined into a single palatal sound category due to their similar behavior in the statistical analysis and to avoid empty cells in the
distribution of the data. Dental and alveolar consonants were also combined into a single category.

4.2.5.1.5 Lexeme language

The language of origin of the lexeme constitutes the fourth linguistic factor included in the analysis. D’Andrés (2002) claims that while a Spanish lexeme can appear with Asturian or Spanish morphology, an Asturian lexeme can only be combined with an Asturian morpheme. Thus, for example, in the case of Spanish “ternero” and Asturian “xatu” (‘calf’), in Asturian Spanish we can find “ternero”, “terneru”, “xatu” but not *“xato”. Even though D’Andrés claimed that the Asturian lexeme/Spanish morpheme combination cannot occur, I found that the Spanish variants of the morphemes can and do appear with Asturian lexemes. Example (46) illustrates this:

(46) “como a las guajas no se metían con ellas…” (GI11-17)
‘since they didn’t mess with the girls…’

The lexical item “guaja” is of Asturian origin (“guaxa” in Asturian, “niña” in Spanish). The plural in Asturian is “guaxes” but the speaker opted for the Spanish morpheme /-as/. Similarly, in (47), the speaker deletes the intervocalic /-d-/ characteristic of Asturian but maintains the Spanish morpheme:

(47) “es como si fueran las fiestas de prao de aquí de Gijón” (GI11-03)
‘it’s as if they were the fairs that we have here in Gijón’
In order to determine what lexemes are Asturian and which Spanish, I followed the guidelines provided by D’Andrés (1993). Apart from the words that have a completely different form, such as “guaxa”, the author considers Asturian lexemes as any item that:

- Contains any of the following phonemes: /ʃ/ (“xatu”/ʃatu/ - ‘calf’), /ʝ/ where Spanish has /x/ (“trabayu”/trabaju/ vs. “trabajo”/trabaxo/ - ‘work’), word-initial /ɲ/ where Spanish has /n/ (“ñariz” vs. “nariz” – ‘nose’), word-initial /ʝ/ where Spanish has /l/ (“llana” vs. “lana” – ‘flat’)
- Contains a pretonic raised vowel (“piquíñín” vs. “pequeñín” – ‘little’)
- Contains the suffix “-iello” instead of Spanish “-illo” (“castiello” vs. “castillo” – ‘castle’)
- Shows deletion of intervocalic /b, d, g/, with the exception of past participles, where /d/ deletion is common in both languages (“bodea” vs. “bodega” – ‘wine cellar’)

It should be noted that there are lexemes that are identical in Spanish and Asturian and, even though those cases were coded separately, they were later combined with the Asturian category for the statistical analysis, since they both had a very similar effect on the variation.

4.2.5.1.6 Frequency

When Konvalinka (1985) talks about the motivations that can lead to cases of interference between Asturian and Spanish, she notes that when interference takes place,
the process develops in several stages, affecting only part of the lexicon in the first of those stages. Dyzmann (2001) points out that the frequency of an element in the discourse might be one of the factors that favor interference between the two languages, such that more frequent elements tend to show interference from Asturian. Thus, the frequency of the lexical item was included as one of the internal factors in the quantitative analysis. Because frequency dictionaries of Asturian utilize data exclusively from written Asturian, I decided to use corpus frequency as a measure of overall frequency of each lexical item in Asturian Spanish. One advantage of this method is that it reflects the use of the lexicon in spoken Asturian Spanish. However, the disadvantage is that the limited range in conversation topics might limit the representativity of this measure.

Instead of using raw frequencies, which can lead to misleading results in the statistical analysis I used a log calculation of frequency, based on Luce’s (1986) Standard Frequency Indices (SFI):

(48) \[
\text{SFI} = 40 + 10 \times (\log(\text{no. of occurrences/corpus size}) + 10)
\]

Figure 12 and Figure 13 show the frequency distribution of the tokens that were included for each variable, where the higher the log frequency the more frequent the token is.
Because the data were not normally distributed in neither of the variables, I decided to not use frequency as a continuous variable. Doing so would have compromised the results of the quantitative analysis, since in regression analyses there exists an assumption of a normal distribution of the data. Instead I opted for a categorical division between low, medium and high frequency. In order to determine the cut off
points I followed the trends outlined in the histograms and divided the data into smaller categories that were normally distributed. Thus, for example, for the /-o/, /-u/ variable, log frequency values between 140 and 141 were coded as low frequency, 141.5 to 143.5 as medium frequency and 144 and up as high frequency. For the /-as/, /-es/ variable, log frequency values between 140 and 142 were coded as low frequency, 143 to 145 as medium frequency and 146 and up as high frequency.

4.2.5.1.7 Lexical category

We have some evidence that the lexical category might play a role in the variation under study. Antón (1995) found that the realization [u] in unstressed positions is more frequent in nouns and adjectives than in verbal forms. Holmquist (1998) claims that, in the Castañer variety in Puerto Rico, final vowel raising is rare outside of the nominal and verbal categories. However, neither of these studies provide an explanation for the results observed.

In her data, Prieto (1991) notices examples like the one in (2), repeated here as (49):

(49) Necesito ochocienta y dos mil pesetas y ni aspiraciones a tenerlas
‘I need eight hundred and two thousand pesetas and I don’t even aspire to having them’
We can see that in (49) the Asturian variant is used in the noun but the Spanish one is used in the clitic that refers to “pesetes”. It could be possible that the variation observed is due to the difference in lexical category.

Since both nominal morphemes appear in slightly different lexical categories, different levels were coded for each linguistic variable. For the variation between /-o/ and /-u/ the following lexical categories were considered: noun, adjective, nominalized adjective, indefinite pronoun, indefinite adjective, clitic, possessive and numeral. All the lexical categories were combined into four groups because of the low number of tokens in some of them (such as numerals and subject pronouns) and the fact that they showed similar trends in the statistical analysis:

1. Nouns, which included nouns, subject and indefinite pronouns.
2. Adjectives, including descriptive adjectives, indefinite adjectives, possessive adjectives and numerals that were used as adjectives (ordinals in all cases) exemplified in (50).

(50) “él subía al segundo piso y ya se encontraba él… hacía dos o tres paradas” (GI11-11)
    ‘he would go up to the second floor and he was already feeling… he would have to stop twice or three times’

3. Nominalized adjectives, also including cases in which the numerals were nominalized, as in (51).

(51) “bueno, el primeru tuvo cuatro moces” (su hijo) (GI11-10)
    ‘well, the first one had four girlfriends’ (his son)
4. Clitics or unstressed pronouns.

For the variation between /-as/ and /-es/ the following lexical categories were coded: determiner (definite and indefinite articles), noun, adjective, subject pronoun, clitic, demonstrative adjective, demonstrative pronoun, indefinite adjective, indefinite pronoun, interrogative “cuantas/es” used as an interrogative pronoun (as in (52)), an interrogative adjective (as in (53)) or in the pronominal phrase “unas cuantas” (as in (54)).

(52) “pienso, ¡cuantes armé yo!” (GI11-10)
‘I think, how much trouble did I get into!’

(53) “¿Cuántas horas de vuelo tienes?” (GI11-20)
‘How many hours is your flight?’

(54) “unes cuantes cases que paso y que conozco” (GI11-16)
‘a few houses that I go by and I know’

As in the /-o/, /-u/ variable, these lexical categories were divided into five groups, based on their distribution and their behavior in the regression analysis:

1. Nouns, which included nouns, indefinite pronouns and the interrogative pronoun when it was used without a noun (as in (52)).
2. Adjectives, including descriptive adjectives, indefinite adjectives and the interrogative when it was used as an adjective (as in (53)) and in pronominal phrases (as in (54)).
3. Determiners, which included definite and indefinite articles
4. Demonstratives, whether they were used as adjectives or pronouns.
5. Clitics or unstressed pronouns.

4.2.5.1.8 Prior form

Previous research has found that speakers have a tendency to repeat structures and sounds. Most of the research in this area has focus on structural priming, or the repetition of syntactic structures (Gries 2005). In Spanish, for example, Cameron and Flores-Ferrán (2004) found that Spreading-Activation Theory (Dell 1986) could account for the patterns observed in subject expression in three varieties of Spanish. They observed that the use of subject pronouns leads to more subject pronouns, while the use of null subjects leads to more null subjects. Linguistic perseveration cannot only be observed in the use of syntactic structures but also in phonetic and phonological processes. Poplack (1981), for instance, found that a prior realization of plural /-s/ in Puerto Rican Spanish triggered more /-s/ realization. Poplack also found that there was a tendency for local redundancy within the noun phrases such that the realization of /-s/ in one of the elements favored the realization of /-s/ in the rest.

Due to the tendencies observed in spreading-activation research, the data was also coded for the linguistic variant that was used in the form preceding the target token. The two categorical factor levels were the two variants for each linguistic variable, as determined in the DAPC and clustering of the tokens according to their formant values.
4.2.5.1.9 Mass/count distinction

This linguistic factor is only applicable to the tokens that alternate between /-o/ and /-u/. As I noted in section 3.2.2.1, /-o/ is morphologically productive in Asturian and used to mark adjectives and pronouns that agree or refer to mass nouns. As was explained, even though masculine singular nouns are invariably marked with /-u/ in Asturian, several cases have been reported in Asturian and Asturian Spanish in which the alternation between the two back vowels could be associated with the semantic properties of that noun; i.e. whether it is a mass or a count noun. In order to examine the role of the mass/count distinction in the variation between back vowels in Asturian Spanish, all the nouns were coded according to whether they could be considered as semantically mass or count. Items like “cuchillo” (‘knife’), “vehiculo” (‘vehicle’) and “hijo” (‘son’) were considered as count nouns, while words like “dinero” (‘money’), “respeto” (‘respect’) and “daño” (‘harm’) were coded as mass.

4.2.5.1.10 Following segment

As well as the mass/count distinction, following segment is also only applicable to the /-o/, /-u/ variable, since the vowel in /-as/ and /-es/ is categorically followed by /s/ and there is no deletion of /-s/ in this variety of Spanish. The coding for this factor was performed in a similar way to that of the preceding segment, so that consonant sounds were grouped according to their point of articulation and vowels were considered independently. The categorization resulted in the following factor levels: bilabial, dental and interdental, alveolar, palatal, velar, /a/, /e/, /i/ and pause. Because I excluded contexts
in which the following vowel was one of the possible linguistic variants, /o/ and /u/ never occurred as the following segment.

4.2.5.1.11 Lexical gender

As I reviewed in section 3.2.3.2, the use of the morpheme /-es/ in Asturian is not restricted to marking feminine plural. Instead, /-es/ can also mark masculine plural in cases in which the corresponding masculine singular form ends in /-a/. Thus, in Asturian, the plural form of the masculine noun “día” (‘day’) is “díes” (“días” in Spanish). Other nouns that fall into this category are: “policías”, when used to refer to male police officers, “temas” (‘themes’), “idiomas” (‘languages’) and “programas” (‘programs’). This category also includes compound nouns that end in /-as/, such as “microondas” (‘microwave’) and “lavavajillas” (‘dishwasher’). Even though second component of the word was originally feminine (“ondas” and “vajillas”), the resulting compound is masculine. As a result all nouns that showed /-as/ or /-es/ were divided into two categories that responded to their lexical gender as explained above: masculine and feminine.

4.2.5.2 External/social variables

The research summarized in section 2.3.3 emphasizes the importance of the social variables in the speakers’ selection of features from Asturian (and the degree to which those features are incorporated into their speech). In addition to those studies, the relationship between social factors and the use of Asturian and Spanish is pointed out in
the sociolinguistic surveys that have been conducted in the region and which were reviewed in section 2.2.3. In this section I detail the social variables that were included in this study.

4.2.5.2.1 Gender

Previous studies have reported contradictory results regarding the role of gender in the use of Asturian or Asturian features in Asturian Spanish. While Antón (1995) concludes that vowel raising is more frequent among men, Amado Rodríguez et al. (1993) find that Asturian features increase in women’s speech. This difference could be due to the fact that Antón analyzes only vocalic variation whereas Amado Rodríguez et al. provide general results for a group of features. It could also be due to geographical differences since the studies focus on different urban areas or to the fact that the sample of speakers included in Amado Rodríguez et al.’s study is homogeneous with regards to social class (they all belong to what he calls the middle/low social class), while Antón’s informants are divided into three different socioeconomic levels. However, other sociolinguistic surveys of usage in Asturian seem to be consistent with Antón’s results. D’Andrés (1996), for example, finds that men claim to use Asturian more than women and also to have a better knowledge of it. Men also maintain a more favorable attitude towards the regional language.
4.2.5.2.2  Education and occupation

Education and occupation were also included in the quantitative analysis. Antón uses these two variables (among others) to determine socioeconomic status but she does not indicate how she operationalizes all the components. She concludes that high vowels are indicators of low socioeconomic status. In his linguistic questionnaire Llera Ramo (1994) reports more Asturian answers among speakers with less education and D’Andrés (1993) states that people that work with the public claim to use more Spanish than those who work in other sectors.

Three different levels of education were determined for the statistical analysis:

- **Low education**: speakers who did not go to school or that have all/some elementary school education (“Educación General Básica” – EGB – or “Educación Primaria”).

- **Middle education**: speakers that have some/all secondary school education (“Bachillerato Unificado Polivalente” – BUP - or “Educación Secundaria Obligatoria” - ESO) or that attended professional school (“Formación Profesional 1 y 2” – FP1 and FP2) without completing any courses in preparation for University education.

- **High education**: speakers that completed the courses in preparation for University education (“Curso de Orientación Universitaria” – COU – or “Bachillerato”) or that have any college education. The highest education level of the speakers was a college degree.
Different occupations were coded separately and then organized based on the different sectors present in the sample and the trends observed in the statistical analysis. The occupations were grouped into three categories:

- Manufacturing and logistic jobs, which includes skilled trades such as miners, metal workers, electricians, seamstresses, drivers, etc.
- Administrative, technical and teaching jobs, including travel agents, administrative assistants, school teachers, doctors, etc.
- Jobs in commerce and in the restaurant/food industry, such as butchers, servers and store clerks.

Because education and occupation are highly correlated, such that the administrative, technical and teaching jobs require post-secondary degrees, only one of these variables should be included in a regression analysis. Whether education or occupation should be included will be determined based on how much of the variance can be explained by each of the factors.

4.2.5.2.3 Age

In her study Antón (1995) mentions that age is the most important factor in the use of raised vowels in unstressed positions, with younger speakers using less raised vowels than older speakers, and she attributes its effect to the higher education levels of younger speakers. Thus, it is possible that education and age are also highly correlated. However, age should also be included in the analysis as an external factor. All prior sociolinguistic studies are unanimous in their conclusions with respect to the effect of
age: while older speakers claim to use more Asturian, to have better knowledge of it and to incorporate more Asturian features in their speech than younger speakers, the latter maintain a more positive attitude towards the autochthonous language than older speakers (D’Andrés 1996, Amado Rodríguez et al. 1993, Llera Ramo 1994). However, Rose Alcorta (2000) states the difference between age groups is not how much Asturian they use but in how many situations they reduce their Asturian features.

Even though age could be considered as a continuous variable, the generational differences noted in other sociolinguistic studies might be overlooked if we do not consider discrete age categorizations. D’Andrés (1996) reports several age classifications that have been used in previous sociolinguistic surveys. Even though different studies might include more or fewer age categories, most of them establish a division at 35 years-old and another at 50. In my analysis establishing three different age categories was not possible due to the reduced number of speakers that constituted my sample. Dividing speakers into three age categories would have created empty cells in the crosstabulations of age and other social factors, making the statistical analysis unreliable. Thus, I opted for dividing the speakers in two groups, those that were younger than 50 and those that were older than 50.

4.2.5.2.4 Social tie

Several sociolinguistic studies on Asturian Spanish, whether they analyzed the occurrence of particular Asturian linguistic features or the usage of the two languages in general, have pointed out the relationship between the choice of language and stylistic variation. In his sociolinguistic survey, Llera Ramo (1994) reports that speakers claim to
use Asturian in interactions with people that they know, such as friends and family, and to talk about personal topics. D’Andrés (1997) makes similar observations, claiming that the use of Asturian features is reserved to informal situations, addressees that are socially equal and for “light” topics of conversation.

The only stylistic factor that I included in the quantitative analysis is the social tie between interlocutors. In order to examine this stylistic factor I considered the different degrees of familiarity that I had with the speakers. In his study on style shifting in AAVE, in which he performed all the interviews himself, Baugh (1979) included familiarity of interviewer and interviewee as a stylistic factor. He operationalized familiarity as a function of the change in his status in the community from the first years of research to the later years, when he had become an active participant in the community. I followed a similar procedure in which different degrees of familiarity between the interlocutors were determined, based on a social network approach (Milroy 2002):

- Strong first-order ties, such as family and close friends.
- Weak first-order ties, such as acquaintances.
- Second-order ties, in which only an indirect relationship through a first-order tie exists.

Strong first-order ties and weak first-order ties were later combined into a single category, since they patterned similarly in the statistical analysis. Because speakers often claim to use Asturian with family and friends we would expect higher rates of the Asturian variants in the speech of speakers with whom I had a first order tie.
4.2.5.3 Summary

Table 2 and Table 3 contain a summary of all the independent variables that were included in the statistical analysis:

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Factor levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the stressed vowel</td>
<td>/i/</td>
</tr>
<tr>
<td></td>
<td>/ç/</td>
</tr>
<tr>
<td></td>
<td>/a/</td>
</tr>
<tr>
<td></td>
<td>/o/</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
</tr>
<tr>
<td>Stress pattern</td>
<td>paroxytone</td>
</tr>
<tr>
<td></td>
<td>proparoxytone</td>
</tr>
<tr>
<td>Preceding segment</td>
<td>/a/ (only for /-o/, /-u/)</td>
</tr>
<tr>
<td></td>
<td>palatal sound</td>
</tr>
<tr>
<td></td>
<td>dental/alveolar cons.</td>
</tr>
<tr>
<td></td>
<td>bilabial cons.</td>
</tr>
<tr>
<td></td>
<td>velar/uvular cons.</td>
</tr>
<tr>
<td>Lexeme language</td>
<td>Asturian/both</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
</tr>
<tr>
<td>Frequency</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>medium</td>
</tr>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Lexical category</td>
<td>nouns</td>
</tr>
<tr>
<td></td>
<td>adjectives</td>
</tr>
<tr>
<td></td>
<td>clitics</td>
</tr>
<tr>
<td></td>
<td>nominalized adj. (only for /-o/, /-u)</td>
</tr>
<tr>
<td></td>
<td>determiners (only for /-as/, /-es/)</td>
</tr>
<tr>
<td></td>
<td>demonstratives (only for /-as/, /-es/)</td>
</tr>
<tr>
<td>Prior form</td>
<td>Asturian</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
</tr>
<tr>
<td>Mass/count</td>
<td>mass</td>
</tr>
<tr>
<td>(only for /-o/, /-u/)</td>
<td>count</td>
</tr>
<tr>
<td></td>
<td>palatal sound</td>
</tr>
<tr>
<td></td>
<td>dental/alveolar cons.</td>
</tr>
<tr>
<td>Following segment</td>
<td>bilabial cons.</td>
</tr>
<tr>
<td>(only for /-o/, /-u/)</td>
<td>velar/uvular cons</td>
</tr>
<tr>
<td></td>
<td>/a/</td>
</tr>
<tr>
<td></td>
<td>pause</td>
</tr>
<tr>
<td>Lexical gender</td>
<td>masculine</td>
</tr>
<tr>
<td>(only for /-as/, /-es/)</td>
<td>feminine</td>
</tr>
</tbody>
</table>

Table 2. Summary of independent linguistic variables and factor levels
Independent variables | Factor levels
---|---
Gender | male
| female
Education | low education
| middle education
| high education
Occupation | manufacturing and logistic jobs
| administrative, technical and teaching jobs
| commerce and restaurant/food industry
Age | younger than 50
| older than 50
Social tie | first order tie
| second order tie

Table 3. Summary of independent social variables and factor levels

4.2.6 Statistical analysis

I tested the effect of the internal and external independent variables on the dependent variables using logistic and linear regression in R (R Development Core Team 2011). For each dependent variable I developed a mixed effects model that included speaker as a random variable, using the lmer function (Bates, Maechler and Bolker, 2011) in R. Including the possible effect of individual speakers provides a more conservative estimate of the effect of the independent variables selected, reducing the possibility of overestimation or Type I errors. As Tagliamonte (2011) points out, “Statisticians argue that there must be statistical validation that the factor groups in an analysis are statistically significant over and above the effect of the individuals that happen to be in the sample” (p. 137). Lexical item was not included as a second random variable in order to be able to include lexical frequency as an independent variable.

For each regression model (linear and logistic), the independent variables were added following a stepwise procedure and nested models were compared using ANOVA. The order in which the variables were added to the model was determined by random
forests, using the \texttt{cforest} function in the \textit{party} package in R (Hothorn et al, 2008).

Random forests evaluate the relative importance of each independent variable, using conditional permutation, a procedure that tests whether any independent variable is a sensible predictor of the response or dependent variable. In order to do so, they follow a process of trial and error working with random samples of the data (Tagliamonte 2011, Tagliamonte and Baayen 2012). The result is a series of conditional inference trees that constitute the random forest.

Using this statistical tool offers some advantages. First, it allows us to easily evaluate what are the most important factors in the variation. Second, because it works with random samples of the data, it is an ideal tool for small samples and large numbers of independent variables. Finally, while collinearity constitutes a problem for regression, the use of conditional permutation and subsampling allow random forests to work with correlated variables. Furthermore, it is an excellent tool to determine which of the possibly correlated variables is most important. One of the disadvantages, however, is that, because random forests work with random samples of the data, the ordering of the predictors can vary slightly from one run to another (especially in cases in which the strength of the effect is very similar). Thus, it is important to ensure that a similar ordering of the factors remains after several runs.

Once a hierarchy of independent variables was created using the random forest, I developed a mixed effects model for each dependent variable, using speaker as a random variable. The role of the fixed and random effects was first tested using the complete data set. The effects of stressed vowel and stress pattern were then added to the independent variables and tested on the group of tokens that could be coded for these two factors; i.e.
all the tokens that were not unstressed pronouns that were in proclitic positions. Finally, the group of nouns was tested separately for each linguistic variable, to explore any possible effect of the mass/count distinction in the case of /-o/ and /-u/ and to analyze the effect of lexical gender in the case of /-as/ and /-es/.

4.3 Results

In this section I review the results from the statistical analyses. For each linguistic variable, I first provide a general description of the distribution of the data and later report the results of the random forests and regression analyses.

4.3.1 /-o/ vs. /-u/

As I mentioned in section 4.2.4.2, the values for the first and second formants of [o] and [u] were used to perform a Discriminatory Analysis of Principal Components, identifying the probabilities of each token of belonging to one of the two predetermined categories, /-o/ or /-u/. All the tokens that had a probability higher than 0.5 of belonging to /-u/ were coded as such, while the rest of the tokens were coded as /-o/. Table 4 shows the overall distribution of the tokens according to the DAPC:

<table>
<thead>
<tr>
<th></th>
<th>/-o/</th>
<th>/-u/</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54.4%</td>
<td>45.6%</td>
<td>1321</td>
</tr>
<tr>
<td></td>
<td>(719)</td>
<td>(602)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Overall distribution of /-o/ and /-u/ as determined by the DAPC
The box plots in Figure 14 and Figure 15 represent the difference in formant values between the two vowels. The difference is statistically significant in both cases:

![Box plots](image)

Figure 14. Distribution of F1 (ERB) frequency by vowel (t=22.3, df=1318.33, p<0.01)

As we can see in Figure 14, /o/ has a significantly higher F1, indicating that the articulation of this vowel is lower than that of /u/.
Figure 15. Distribution of F2 (ERB) frequency by vowel (t=27.5, df=1312.11, p<0.01)

The plot in Figure 15 shows that the F2 of /u/ is significantly lower than that of /o/, indicating that /u/ is articulated further back in the vocal tract than /o/. As we can see in the box plots, there exists more variation in the values of the F2 than in those of the F1. However, because the difference between the means of both formants is statistically significant, it is important to include both continuous variables in the regression analysis.

In the remainder of this section I report the results of the statistical analyses, including the results of the analysis that evaluates the effect of the independent variables on categorical choice between vowels (categorical analysis) and the results of the analysis that examines the effect of the same independent variables on the continuous measure of the vocalic formants (continuous analysis).
4.3.1.1 Categorical analysis

The first step in testing the effect of the independent variables on the speakers’ election of /-o/ or /-u/ was to develop a random forest that depicts the relative importance of each linguistic and social variable. The results of the random forest are included in Figure 16:

Figure 16. Random forest with all predictors of /-o/ and /-u/ alternation

As I mentioned in section 4.2.5.2.2, I expect Education and Occupation to be highly correlated, since there are no speakers in the low education level that work in administrative, technical or teaching positions. Similarly none of the speakers that have a high education level work in manufacturing jobs. Table 5 shows the distribution of speakers in both categories:
Random forests can work around collinearity and are the ideal tool to elucidate which one of the correlated variables can best account for the distribution of the data. In this case, Occupation is not only more important than Education, it is also the most important factor in the variation between /-o/ and /-u/ after taking into account the effect of the individual speaker (which was included as a random variable). Because of this, only Occupation was included in the regression analysis, while the possible effects of Education are considered in an alternative model in which Education substitutes for Occupation.

The results of the mixed effects model on the complete data set using speaker as a random effect show that Occupation, Preceding segment, Gender, Vowel duration, Following segment, Prior form and Social tie are significant predictors of the choice between /-o/ and /-u/. Table 6 below summarizes the results of the regression analysis for the use of /-u/. The factor levels that have a significant effect on the variation are bolded:
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.30</td>
<td>0.40</td>
<td>-0.75</td>
<td>0.45</td>
</tr>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>1.79</td>
<td>0.38</td>
<td>4.74</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1.38</td>
<td>0.38</td>
<td>3.66</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preceding segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>-2.54</td>
<td>0.40</td>
<td>-6.34</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Palatal sound</td>
<td>-2.35</td>
<td>0.40</td>
<td>-5.18</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>-1.70</td>
<td>0.33</td>
<td>-6.60</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Velar</td>
<td>0.36</td>
<td>0.38</td>
<td>0.95</td>
<td>0.34</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.27</td>
<td>0.36</td>
<td>3.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>3.17</td>
<td>1.34</td>
<td>2.37</td>
<td>0.02</td>
</tr>
<tr>
<td>Following segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>-1.08</td>
<td>0.31</td>
<td>-3.55</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Palatal sound</td>
<td>-0.75</td>
<td>0.22</td>
<td>-3.40</td>
<td>0.01</td>
</tr>
<tr>
<td>Pause</td>
<td>-0.70</td>
<td>0.24</td>
<td>-2.94</td>
<td>0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>-0.64</td>
<td>0.21</td>
<td>-3.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Velar</td>
<td>0.16</td>
<td>0.25</td>
<td>0.62</td>
<td>0.53</td>
</tr>
<tr>
<td>Prior form (reference level is /-o/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/-u/</td>
<td>0.35</td>
<td>0.14</td>
<td>2.50</td>
<td>0.01</td>
</tr>
<tr>
<td>Social tie (reference level is First order tie)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second order tie</td>
<td>-0.68</td>
<td>0.32</td>
<td>-2.16</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 6. Linguistic and social factors contributing to the use of /-u/ in the complete data set.

According to the statistical analysis, speakers that work in commerce, restaurant and food industries and those who have a manufacturing job are significantly more likely to use /-u/ than those who work in administrative, technical and teaching positions.
However, the probability difference between the commerce and the manufacturing
groups is not statistically significant. Male speakers are also significantly more likely to
use the Asturian variant. In contrast to this, speakers that had a second order social tie
with their interlocutor are significantly more likely to use the Spanish vowel.

The use of /-u/ is also significantly less likely when /a/, a palatal sound, alveolar
or dental consonants precede the target vowel, as compared to a preceding bilabial
consonant. Because I did not have a null or neutral context to which I could compare the
effects of the preceding segment, I chose the bilabial context as the reference level. This
decision was based on the findings of Hillenbrand, Clark and Nearey (2001), who
analyzed the effect of consonant environment on the frequencies of formant values. In
their results, vowels in labial contexts patterned the closest to those in null contexts, in
which there was not a neighboring consonant. In order to get a complete picture of the
role of the preceding segment I also compared the effect of a preceding velar to the rest
of the contexts, since a preceding velar is expected to favor /-u/ the most (due to the
proximity in the place of articulation between the two sounds). The analysis reveals that
when the preceding context is a velar the use of /-u/ is significantly more likely than in
any other context, except for the bilabial one.

A similar effect to that of preceding segment is observed for the following
context, since a following /a/, palatal sound, dental/alveolar consonant and pause
significantly decrease the odds of using the Asturian variant (in comparison to a
following bilabial consonant). As was the case for the preceding segment, when a velar
consonant follows, the use of /-u/ is significantly more likely than when /a/, a palatal or al
alveolar/dental sound follow. Vowel duration is also significantly correlated with the use
of /-u/, such that the longer the duration of the vowel, the more likely it is for this vowel to be /-u/. Finally, an effect of prior form was observed, with a prior /-u/ priming more use of /u/.

Due to the collinearity between Occupation and Education, the effect of Education was tested independently. As I mentioned above, despite the fact that the relative importance of Occupation is higher than that of Education, the latter is still ranked third in the random forest. Thus, it is important to explore this independent variable to see what educational groups are more likely to use the Asturian variant. Table 7 shows the effect of education on the overall data set.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.18</td>
<td>0.43</td>
<td>-0.42</td>
<td>0.68</td>
</tr>
<tr>
<td>Education (reference level is High education)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle education</td>
<td>1.42</td>
<td>0.40</td>
<td>3.53</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Low education</td>
<td>1.23</td>
<td>0.39</td>
<td>3.15</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 7. Effect of Education on the use of /-u/ in the complete data set.

As we can see, the odds of using /-u/ are significantly higher in the low and middle education groups than in speakers with high education. The difference between low and middle education, however, is not statistically significant. The stronger effect that middle education has on the use of /-u/, as compared to the one of the low education group, is an unexpected result. This outcome could be due to the gender distribution among education levels, since there are seven male speakers and only two female speakers in the middle education group. Because men are significantly more likely to use
/-u/, it is possible that the higher number of male speakers in the middle education group is increasing the effect of middle education. In contrast with this, the low education group is evenly distributed with regards to gender (four females and three males). This issue is further explored in the discussion section.

4.3.1.2 Continuous analyses

The value of the first two formants of each vowel were also considered as continuous dependent variables and linear regression models were developed to test the effect of the independent variables on the formant frequencies. In Asturian Spanish, as well as in Spanish, /o/ shows higher formant values than /u/. The F1 is associated with vowel height, such that the higher the formant frequency the lower the vowel. The F2 is associated with place of articulation of the vowel, such that the lower the frequency of the formant, the further back it is articulated. The results for each formant are explored separately in the following sections.

4.3.1.2.1 First formant (F1)

As was done in the categorical analysis, the first step in testing the effect of the independent variables on the frequency of the F1 was to develop a random forest that indicates the relative importance of each predictor. The results of the random forest are included in Figure 17:
If we compare the results of the random forest for the categorical response to those obtained for the F1 we observe some differences. First, the effect of Preceding segment is stronger than that of Occupation. Second, the relative importance of Following segment, Age, Vowel duration and Social tie is different in both models.

The ranking of the predictors was used to determine the order in which each independent variable was added into a linear regression model that combined fixed and random effects (Speaker). The best-fit model included only Preceding segment, Occupation, Gender, Following segment and Vowel duration as significant predictors of a higher F1. The results of the linear mixed model are summarized in Table 8:

Figure 17. Random forest with all predictors of higher F1 frequency of /-o/ and /-u/.
The results of the linear mixed model indicate that the frequency of the F1 is significantly higher when the preceding segment is /a/, a velar, a dental or an alveolar consonant than when the preceding segment is a bilabial consonant. The difference in F1 frequency between the velar context and a palatal context or /a/ is also statistically significant. As for the following context, the results show that a following /a/ and pause have a significant effect on formant frequency, such that the frequency is higher in these

Table 8. Linguistic and social factors contributing to higher F1 frequency of /-o/ and /-u/ in the complete data set
contexts than when the adjacent sound is a bilabial consonant. No other following contexts are significantly different from a following bilabial consonant. When a following segment is a velar, the frequency of the F1 is only significantly lower than when /a/ follows.

The effect of vowel duration seems to contradict the findings of the categorical analysis. The linear model indicates that vowel duration and F1 frequency are positively correlated; i.e. the longer the duration of the vowel, the higher the value of the F1 (and the lower the vowel).

The effects of the social variables that were found to significantly predict the value of the F1 are parallel to the effects observed in the categorical analysis. The frequency of the F1 is significantly lower when the speaker works in commerce, restaurant or food industries than when the speaker has a teaching, administrative or technical job. A similar, but weaker, effect was found for manufacturing jobs. Finally, formant frequency is also significantly lower in male speakers than in female speakers, replicating the effect observed in the categorical analysis.

The variable Education was later substituted for Occupation, to investigate the effect of the two collinear variables separately. The results for Education are included in Table 9:
The results indicate that the frequency of the F1 is significantly lower in the low and middle education groups than in the high education group. As was the case in the categorical analysis, the difference between low and middle education is not statistically significant. We also notice that, contrary to the results shown in the categorical analysis, the effect of low education is stronger than that of middle education. The seemingly inconsistent result obtained in the categorical analysis could be attributed to a possible interaction between social tie and education, since that is the only social variable that has a significant effect on the use of /-u/ but not on the frequency of the F1. This, together with the larger presence of male speakers in the middle occupation group, could explain why the effect of middle education is stronger than that of low education in the categorical analysis. These issues are further addressed in section 4.4.2.

4.3.1.2.2 Second formant (F2)

A random forest was also developed for the analysis of the effect that the independent variables have on the frequency of the F2. The results of the random forest are included in Figure 18:
Figure 18. Random forest with all predictors of higher F2 frequency of /-o/ and /-u/.

The main difference between the results of this random forest and the one developed for the F1 is that Preceding segment is the most important predictor of the frequency of the F2, even over the random effect of speaker. Vowel duration also has a higher relative importance in comparison to what was observed in the categorical analysis and in the analysis of the F1. Finally, Occupation and Education are relegated to the bottom of the scale.

As in previous analyses, the relative importance of the independent variables was used to determine the order in which the factors were added to the linear model. However, not all the predictors were included in the regression analysis, since only Preceding segment, Vowel duration, Gender, Following segment and Prior form were
found to be significant in the stepwise comparison of the models. Table 10 summarizes the results of the best fit linear mixed model.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>17.60</td>
<td>0.24</td>
<td>73.51</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preceding segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatal sound</td>
<td>2.15</td>
<td>0.18</td>
<td>12.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>1.14</td>
<td>0.17</td>
<td>6.92</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>/a/</td>
<td>0.51</td>
<td>0.20</td>
<td>2.53</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Velar consonant</td>
<td>-0.13</td>
<td>0.19</td>
<td>-0.68</td>
<td>0.50</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>-4.04</td>
<td>0.59</td>
<td>-6.79</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.70</td>
<td>0.20</td>
<td>-3.40</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Following segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>0.71</td>
<td>0.11</td>
<td>6.32</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Palatal sound</td>
<td>0.66</td>
<td>0.12</td>
<td>5.63</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>/a/</td>
<td>0.32</td>
<td>0.16</td>
<td>1.99</td>
<td>0.05</td>
</tr>
<tr>
<td>Pause</td>
<td>0.28</td>
<td>0.12</td>
<td>2.26</td>
<td>0.02</td>
</tr>
<tr>
<td>Velar consonant</td>
<td>0.14</td>
<td>0.13</td>
<td>1.02</td>
<td>0.31</td>
</tr>
<tr>
<td>Prior form (reference level is /-o/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/-u/</td>
<td>-0.16</td>
<td>0.08</td>
<td>-2.14</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 10. Linguistic and social factors contributing to higher F2 frequency of /-o/ and /-u/ in the complete data set

The results show that the effect of preceding and following segments is similar to that observed in the categorical analysis. The frequency of the F2 is significantly higher when the preceding segment is a palatal sound, a dental or alveolar consonant or the low vowel /a/. The frequency of the F2 is also significantly lower when the preceding segment is a velar than in the other contexts, with the exception of a preceding bilabial.
The statistical analysis shows that the phonological contexts in which the F2 has a significantly higher frequency are the same as those in which /-u/ was disfavored in the categorical analysis. A significant effect of a following /a/, pause, dental and alveolar consonant or palatal sound was found for the frequency of the F2. In all these contexts the value of the F2 is also significantly higher than when the following context is a bilabial consonant. Comparing the effect of a following velar to the other contexts reveals a significant difference between this context and a following dental/alveolar consonant or a palatal sound, where the F2 shows higher frequencies.

The results also indicate that vowel duration and the frequency of the F2 are negatively correlated; i.e. the higher the frequency of the F2, the shorter the duration of the vowel. This correlation is statistically significant and indicates that shorter vowels are associated with realizations of [o]. This result parallels the observations made in the categorical analysis. The effect of Prior form is also similar to that observed in the categorical analysis, in that the frequency of the F2 is significantly lower when the prior form is /-u/.

The only social variable that was selected as significant in the regression analysis was Gender, showing that male speakers produce significantly lower frequencies in the F2 (i.e. realizations closer to /u/). This result also parallels the findings of the categorical analysis and the results of the continuous analysis of the F1.
4.3.1.3  **Effect of stressed vowel and stress pattern**

The effect of stressed vowel and stress pattern on the three dependent variables was tested excluding all the unstressed pronouns that appeared in proclitic position, since a stressed vowel could not be determined in these tokens. The results of the categorical analysis show that while stress pattern did not have a significant effect on the use of /-u/, the quality of the stressed vowel had a significant effect on the choice between the two variants. The results of this analysis are included in Table 11:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>Z Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.04</td>
<td>0.44</td>
<td>-0.08</td>
<td>0.93</td>
</tr>
<tr>
<td>Stressed vowel (reference level is /a/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u/</td>
<td>0.56</td>
<td>0.26</td>
<td>2.16</td>
<td>0.03</td>
</tr>
<tr>
<td>/o/</td>
<td>0.52</td>
<td>0.27</td>
<td>1.95</td>
<td>0.06</td>
</tr>
<tr>
<td>/e/</td>
<td>-0.12</td>
<td>0.20</td>
<td>-0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>/i/</td>
<td>-0.09</td>
<td>0.22</td>
<td>-0.44</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Table 11. Effect of Stressed vowel on the use of /-u/

As we can see, when the stressed vowel is /u/ the odds of having /-u/ are significantly higher than when the stressed vowel is a front vowel. No other factor levels differ significantly from a stressed /a/. The effect of stressed /u/ was also significantly different from that of /e/ and /i/ but not from that of /o/.

The continuous analyses also show that the quality of the stressed vowel has a significant effect on the frequency of the two formants, while the stress pattern does not. The results for the F1 indicate that the frequency of this formant is lower when the stressed vowel is /o/ than when /a/ is in this position. Table 12 summarizes the results:
These results differ from what we observed in the categorical analysis, where the same effect was reported but for stressed /u/ instead of stressed /o/, even though the difference in the effect of the two stressed vowels was not statistically significant. In the discussion section I address this issue and I explore the possible correlations and data distributions that might explain the results. The effect of stressed /o/ is also significantly different from that of stressed /i/.

Finally, the continuous analysis of the F2 reveals that the formant frequency is significantly higher when the stressed vowel is /e/ than when /a/ is the vowel in this position. The effect of stressed /e/ is also significantly different from that of stressed /o/ and /u/. The behavior of the two back vowels in stressed position does not differ significantly.
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>16.90</td>
<td>0.25</td>
<td>67.61</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Stressed vowel (reference level is /a/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>0.24</td>
<td>0.10</td>
<td>2.36</td>
<td>0.02</td>
</tr>
<tr>
<td>/i/</td>
<td>0.15</td>
<td>0.11</td>
<td>1.34</td>
<td>0.18</td>
</tr>
<tr>
<td>/o/</td>
<td>-0.14</td>
<td>0.14</td>
<td>-0.99</td>
<td>0.32</td>
</tr>
<tr>
<td>/u/</td>
<td>-0.18</td>
<td>0.13</td>
<td>-1.39</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 13. Effect of Stressed vowel on higher F2 frequency of /-o/ and /-u/.

It should be noted that the relative importance of the effect of stressed vowel is lower in each case than that of the significant predictors determined in the overall categorical analysis.

4.3.1.4 Mass/count distinction in nouns

The class of nouns was tested separately to investigate whether the semantic differentiation between mass and count nouns has an effect on the choice between /-o/ and /-u/. Previous research on Asturian and Asturian Spanish has proposed that, even though masculine singular nouns are invariably marked with /-u/ in Asturian, the alternation between the two back vowels in nouns could be associated with the semantic properties of that noun; i.e. whether it is a mass or a count noun (D’Andrés 1994, 1998, García Arias 2003, Prieto 1991). However, whether the noun was semantically mass or noun was not selected as a significant predictor in any of the statistical analyses.
4.3.2 /-as/ vs. /-es/

The values of the first and second formants of [a] and [e] were used to perform a Discriminatory Analysis of Principal Components, which determined the probabilities of each token of belonging to the /a/ or /e/ categories. All the tokens that had a probability higher than 0.5 of belonging to /e/ were coded as such and the rest were coded as /a/.

Table 14 shows the overall distribution of the tokens in the data based on the DAPC:

<table>
<thead>
<tr>
<th>/-as/</th>
<th>/-es/</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>62%</td>
<td>38%</td>
<td>813</td>
</tr>
<tr>
<td>(504)</td>
<td>(309)</td>
<td></td>
</tr>
</tbody>
</table>

Table 14. Overall distribution of /-as/ and /-es/ as determined by the DAPC

As we can see, while the frequency of /-o/ and /-u/ was evenly distributed (as shown in Table 4), the most frequent variant is /-as/ constituting 62% of the data.

The plots in Figure 19 and Figure 20 represent the difference in formant values between /a/ and /e/, which is statistically significant in both cases:
Figure 19 shows that the frequency of the F1 of /e/ is significantly lower than that of /a/, indicating that /e/ is produced higher than /a/.

Figure 20. Distribution of F2 (ERB) frequency by vowel (t=-21.4, df=634.087, p<0.01)
In Figure 20 we see that the F2 of /e/ is significantly higher than that of /a/, indicating that /e/ is articulated further to the front than /a/.

In the remainder of this section I report the results of the categorical and the continuous analyses for the /-as/, /-es/ variable.

4.3.2.1 Categorical analysis

As I did for the /-o/, /-u/ variable, the first step in testing the effect that the independent variables have on the choice between /-as/ and /-es/ was to develop a random forest that provided us with a ranking of the predictors based on their relative importance in accounting for the variation. The random forest is shown in Figure 21:

Figure 21. Random forest with all predictors of /-as/ and /-es/ alternation
While Speaker and Occupation were the most important predictors of the use of /-u/, Preceding segment is the most important predictor in the case of /-as/ and /-es/, even over the random effect of speaker. Furthermore, Education is now ranked higher than Occupation. Finally, the relative importance of Gender is also much lower in the /-as/, /-es/ variable than in the /-o/, /-u/ variable and the opposite is true for Prior form.

Table 15 reflects the differences in the ranking of the external and internal predictors in both linguistic variables (excluding Following segment, which was particular to the /-o/, /-u/ variable). The bolded predictors are those that were found to be significant in each regression analysis:

<table>
<thead>
<tr>
<th>/-o/, /-u/</th>
<th>/-as/, /-es/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>Preceding segment</td>
</tr>
<tr>
<td>Occupation</td>
<td>Speaker</td>
</tr>
<tr>
<td>Preceding segment</td>
<td>Education</td>
</tr>
<tr>
<td>Education</td>
<td>Prior form</td>
</tr>
<tr>
<td>Gender</td>
<td>Vowel duration</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>Occupation</td>
</tr>
<tr>
<td>Prior form</td>
<td>Age</td>
</tr>
<tr>
<td>Social tie</td>
<td>Lexical category</td>
</tr>
<tr>
<td>Age</td>
<td>Social Tie</td>
</tr>
<tr>
<td>Frequency</td>
<td>Gender</td>
</tr>
<tr>
<td>Lexeme language</td>
<td>Frequency</td>
</tr>
<tr>
<td>Lexical category</td>
<td>Lexeme language</td>
</tr>
</tbody>
</table>

Table 15. Comparison of the ranking of predictors in the two linguistic variables

A mixed effects model that uses speaker as a random variable was developed to fit the data. The best fit model only includes Preceding segment, Education, Prior form and Vowel duration as significant predictors of the use of /-es/. Table 16 summarizes the results of the best fit model that predicts the choice between /-as/ and /-es/:
According to the regression analysis, when the final vowel is preceded by a palatal sound the odds of this vowel being /-e/ are significantly higher than in a bilabial context. The effect of a preceding palatal is also significantly different from that of any other context, favoring the use of /-es/ more than any other phonological context. Vowel duration is also significantly correlated with the probability of having /-es/, such that the longer the duration of the vowel the less likely it is for this vowel to be /e/. A priming effect similar to that reported for the /-o/, /-u/ variable was also observed in this case, such that a prior /-es/ primes more /-es/.

As I mentioned in the description of the random forest, Education was given a higher relative importance than Occupation, which contrasts with what we observed for the /-o/, /-u/ variable. Thus, the regression model was first developed using Education, and Occupation was tested later in place of Education, to avoid effects of collinearity.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.51</td>
<td>0.84</td>
<td>-2.99</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preceding segment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatal sound</td>
<td>3.34</td>
<td>0.76</td>
<td>4.41</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>1.06</td>
<td>0.74</td>
<td>1.44</td>
<td>0.15</td>
</tr>
<tr>
<td>Velar consonant</td>
<td>-0.87</td>
<td>0.82</td>
<td>-1.06</td>
<td>0.29</td>
</tr>
<tr>
<td>Education (reference level is High)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>2.43</td>
<td>0.55</td>
<td>4.45</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Middle</td>
<td>0.83</td>
<td>0.51</td>
<td>1.61</td>
<td>0.10</td>
</tr>
<tr>
<td>Prior form (reference level is /-as/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/-es/</td>
<td>0.79</td>
<td>0.20</td>
<td>4.03</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>-16.85</td>
<td>3.59</td>
<td>-4.70</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 16. Linguistic and social factors contributing to the use of /-es/ in the complete data set
The results for Education (in Table 13) are similar to those observed in the variation between /-o/ and /-u/. Speakers with a low education level are significantly more likely to use /-es/ than those with a high education level. The difference between the two linguistic variables with respect to Education is that in the case of /-as/ and /-es/, low and middle education also differ significantly: the odds of /-es/ are significantly lower in the middle education group than in the low education one. Table 17 reflects the results for Education taking low education as the reference level:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.79</td>
<td>0.59</td>
<td>1.35</td>
<td>0.18</td>
</tr>
<tr>
<td>Education (reference level is Low)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>-2.43</td>
<td>0.55</td>
<td>-4.45</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Middle</td>
<td>-1.60</td>
<td>0.52</td>
<td>-3.10</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 17. Effect of Education on the use of /-es/ in the complete data set.

Because Occupation was still ranked high in the random forest, its possible effect on the data was tested independently, substituting Education in a separate logistic model. The results of this analysis are included in Table 18:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.91</td>
<td>0.63</td>
<td>1.45</td>
<td>0.14</td>
</tr>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>2.14</td>
<td>0.63</td>
<td>3.39</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1.20</td>
<td>0.56</td>
<td>2.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 18. Effect of Occupation on the use of /-es/ in the complete data set.
This analysis yields results similar to those obtained for the /-o/, /-u/ variable. The use of /-es/ is significantly more likely among speakers that work in commerce or in the restaurant and food industries than among those who work in administrative, teaching and technical jobs. Similarly, but to a lesser extent, working in manufacturing jobs also increases the odds of using /-es/. As was the case in the /-o/, /-u/ variable, the probability difference between manufacturing and commerce, restaurant and food industries is not statistically significant.

Table 19 below compares the effect of social predictors on each of the linguistic variables:

<table>
<thead>
<tr>
<th></th>
<th>/-o/, /-u/</th>
<th>/-as/, /-es/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>1.79</td>
<td>2.14</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1.38</td>
<td>1.20</td>
</tr>
<tr>
<td>Education (reference level is High)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1.23</td>
<td>2.43</td>
</tr>
<tr>
<td>Middle</td>
<td>1.42</td>
<td>0.83</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.27</td>
<td>N/S</td>
</tr>
<tr>
<td>Social tie (reference level is First order tie)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second order tie</td>
<td>-0.68</td>
<td>N/S</td>
</tr>
</tbody>
</table>

Table 19. Comparison of the effects of social predictors on the categorical choice between linguistic variants.

As we can see, while the effects of Occupation are similar in both variables, the role of Education differs between the two, especially in the behavior of the middle
education group, such that while middle education has the strongest effect on the use of /-u/, the low education has the strongest effect on the use of /-es/. Finally, it is important to note that while Social tie and Gender were found to be significant in the variation between /-o/ and /-u/, these predictors did not have a significant effect on the choice between /-as/ and /-es/.

4.3.2.2 Continuous analyses

The value of the first two formants of each vowel were also considered as numeric dependent variables and linear regression models were developed to test the effect of the independent variables on the frequencies of these formants. In Asturian Spanish, as well as in Spanish, /a/ shows a higher F1 but a lower F2 than /e/. The higher F1 of /a/ corresponds to its lower articulation while the lower F2 is a consequence of its central articulation. As I mentioned in section 4.3.1.2, the higher the frequency of the F2, the further front it is articulated. The results for each formant are explored separately in the following sections.

4.3.2.2.1 First formant (F1)

A random forest was developed for the analysis of the effect that the independent variables have on the frequency of the F1. The results of the random forest are included in Figure 22:
The greatest difference between this random forest and the one developed for the categorical distinction between /-as/ and /-es/ is the relative importance given to Preceding segment in the continuous analysis of the F1, which is now below the effect of prior form. The lower relative importance of this factor in the F1 is not surprising since, as I explain in section 4.4.1.1, the place of articulation of the preceding segment is expected to have a stronger effect on vowel fronting (F2) than on vowel height (F1).

The linear mixed model only included Education, Gender, Vowel duration, Prior form, Preceding segment and Lexical category as significant predictors of the frequency of the F1. The effect of Occupation was tested in an alternative model that excluded the
Education variable, in order to avoid problems of collinearity. The results of the linear mixed model are included in Table 20:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>12.63</td>
<td>0.24</td>
<td>53.17</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Education (reference level is High)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>-1.02</td>
<td>0.20</td>
<td>-5.06</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.71</td>
<td>0.21</td>
<td>3.40</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.77</td>
<td>0.18</td>
<td>-4.42</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>9.36</td>
<td>0.93</td>
<td>10.05</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Prior form (reference level is /-as/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/-es/</td>
<td>-0.33</td>
<td>0.06</td>
<td>-5.22</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preceding segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatal sound</td>
<td>-0.51</td>
<td>0.17</td>
<td>-2.95</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>-0.35</td>
<td>0.17</td>
<td>-2.08</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>Velar consonant</td>
<td>-0.12</td>
<td>0.20</td>
<td>-0.58</td>
<td>0.57</td>
</tr>
<tr>
<td>Lexical category (reference level is Noun)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determiner</td>
<td>0.37</td>
<td>0.07</td>
<td>5.03</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Clitic</td>
<td>0.04</td>
<td>0.17</td>
<td>0.25</td>
<td>0.80</td>
</tr>
<tr>
<td>Adjective</td>
<td>0.04</td>
<td>0.10</td>
<td>0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>Demonstrative</td>
<td>-0.10</td>
<td>0.14</td>
<td>-0.70</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 20. Linguistic and social factors contributing to higher F1 frequency of /-as/ and /-es/ in the complete data set

Vowel duration and Prior form have the same effect in the categorical analysis and the continuous analysis of the F1. The longer the duration of the vowel the higher the frequency of the F1; i.e. the lower the articulation of the vowel. As expected, having a prior /-es/ significantly lowers the value of the F1.
In the continuous analysis of the F1 we also observe a similar role of the preceding segment. In this case having a preceding palatal sound, a dental or an alveolar consonant significantly decreases the frequency of the F1, whereas only an effect of a palatal sound was found to be significant in the categorical analysis. As in the categorical analysis, the effect of a preceding palatal is also significantly different from that of any other context. Finally, Lexical category was also selected as a significant predictor of F1 frequency. This contrasts with the results of the categorical analysis of /-as/ and /-es/ and also with the results obtained for the /-o/, /-u/ variable, since Lexical category did not have a significant effect in any of them. Here we observe that the frequency of the F1 is significantly higher when the token belongs to the determiner category.

The results for education are similar to those observed in the categorical analysis. The low education group shows a significantly lower frequency of the F1. A similar but weaker effect is found for the middle education group. Even though in the categorical analysis low and middle education differ significantly in their use of /-es/, these two groups are not significantly different with regards to the frequency of the F1.

When Occupation is substituted for Education, we obtain the results shown in Table 21, which are comparable to those obtained in the categorical analysis:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>12.72</td>
<td>0.19</td>
<td>65.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>-1.05</td>
<td>0.20</td>
<td>-5.14</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>-0.69</td>
<td>0.21</td>
<td>3.35</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 21. Effect of Occupation on F1 frequency of /-as/ and /-es/ in the complete data set.
The results indicate that the frequency of the F1 is significantly lower among speakers that work in commerce, restaurant and food industries. The same but weaker effect is found for manufacturing jobs.

4.3.2.2  Second formant (F2)

Figure 23 reflects the results of the random forest that was developed to evaluate the relative importance of the effect of each independent variable on the frequency of the F2.

![Random forest with all predictors of higher F2 frequency of /-as/ and /-es/.](image)

Figure 23. Random forest with all predictors of higher F2 frequency of /-as/ and /-es/.
The main difference between this ranking and the rest of the results obtained in the analysis of the /-as/, /-es/ variable is that in the case of the F2, Vowel duration is relegated to the bottom of the scale. The effect of the preceding segment is also more important than in the analysis of the F1. The rest of the ranking is similar to the one provided for the categorical analysis and the continuous analysis of the F1.

The best fit linear mixed model only includes Gender, Preceding segment, Lexical category and Prior form as significant predictors of F2 frequency. The results are summarized in Table 22:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>19.01</td>
<td>0.25</td>
<td>75.25</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-1.09</td>
<td>0.25</td>
<td>-4.40</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preceding segment (reference level is Bilabial consonant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatal sound</td>
<td>1.52</td>
<td>0.18</td>
<td>8.49</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dental/Alveolar consonant</td>
<td>0.77</td>
<td>0.17</td>
<td>4.42</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Velar consonant</td>
<td>0.70</td>
<td>0.21</td>
<td>3.39</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Lexical category (reference level is Noun)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determiner</td>
<td>-0.44</td>
<td>0.07</td>
<td>-5.91</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Adjective</td>
<td>-0.29</td>
<td>0.10</td>
<td>-2.78</td>
<td>0.80</td>
</tr>
<tr>
<td>Demonstrative</td>
<td>-0.15</td>
<td>0.14</td>
<td>-1.08</td>
<td>0.28</td>
</tr>
<tr>
<td>Clitic</td>
<td>0.03</td>
<td>0.17</td>
<td>0.15</td>
<td>0.88</td>
</tr>
<tr>
<td>Prior form (reference level is /-as/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/-es/</td>
<td>0.23</td>
<td>0.06</td>
<td>3.33</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 22. Linguistic and social factors contributing to higher F2 frequency of /-as/ and /-es/ in the complete data set
The results indicate that the frequency of the F2 is significantly lower when the speaker is male. Since lower frequency of the F2 is associated with a less fronted vowel, the result for Gender contradicts what has been found for the /-o/, /-u/ variable and in the F1 analysis of /-a/ and /-e/. This issue will be addressed in the discussion section of this chapter. No other social variables were selected as significant predictors of the variation.

The results show that a preceding palatal sound, an alveolar, dental or velar consonants significantly raise the frequency of the F2, with the palatal sound having the strongest effect. In fact, the effect of a preceding palatal is significantly different from all the other contexts. As was the case in the analysis of the F1, if the token is a determiner, the frequency of the F2 is significantly lower. A parallel but weaker effect is also observed when the lexical item is an adjective. Finally, the presence of a prior /-es/ significantly raises the frequency of the F2.

4.3.2.3 Effect of stressed vowel and stress pattern

All the tokens that were not unstressed pronouns in proclitic position were tested separately to include the effect of the quality of the stressed vowel and the stress pattern of the word. As we saw in sections 4.3.1.3 and 4.3.2.3, the quality of the stressed vowel had a significant effect on the choice between /-o/ and /-u/. However, no effect of either of these variables was found on any of the dependent variables for the /-as/ and /-es/ variation.
4.3.2.4 Effect of lexical gender

The group of nouns was also tested separately to explore a possible effect of the lexical gender of the word. If we recall from section 3.1.3.2, the use of the morpheme /-es/ in Asturian can also mark masculine plural in cases in which the corresponding masculine singular form ends in /-a/. Thus, nouns were further coded for lexical gender. However, no effect of this linguistic factor was found in the data.

4.3.3 Summary

Table 23 includes a comparison of the linguistic factors that were found to be significant predictors of all or some of the dependent variables. The presence of a check mark indicates that a particular predictor had a significant effect on the corresponding dependent variable:

<table>
<thead>
<tr>
<th></th>
<th>-o/ vs. -u/</th>
<th>-as/ vs. -es/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat.</td>
<td>F1</td>
</tr>
<tr>
<td>Preceding segment</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Following segment</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Prior form</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Stressed vowel</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Lexical category</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 23. Significant linguist predictors across the six dependent variables

Table 24 compares the effect of significant social factors on the dependent variables:
<table>
<thead>
<tr>
<th></th>
<th>/-o/ vs. /-u/</th>
<th>/-as/ vs. /-es/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat.</td>
<td>F1</td>
</tr>
<tr>
<td>Occupation</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Education</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Gender</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Social tie</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Table 24. Significant social predictors across the six dependent variables

Despite the fact that the effect of the individual factor levels can vary in some dependent variables, this summary provides us with the relative importance of the linguistic and social factors across all the dependent variables explored in the statistical analysis. The comparison shows that the phonological context and the prior form are the most important linguistic predictors of the variation between the Asturian and Spanish vowels. It also indicates that occupation, education and gender are the most important social correlations. In the following section I explore in detail each of the effects observed.

4.4 Discussion

In this section I discuss the results of the quantitative analysis described in the previous section. The discussion is divided in two parts: first, I analyze the effect that the linguistic independent variables have on the categorical choice of each vowel and on the frequency of the first two formants; then I discuss the observed correlations between social predictors and each of the dependent variables.
4.4.1 Internal/linguistic variables

The summary presented in Table 23 indicates that the phonological context in which each vowel occurs plays the most important role among the linguistic variables in the variation between Spanish and Asturian vowels. Furthermore, in the majority of cases the effect of the preceding segment is stronger than that of social variables. Vowel duration is the second most important linguistic factor, being significant for all the dependent variables except for the F2 of /-as/ and /-es/. Because vowel duration is intrinsic to the quality of the vowel, it is impossible to determine the direction of the effect, i.e. if the vowel duration is a result of the different quality of the vowel or vice versa. As a consequence its role in the variation is not as significative as that of other linguistic factors. A priming effect of the prior form was also observed for most of the variables. Finally, the effect of the quality of the stressed vowel was only observed for the /-/o/,-/u/ variable while lexical category was a significant predictor only for the /-/as/, /-/es/ variable.

The remaining linguistic factors were not found to have a statistically significant effect on the variation. This group includes Frequency, Stress pattern, Lexeme language, Mass/count distinction and Lexical gender.

4.4.1.1 Preceding segment

Previous research on the production of vowels has shown that neighboring sounds have an effect on the acoustic properties of vowels. With respect to consonants, it has been found that place of articulation can affect the frequency of the first two formants.
Stevens and House (1963) and Hillebrand, Clark and Neary (2001) found that, for English, the effect of neighboring consonants on formant frequency is greater in the F2 than in the F1. Stevens and House also point out that, while the manner and voicing of the consonants also have an effect on formant frequencies (measured at the vowel’s midpoint), place effects were the most important. Fowler and Saltzman (1993:190) explain these coarticulatory effects as resulting from two sources: “the time courses of the activation waves for the gestures in the utterance (the “speech plan”), and the manner in which the coordinative constraints of temporally overlapping (coproduced) gestures blend or interact with one another”.

Figure 24 summarizes the effects of preceding context for the /-o/, /-u/ variable:

![Figure 24. Effect of preceding segment on the F1 and F2 of [o] and [u]](image-url)
In Figure 24, each point represents the average frequency of the F1 (square points) and the F2 (circular points) in a particular phonological context. For example, the average frequency of the F1 in a bilabial context is approximately 500 Hz, while the average frequency of the F2 in the same context is approximately 1150 Hz. The differences observed in the F2 are also greater than those reported for the F1. The greatest effect observed on the frequency of the F1 is that of vowel /a/. Because /a/ is articulated in a lower position, it has a higher F1 value than any of the two back vowels. The influence of this higher F1 significantly raises the frequency of the corresponding formant in the back vowels that come in contact with /a/.

The frequency of the F2 is significantly higher when the preceding sound is a dental or alveolar consonant or a palatal sound. The first of these results replicates the findings made by Hillebrand et al. for English. The authors found significant upward shifts in F2 for alveolar environments. Apart from the dental/alveolar context, a preceding palatal also raises significantly the frequency of the F2. Because the F2 reflects the fronting of the vowel, such that more fronted vowels have a higher F2, it is expected that the immediate presence of consonants in which the tongue is an active articulator in the front of the vocal tract trigger higher F2 frequencies (and fronter articulations) in a neighboring vowel. In contrast with this, the central articulation of /a/, and its consequent lower F2 frequency, significantly lowers the F2 of the back vowels. The effect of a preceding velar is also significantly different from that of all other contexts, with the exception of the bilabial one. In this context we find the most /u/-like realizations according to the frequency of both formants. This is expected, since the point of articulation of /u/ is closest to that of velar and uvular consonants.
Figure 25 shows the effects of preceding context on formant frequencies for the /- as/, /-es/ variable:

![Figure 25. Effect of preceding segment on the F1 and F2 of [a] and [e]](image)

According to the results of the regression analysis, the frequency of the F1 is significantly lower in dental/alveolar and palatal contexts than when the preceding sound is a bilabial consonant. Velar and bilabial contexts do not differ significantly. The results for the dental/alveolar context are in accordance with findings of Hillebrand et al. for front and central vowels in English and the effect of a palatal sound is expected since their articulation approximates that of /e/. Once again, the effects of the preceding segment are greater in the frequencies of the F2, where we see a significant upward shift for alveolar/dental and palatal environments.

To summarize, the vowels under study show some coarticulation effects that have been found to significantly alter formant frequencies in other languages like English. All
of the effects of neighboring sounds are explained if we compare the place of articulation of the preceding segment to that of the target vowel. Final vowels tend to approximate the articulation of the preceding sound. If we now consider the results of the categorical analysis we see that the preceding context does not only affect the phonetic realization of a particular vowel, but it also constrains in some cases what vocalic sound the speakers choose. A preceding dental, alveolar, palatal and /a/ significantly increase the odds of having /-o/ as the final vowel. This result reflects the expected stronger effect that the preceding context has on the frequency of the F2; i.e. it is the backness or frontness of the vowel that is influenced by the preceding context and not vowel height, since consonants would always favor higher vowels as a result of coarticulation. Similarly to what we observed in the back vowels, a preceding palatal sound significantly increases the probabilities of having /-es/; however, the effect of preceding alveolar, dental and velar sounds is not strong enough to result in a categorical choice of /-es/ over /-as/. Instead, what we find in the latter contexts is more fronted realizations of [a] and [e] respectively.

4.4.1.2 Following segment

Anticipatory coarticulation effects have been found to be weaker than carryover effects like the ones described in the previous section (Fowler 1981). We also observe this tendency in the Asturian Spanish data illustrated in Figure 26:

---

7 It is interesting to note that, even though this is the case in the data analyzed for this dissertation, this tendency is the opposite to what has been found for Spanish, where anticipatory assimilatory effects are stronger than carry-over...
Even if the anticipatory coarticulation effects are similar to the results observed for the preceding contexts in terms of the direction of the effects, the strength of these effects is weaker in this case. This is especially noticeable in the F2, where the effects of a palatal sound and /a/ are much stronger in the preceding context.

4.4.1.3 Vowel duration

Vowel duration was significantly correlated with the frequency of the F1 in the two linguistic variables. Previous research has found that, cross-linguistically, low vowels tend to be realized with longer duration than high vowels, provided vowel length is not phonologically contrastive (Chladkova, Escudero and Boersma 2011). In Spanish, Marín Gálvez (1994-1995) found that among Spanish vowels, [a] had the longest
duration and [i] the shortest. The tendency of low vowels to have a longer duration also held for the back vowels so that [o] was articulated with longer duration than [u].

In the Asturian Spanish data, for the /-as/, /-es/ variable, I reported that the longer the duration of the vowel the higher the frequency of the F1 and the lower the vowel. The categorical analysis also indicated that the probabilities of having /-es/ increased as the duration of the vowel decreased. A similar effect for the F1 was found for the /-o/, /-u/ variable, such that vowel duration was positively correlated with the frequency of the formant. Thus, the correlations between vowel height and vowel duration found in this analysis go in accordance to previous observations in Spanish and other languages.

4.4.1.4 Structural priming and spreading activation

Previous research has found that in discourse speakers tend to repeat linguistic forms or structures. This repetition can occur at the morphosyntactic level (structural priming – Bock and Griffin 2000, Gries 2005) or at the phonological one. They all constitute cases of linguistic perseveration (Dell 1986, Dell, Burger and Svec 1997), by which the use of a particular form or structure triggers further use of that form or structure. To explain this effect Dell (1986) developed the Spreading Activation Theory. According to this theory, the production of an utterance involves the construction of its mental representation before the message is uttered. Within each linguistic level the choice of one element over an alternative depends on the level of activation of those options. The higher the activation, the more likely a particular element is to be chosen. This spreading activation can occur in any direction, left to right and right to left.
In the results of the quantitative analysis we observed that the use of a prior /-u/ triggers more use of /-u/ and a similar effect was found for /-es/. This goes along with the prediction of Spreading Activation Theory and with the findings of other variationist research that explores the effects of linguistic perseveration. In Spanish, for example, Poplack (1981) investigated the realization of plural /-s/ in Puerto Rican Spanish and found that the use of /-s/ leads to more /-s/ while deletion of this phoneme leads to more deletion. Furthermore, she also observed a tendency towards local redundancy such that marking of /-s/ occurred either in every member of an NP or in none of them. A similar tendency is observed in the Asturian Spanish data, where the use of /-es/ or /-as/ in the determiner favors the use of the same form in the subsequent nouns and adjectives. Thus, examples like “les mochilas” (‘the backpacks’) are very rarely found in the data.

4.4.1.5 Quality of the stressed vowel

The effect of the quality of the stressed vowel was only observed in the /-o/, /-u/ variable. As we saw in section 4.3.1.3, an effect of different stressed vowels was found to be significant in the categorical analysis and the two continuous analyses. In the categorical analysis, a stressed /u/ significantly increases the odds of having /-u/ as the response variant. In contrast with this, the results of the continuous analyses indicate that it is a stressed /o/ and not /u/ the factor level that significantly lowers the frequency of the F1 (yielding a higher vowel), while a stressed /e/ significantly raises the frequency of the F2 (resulting in a more fronted vowel). The divergence in the results calls for a closer look at this variable.
The distribution of stressed vowels across the possible segments preceding the final vowel reveals that not all stressed vowels occur in all possible contexts. Table 25 shows this distribution:

<table>
<thead>
<tr>
<th>Preceding segment</th>
<th>Stressed vowel</th>
<th>Bilabial consonant</th>
<th>Dental/alveolar consonant</th>
<th>Palatal sound</th>
<th>Velar consonant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/a/</td>
<td>8</td>
<td>102</td>
<td>147</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>31</td>
<td>2</td>
<td>232</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>/i/</td>
<td>15</td>
<td>0</td>
<td>106</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>/o/</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>9</td>
<td>0</td>
<td>139</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 25. Distribution of tokens of /-o/ and /-u/ across Stressed vowel and Preceding segment

As we can see, there are several empty cells when the preceding segment is a bilabial consonant or the vowel /a/. Furthermore the majority of the tokens of stressed /u/ occur in the dental/alveolar environment. Because of this distribution, I tested for the effect of stressed vowel in the only context in which the distribution of tokens with regards to the quality stressed vowel is more even, when a dental or an alveolar consonant precedes the vowel. The results show no effect of the quality of stressed vowel on the final vowel.

The distribution of the independent linguistic variables makes any generalization regarding the effect of the quality of the stressed vowel on the alternation between /-o/ and /-u/ impossible. In order to determine the role of this linguistic factor on the variation, a study that tests for stressed vowel in a controlled phonetic environment is necessary. It should be noted, however, that any effect that the stressed vowel could have
in the data of this study is weaker than that of the phonological context, the duration of
the vowel and the prior form.

4.4.1.6 Lexical category

The lexical category of the tokens only had a significant effect in the variation
between /-as/ and /-es/. The results indicate that the frequency of the F1 was significantly
higher when the token was a determiner than when it was a noun. The results also show a
significant downward shift of the F2 when the lexical category was a determiner than
when it was a noun. Despite these results, we do not see a significant categorical effect of
this independent variable.

When examining the relationship between preceding context and lexical category
we find that in determiners the target vowel is invariably preceded by an alveolar sound.
Furthermore, as we saw in the case of the stressed vowel, the only phonological context
that occurs in all lexical categories is a preceding alveolar or dental consonant. The
distribution of the two variables is represented in Table 26:

<table>
<thead>
<tr>
<th>Lexical category</th>
<th>Preceding segment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bilabial consonant</td>
</tr>
<tr>
<td>noun</td>
<td>20</td>
</tr>
<tr>
<td>adjective</td>
<td>4</td>
</tr>
<tr>
<td>clitic</td>
<td>0</td>
</tr>
<tr>
<td>demonstrative</td>
<td>0</td>
</tr>
<tr>
<td>determiner</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 26. Distribution of tokens of /-as/ and /-es/ across Lexical category and Preceding segment
The effect of lexical category was tested on the group of tokens in which the preceding segment is an alveolar or a dental consonant. The results are summarized in Table 27. The effects of the other linguistic and social factors were also included in the analysis and remain significant when only the dental and alveolar contexts are considered:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>zValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.79</td>
<td>0.59</td>
<td>-3.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Lexical category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determiner</td>
<td>-0.90</td>
<td>0.29</td>
<td>-3.10</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Adjective</td>
<td>0.03</td>
<td>0.45</td>
<td>-0.08</td>
<td>0.94</td>
</tr>
<tr>
<td>Clitic</td>
<td>0.28</td>
<td>0.53</td>
<td>0.52</td>
<td>0.60</td>
</tr>
<tr>
<td>Demonstrative</td>
<td>0.43</td>
<td>0.50</td>
<td>0.87</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table 27. Effect of Lexical category on the use of /-es/ in tokens that have an alveolar or a dental consonant as the preceding segment.

The results show the same effect of determiner in the frequencies of the F1 and the F2 as when the entire data set was analyzed; i.e. we find significantly higher F1 frequencies and significantly lower F2 frequencies (indicating more /a/-like articulations) when the lexical item is a determiner. In this case we also observe a significant result in the categorical analysis. As expected, the probabilities of having /-es/ are significantly lower when the item is a determiner.

Because the target vowel in determiners is always preceded by an alveolar consonant one could attribute the effect of lexical category to the effect of the preceding segment. However, the effect of an adjacent alveolar sound on the vowel is the opposite of that reported for the determiner class; i.e. a preceding alveolar or dental consonant
significantly lowers the frequency of the F1 and raises the frequency of the F2 (favoring the mid front vowel). Furthermore, we should be able to see any effect caused by the preceding segment also in the clitics category, but this category did not have a significant effect on either of the two formants.

Another possibility is that the effect of lexical category is reflecting the effect observed for vowel duration; i.e. that lexical category and vowel duration are correlated. In fact, the unstressed nature of determiners and clitics would lead us to expect shorter duration in their vowels, since stress is the most important factor contributing to vowel duration in Spanish (Marín Gálvez, 1994-1995). Unstressed vowels are shorter than stressed vowels and the proximity of an unstressed syllable also has an effect on vowel duration, such that vowels are shortened by preceding and following unstressed syllables (Fowler 1981). Due to the abundance of paroxytones in Spanish, unstressed lexical items like determiners and clitics are more likely than other lexical categories to be adjacent to other unstressed syllables. When we examine the distribution of vowel duration across lexical categories we find that, in fact, the target vowels that appear in determiners and unstressed pronouns are shorter than those that appear in any other lexical category. The plot in Figure 27 illustrates this distribution.
However, the effect of duration is the opposite of the effect recorded for determiners; i.e. while determiners favor the use of /-as/, shorter duration favors the use of /-es/. In this case, we would also expect to see a similar trend in the clitics category, since their duration is very close to that of determiners.

Because the effect of vowel duration and a preceding alveolar sound are the opposite of that of determiners, we can conclude that the effect of lexical category is independent of that of other linguistic factors. Instead, I propose that determiners are more likely to maintain the Spanish variant because of their status as function words. Code-switching research has found that determiners, as well as other function words, are not frequently borrowed. The language of the determiner has also been used in some studies as an indicator of the language of the noun phrase (Sankoff et al. 1990, Torres
Cacoullos and Aaron 2003). If we compare the rates of use of /-es/ across lexical categories, we observe that, with the exception of clitics, closed-class items show the lowest rates of /-es/, in contrast with nouns and adjectives:

<table>
<thead>
<tr>
<th>Category</th>
<th>Forms with /-es/</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clitics</td>
<td>50% (12)</td>
<td>24</td>
</tr>
<tr>
<td>Nouns</td>
<td>43.41% (201)</td>
<td>463</td>
</tr>
<tr>
<td>Adjectives</td>
<td>42.02% (29)</td>
<td>69</td>
</tr>
<tr>
<td>Demonstratives</td>
<td>37.83% (14)</td>
<td>37</td>
</tr>
<tr>
<td><strong>Determiners</strong></td>
<td><strong>24.09% (53)</strong></td>
<td><strong>220</strong></td>
</tr>
</tbody>
</table>

Table 28. Use of /-es/ by lexical category

The behavior observed in clitics can be explained if we compare the pronominal system of Asturian to that of Spanish. “Les” exists as a dative unstressed pronoun in Spanish. However, in regions where ‘leísmo’ is frequent, such as the north of Spain, “les” is variably used as the accusative pronoun. The possible use of “les” in Spanish as a direct and indirect object pronoun makes the presence of /-es/ in the clitic category less marked than as a determiner, where the variant with the raised vowel is not possible in Spanish, only in Asturian.

### 4.4.1.7 Non-significant factors

The results of the statistical analysis indicate that Frequency, Stress pattern, Lexeme language, Lexical gender and the semantic categorization of the noun are not significant predictors of any of the dependent variables.
The fact that frequency was not selected as a significant factor could be attributed to the way in which frequency rates were determined. As I explained in section 4.2.5.1.6, because the only available frequency dictionary is based on written data, I used the lexical frequency of each item in the corpus under study. However, this corpus might be too small to obtain a reliable measure of frequency. Using a larger corpus in future research could help determine whether frequency is a significant predictor in the use of Asturian and Spanish variants.

The results for lexeme language suggest that any language combination of lexeme and morpheme are possible, contradicting the observations made by D’Andrés (2002). He stated that while Spanish lexemes could appear with morphemes in either language, Asturian lexemes are only possible with Asturian morphemes. However, the use of Asturian lexemes does not constrain the language of the morpheme in my data.

4.4.1.7.1 Mass/count distinction, lexical gender and morpho-phonological variation

One of the objectives in testing the role of lexical gender in the /-as/, /-es/ variable and the possible effect of the mass/count distinction in nouns in the /-o/, /-u/ variable was to test whether the distinction between the vowels was morphologically productive. If the use of /-o/ was significantly higher in mass nouns, we could conclude that speakers are using the morphological distinction between mass and count nouns present in Asturian and applying it to masculine singular nouns, a category that is marked invariably with /-u/ in prescriptive Asturian, marking masculine singular mass nouns with /-o/ and masculine singular count nouns with /-u/. The effect of the mass/count distinction is significant when this independent variable is considered in isolation. However, once other linguistic
variables are added to the model, the effect is lost. The same occurs in the analysis of lexical gender.

The results observed for the /-o/, /-u/ variable raise the question of whether the variation observed in Asturian Spanish vowels is morpho-phonological or phonetic. The fact that the variation is restricted to a limited number of morphological environments (described in section 3.2.2.2 and section 3.2.3.2) and that the variants are used to mark a number of morphological categories supports the former. However, the acoustic similarities between the realizations of [o] and [u] leads us to question whether speakers of Asturian Spanish make a phonetic distinction between the two back vowels.

In order to test this, I evaluated the degree of merger between the two vowels, using the Pillai-Barlett statistic (Pillai score) (Hay et al. 2006, Hall-Lew 2010) in R. The Pillai score is used to calculate the degree of merger between two vowel clusters. As Hall-Lew (2010:2) explains, the resulting measure is “an output of a Multivariate Analysis of Variance (MANOVA), which represents the proportion of one variance that can be predicted by another variance, given any known conditioning”. Pillai scores normally range from 0 to 1 and the higher the score the lower the degree of overlap between two vowel clusters in the vowel space.

There are some advantages to using Pillai scores over other methods (such as the Euclidean distance in F1/F2 space). First, Pillai scores are calculated without relying on averages, taking, instead, individual formant values as input, which allows us to account for the variability within a vowel cluster. Second, Pillai scores can account for phonological environment and address the variation without limiting the observations to a
single phonological context. This allows us to use spontaneous linguistic data, where there is an imbalance with regards to phonological environment.

Pillai scores were calculated for the /-o/, /-u/ variable for each speaker of the sample. The resulting Pillai scores range from 0.4475 to 0.8019. A first look at the results indicates that the differentiation between [o] and [u] is statistically significant for all speakers. This result, however, was expected because of the method used to cluster the tokens into two vowel categories (DAPC). Most interesting are the correlations observed between degree of merger (Pillai score), rates of /-u/ usage and the social variables explored in this chapter. Figure 28 compares the Pillai score to the probability of each speaker using /-u/. Each point in the plot represents a single speaker. Each speaker’s rate of /-u/ usage is represented on the x axis, such that the further right on the x axis the more probable the speaker is to use /-u/. Pillai scores are represented on the y axis such that the higher the position on the y axis the more distinct /-u/ is from /-o/.
Figure 28. Correlation between Pillai score and usage of /-u/. R=0.1442, p<0.01

Figure 28 shows that in general the more a speaker uses /-u/ the higher the Pillai score and the greater the distance between [o] and [u], while speakers that have lower rates of /-u/ usage are more likely to merge the two realizations.

Because there exists a significant correlation between the use of /-u/ and Pillai score, we expect for the latter to also be correlated with the same social variables that had a significant effect on the choice between /-o/ and /-u/. In order to determine whether these correlations exist, I developed a linear mixed model using Pillai score as the dependent variable, the social factors as independent variables and speaker as a random effect. The results of this analysis are summarized in Table 29:
Table 29. Effect of social predictors on Pillai score for the /-/o/, /-u/ variable

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.58</td>
<td>0.00</td>
<td>144.14</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>0.14</td>
<td>0.01</td>
<td>23.82</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.12</td>
<td>0.01</td>
<td>21.36</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.06</td>
<td>0.01</td>
<td>-12.64</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

The results indicate that speakers that work in commerce, restaurant or food industries have significantly higher Pillai scores than those who work in administrative, teaching or technical jobs. The same but weaker effect is observed for manufacturing jobs. In contrast to this, while male speakers favor the use of /-u/, they have a significantly lower Pillai score than female speakers, indicating that they are closer to merging the two back vowels. When Education is substituted for Occupation we obtain the following results:

Table 30. Effect of Education on Pillai score for the /-/o/, /-u/ variable

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.18</td>
<td>0.06</td>
<td>18.67</td>
<td>0.18</td>
</tr>
<tr>
<td>Education (reference level is High)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.23</td>
<td>0.09</td>
<td>2.67</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Middle</td>
<td>0.27</td>
<td>0.09</td>
<td>2.64</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
The results show that speakers with low and middle education levels have significantly higher Pillai scores than those who have a high education level. The low and middle education groups are not statistically significant.

Thus, we see that the same social factors that had a significant effect on the choice between /-o/ and /-u/ also have an effect on the Pillai score or how distinct the realizations of the back vowels are. The same direction of effect is also observed for Education and Occupation. We see that for some speakers the distinction between the two back vowels is clearly maintained. However, for speakers who do not use the Asturian variant as often, the two realizations have started to merge.

Even though the effect of the age of the speaker was not strong enough to significantly improve the regression model with the addition of the other (more important) social variables, the significant correlation between Age and Pillai scores gives us some insight into the direction of the change. In Figure 29 the age of each speaker is represented on the x axis, such that the further right on the x axis the older the speaker. Pillai scores are represented on the y axis such that the higher the position on the y axis the more distinct /-u/ is from /-o/.
As we can see, while older speakers are more likely to maintain the distinction between the two back vowels, the two linguistic variants are closer in the younger generations. I hypothesize that a change in progress is taking place and that, instead of completely dropping the use of one of the variants (in this case the Asturian /-u/) speakers that present lower rates of /-u/ are opting for a realization that is less distinct from /o/ and less marked as Asturian.

Qualitative evidence also indicates that there is still a phonetic contrast between the two back vowels in final position. For example, in written representations of the linguistic variety it is common to see speakers of Asturian Spanish use different orthographic representations for /o/ and /u/. The following examples are taken from online social media and reflect the spoken use of the two linguistic variants:

Figure 29. Correlation between Pillai score and age. R=0.2011, p<0.01
Utterances like the ones in (55) and (56) are frequent among speakers of Asturian Spanish in social media and show that “u” and “o” are both used in written manifestations of the language. When speakers of Asturian Spanish are asked about what characterizes their speech they mention the use of /-u/ as a salient feature:

(57) “Muy de Gijón y casi todo teminado en u” (PER12-625)
‘Very Gijón-like and almost everything ending in u’

(58) “Hablo castellano pero también uso palabras o expresiones típicas de Asturias y las terminaciones en u” (PER12-195)
‘I speak Castilian but I also use words or expressions that are typical of Asturias and endings in u’

A perception experiment that tests how well speakers of this variety discriminate between the two final vowels would be necessary to corroborate these results.

Similar results were obtained for the /-as/, /-es/ variable. Pillai scores for /a/ and /e/ ranged from 0.3598 to 0.9166. If we compare the scores for each vowel pair, we see
that the mean Pillai score for [a] and [e] is higher than that of [o] and [u] and that the
difference between them is statistically significant:

<table>
<thead>
<tr>
<th>[a], [e]</th>
<th>[o], [u]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7133</td>
<td>0.6317</td>
</tr>
</tbody>
</table>

Table 31. \(t=15.90, \text{df}=2127, p<0.01\)

Speakers seem to differentiate [a] and [e] better than [o] and [u]. In the case of the
front and central vowels there is also a correlation between usage of /-es/ and degree of
differentiation (Pillai score), but even though it is significant, the correlation is not as
strong as the one reported for the use of back vowels. In Figure 30 each speaker’s rate of
/-es/ usage is represented on the x axis, such that the further right on the x axis the more
probable the speaker is to use /-es/. Pillai scores are represented on the y axis such that
the higher the position on the y axis the more distinct /-a/ is from /-e/:
As we saw for the back vowels, the Pillai score is also correlated here to some of the social variables included in the analysis of the variation between /-as/ and /-es/. The results, summarized in Table 32, indicate that speakers that work in commerce and male speakers have significantly higher Pillai scores than those in the reference levels (Female and Administrative, Teaching and Technical, respectively):

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.64</td>
<td>0.01</td>
<td>69.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Occupation (reference level is Administrative, Technical and Teaching)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce, Restaurant and Food</td>
<td>0.06</td>
<td>0.01</td>
<td>4.77</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>-0.03</td>
<td>0.01</td>
<td>-1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>Gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.11</td>
<td>0.01</td>
<td>9.43</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 32. Effect of social predictors on Pillai scores for the /-as/, /-es/ variable
In contrast to what we saw for [o] and [u], the Pillai scores for [a] and [e] are not significantly correlated with Education, which was the main social effect in the /-as/-/-es/ variable. The correlation between Pillai scores and Age also lacks statistical significance. Thus, while the degrees of phonetic distinction between /o/ and /u/ are socially constrained, that is the case only to a certain extent for /a/ and /e/. I attribute this differentiating effect to two reasons: The greater proximity between the back vowels in Asturian and Asturian Spanish and the different social status of the two linguistic variables, which will be addressed in section 4.4.2.1.

4.4.2 Social variables

In this section I discuss the effect that the social factors included in the quantitative analysis had on the two linguistic variables. As we saw in Table 24, the effect of social predictors is most noticeable on the categorical choice between the two possible vowels and the frequencies of the F1. The only social factor that had a significant effect on the frequency of the F2 in both variables was gender. We also saw that Social tie was only significant in the categorical choice between /-o/ and /-u/ and that the relative importance of Occupation and Education was different in each linguistic variable. While Occupation was the most important social factor in all the dependent variables that measured the use of /-u/, Education had the highest relative importance in the use of /-es/.

Before discussing the individual effects of the social factors, I explore the possible interactions between them. In order to do so, I developed Conditional Inference
Trees, using the `cforest` function in the `party` package in R (Hothorn et al., 2008).

Conditional inference trees evaluate whether an independent variable is a useful predictor of the two possible responses of the dependent variable and divides each of the predictors into subsets that are further evaluated, considering the effect of subsequent factors.

Inference trees allow us to see how different independent variables interact with each other. Figure 31 contains a conditional inference tree that represents the effects and interactions that were found to significantly predict the use of /-/u/:

![Conditional inference tree](image)

Figure 31. Conditional inference tree of social factors that significantly predict the use of /-/u/
The inference tree shows that the most important social factor in the variation is Occupation and that manufacturing jobs and jobs in commerce (including jobs in the restaurant and food industries) pattern alike. Gender is a significant predictor but only for speakers that work in manufacturing and commerce; i.e. there is a significant interaction between Gender and Occupation and gender has a differential effect based on the occupation of the speaker, such that it is only a significant predictor of /-u/ when the speakers work in manufacturing and commerce. In general, male speakers use /-u/ more often. If we continue to work our way down the inference tree we observe that Social Tie has a differential effect with regards to Gender, showing that only male speakers behave differently based on the type of social tie they have with their interlocutor. This could be attributed to the fact that male speakers in general use more Asturian features than female speakers. If we look at the inference tree in Figure 31, we see that the rates of /-u/ usage in women in any situation are still lower than those of men in contexts in which they have a second order tie with their interlocutor. Thus, I propose that because male speakers use /-u/ more, they feel the necessity of reducing the use of the Asturian variant in what is perceived as a formal situation. Finally, Education has the lowest relative importance and it only has an effect on the group of male speakers that work in manufacturing/commerce and that have first order ties with their interlocutor. In this group, speakers with middle education use /-u/ more frequently than those with a low education level. However, further exploration of the data revealed that there was only one speaker that had low education, with whom I had a first order tie, that was also male and worked in manufacturing or commerce. Because the effect of low education is then attributed to a single speaker, we cannot make any generalizations regarding the effect that this social
factor has on the variation. We can also conclude that the differential effect of the middle education group is simply a reflection of the effect of social tie.

Figure 32 represents the conditional inference tree for the /-as/, /-es/ variable:

![Conditional inference tree](image)

Figure 32. Conditional inference tree of social factors that significantly predict the use of /-es/

If we compare the results in Figure 32 to those in Figure 31, we observe that, in the case of final /-as/ and /-es/ the most relevant factor is Education (recall that only Education and Occupation were significant in the distribution of this linguistic variable). In contrast to what was reported for /-o/ and /-u/, here middle and high education pattern
together. Occupation only has an effect on these two educational groups, such that people that work in commerce have higher rates of /-es/.

4.4.2.1 Occupation and Education

Even though there are differences in the effect that Occupation and Education have on the two linguistic variables, there are some general trends observed in the data. In both cases the occupational group that uses the Asturian variants the most is the one whose members work in commerce and in the restaurant and food industries. This group includes mainly servers that work in bars and speakers that work in small specialty markets or stores. While this group and the manufacturing group show similar effects on the variation (with the effect of commerce being stronger), the reasons why they both favor the Asturian variants may differ.

The difference between manufacturing and commerce jobs is that in the latter, speakers are in frequent interaction with the public and the success of their businesses depends highly on relationships with customers. All the participants that were grouped into this category worked in small stores or bars that rely heavily on a regular clientele with whom they try to establish a familiar relationship. The objective is to create a comfortable atmosphere that allows the workers in these businesses to earn the customers’ trust and loyalty. There are many ways in which this relationship is built, from remembering the customers’ preferences in their orders to exchanging personal information on a regular basis. Another method involves the use a particular linguistic variety. Research on Asturian sociolinguistics is unanimous regarding the stylistic use of
Spanish. Sánchez Álvarez (1979), for example, observes that in Asturian Spanish, Asturian features, whether they are lexical, phonological or morphosyntactic, are not used randomly. Instead, they are granted an “affective” and “expressive” value on the part of the speakers that make them irreplaceable by their Spanish equivalents, which she qualifies as having a neutral affective value. According to Sánchez Álvarez this is mostly observable in colloquial styles and in familiar situations. Prieto (1991) makes similar observations saying that the presence of Asturian features normally increases in relaxed situations. She also reports the case of a few immigrant vendors in the market she studied whose native language is Galician. These speakers use Asturian Spanish (instead of Spanish) as their variety of choice in the market setting. According to Prieto, this behavior responds to a desire to match their customers’ speech in an attempt to attract them to their products. D’Andrés (1997) makes similar observations, claiming that the use of Asturian features is reserved for informal situations and addressees that are socially equal. In the analysis of the data for this study I found evidence that speakers use the alternation between Asturian and Spanish features as part of style shifting strategies. For example, one of the speakers, a 49 year-old female medical assistant, rarely included Asturian features in her speech during the sociolinguistic interview. At a certain point in the interview her cell phone rang, and she instructed her husband, who was also present, to answer it for her, using the utterance in (59):

(59) “Oscar, ¿cógesmelu?” (GI11-26)
‘Oscar, can you answer it for me?’
In the example above the speaker uses two linguistic features from Asturian, the enclisis of unstressed pronouns and the use of /-u/ in the clitic. While the use of /-u/ was recorded sporadically for this speaker, this is the only instance in which she uses the pronoun enclisis characteristic of Asturian. After finishing interviewing a 25 year-old male that works in transportation, and after I had turned off the recorder, I explained what the goals of my study were and I mentioned how few Asturian features he used in his speech. He reacted to my explanation saying that had he known what I was looking for he could have used much more Asturian, as he does when he is interacting with his coworkers. When I asked him when he normally uses Asturian features, he claimed to use them more at work in order to fit in better with the rest of the employees in the company, since they all use Asturian features regularly. He also added that occasionally he uses them with friends, but to a lesser degree.

I propose that speakers that belong to the commerce occupational group increase their use of Asturian features as a mechanism to create a particular relationship with their customers and a social setting that resembles that of the family and close friendship. Even though the sociolinguistic interview could be interpreted by some participants as a formal situation, 4 out of the 6 interviews with speakers that belong to this occupational group took place at their workplace, making it possible for them to reinterpret the situation as an interaction with an actual or potential customer.

In her discussion of intra-speaker variation, Schilling-Estes (2002:378) points out how speakers “use their speech to help shape and re-shape the external situation (whether the immediate interactional context or wider societal forces), as well as their interpersonal relationships and, crucially, their personal identities”. As an example,
Johnstone (1999) mentions how Southern women from Texas use stylistic shift to a Southern sounding variety in a strategic and planned manner. One of the speakers she analyzes, a salesperson, claims that switching to this variety, especially with male customers, increases her profits greatly. It is possible that a similar behavior is observed in the Asturian Spanish of speakers that belong to the commerce occupational group. The fact that Occupation interacts with Education in the /-as/, /-es/ variable indicates that, while the correlation between both social factors is still important, the effect of Education cannot fully account for the observed linguistic behavior in different occupational groups.

In contrast with the commerce group, the use of Asturian variants by speakers that work in manufacturing might be more directly associated with both a lower educational level and the masculine dominance in this field (only 2 out of 10 speakers that were classified in this occupational group were female). This majority of male speakers could explain why in the linguistic variable in which gender does not have a significant effect (/-as/, /-es/) the linguistic behavior of speakers in manufacturing is similar to that of speakers in the administrative, teaching and technical group.

At this point, one wonders why the effect of Occupation is stronger in the /-o/, /-u/ variable than in the variation between /-as/ and /-es/. In the discussion of the different degrees of merger between the two pairs of vowels I suggested that the two linguistic variables have a different social status. Also, in the description of the hierarchical scale of differentiating features associated with Asturian Spanish, I noted that /-u/ is one of the few features that is used by all speakers of the variety. Furthermore, the quotes included in (57) and (58) show that /-u/ is a salient feature of Asturian, recognized by speakers as such, while /-es/ is never cited as an example of Asturian features. Trudgill (1986)
explains that such highly noticeable features can be subject to hyperdialectalism; i.e. their use can be overgeneralized into environments where variation is not expected. In the Asturian Spanish data, this type of overgeneralization is only observed for the use of /-u/ and the only speaker that showed hyperdialectalism belongs to the commerce occupational category. An example of this phenomenon is illustrated in (60), where the speaker uses /-u/ in a preposition that has /-o/ in standard Asturian:

(60) “bajo a un supermercado debaju casa” (GI11-22)
‘I go down to a market that’s nearby my house’

I propose that the different effect of Education and Occupation observed in the two linguistic variables stems from the importance that /-u/ has as a marker of Asturianness. This hypothesis will be tested in Chapter 5, in which I analyze the social perceptions of each linguistic variable.

The effect of Education was the expected one, based on the previous literature on the social status of the language. Asturian was not present in the education curriculum until the 1980s and still today is not part of a bilingual curriculum but is treated as an optional additional course. Nowadays, even with the inclusion of Asturian as a subject in some schools, the view of the regional language as an imperfect variety of Spanish still persists. Thus, Spanish is reinforced as the prestige language in society. The longer the time students spend in the educational system, the more they are exposed to this standard variety of the language and the more likely they are to internalize the stigmatization of Asturian. As a consequence, Spanish features are more readily available in the linguistic
system of speakers that have higher education levels. Even if their exposure to Asturian is the same as that of speakers with low education, their higher competence in Spanish allows them to have better control over the amount of features from the national and regional languages that are present in their speech according to the communicative situation. In contrast to this, I propose that speakers with a low education level have less control over the Spanish variants and tend to show less intra-speaker variation. To test these hypotheses, however, it would be necessary to conduct a study that analyzes the speech of participants in different communicative situations, manipulating the social variables that were included in this study.

4.4.2.2 Gender

Previous research has obtained contradicting results with regard to gender and the use of Asturian. However, most researchers agree that male speakers normally claim to use more Asturian and have more positive attitudes towards it. Because Asturian is commonly associated with the rural world and this world is overwhelmingly masculine, in that men normally do the work outside while women stay in the house, it is expected that men use the language more than women, since they have more frequent interactions with other speakers. While the association of Asturian with rural areas and masculinity is tested in the perception experiment reported in Chapter 5, the results of the quantitative analysis seem to support this hypothesis. In the analysis of the variation between /-o/ and /-u/ I reported that the effect of gender is statistically significant in the categorical choice of /-u/ and in the frequency of the first two formants. Male speakers show significantly
higher rates of /-u/ and lower frequencies in the F1 and the F2. The same is observed in the use of /-as/ and /-es/, except for the observations on the categorical choice between the two variants, where the effect of gender is not statistically significant. Because lower F1 and F2 values correspond to realizations of [u], it is difficult to determine whether the effect observed is due to the role of gender as a social factor or to the physiological differences between men and women that could still persist (to a lower degree) after normalizing the formant values. The hypothesized difference in the social nature of the two linguistic variables could also explain why an effect of gender is only seen in the variable that is more saliently Asturian. The results of the perception experiment will help clarify these questions.

4.4.2.3 Social tie

An effect of Social tie was only observed in the /-o/, /-u/ variable. As we saw in the inference tree, this effect is restricted to male speakers in manufacturing and commerce occupations; i.e. the group of speakers that has the highest rates of use of the Asturian variable. Within this group of speakers, those with whom I had a first order social tie have significantly higher probabilities of using /-u/ than those with whom I had a second order tie. In his study of stylistic variation, Bell (1984) developed a framework that revolves around the role of the addressee which he coined “audience design”. According to the audience design model, speakers respond to their interlocutors in order to design their speech. Even though later research has criticized the reactive nature of Bell’s model, the role of the addressee is still considered a relevant factor in style shifts.
According to Coupland (2007), both self-identity and audience design are present in style variation. Previous research on Asturian has found that speakers tend to use more Asturian in interactions with people that they know (Llera Ramo 1994, D’Andrés 1997). Thus we expect that in speakers that have high rates of /-u/, the formality of the interview context was not sufficient for them to avoid the use of Asturian features, since they had a direct relationship with their interlocutor. It should be noted that, because I am a native speaker of the Asturian Spanish variety, I also used Asturian features in my interactions with participants. However, my use of Asturian features alone was not sufficient for speakers to overcome the formality of the communicative situation in the cases in which there was a second order tie.

The fact that we only find an effect of social tie in the use of the back vowels and that this effect is restricted to a particular social group makes us wonder how relevant familiarity with the interlocutor is and whether the effect could be explained by the distribution of the data. However, no collinear effects were found for social tie and the restriction of the effect to the variation between /-o/ and /-u/ could be explained by the different social nature of this variable, which has been addressed above. The results of the perception experiment will also contribute to finding answers to these questions.

4.4.2.4 Non-significant factors

The only social variable that was not found to significantly predict the outcome of the variation in either of the linguistic variables was Age. This is a surprising result since research on Asturian sociolinguistics and the use of Asturian features in hybrid varieties
of the language is unanimous regarding the effect of age: older generations maintain the
use of Asturian and Asturian features while younger speakers are starting to assimilate
completely to Spanish. Furthermore, this effect of age is recognized by the speakers
themselves. In (61), I included a quote from one of the speakers that participated in the
study.

(61) "eso [Asturian features] ahora puedes oirlo por les sideries, con los
paisanos de 60 años, que ye prestosu de la de dios, pero a un guaje de menos de
25 años, imposible. De 25 a 35 alguno hay, pero de 25 pa' bajo ni dios" (GI11-25)
‘you can hear it [Asturian] in the bars, from men that are 60 years old, and I really
enjoy that, but from a kid that is younger than 25, it is impossible. From 25 to 35
years old, there are some, but younger than 25 there is no one’

When we test the correlation between the rates of use of the two Asturian variants
and age as a continuous measure, we obtain significant but very weak correlations. Figure
33 and Figure 34 show this correlation. In these figures each speaker’s rate of /-u/ or /-es/
usage respectively is represented on the y axis, such that the higher on the y axis the more
probable the speaker is to use /-u/ or /-es/. Age is represented on the x axis such that the
higher further right on the x axis the older the speaker is:
Figure 33. Correlation between Age and usage of /-u/. $R = 0.04087$, $p < 0.01$

Figure 34. Correlation between Age and usage of /-es/. $R = 0.04076$, $p < 0.01$
The information in the plots and the proximity of the R-squared values to 0 indicate that, while there is a significant trend that points to an increased use of Asturian features with age, this correlation is very weak and it is understandable that it would be shadowed by other more important social variables.

In the description of the hierarchical scale of features in Section 2.3.2 I pointed out that there is a decrease in the number of Asturian features in the speech of younger speakers. It is possible that age constrains how many and what features are included in speech but that, once a feature is present, age has a much weaker effect on how much that particular feature is used. Instead, the rates of use of a particular feature depend on the other social factors that I have examined in this study.

4.5 Conclusions

In this chapter I have reported and discussed the results of the quantitative analysis of production data. The results obtained show that the choice between Spanish and Asturian linguistic variants is linguistically and socially constrained. These effects are observed in the categorical choice between the linguistic variants and, in the majority of the cases, also in the frequency of the formants. For both variables, the most important linguistic predictor is the segment that precedes the target vowels, followed by vowel duration and prior form. The effect of the preceding segment follows expected tendencies of coarticulation that have been outlined in previous research in Spanish and other languages, such that the place of articulation of each vowel approximates that of the previous segment. In some cases the influence of the previous segment results in different
realizations within the same vowel category but in other cases, when the effect of preceding segment is strong enough, the point of articulation of the previous sound constraints the categorical choice between the two possible vowels. Final /-u/ is less likely to be used when a palatal sound, a dental or alveolar or /a/ precedes the final vowel and the odds of /-es/ increase in a palatal context. Anticipatory coarticulation is also registered for the back vowels but its effect is weaker than that of the preceding segment. Vowel duration also constraints the choice between the Asturian so that longer duration corresponds with lower or more backed vowels, respectively. Finally, a tendency to linguistic perseveration, such that speakers tend to repeat the variant used, was also observed.

The results also show that the Asturian variant /-es/ is significantly less likely to be used in determiners than in other open-class lexical categories. I attribute this behavior to their status as function words. The reason why we observe this effect in the /-as/, /-es/ variable and not in the /-o/, /-u/ variable is that neither of back vowels appear in determiners. The rest of the linguistic factors explored in the statistical analysis were either found to be non-significant predictors or their effects could not be generalized due to the distribution of the data and the correlation with other linguistic factors. Further analysis is necessary to confirm, for example, the effects of the quality of the stressed vowel.

The observed role of linguistic predictors shows that the variation registered in Asturian Spanish follows patterns observed in other linguistic phenomena and contributes to understanding general mechanisms that guide language variation and change. It provides further evidence of the effect that neighboring sounds and vowel duration have
on vocalic segments and it also contributes to theories on how language processing occurs.

As hypothesized, the variation between Asturian and Spanish features is also socially constrained. The analysis of the effect that social factors have on the choice between linguistic variants showed that there are differences in the social behavior of the two linguistic variables. While the choice between /-o/ and /-u/ is constrained by occupation, gender, education and social tie, with occupation being the most important factor, only education and occupation have a significant effect on the variation between /-as/ and /-es/, with education being the most important predictor in this case. These differences stem from the fact that the two linguistic variables appear to do different “social work”, such that /-u/ is a stronger marker of “Asturianness” than /-es/. Qualitative and quantitative data support this hypothesis, which is tested in chapter 5. The effect that occupation has on the variation between back vowels points to a strategic use of the Asturian variants on the part of the speaker and provides further proof that /-es/ and /-u/ are socially different.

Another important finding in the analysis of production data is the fact that speaker do not only vary in their choice of vowels but also in the distinction maintained between them. I found that speakers that show higher rates of /-u/ and /-es/ also tend to keep them from merging with the Spanish variants (with the correlation being stronger in the case of back vowels). I argue that speakers that only use /-u/ occasionally are opting for a realization of this vowel that is less distinct from /-o/ and, thus, less marked as Asturian. Even though there are some trends with regards to the effect of age on the variation, these observations are not sufficient to prove that a change in progress is taking
place. Further research and a larger sample of speakers would be necessary to be able
to generalize any effect of age. Based on the data analyzed in this part of the dissertation, I
propose that age has a stronger effect on how many and what features from Asturian
speakers incorporate in their speech and a much weaker one on how often one particular
feature is used.

While there is evidence that points to the strategic use of Asturian features and to
the speakers’ awareness of the presence and purpose of these features in their speech, the
results of quantitative analysis only allow us to establish the existence of certain
correlations between linguistic use and socio-demographic categories. The perception
experiment reported in the next chapter not only contributes to better understanding these
correlations but it also determines whether these correlations are meaningful and whether
the choice of one variant over the other is used by listeners to evaluate the identity of
speakers.
CHAPTER 5: SOCIAL PERCEPTIONS OF THE LINGUISTIC VARIABLES

5.1 Sociolinguistic perception

Even though the quantitative description of the variation is a necessary first step in the sociolinguistic description of the urban variety of Asturian Spanish, using this type of analysis we can only discover whether there is a correlation between the social categories explored in chapter 4 and the occurrence of a particular linguistic variant. In the analysis of the Asturian Spanish production data I discovered strong correlations between the use of Asturian /-u/ and /-es/ and social categories such as occupation, education, gender and social tie. Adopting the “orders of indexicality” proposed by Silverstein (2003), Johnstone et al. (2006) and Johnstone and Kiesling (2008) refer to such correlations between social categories and particular linguistic variants as first order indexicality, which corresponds to what Labov (1972) coined as “indicators”. This kind of correlation can be recognized by out-group speakers; however, speakers that belong to that particular social category are unaware of the correlation. First order indexicals (or indicators) can become second order indexicals (or markers in Labov’s terms). As Johnstone et al. (2006:83) point out, “Second order indexicality occurs when people begin to use first-order correlations to do social work, either interpretive or performative”. When this shift takes place, speakers that have been exposed to different
variants of the same linguistic variable are aware of the connection between a particular variant and a specific social category. Furthermore, when speakers are aware of these correlations, they tend to show intra-speaker variation, choosing a variant over the other when they are trying to associate themselves with a particular social category.

One of the objectives of this dissertation is to determine whether speakers use the variation between Asturian and Spanish features to “do social work”, i.e. whether the Asturian variants of the morphemes under study have second order indexicality. Because the existence of a correlation does not necessarily imply that the linguistic feature is a marker of these social identities (Johnstone and Kiesling, 2008), we need further evidence to prove that the correlations observed are meaningful. One way of obtaining such evidence is by analyzing the social perceptions that the listeners have of the variants under study. Furthermore, exploring the alignment between what was observed in the analysis of production data and the social perceptions of the linguistic variables will contribute to understanding the nature of the correlations described in Chapter 4.

5.1.1 Social meaning

Recent research on sociolinguistic variation has been increasingly concerned with how linguistic behavior relates to social meaning. As Campbell-Kibler (2009:136) explains, one of the main assumptions behind the concept of social meaning is that “linguistic variation not only reflects social differences, but it is also used by speakers to position themselves within the social world, and through such position, to build and rebuild that world”.

189
Some researchers have focused on speakers’ performances to explore how linguistic variants correlate to locally defined social categories (Eckert 2000). Other scholars have looked at how social perceptions of a speaker change through the manipulation of individual linguistic variables. For example, Fridland, Bartlett and Kreuz (2004) examined the relationship between different vocalic realizations and perceived Southern accent in Memphis, TN. They found that shifted mid-front vowels were salient markers of local and Southern identities, while back vowel shifts did not have strong salience as regional markers. The authors conclude that there is a match between production and perception, such that the vowel shifts that are present in the Southern Memphis linguistic system are more easily identified as Southern than other shifts. Plichta and Preston (2005) investigated the speakers’ perception of the monophthongization of (ay) by synthetically manipulating the realization of the sequence. They found that listeners identified higher degrees of monophthongization as more Southern speech, by choosing from nine possible cities located in a continuum from North to South. They also discovered that male and female voices were rated differently, with female speakers being identified as further North than male speakers. These studies show that listeners not only rely on linguistic cues to relate the speaker to certain social categories; it also shows that they use phonetic detail, such as the difference in formant frequencies or a gradient degree of monophthongization, in evaluating speakers’ identities.

Campbell-Kibler (2008) emphasizes the fact that the social meaning attached to particular linguistic variants can vary from one individual to another. As Fridland et al. pointed out, the exposure to and the use of a particular linguistic variable can determine
the social meaning that listeners associate with linguistic variants. Other factors that have been found to affect the listeners’ social evaluations of linguistic variation are the socio-demographic traits of the listener, such as sex, age and regional background (Paltridge and Giles 1984) or even differences in interpersonal factors, such as mood, personality or situational goals (Campbell-Kibler 2008). This research indicates that there is a necessary link between production and perception and it is the joint consideration of both types of data that will shed light on the social meaning of the Asturian Spanish linguistic variables that I am studying.

5.1.2 Social perceptions of Asturian

The majority of studies that address the social perceptions of Asturian or Asturian Spanish in any way use sociolinguistic surveys as a means to collect their data. Some examples of such studies were reviewed in section 2.2.3. The questionnaires that are used are normally intended to elicit information about the social distribution of the linguistic varieties, the domains in which each language is used and the attitudes towards them. To my knowledge, the only study that evaluates listeners’ perceptions of Spanish, Asturian and Asturian Spanish is that of Rose Alcorta (2000). The author investigates the use of the linguistic varieties and the perception of each in the coastal town of Candás combining sociolinguistic interviews, analysis of spontaneous speech and a perception experiment. In her perception experiment with 40 participants that lived in the area, she used stimuli in which the same speaker read a paragraph in 4 different varieties: Asturian, Spanish, Asturian Spanish and Catalan. The author explains that Catalan was included in
the study because it is a minority language that has more prestige than others spoken in the Peninsula (such as Galician) and that can be understood by Asturian speakers better than other regional languages like Basque. Rose Alcorta did not explore the social meaning of a single linguistic variable but the use of the language variety in general.

She found that higher ratings are given to Spanish in status categories, such as intelligence, efficiency, etc. However, she found a differential effect of gender, such that women give higher ratings to Asturian and Catalan in intelligence. In contrast to this result, Asturian and Asturian Spanish are rated higher in solidarity dimensions such as trustworthiness, kindness, friendliness, etc. Here there is a differential effect of age, with younger speakers giving higher ratings to the two local varieties than older speakers, who also gave the highest ratings to the mixture but who rated Spanish higher than Asturian. Thus, there exists some evidence that points to different evaluations on the part of the listeners of the language varieties under study. Rose Alcorta’s study also shows that the factors associated with the listener can also influence their social evaluations of the speakers.

5.2 Methodology

In order to evaluate the social meaning of Spanish and Asturian in the community under study, I created a perception experiment designed to uncover what types of social characteristics speakers associate, consciously or unconsciously, with specific linguistic variants. In this section I explain the methodology used in the design and implementation of this experiment.
5.2.1 Matched Guise Technique

The perception experiment was designed using a modified form of the Matched Guise Technique. This procedure was first introduced by Lambert, Hodgson, Gardner and Fillenbaum (1960) to explore social perceptions of French and English-Canadian speakers among the two groups. In their experiment, Lambert et al. recorded the same bilingual speaker reading two passages in French and English. Participants were then asked to judge the speakers based on the utterances they heard and according to a series of social characteristics and personal traits. This technique has continued to be in use in later studies that explore listeners’ attitudes to a linguistic variety or to a specific linguistic feature. Thus, in some cases they compare reactions to different languages, like Rose Alcorta (2000) did for Asturian, Spanish, Catalan and Asturian Spanish, but in other cases they investigate the judgments of specific linguistic variables (Campbell-Kibler 2008, 2009). Most studies that have employed this technique have found that the same speaker is rated differently on social or personal characteristics depending on which language variety or linguistic variant is used.

5.2.2 Stimuli

There are different ways in which stimuli can be created for this type of experiment. Studies that compare language varieties, without focusing on any individual linguistic variable, often elicit data by instructing bilingual speakers to read a paragraph in different languages. Studies that focus on eliciting judgments on a particular linguistic variable have also used the same method, asking the speaker to shift between variants, or
have employed digitally manipulated speech to control what variant is used in each version of the utterance. Digital manipulation ranges from alterations of formant values in vowels to copying and pasting particular segments. One of the advantages of digital manipulation is that it allows the researcher to work with spontaneously produced data instead of read utterances and that it controls speech so that the speaker does not change any other linguistic features in their speech production.

In the perception experiment reported here, the use of spontaneous speech was problematic because spontaneous utterances often contained multiple sites in which variation between Asturian and Spanish features was possible. Because in my analysis I have found that not all linguistic features from Asturian are equally frequent in speech or equally salient as being Asturian, in order to accurately evaluate the social perceptions of the variables under study, it was necessary to ensure that no other features were or could be present in the utterances. Thus, I opted for eliciting read speech from native speakers of Asturian Spanish.

The utterances used in the stimuli were short sentences that lasted between 3 and 5 seconds. The sentences were adapted from utterances that were recorded in the sociolinguistic interview, with the goal of maintaining linguistic structures that occurred in spontaneous speech. However, they were later modified to eliminate any context in which the use of an Asturian feature other than /-u/ and /-es/ was possible and then read by native speakers of Asturian Spanish. Each utterance contained at least 2 occurrences of one of the variables, in order to ensure that the listener had more than one chance to notice them. The variables appeared in a variety of nouns, determiners and adjectives, sometimes in the same NP and in some occasions in different NPs. This distribution was
selected to ensure that the effects observed were not constrained to a particular lexical
item, lexical category or construction.

A total of 8 pairs of sentences, which are included in Appendix A, were used to
elicit data from 8 speakers, so that each matched pair of sentences was read by a different
person. All the speakers were native speakers of Asturian Spanish and their ages ranged
from 28 to 57 years old. Two male and two female speakers were selected for each
variable. Two of them were under 40 years old and the other two, older than 40.
Coaching was provided to ensure that speakers read the sentences as naturally as possible
and that they only altered the features under study. Each speaker was also given some
time to practice before recording the utterance. All the recordings were made in a quiet
place onto a Sony IC digital recorder (SX712) with a sampling rate of 44.1kHz. Each
speaker recorded the pair of utterances multiple times and the most natural sounding ones
were selected to include as stimuli in the perception experiment. Each token was
analyzed acoustically to ensure that the speakers’ articulation of the different vowels in
the matched items corresponded to a difference in the formant frequencies of those
vowels.

5.2.3 Questionnaire

Once the stimuli were created, a survey was designed and implemented on the
web survey platform SurveyGizmo (Vanek and McDaniel 2006). On each page of the
survey, the participants were introduced to a speaker and were told to listen to a digital
recording uttered by that speaker. After listening to each utterance, participants had to
respond to a series of questions about the speaker they had just listened to. The complete questionnaire, which was adapted from Campbell-Kibler (2008), is included in Appendix B. First, they had to rate the speaker according to certain social, personal and situational characteristics, such as intelligence, masculinity/femininity, formality, urban/rural origin, economic status and familiarity with their interlocutor. The choice of characteristics was based on the findings of previous research and on the results obtained in the analysis of production data. For each of them, participants had to provide a rating on a Likert scale from 1 to 6. The scale was divided into an even number of points (6) to avoid neutral responses. Participants also had to select an answer for 4 multiple-choice questions regarding the perceived education of the speaker, their age, occupation and city or place of origin. The number of answers that they could choose from ranged from 5 to 7 and the selection of only one checkbox was allowed. At the end of each page, participants were also asked to provide any supplemental information about the speaker in a box for open comments. At the end of the survey participants were also asked to provide information regarding their gender, education, occupation, age, whether they considered themselves speakers of Asturian and whether they believed that Asturian should be an official language in the region. Finally, they were asked to describe their own speech in an optional section for open comments.

Even though the 16 utterances, including both versions of each matched pair, were included in the questionnaire, the pages of the survey were branched so that each participant was randomly assigned to only one utterance for each of the 8 speakers and each matched pair. Thus, participants listened to a total of 8 recordings, which included 2 utterances that contained each linguistic variant (/-o/, /-u/, /-as/ and /-es/). The order in
which the pages of the survey were presented was also randomized to avoid effects of a particular order on the presentation of the stimuli. Fillers were not used in the survey, thereby allowing participants to complete it in less than 15 minutes.

5.2.4 Implementation and participants

The survey was implemented online. Participants were provided with a link that led them directly to the survey. This link was promoted on social media (Facebook) in order to recruit male and female participants of a variety of ages, education levels and occupations that were from Gijón or had lived in the city for the majority of their lives. At the beginning of the survey, participants were asked whether or not they were from Gijón and, if the answer was negative, whether they lived in Gijón at that moment and for how long they had been living there. Participants who were not from Gijón and who had not lived in the city for more than 30 years were not allowed to participate in the study. 30 years were chosen as the minimum time spent in Gijón to ensure that participants had been in contact with the variety under study for more or less half of their life in the case of the oldest participants.

A total of 192 subjects participated in the study. Their ages ranged from 20 to 62 years old. Table 33 shows the distribution of participants by gender and education level:
<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary education</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Secondary education</td>
<td>12</td>
<td>22</td>
<td>34</td>
</tr>
<tr>
<td>Post-secondary education</td>
<td>16</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Professional school</td>
<td>11</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>College/Graduate degree</td>
<td>31</td>
<td>56</td>
<td>87</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>73</td>
<td>119</td>
<td>192</td>
</tr>
</tbody>
</table>

Table 33. Distribution of participants by gender and education level

As we can see, with the exception of primary and college education, the participants were well distributed across education levels and gender. Of the 192 participants, 100 of them considered themselves to be speakers of Asturian and 120 believed that Asturian should be an official language in the region. Table 34 shows the distribution of speakers regarding the answers to these two questions:

<table>
<thead>
<tr>
<th>Should Asturian be an official language?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>75</td>
<td>45</td>
</tr>
<tr>
<td>No</td>
<td>25</td>
<td>47</td>
</tr>
</tbody>
</table>

Table 34. Distribution of participants according to whether they speak Asturian and whether they believe that Asturian should be an official language in the region

The distribution indicates that while participants that do not consider themselves to be speakers of Asturian are equally distributed with regards to the official status of the language, 75% of those that characterize themselves as speakers of Asturian believe that the language should be official, indicating that both characteristics could be correlated.

---

8 Note that speakers of Asturian Spanish might consider themselves to be speakers of Asturian if they include all or some of the differentiating features.
5.2.5 Statistical analysis

In the first part of the statistical analysis I explored whether the speakers were rated differently when they used the Spanish or the Asturian variants. Thus, I developed separate linear mixed models to test what effect the language of the feature used in the utterances (i.e. whether the speaker used the Asturian variants /-u/ and /-es/ vs. the Spanish variants /-o/ and /-as/) and the social characteristics of participants and speakers had on the rating of each personal, social and situational trait, using each rating as a separate continuous dependent variable. The six dependent variables were: intelligence, masculinity/femininity, formality, rural/urban, economic status, and familiarity. I also included participant as a random variable, in order to account for the influence that individual participants have on the data. I considered a total of 7 independent variables:

1. Feature language, treated as a binary division between Asturian or Spanish
2. Participant age, treated as a continuous variable
3. Participant gender, treated as a binary distinction between male and female.
4. Participant education. Even though participants could choose from seven educational levels, these levels were further collapsed to match the educational categories used in the analysis of production data, resulting in three levels: low, middle and high education.
5. A binary choice between whether the participants considered themselves as speakers of Asturian or not.
6. A binary choice between whether the participants believed that Asturian should be an official language or not.
7. Speaker gender, treated as a binary distinction between male and female.
I tested the effect of these predictors on the 6 dependent variables considering the two linguistic variables (/-o/, /-u/ and /-as/, /-es/) separately, in order to easily register any differential results according to linguistic variable.

The order in which the independent variables were added into the model was determined using the step function in R. This method was used instead of the random forests because of the experimental nature of the data and the objectives of the analysis. When we explore linguistic data from a variationist perspective, as I did in Chapter 4, the objective is to uncover how social and linguistic factors work together in constraining the variation and to discover what are the best predictors of the variation and how they interact. Furthermore, and as Tagliamonte (2011:150) explains, “comparison across constraint hierarchies and the relative strength of factors is critical to address questions relating to language change, grammaticalization, acquisition, contact, ancestry, etc.” Random forests are a great tool to elucidate these relationships and to construct a hierarchy of constraints.

In contrast to this, in experimental research, such as that reported in this chapter, the objective is to find the best model that can account for the variation observed using the least possible amount of predictors (principle of parsimony). Thus, constraint hierarchies taken from random samples of the data are not as useful in this case. Instead, the step function provides us with an objective and rigorous method of determining the most parsimonious model (the simplest model that can account for most of the variation). In order to do this, the step function uses the Akaike Information Criterion, which evaluates the value of each independent variable using a log-likelihood measure of how well each model fits the data. This method considers nested models adding one variable
at a time and “penalizing” the addition of each new predictor. In order for predictors to overcome this “penalty” they have to greatly improve the predictions made by the model; i.e. small improvements are not worth the complexity of adding a new variable and are rejected (Johnson 2008). As Johnson (2008:94) explains, “The idea is that we are seeking a stable model that produces a good fit to the data, and the AIC approach scores stability in terms of the information gain, per coefficient added to the model. This tends to keep out weak, unstable coefficients so that we avoid modeling the randomness of the data set and instead only capture robust relationships”. Once the nested linear mixed models were developed, they were compared using ANOVA. Conditional inference trees were also developed to examine interactions between independent variables.

In the second part of the statistical analysis I tested whether a particular choice among the ones presented in the checkbox variables was more likely to be chosen when the feature language was Asturian (i.e. when the speaker used the Asturian variants /-u/ and /-es/ vs. the Spanish variants /-o/ and /-as/), using each of the checkbox variables as a separate dependent variable and feature language as the independent variable in each analysis. Again, the two linguistic variables were tested independently to observe differential effects with regards to the variable used. For this part of the analysis I employed multinomial logistic regression, using the mlogit function from the mlogit package in R (Croissant 2012). Multinomial logistic regression works similarly to the binomial logistic regression used in Chapter 4. The difference between them is that while in binomial logistic regression the dependent variable is a categorical choice between two levels, in multinomial logistic regression the dependent variable can have more than two levels. There are two disadvantages to using this method. First, there is no
straightforward and accurate way to evaluate the resulting regression models. Thus, I limited the analysis to testing the effect that the language of the feature had on the categorical classification, making feature language the only independent variable.

Second, it is not possible to include random variables so the individual effect of participant could not be included.

I considered a total of four dependent variables:

1. Education level of the speaker. Participants were initially given seven choices to rate each speaker according to their perceived education level: no education, primary education, secondary education, university preparatory course (Bachillerato or COU), professional school (Modulo or FP), three-year college degree (Diplomatura) and five-year college degree (Licenciatura). Because of the distribution of the responses and the similarity between some of these categories, they were later combined into four groups:
   - Elementary, including no education and primary education
   - Secondary
   - Post-secondary, which included professional school and university preparatory courses
   - University, including three and five-year degrees

2. Age of the speaker. Participants were given five choices: less than 30, 30 to 40, 40 to 50, 50 to 60 and 60 or more. The last two groups were combined due to the low number of responses for each of those categories, resulting in 4 age groups.
3. Occupation of the speaker. Participants in the survey could choose among six different professions: bar or restaurant employee, construction worker, farmer, store clerk, teacher, doctor or lawyer.

4. Speaker’s place of origin. Participants could choose among five different areas which they believe the speaker came from: a rural area, the mining area (Cuencas), Gijón, Oviedo or outside of Asturias.

5.3 Results

In this section I report the results obtained in the statistical analysis. First, I examine the results for the rating variables, including the output of the linear mixed models and the conditional inference trees. Then, I report the results obtained in the multinomial logistic regression models.

5.3.1 Rating variables

Each speaker was evaluated according to six rating variables: intelligence, masculinity/femininity, formality, rural/urban, economic status, and familiarity. In this section I report the results of the linear mixed models in which each of these ratings are dependent variables.
5.3.1.1 Intelligence

A linear mixed model was developed using Intelligence rating as the dependent variable. The best fit model includes feature language as the only significant predictor of intelligence rating for the /-o/, /-u/ variable. The coefficients are shown in Table 35:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.70</td>
<td>0.05</td>
<td>80.93</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Feature language (reference level is Spanish)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asturian</td>
<td>-0.08</td>
<td>0.02</td>
<td>-4.00</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 35. Effect of feature language on intelligence rating of utterances that contain the /-o/, /-u/ variable.

As we can see, when speakers use the variant /-u/ they are rated as significantly less intelligent than when they use /-o/. No significant effects were observed for any other independent variable; i.e. the social characteristics of the participants and the gender of the speaker did not have a significant effect on variable rating.

None of the independent variables were selected as significant predictors of intelligence rating for the /-as/, /-es/ variable over the random effect of participant.

5.3.1.2 Masculinity/Femininity

Speakers were also rated on a masculinity/femininity scale. When the speaker was a woman the scale ranged from less feminine to more feminine, and when the speaker was a man the scale ranged from more masculine to less masculine, so that an increase in the scale implied an increase in femininity or non-masculinity. Thus high scores in this
variable rating corresponded with low masculinity if the speaker was a man and high femininity if the speaker was a woman. Similarly, low scores indicated high masculinity when the speaker was a man and low femininity when the speaker was a woman. For the /-o/, /-u/ variable, the best-fit model includes speaker gender, feature language and whether the participants considered themselves to be a speaker of Asturian. The results are summarized in Table 36:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.85</td>
<td>0.07</td>
<td>73.07</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Speaker gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.81</td>
<td>0.03</td>
<td>-29.04</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Feature language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asturian</td>
<td>-0.20</td>
<td>0.03</td>
<td>-7.12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Asturian speaker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.19</td>
<td>0.09</td>
<td>2.21</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 36. Factors contributing to masculinity/femininity rating of utterances that contain the /-o/, /-u/ variable.

The results show that male speakers are given significantly lower scores in the masculinity/femininity scale than women (regardless of whether the speaker used Asturian or Spanish variants). When the utterance includes the Asturian variant /-u/ speakers are rated as significantly less feminine/more masculine. Finally, when participants consider themselves to be speakers of Asturian, they rate the speakers as significantly more feminine or less masculine.

In order to explore the interactions between the independent variables, I developed a conditional inference tree that shows the differential effects of each of them.
For better visualization, the mean rating for each group was included under each final node:

![Conditional inference tree of factors that have a significant effect on masculinity/femininity rating of utterances that contain the /-o/, /-u/ variable.]

Figure 35 shows that feature language is only a significant predictor of masculinity/femininity when the sentence is uttered by a female speaker. While male speakers are not judged differently when they use Spanish or Asturian variants, female speakers are rated as more feminine when they use Spanish /-o/ than when they use Asturian /-u/. However, when they use the Asturian variant, they are rated as significantly more feminine when the participant considers himself or herself a speaker of Asturian. In
this case, the femininity rating is as high as when female speakers are using the Spanish variant.

When we analyze the data for /-as/, /-es/ variable we find that the gender of the speaker is the only predictor that has a significant effect on the masculinity/femininity rating, regardless of whether /-as/ or /-es/ is used. The results for this analysis are included in Table 37:

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.94</td>
<td>0.07</td>
<td>73.63</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 37. Effect of speaker gender on masculinity/femininity rating of utterances that contain the /-as/, /-es/ variable.

5.3.1.3 Formality

The formality rating shows the greatest influence of factors related to the participant. The age of the participant, the feature language, the gender of the speaker and the education level of the participant have a significant effect on the score given to the utterances that contain the /-o/, /-u/ variable. The effects are summarized on Table 38:
The results of the linear mixed model indicate that as the age of the participant increases, so does the formality score. When the utterance contains items with Asturian /-u/ instead of /-o/ the rating is significantly lower. The gender of the speaker also has an effect, such that male speakers are rated as significantly more formal than female speakers. Finally, participants with a low education level give significantly higher ratings than those with a high education level (the difference between middle education and high education is not statistically significant). The inference tree included in Figure 36 shows the differential effects of these variables:
The inference tree first indicates that participants that are older than 53 behave significantly different, showing higher formality scores. However, because only 3 participants fall into this category, the statistical significance of participant age is compromised. If we consider the group of speakers younger than 53 we see that feature language has a differential effect on their formality scores, such that significantly higher ratings are given to speakers that use Spanish /-o/ instead of /-u/. When the Asturian variant is used female speakers are rated as significantly less formal than male speakers. It should be noted that while the education level of the participant was selected as a
significant predictor in the regression model, it does not have a significant effect in the conditional inference tree.

When we examine the results for the utterances that contain the /-as/, /-es/ variable we observe the same effects of participant and speaker factors as for the /-o/, /-u/ variable. However, the best fit model does not include feature language as a significant predictor of formality rating. Thus, the effects summarized in Table 39 apply regardless of what variant of the feature is used, showing that speakers are not reacting to whether /-as/ or /-es/ is used. The results of this analysis are reported to highlight the differences in the perception of the two linguistic variables:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.57</td>
<td>0.30</td>
<td>8.44</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Participant education (reference level is High)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.64</td>
<td>0.31</td>
<td>2.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Middle</td>
<td>0.30</td>
<td>0.13</td>
<td>2.42</td>
<td>0.02</td>
</tr>
<tr>
<td>Speaker gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.34</td>
<td>0.08</td>
<td>4.05</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Participant age</td>
<td>0.02</td>
<td>0.01</td>
<td>2.11</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 39. Factors contributing to formality rating of utterances that contain the /-as/, /-es/ variable.

The effect of these factors on formality scores given to utterances that contain /-as/ or /-es/ is similar to the one reported for utterances with /-o/ and /-u/. However, as I mentioned above, their effect in this case is independent of whether the speaker uses Asturian or Spanish features, since feature language was found not to have a statistically significant effect on formality rating.
5.3.1.4 Rural/Urban

Participants were also asked to rate speakers on a rural/urban scale. In this case, the higher the rating the more urban they considered the speaker to be. The results of the statistical analysis reveal that feature language and the gender of the speaker are significant predictors of urban rating when using either of the linguistic variables. However, while feature language is the most important factor in the urban score in utterances that contain /-o/ or /-u/, the gender of the speaker is more important in the rating associated with the /-as/, /-es/ variable. The results are reported in Table 40 and Table 41:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.67</td>
<td>0.09</td>
<td>42.28</td>
</tr>
<tr>
<td>Feature language (reference level is Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asturian</td>
<td>-0.75</td>
<td>0.09</td>
<td>-8.21</td>
</tr>
<tr>
<td>Speaker gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.31</td>
<td>0.09</td>
<td>3.36</td>
</tr>
</tbody>
</table>

Table 40. Factors contributing to rural/urban rating of utterances that contain the /-o/, /-u/ variable.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.58</td>
<td>0.09</td>
<td>41.15</td>
</tr>
<tr>
<td>Speaker gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.28</td>
<td>0.10</td>
<td>2.70</td>
</tr>
<tr>
<td>Feature language (reference level is Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asturian</td>
<td>-0.24</td>
<td>0.10</td>
<td>-2.31</td>
</tr>
</tbody>
</table>

Table 41. Factors contributing to rural/urban rating of utterances that contain the /-as/, /-es/ variable.
In both cases male speakers are rated as significantly more urban than female speakers and the utterances that contain the Asturian variables receive significantly lower scores (i.e. more rural). However, the inference tree for the /-as/, /-es/ variable does not find feature language to be a significant predictor of urban rating.

The interaction between feature language and speaker gender in the urban rating of utterances that contain the /-o/, /-u/ variable is depicted in Figure 37:

![Figure 37](image)

Figure 37. Conditional inference tree of factors that have a significant effect on urban rating of utterances that contain the /-o/, /-u/ variable.

The inference tree confirms that speakers that use /-u/ are judged as significantly more rural than those who use /-o/. It also shows that gender has a differential effect only
when the speakers use the Asturian variant. In these cases male speakers are judged as significantly more urban than female speakers. In general, the lowest score on the scale (the most rural rating) is given to women that use /-u/.

### 5.3.1.5 Economic status

Participants also rated speakers according to their perceived economic status. The Likert scale used in this case ranged from poor to rich. The statistical analysis of the ratings given to utterances that contain /-o/ or /-u/ indicates that feature language and the gender of the speaker are the only factors that significantly predict the rating in this category. Table 42 summarizes the results:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.48</td>
<td>0.05</td>
<td>71.57</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Feature language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asturian</td>
<td>-0.24</td>
<td>0.04</td>
<td>-5.40</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Speaker gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.14</td>
<td>0.04</td>
<td>3.14</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 42. Factors contributing to rural/urban rating of utterances that contain the /-o/, /-u/ variable.

The results show that when the speakers use the Asturian variant they are rated as having a lower economic status than when they use the Spanish one. With regards to the gender of the speaker, men are perceived as having a significantly higher economic status than women. The inference tree in Figure 38 shows how feature language and speaker gender interact in the rating of economic status:
The inference tree indicates that, while there is no significant effect of speaker gender when the utterance contains /-o/, when the Asturian variant /-u/ is used male and female speakers are rated differently, with men receiving higher ratings for economic status than women. Once again, the lowest rating is given to female speakers when using the Asturian variant.

As was the case for most of the other dependent variables, the language of the linguistic features under study did not have a significant effect when the variable used was /-as/, /-es/. In this case, only the gender of the speaker was a significant predictor of the variable rating. The results of the linear mixed model are shown in Table 43:
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.48</td>
<td>0.05</td>
<td>71.60</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Speaker gender (reference level is Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.19</td>
<td>0.06</td>
<td>3.49</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 43. Effect of speaker gender on economic status rating of utterances that contain the /-as/, /-es/ variable.

The results indicate that male speakers are rated as having a significantly higher economic status than female speakers.

5.3.1.6 Familiarity

For the last variable rating, participants were asked to evaluate how well speakers knew their interlocutor. High values on this scale indicate high familiarity between the speaker and the person they were addressing in the stimuli. The effects of independent variables on familiarity ratings present unexpected results if we consider what has been reported for the other variable ratings. In this case, the only factor that has a significant effect on the ratings of utterances that contain the /-o/, /-u/ variable is the gender of the speaker and feature language is a significant predictor of familiarity rating only when the utterances contain the /-as/, /-es/ variable. The results for each linguistic variable are reported in Table 44 and Table 45:
Table 44. Effect of speaker gender on familiarity rating of utterances that contain the /-o/, /-u/ variable.

Table 45. Effect of speaker gender on familiarity rating of utterances that contain the /-as/, /-es/ variable.

The results in Table 44 show that, when using the /-o/, /-u/ variable, male speakers are rated as significantly more familiar than female speakers, regardless of whether they used the Asturian or the Spanish variant. Table 45 contains the results for familiarity rating when the variable used is /-as/ vs. /-es/, indicating that when speakers use /-es/ they are rated as significantly more familiar with their interlocutor than when they use /-as/.

5.3.2 Checkbox variables

Participants in the experiment were also asked to choose a categorical classification of the speaker according to their perceived education, age, occupation and place of origin. Each participant could choose only one of the options provided for each category. In this section I explore the distribution of responses to each checkbox variable.
according to the feature language being used by the speaker and I report the results of the multinomial logistic regression. In every statistical analysis, I selected as reference level the category that was expected to be chosen most frequently when the utterance contained the Asturian variants, based on the analysis of production data and the observations made by previous research. The combination of graphical information and statistical data provides a clear picture of the trends observed in the responses and the differences between the two variables under study.

5.3.2.1 Education

As I mentioned in section 5.2.5, the seven educational levels from which the participants could choose were collapsed into four different categories: elementary, secondary, post-secondary and university education. The graphs in Figure 39 show what proportion of the utterances that were assigned to each education level contained Asturian or Spanish features. The graph on the left illustrates the results for the /-o/, /-u/ variable and the one on the right, the results for the /-as/, /-es/ variable.
Figure 39. Bar plots illustrating what proportion of the utterances that were assigned to each education level contained Asturian or Spanish features.

The graph on the left, for example, shows that, out of all the sentences that were judged as being uttered by a speaker with elementary education, roughly 70% of them contained /-u/ and 30% contained /-o/. We can see that the proportion of utterances with the Asturian variant decreases as the education level increases. Of the sentences that were classified as uttered by a speaker with university education only around 30% contained an Asturian variant. Even though the same trend is observed when the features used are /-as/ and /-es/, the graph on the right side shows that the difference between educational levels is not as pronounced in this case. This differential behavior of each linguistic variable is corroborated by the results of the multinomial logistic regression model, which shows that while feature language has a significant effect on what educational level is chosen by the participant when the variable being used is /-o/, /-u/, this effect is not statistically significant when the utterances contain the /-as/, /-es/ variable. The results for the former are reported in Table 46:
<table>
<thead>
<tr>
<th>Reference level is Elementary education</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>University education: (Intercept)</td>
<td>0.89</td>
<td>0.19</td>
<td>4.71</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Post-secondary education: (Intercept)</td>
<td>1.30</td>
<td>0.18</td>
<td>7.30</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Secondary education: (Intercept)</td>
<td>0.91</td>
<td>0.19</td>
<td>4.90</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Feature language (reference level is Spanish)

| University education: Asturian         | -1.50    | 0.26| -5.76  | <0.01   |
| Post-secondary education: Asturian     | -0.90    | 0.23| -3.99  | <0.01   |
| Secondary education: Asturian          | -0.59    | 0.23| -2.54  | 0.01    |

Table 46. Effect of feature language on the selection of educational level in utterances that contain the /-o/, /-u/ variable.

The results reported above indicate that the odds of the speaker being classified as having a secondary education level over elementary education level decrease significantly when the utterance evaluated contains /-u/. The odds of the speaker being classified as having a post-secondary or university education over elementary education are also lower when /-u/ is present. Thus, higher education levels are increasingly less likely to be selected when the Asturian variant is used. All the education levels behave significantly differently from each other except for the case of secondary and post-secondary education; i.e., the use of Asturian or Spanish features does not have a significant effect on the choice between secondary and post-secondary education.

5.3.2.2 Age

Participants were also asked to choose an age range for each speaker that they heard. The graphs in Figure 40 shows the distribution of the responses according to what
proportion of the utterances assigned to a particular age range contained Asturian or Spanish features:

![Bar plots illustrating what proportion of the utterances that were assigned to each age range contained Asturian or Spanish features.](image)

Figure 40. Bar plots illustrating what proportion of the utterances that were assigned to each age range contained Asturian or Spanish features.

The bar plots show very different effects of feature language for each linguistic variable. While the proportion of utterances that contain Asturian variants increases as the responses move from younger to older age when the /-as/, /-es/ variable is used, an almost reverse effect is observed for the /-o/, /-u/ variable. This is an unexpected result and raises some questions with regards to the relationship between the use of Asturian and age, which will be addressed in section 5.4.1.4.

The statistical analysis elucidates whether feature language is a significant predictor of age range assignment. In Table 47 I report the results for the utterances that contain the /-as/, /-es/ variable.
Table 47. Effect of feature language on the selection of age range in utterances that contain the /-as/, /-es/ variable.

<table>
<thead>
<tr>
<th>Reference level is more than 50</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 30: (Intercept)</td>
<td>0.18</td>
<td>0.17</td>
<td>1.04</td>
<td>0.30</td>
</tr>
<tr>
<td>30 to 40: (Intercept)</td>
<td>1.11</td>
<td>0.15</td>
<td>7.51</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>40 to 50: (Intercept)</td>
<td>0.36</td>
<td>0.17</td>
<td>2.13</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Feature language (reference level is Spanish)

| Less than 30: Asturian         | -0.77    | 0.25| -3.10  | 0.01    |
| 30 to 40: Asturian             | -0.61    | 0.20| -3.05  | <0.01   |
| 40 to 50: Asturian             | -0.39    | 0.23| -1.72  | 0.08    |

The odds of participants choosing the “less than 30” category over the “more than 50” one are significantly lower when the speaker uses the Asturian variant. The odds of choosing the “30 to 40” category are also significantly lower when the variant used is /-es/ but the difference in probabilities between that category and the “more than 50” one is smaller in this case. Finally, even though choosing “40 to 50” over “more than 50” is less likely when the Asturian variant is used, the difference is not statistically significant. The difference between adjacent age categories is also not statistically significant. Thus, the only significant effect is the choice of “more than 50” over the two youngest generations.

In contrast to the observations made for the /-as/, /-es/ variable, the multinomial logistic regression model for the utterances that contain /-o/ or /-u/ indicates that the probabilities of choosing any of the age categories over the 50 or more choice are not significantly different when speakers use /-u/ instead of /-o/. The only significant effect observed for this variable is the fact that when /-u/ is used, the choice of “less than 30” is significantly more likely than that of “40 to 50”.

221
5.3.2.3 Occupation

Participants in the experiment were also asked to select an occupation for each speaker. The options provided included jobs in different sectors, such as manufacturing, commerce and education. The graphs in Figure 41 shows the distribution of the responses according to what proportion of the utterances assigned to a particular occupation contained Asturian or Spanish features:

Figure 41. Bar plots illustrating what proportion of the utterances that were assigned to each occupation contained Asturian or Spanish features.

The information in the graphs was organized so that the occupations that were expected to be most frequently assigned to utterances that used the Asturian variants appear on the left side, while occupations that were more likely to be chosen when speakers used the Spanish variants were placed on the right side of the graph. These hypotheses were based on the results obtained in the analysis of production data. The graph shows that the expected ordering of the different occupations is mostly maintained.
when participants heard /-o/ or /-u/, with the exception of the category “farmer”. Thus, of all the utterances that were judged to be spoken by someone that works in a bar or in a restaurant, about 60% contained /-u/. On the other side of the spectrum, of the utterances that were judged to be spoken by a teacher, only 25% contained the Asturian variant.

The distribution of responses when speakers use the /-as/, /-es/ variable, which does not match the expectations based on the analysis of production data, is very different from the results for /-o/ and /-u/. In this case, utterances that were judged as being spoken by someone who works in construction showed the highest proportion of sentences containing the Asturian variant, with the teacher category showing the second highest percentage of utterances containing /-es/. The results of the regression analysis reflect the differences observed in the graphical representation:

<table>
<thead>
<tr>
<th>Reference level is Bar/Restaurant</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher: (Intercept)</td>
<td>-0.24</td>
<td>0.19</td>
<td>-1.24</td>
<td>0.21</td>
</tr>
<tr>
<td>Doctor/Lawyer: (Intercept)</td>
<td>-0.93</td>
<td>0.24</td>
<td>-3.87</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Store clerk: (Intercept)</td>
<td>0.74</td>
<td>0.16</td>
<td>4.82</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Construction: (Intercept)</td>
<td>0.45</td>
<td>0.16</td>
<td>2.77</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Farmer: (Intercept)</td>
<td>-0.85</td>
<td>0.23</td>
<td>-3.64</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Feature language (reference level is Spanish)

| Teacher: Asturian                | -1.55    | 0.33| -4.68  | <0.01   |
| Doctor/Lawyer: Asturian          | -1.23    | 0.40| -3.09  | <0.01   |
| Store clerk: Asturian            | -0.59    | 0.21| -2.81  | <0.01   |
| Construction: Asturian           | -0.20    | 0.21| -0.93  | 0.35    |
| Farmer: Asturian                 | -0.53    | 0.32| -1.63  | 0.10    |

Table 48. Effect of feature language on the selection of occupation in utterances that contain the /-o/, /-u/ variable.
As we can see in Table 48, the odds of the participant selecting teacher over bar/restaurant are significantly lower when the feature language used in the utterance is Asturian. The same trend but with a smaller effect size is observed in the selection of doctor/lawyer and store clerk over the bar/restaurant category. Feature language, however, does not have a significant effect on the selection of the construction or farmer categories over the bar/restaurant option. Further exploration of the data reveals that there is not a significant difference between these categories but that bar/restaurant and construction are significantly more likely to be chosen over store clerk, doctor/lawyer and teacher when the Aturian features are used. Store clerk, doctor/lawyer and teacher do not behave significantly different.

When the variable used is /-as/ vs. /-es/, the only category whose probabilities of being selected are significantly different from the bar/restaurant choice is the construction group. Table 49 summarizes the results of the multinomial logistic regression model using bar/restaurant as the reference level:
The results show that while choosing doctor/lawyer or store clerk over bar/restaurant is less likely when the feature language is Asturian, this difference is not statistically significant. In this case, participants are more likely to choose teacher or farmer over bar/restaurant when the utterance contains the Asturian variant, but the difference is not statistically significant either. The construction category is the only one that is significantly more likely to be chosen over bar/restaurant when the speaker uses /-es/ instead of /-as/. The probabilities of selecting construction when the Asturian features are used are also significantly higher than over store clerk and the doctor/lawyer options. The rest of the occupation choices do not show significant differences.
5.3.2.4 Place of origin

The last categorical choice that participants had to make was regarding the perceived place of origin of the speaker. Participants could choose from five areas: a village (referring in general to rural Asturias), Cuencas (the mining area), Gijón, Oviedo and from outside of Asturias. The graphs in Figure 42 show the distribution of the responses according to what proportion of the utterances assigned to a particular location contained Asturian or Spanish features:

![Figure 42. Bar plots illustrating what proportion of the utterances that were assigned to each place of origin contained Asturian or Spanish features.](image)

In both graphs we observe a similar distribution of the responses. While most of the utterances assigned to the village and Cuencas categories contained Asturian features, the opposite was true for the sentences that were judged as spoken by someone from Oviedo or outside of Asturias. In both cases Gijón occupied the middle of the spectrum. The only difference between the two linguistic variables is one of degree, such that the
changes in the proportions from one category to the next are less pronounced. The results of the regression analysis, reported in Table 50, confirm these observations:

<table>
<thead>
<tr>
<th>Reference level is Village</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside of Asturias: (Intercept)</td>
<td>-1.01</td>
<td>0.24</td>
<td>-4.24</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Oviedo: (Intercept)</td>
<td>0.25</td>
<td>0.16</td>
<td>1.54</td>
<td>0.12</td>
</tr>
<tr>
<td>Gijón: (Intercept)</td>
<td>0.77</td>
<td>0.15</td>
<td>5.14</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cuencas: (Intercept)</td>
<td>0.02</td>
<td>0.17</td>
<td>0.09</td>
<td>0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature language (reference level is Spanish)</th>
<th>Estimate</th>
<th>SE</th>
<th>tValue</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside of Asturias: Asturian</td>
<td>-1.89</td>
<td>0.48</td>
<td>-3.91</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Oviedo: Asturian</td>
<td>-1.54</td>
<td>0.26</td>
<td>-5.86</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gijón: Asturian</td>
<td>-0.85</td>
<td>0.20</td>
<td>-4.19</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cuencas: Asturian</td>
<td>0.23</td>
<td>0.22</td>
<td>1.06</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 50. Effect of feature language on the selection of place of origin in utterances that contain the /-o/, /-u/ variable.

The results indicate that the odds of selecting outside of Asturias, Oviedo or Gijón over village as the place of origin of the speaker are significantly lower when the utterances contain the Asturian variant, with Gijón being the category that behaves most similarly to the village choice. In contrast with this, the odds of choosing Cuencas over village are not significantly different when /-u/ is used. Further exploration of the data reveals that Gijón behaves significantly different from all the other categories. I also discovered that the odds of choosing Oviedo over outside of Asturias are not significantly higher when using Asturian variants. The fact that sentences that contain Asturian variants can be judged as being uttered by speakers from outside of Asturias and that this category does not behave significantly different from the Oviedo choice is surprising. I
attribute this to the fact that the utterances were read by the speakers and that the speech mode used in the stimuli could have a stronger effect than the presence of Asturian features for some of the participants.

The same model was developed for the /-as/, /-es/ variable, however, in this case feature language only had a significant effect on the choice of Cuencas over Gijón, Oviedo and outside of Asturias, such that the presence of /-es/ significantly increased the odds of Cuencas being chosen over the other three locations.

In the following section I discuss the results obtained in the perception experiment, analyzing the effect that each of the independent variables has on the variable ratings of the speaker and discussing the observed correlations between the use of Asturian or Spanish features and the selection of particular categories for the checkbox variables.

5.4 Discussion

The results of the statistical analyses indicate that, across all the characteristics on which speakers were rated, the two most important predictors of variable rating are the language of the feature and the gender of the speaker. We also see that, when the variable used is /-o/ vs. /-u/, the language used in the utterance has a significant effect on all the variable ratings except for familiarity. The opposite is true for the /-as/, /-es/ variable, in that whether the speaker uses Asturian or Spanish has a significant effect on the familiarity rating and not on the other characteristics, with the exception of the rural/urban rating, where the role of feature language is significant for both linguistic
variables. Finally, the age and education of the participant only have a significant effect on the formality rating and whether participants consider themselves to be speakers of Asturian is only a significant predictor of the rating given in the masculinity/femininity category. Table 51 summarizes these findings:

<table>
<thead>
<tr>
<th>Feature language</th>
<th>Speaker gender</th>
<th>Participant age</th>
<th>Participant education</th>
<th>Asturian speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>-o/ vs. -u/</td>
<td>Intelligence</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masculinity</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Formality</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Rural/Urban</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economic st</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature language</th>
<th>Speaker gender</th>
<th>Participant age</th>
<th>Participant education</th>
<th>Asturian speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>-as/ vs. -es/</td>
<td>Intelligence</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masculinity</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formality</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Rural/Urban</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economic st</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 51. Significant predictors across the six variable ratings

Because the effect of participant age, education and being an Asturian speaker is only observed in one category respectively, the discussion in this section focuses on the effect of feature language and the gender of the speaker.

For the checkbox variables, I found that the choices of the participant are, in general, constrained by the language of the features included in the utterance. This effect is more robust when the variable used is -o/ vs. -u/ than when -as/ and -es/ are present in the stimuli. When -o/ and -u/ are used, with the exception of the Age variable, the distribution of the choices in relationship to the language of the feature is normally
hierarchical, with each (or most) of categories differing significantly from each other. In contrast with this, when the variable used is /-as/ vs. /-es/, it is normally one single category that differs significantly from the rest. For example, the presence of /-es/ significantly increases the odds of choosing Cuencas over Gijón and over Oviedo; however, it does not make a difference in the choice of Gijón over Oviedo. In contrast to this, the presence of /-u/ significantly increases the odds of the participant selecting Cuencas over Gijón and over Oviedo, but it also significantly increases the probabilities of choosing Gijón over Oviedo.

5.4.1 Effect of feature language

The results of the multiple linear mixed models indicate that, in general, when speakers use Asturian features, they are given lower ratings on intelligence, formality and economic status. They are also rated as more masculine and more rural and are given higher scores when judging how familiar the speaker is with his or her interlocutor. The graphs in Figure 43 illustrate the different ratings according to feature language and divided by linguistic variable. The y axis represents the rating of each variable and the x axis contains the language of the linguistic features under study. Each rating variable is represented by a different shape:

---

Footnote: Familiarity was excluded from the graphs because the scale is reversed here (higher scores are normally associated with Asturian and not Spanish) so it impeded clear visualization of the graph. The effect of language on this variable will be addressed later in the section.
The graphical comparison of the effects of feature language on each variable rating reflects the differences observed in the statistical analysis. As we can see on the graph on the left side, the difference in ratings between utterances that only vary on the language used in the features under study is much greater when the variants compared are /-o/ and /-u/. In fact, the differences observed for this variable are statistically significant in every category except for familiarity. In contrast with this, when the variable used is /-as/ vs. /-es/ the ratings given to each variant are only significantly different in the case of the urban/rural score (and familiarity, which will be discussed later). Furthermore, when listeners rate speakers on intelligence, economic status and formality, they receive slightly higher scores when using /-es/. However, because this difference is not statistically significant, we cannot make any generalization regarding this apparently contradictory result.
In the statistical analysis of the checkbox variables I also discovered different tendencies in the two linguistic variables. With the exception of the age category, I found that the effect of feature language is weaker when the variable used is /-as/ vs. /-es/. The results of the analysis of both types of responses shows that the two variables under study are perceived differently. The fact that /-o/ and /-u/ are closer in their articulation than the front and central vowels (see discussion in section 4.4.1.7.1) is especially meaningful, since we would expect participants to have more difficulty in discriminating between those two sounds than between /-as/ and /-es/. Instead, it is the alternation in back vowels that has a stronger effect on the variable ratings and checkbox questions. This finding also contributes to the idea that listeners can perceive small differences in acoustic detail and associate it with certain social categories.

In section 4.3.2, I proposed that the two variables under study have different social status. This hypothesis was based on the evaluation of the effect that external or social variables have on the alternation between Asturian and Spanish variants. We saw that the effect of Occupation is a stronger predictor of the variation between /-o/ and /-u/ than between /-as/ and /-es/. I also noted that the use of /-u/ is not only widespread among most speakers of Asturian Spanish but it is also recognized by speakers as a salient feature of Asturian. The results of the perception experiment provide further evidence that /-u/ is a stronger marker of the social categories explored in this experiment than /-es/ is.
5.4.1.1 Education and intelligence

In general, participants associate utterances that contain Asturian /-u/ with lower education levels and lower intelligence. This confirms the results obtained in the analysis of production data, according to which, speakers with lower education levels are more likely to use the two Asturian variants. Even though speakers with lower education also show higher rates of /-es/, this correlation is not perceived by the participants in the perception experiment. The fact that listeners associate /-u/ with low education and low intelligence confirms the stigmatization of the regional language and its common characterization as an imperfect variety of Spanish. For many speakers of Asturian Spanish, including Asturian features in their speech is equivalent to speaking incorrectly.

The quote reproduced in (62), uttered by one of the speakers that I interviewed for the analysis of production data, illustrates this stigmatization of the Asturian features:

(62) “en mi pueblo no se habla así tan mal… pero a ver, por ejemplo, en el pueblo de la mi cuñada, fatal, fatal hablan, y son cerca de nosotros y hablan fatal. Y vienen pa'ya y no se les corrige, no corrigen. Chica, esas palabras tan… tan de pueblo... yo... parezme que no tengo, no sé. Yo vine con diez años también pero bueno” (GI11-18)

‘in my village they don’t speak so badly… but, I mean, for example, in my sister-in-law’s village, terribly, they speak terribly, and they are from close to us and they speak terribly. And they come here (to Gijón) and nobody corrects them, they don’t correct. Girl, those words that are so, so rural… I… I don’t think I have them, I don’t know. I came here when I was ten but anyway’
Note that despite the negative attitudes towards the use of Asturian features and claiming that she does not have the characteristic “rural features” that she is talking about, her speech still contains features from Asturian, bolded in the example. When the same speaker is asked how many languages she speaks she responds with the quote included in (63):

(63) "Bien, ninguna. O sea, el castellano mal, pero bueno." (GI11-18)
‘I don’t speak any (language) well. I mean, I speak Castilian badly, but anyway’

When the participants in the perception experiment were asked to describe their speech, some of them also refer to the standard variety of Spanish as the correct way of speaking. Their descriptions are included in examples (64), (65) and (66):

(64) “Normal, se ve que soy asturiana pero intento hablar lo mas correcto posible.” (PER12-238)
‘normal, you can tell I’m Asturian but I try to speak as correctly as possible’

(65) “Castellano muy pobre” (PER12-404)
‘Very poor Castilian’

(66) “Procuro hablar bien, aunque se mezclan castellano y palabras sueltas en asturiano” (PER12-157)
‘I try to speak correctly, even though I mix Castilian and isolated words in Asturian’
5.4.1.2 Occupation and economic status

As I mentioned in Chapter 4, education is correlated with occupation but also is economic status, since the average income is different in each profession and listeners are aware of these differences. Speakers that are judged as having high education degrees and those that are perceived to be doctors, lawyers or teachers are also given significantly higher scores in the economic status category.

In the responses for the occupation question we register an effect of feature language similar to the one reported in the education category. The relationship between the speakers’ use of /-u/ and the selection of particular professions reflects the correlations found in the analysis of production data. Speakers that work in commerce and labor show higher rates of /-u/ and speakers that use this variant are also perceived as having jobs in these fields. Even though the correlation between use of /-es/ and occupational group was also significant in the analysis of production data, listeners do not clearly associate /-as/ or /-es/ with particular occupations and the trends observed in the perception experiment do not match the results obtained in the exploration of production data, where speakers that work in commerce (including bars and restaurants) show the highest rates of /-es/. Instead listeners are more likely to judge speakers that use the Asturian variant as construction workers or teachers. Furthermore, speakers that use /-u/ are judged as having a lower economic status, while speakers that use /-es/ are not judged significantly differently from speakers that use /-as/. I propose that these differences, together with the lack of alignment between production and perception in the occupation category observed in the analysis of /-as/ vs. /-es/, also stem from the different social status of the two linguistic variables under study.
5.4.1.3 Formality and familiarity

Utterances that contain /-u/ are judged as being significantly more informal than those which contain the Spanish variant. This difference is not statistically significant in the case of /-as/ vs. /-es/. However, when speakers use /-es/ they are given significantly higher scores in the familiarity rating, indicating that listeners perceived them to know their interlocutor better than when they used /-as/. In contrast with this result, feature language did not have a significant effect on the familiarity rating when the /-o/, /-u/ variable was used.

These results are surprising for several reasons. First, I expected familiarity and formality to pattern similarly with respect to feature language so that if the use of Asturian variants is judged as less formal, speakers that use them would also be judged as more familiar with their interlocutor. This is the case for the /-o/, /-u/ variable; however, when speakers use /-es/ they are rated as more formal but also more familiar. Second, because in the analysis of production data social tie only had a significant effect on the variation between /-o/ and /-u/, it is surprising that familiarity is only significant when speakers use the /-as/, /-es/ variable. Further exploration of the results of the perception experiment reveals that one speaker, Oscar, is driving these results, since he is the only one that is evaluated differently with respect to familiarity when using the Asturian and Spanish variants. In the utterance he recorded he talks about working long hours in the past and not having sufficient time to spend with his family. Among all the utterances evaluated by the listeners, this is the only case in which the speaker talks about a personal aspect of his life. Even if Oscar does not have the highest familiarity score among speakers, it is possible that the difference in the content of the utterance is responsible for
the differential effect of feature language. Further research would be necessary to make any generalizations regarding the role of familiarity.

The results obtained for the formality rating align better with the production data, which indicated that speakers with whom I had a first order social tie showed higher rates of /-u/. This effect was only observed in the group of speakers that were already more likely to use /-u/ (male speakers that work in commerce or manufacturing jobs). In section 4.4.2.3 I hypothesized that the direct relationship with their interlocutor decreased the formality of the interview contexts and resulted in higher use of Asturian /-u/. In the perception experiment, utterances that contained /-u/ were evaluated as being uttered in a more informal communicative situation. This result together with the findings from the analysis of production data provides further evidence of the importance of style in the variation between Spanish and Asturian features.

When participants were asked to characterize their speech they frequently alluded to their use of the two languages in relation to the formality of the situation. The quotes in (67), (68) and (69) illustrate this:

(67) “En ámbitos formales hablo un castellano casi perfecto. En ámbitos más desenfadados lo que llamamos “amestao”, una mezcla de castellano y asturiano.” (PER12-375)
     ‘In formal situations I speak almost perfect Castilian. In more relaxed situations, what we call “amestao”, a mixture of Castilian and Asturian’
“Amestao cuando estoy en casa o con los amigos en modo informal y más castellano o más asturiano dependiendo del contexto” (PER12-162)

“Amestao” when I’m at home or with my friends in an informal situation and more Castilian or more Asturian depending on the context.

“Uso el español adecuando mi registro a cada situación. A veces mezclo palabras o estructuras asturianas si estoy hablando de forma natural en un contexto relajado.” (PER12-14)

I use Spanish changing my register according to each situation. Sometime I mix Asturian words or structures if I’m speaking in a natural way in a relaxed context.

As we can see in the examples above, speakers are aware of the fact that they shift styles depending on the communicative context and that style shift involves increasing or decreasing the amount of Asturian features. The situations in which participants claim to use Asturian features (or amestao) are characterized as relaxed, natural and informal. The use of the word “natural” is particularly informative, indicating that Asturian Spanish is their default and preferred linguistic variety, while the use of more Spanish features is seen as something “unnatural” that they are required to do in particular communicative situations. The participant’s awareness of their own stylistic shift supports the hypothesis presented in section 4.4.2.1, in which I proposed that some speakers strategically use Asturian features with an economic purpose. Further research is necessary to make conclusive claims regarding the effects of addressee, setting, topic and stylistic factors that are related to the speaker’s agency, such as purpose, key and setting. However, the
evidence presented here and in chapter 4 points to the relevance of these factors on intra-speaker variation.

5.4.1.4 Age

In section 4.4.2.4 I hypothesized that, based on the results of the hierarchical scales and the quantitative analysis of the production data, it is possible that age constrains how many and what features are included in speech but that, once a feature is present in all generations, age has a much weaker effect on how much that particular feature is used. The results of the perception experiment corroborate this observation. Feature language does not have a significant effect on the selection of an age category for the speaker when the variable used is /-o/ vs. /-u/. However, when /-as/ or /-es/ are used, the language used has a significant effect on the age assigned to the speaker, with older age ranges being more likely when the Asturian variant is used. Because /-u/ is present in the speech of speakers of all ages, its use is not perceived as a particular marker of any age category. However, the use of /-es/ is not as widespread among younger speakers as /-u/ is and, as a consequence, its presence in the stimuli could be more easily associated with the speech of older speakers.

5.4.1.5 Rural vs. urban

Feature language has the strongest effect on the rural/urban rating of the speaker. Not only is the effect stronger than in other variable ratings, but it is also the only category in which the role of feature language is significant in the two linguistic
variables. This supports the idea that the use of Asturian or Spanish is strongly connected with place and the construction of local (urban) and regional identities introduced in section 2.4.

The results indicate that speakers are rated as significantly more rural when they use Asturian variants and the effect is stronger when the linguistic variable used is /-o/ vs. /-u/. The common association of Asturian with the rural world is also observed when we examine the responses of the participants to the question of how they would describe their speech. Two of those responses are included in (70) and (71):

(70) “Bastante de pueblo, como me dicen los que me conocen” (PER12-109)
‘Pretty rural, as people that know me say’

(71) “Entremezclando entre el bable y el castellano, como se suele decir, algo pueblerino” (PER12-454)
‘A mixture of Asturian and Castillian, as people say, a little rural’

When we examine the responses to the checkbox variables we see that different degrees of “Asturianness” are associated with different areas within the region. Speakers that use Asturian variants are most often judged as being from a village or from the mining areas while speakers that use the Spanish alternatives are most frequently thought to be from Oviedo or from outside of Asturias. More importantly, Gijón is in the middle of the Asturian/rural and Spanish/urban scale. Even though Gijón and Oviedo are both cities (and in fact Gijón is larger than Oviedo), the difference between them is that Gijón is an industrial city, while Oviedo is not, and has attracted much more population from
rural areas than the capital Oviedo, which constitutes the cultural, political and administrative center of the region. As a result, what we find in Gijón is a conflict between rural and urban identities. Speakers from this city want to distance themselves from the rural world but at the same time they do not want to sound too Spanish, since there is a strong sense of regional identity. The search for this balance is also perceived in the descriptions that participants provide of their own speech. Some examples are shown in (72), (73) and (74):

(72) “Con acento asturiano, pero no de pueblo” (PER12-140)
‘With an Asturian accent but not too rural’

(73) “Con acento Asturiano sin ser cerrado” (PER12-153)
‘With an Asturian accent but not too strong’

(74) “Con cierto acento gijonés, pero sin usar demasiadas palabras en asturiano” (PER12-231)
‘With a “gijonés” accent but without using too many words in Asturian’

These descriptions show the desire of sounding as an Asturian but not rural and illustrate the conflict between urban and regional identity.

Even though the use of Asturian is clearly stigmatized, the use of Spanish is also stigmatized in Gijón and speakers that try to erase any interference from Asturian are referred to as “fancy speakers” (Prieto 1991). Some of the listeners in the experiment qualify speakers that use Spanish variants as fancy (“fino”), uptight (“estirada”) or arrogant (“prepotente”). It is this partial stigmatization of both languages that governs the
variation between Spanish and Asturian features in Gijón and motivates the use of a mixed system, the variety that appears to be less stigmatized.

5.4.2 Masculinity/Femininity and the effect of gender of the speaker

The results of the experiment indicate that listeners judge as significantly more masculine/less feminine speakers that use /-u/ instead of /-o/. Even though a similar trend is observed in the /-as/, /-es/ variable, the effect of feature language is not statistically significant in this case. The rural world is traditionally viewed as a masculine one, since men were the ones dedicated to farming while women worked in the house. The tendencies observed in the analysis of production data with regards to the effect of gender are supported by the findings of the experiment, which strengthen the idea that gender has a role in the variation since there is alignment between production and perception.

Not only is Asturian associated with masculinity but also the gendering of the language results in women being judged differently from men when using Asturian features. The gender of the speaker was the second most important predictor in the variable rating of speakers. In general the results indicate that men are rated more positively than women on formality, economic status, familiarity and on the rural vs. urban scale. However, when feature language is also a significant predictor of the variable rating (rural vs. urban, formality and economic status for the /-o/, /-u/ variable) the gender of the speaker only has a significant effect in the utterances that contain Asturian variants, indicating that women are judged differently than men when they use /-u/, or vice versa.
These results show that women suffer greater consequences than men for choosing a linguistic variant over the other. In the community under study female speakers are expected to conform to the prestige variety of the language to a greater degree than men and are penalized for choosing a vernacular form by rating them as less urban, less formal, more familiar and as having a lower economic status, while the men’s use of Asturian variants is expected.

5.5 Conclusion

The results of the perception experiment reveal that listeners associate the use of Asturian variants with certain social categories. The most important effect reported is that of feature language on urban/rural rating, indicating that speakers use the linguistic variables under study to mark place (Johnstone et al. 2006). The relationship between language and place is not a straightforward one, since there are other social variables that add complexity to this relationship, such as gender. Furthermore, Asturian features are not only markers of place and their use is not limited to constructing regional identity. They are also associated with low education, low intelligence, lower economic status, etc., making it impossible to separate place and ethnicity from social and cultural identity. One of the most important findings, however, is the fact that different variables contribute in different ways to this purpose, corroborating the fact that /-u/ and /-es/ mark “Asturianness” differently.

The findings of the perception experiment indicate that the linguistic variants under study carry social meaning and hearers use this association of linguistic form with
social categories to situate speakers in the social world. Even though further research is necessary to confirm this correlation, the listeners’ association of particular social information with the use of Asturian vowels, together with the results obtained in the analysis of production data and the observations made by previous research on Asturian sociolinguistics, suggest that speakers of Asturian Spanish may strategically increase or decrease the amount of Asturian features according to how Asturian they want to sound. These findings contribute to the idea that linguistic variation is not just a reflection of pre-existing social categories and that speakers actively use variation to position themselves in the social world.
CHAPTER 6: CONCLUSIONS

6.1 Summary of findings

In this dissertation I analyzed the variation between Asturian and Spanish features in the variety of Asturian Spanish spoken in the city of Gijón, which characterizes the speech of urban areas in the region of Asturias. The analysis focused on the alternation between raised and non-raised vowels in the masculine singular morpheme (/-o/ vs. /-u/) and the feminine plural morpheme (/-as/ vs. /-es/). The first objective of the dissertation was to explore quantitatively the distribution of each linguistic variant in the speech of native speakers of Asturian Spanish in order to uncover the linguistic constraints that govern the variation and the correlations between their use and particular socio-demographic categories.

The quantitative analysis of production data was reported in Chapter 4. In this chapter I performed a statistical analysis that evaluated the effect that linguistic and social predictors have on the categorical choice between the vowels under study and also on the frequency of the first two formants of each vowel, in order to capture the gradient nature of the vocalic variation. The results of the statistical analysis show that the choice between Spanish and Asturian variants and the differences in formant frequencies are linguistically and socially constrained. Among the linguistic constraints, I found that, in
general, the effect of the adjacent segment was the strongest, followed by that of the prior form and that of the lexical category to which each token belonged. While the effect of lexical category was restricted to the variation between /-as/ and /-es/, the other two factors had a significant effect on the two linguistic variables. The results revealed that final /-u/ was less likely to be used in contexts in which the adjacent segment was articulated towards the front of the vocal tract (when a palatal sound, dental or alveolar precedes the vowel) and before the low central vowel /a/. Similarly, the use of /-es/ is favored by a preceding palatal sound, showing expected coarticulation tendencies that have been found in other linguistic varieties. A tendency to linguistic perseveration was also found in the two linguistic variables, such that the use of a particular variant in the previous form resulted in repeated use of that same variant. Finally, /-es/ was significantly more likely to occur in determiners than in other lexical categories but the effect of this linguistic factor was not significant for the /-o/, /-u/ variable.

The statistical analysis also showed that variation between Asturian and Spanish features is socially constrained. The choice between /-o/ and /-u/ is correlated with the occupation, gender and education of the speaker, and also with the social tie that the speaker had with his or her interlocutor, with occupation being the most important predictor. The alternation between /-as/ and /-es/, however, was only constrained by the education level and occupation of the speaker, with education being the most important predictor in this case. Speakers are significantly more likely to use the Asturian variants if they have a lower education level and if they work in manufacturing or commerce. In the case of /-o/ vs. /-u/ men are also significantly more likely to use /-u/, together with speakers that were in interactions with an interlocutor with whom they had a second order
social tie. One of the most significant findings in this part of the analysis are the differences observed in the social correlations of the two linguistic variables; i.e. different social constraints govern the choice between Asturian and Spanish features in the two linguistic variables and the relative importance of these social predictors also varies depending on the linguistic variables. Based on these observations, I hypothesized that /-u/ is a stronger marker of “Asturianness” than /-es/. This hypothesis is corroborated by the results presented in chapter 5.

The second objective of this dissertation was to discover whether the social correlations explained above are perceived by listeners that belong to the same speech community, in order to evaluate whether the linguistic variants have social meaning and what this social meaning is. This goal was achieved by conducting a perception experiment in which participants were asked to evaluate the identity of speakers based on the sentences they uttered, using a Matched Guise Technique. The results from this experiment were reported in chapter 5. The statistical analysis of the responses revealed that listeners associate the use of the Asturian variants with certain social categories. The use of /-u/ was associated with low education, low intelligence and education level, lower economic status, low formality and low femininity. In contrast to this, the use of /-es/ was only associated with high familiarity between the interlocutors and with older speakers. Both variables were also associated with the perceived place of origin of the speaker, such that Asturian /-u/ and /-es/ received significantly lower ratings on the rural/urban scale and speakers that use them were more likely to be perceived as being from rural and mining areas of the region. These results confirm that the social correlations that were found to be significant in the quantitative analysis of production data are meaningful and
used by listeners to evaluate the identity of other speakers. One of the most important findings was the fact that /-u/ is a stronger marker of the social categories explored in this experiment than /-es/, contributing to the idea that the two linguistic variables under study do different “social work”.

6.2 Contributions

This study constitutes the first rigorous empirical analysis of the urban varieties of Asturian Spanish, which contrasts with the overwhelmingly dialectological tradition that has focused on the study of “pure” Asturian and has consequently ignored the outcomes of language contact between Spanish and Asturian. As I mentioned in chapter 2 and chapter 3, previous work addressing the variation between Asturian and Spanish features is very limited. Most of these studies were qualitative in nature and the majority of the observations made were merely impressionistic. Furthermore, the role of linguistic factors on the variation was underexplored, ignoring the possible effect of internal variables that could account for intra-speaker variation and emphasizing the relationship between the use of Asturian features and particular social categories. The results obtained in the analyses of production data and in the perception experiment provide us with a more complete characterization of the linguistic and social distribution of the Asturian Spanish variants under study.

The effects of linguistic factors discovered in the analysis of production data are essential to understanding the nature of the variation in Asturian Spanish. First, the existence of these effects partially explains why intra-speaker variation occurs in the
variety under study. These findings also contribute to our knowledge about the general linguistic mechanisms that guide language variation. The findings regarding the effect of the phonological context, for instance, corroborate the effect that coarticulation has on vocalic variation, showing that the articulation of vowels approximates that of the segment that is adjacent to them. Similarly, the significant tendency for linguistic perseveration observed in the Asturian Spanish production data contributes to our understanding of how language processing takes place. The fact that a previous /-u/ and /-es/ triggers more use of /-u/ and /-es/, respectively, supports the idea that the use of a particular linguistic form triggers its activation in the mental representation of following utterances, and that the higher the level of activation, the more likely it is for that same form to be chosen in subsequent linguistic production (spreading activation). Finally, the effect that the lexical category has on the variation between /-as/ and /-es/ contributes to observations made by research on code-switching, according to which function words are less likely to be borrowed than open-class items.

In the analysis of the role that social factors have on the variation between Asturian and Spanish features, I found that the most important factor constraining the variation between /-o/ and /-u/ is the occupation of the speaker. One interesting result from this part of the analysis was the fact that the occupational group that was significantly most likely to use Asturian /-u/ was the one that included speakers who work in bars, restaurants and small specialty stores. I propose that these speakers use the variation strategically, including more Asturian features as a way to create a particular relationship with their customers and a social setting that resembles that of the family and close friendship. This ensures the loyalty of their customers and, with it, the success of
their businesses. Other research, mainly on variation in English, has found that speakers may use certain linguistic variants in a strategic and planned manner (Johnstone 1999). These observations also help explain the existence of intra-speaker variation and are related to the fact that stylistic factors seem to constrain the choice between Spanish and Asturian features. In the analysis of production data I found that speakers are significantly more likely to use /-u/ when they have a first order tie (a direct relationship) with their interlocutor. I also found qualitative evidence of speakers’ awareness of their own style shifts between varieties that include more or less Asturian features. Even though further research is necessary to fully account for the effects that style-shift has on the variation between Spanish and Asturian features, the results obtained for the speakers that work in commerce and food industries suggest that stylistic variation is not only the result of the speaker’s reaction to certain elements present in the communicative situation, but an active and creative process of presenting identity (Schilling-Estes 2002).

As I have already mentioned above, in addition to discovering the correlations between linguistic use and socio-demographic categories, this dissertation has also contributed to uncovering the social meaning that is associated with each of the features under study. I have shown that at least one of the linguistic variables under study (/-o/ vs. /-u/) has second order indexicality and that listeners use the presence of Asturian or Spanish vowels to evaluate the identity of other speakers. This discovery supports the idea that linguistic variants have social meaning, that speakers are aware of the correlations between linguistic choices and social traits and that they use these correlations to construct their own identity.
In the case of Asturian Spanish, this process is necessarily linked to how speakers relate to place and the conflict between regional and urban identity. In fact, as I mentioned in the discussion of the results of the perception experiment reported in chapter 5, the rating variable on which the language of the feature had the strongest effect was the rural/urban rating. In the discussion of the results for the checkbox variables I explained that Gijón is perceived as being in the middle of the rural/urban and Asturian/Spanish continuum, showing that the use of Asturian features and Spanish ones are equally associated with the city. I propose that it is the conflict between speakers’ desire to sound Asturian but also urban that motivates the variation between Asturian and Spanish features and the use of a mixed system, in which a particular set of linguistic features are not used categorically in one or the other language.

These results contribute to our understanding of how speakers use linguistic variation to relate to place, complementing recent work in sociolinguistics (Eckert 2004, Johnstone, Andrews and Danielson 2006, Myers 2006, Johnstone and Kiesling 2008, Johnstone 2010). The findings also show that this relationship is not a straightforward one; instead, regional identity or ethnicity is intricately related to other social variables such as gender, education, occupation and economic status. Speakers of Asturian Spanish are aware of how speech in Asturias differs from that outside of the region and recognize Asturian features as markers of Asturian identity. However, the stigmatization of Asturian and its association with uneducated speech and lower economic status complicates the relationship between linguistic use and place for speakers in urban areas, since they want to distance themselves from those social categories. Thus, the conflict
between urban and regional identity stems from the fact that Asturian features are not only markers of place but also markers of low socio-economic status.

One of the innovations of this dissertation is that it addresses the relationship between linguistic use and social meaning in a situation in which the choices available to the speakers originated historically in the contact between two different languages. Because of this, the research presented here also constitutes a novel contribution to the field of contact linguistics, showing how elements from the two languages contribute differently to constructing social identity and how linguistic competition between two systems leads to social competition in the use of variants that belong to one language or the other.

6.3 Future directions

In the discussion of the effect that linguistic factors had on the variation between raised and non-raised vowels, I explained that on many occasions the distribution of the data obscured the results obtained and generalizations regarding the effect of some of those internal variables were not possible. This was the case for the effect of the stressed vowel. Because we find instances of vowel harmony in certain varieties of Asturian and in varieties of Spanish spoken in areas close to Asturias, it would be worth further exploring the effect that the stressed vowel has on the choice between Asturian and Spanish vowels. In order to do so, a study that tests for stressed vowel in a controlled phonetic environment is necessary. Similarly, the frequency of the token was not found to significantly constrain the variation. However, this result could be attributed to the
limitations of the corpus and the way frequency was determined. Using a larger corpus would allow us to have a more reliable measure of frequency and further explore its effect on the choice between Asturian and Spanish variants.

After discussing the role that linguistic factors had on the variation, one of the questions that arose was whether the speakers’ realizations of [o] and [u] were clearly different. Even though the results of the perception experiment indicate that listeners are able to hear the differences between the two back vowels, it would be interesting to conduct an experiment that tests what realizations speakers identify as [o] and which as [u]. The goal of such an experiment would be to create parameters for auditory coding that are based on the perceptions of speakers of the same linguistic variety. This type of auditory coding would eliminate the overlap that results from auditory coding based solely on the perceptions of the researcher and, at the same time, avoid the relative arbitrariness of the Discriminant Analysis of Principal Components as a method to assign the individual tokens to a particular vowel.

Among the social factors that were not found to have a significant effect on the variation is the age of the speaker. Previous research on Asturian sociolinguistics has been unanimous with regard to the correlation between age and the use of Asturian or Asturian features: older speakers use more Asturian than speakers in younger generations. Despite these observations, I did not find a significant effect of age in the quantitative analysis of production data. In the results of the perception experiment we also observed that speakers’ assignment to different age categories depending on the feature language used in the utterances was only significant for the /-as/, /-es/ variable. Based on these results I hypothesized that while age constrains the appearance of a
particular Asturian feature, for features that are widespread among speakers (such as /-u/), age does not constrain how much that feature is used. I would be interested in further testing the effect of age by using a larger sample of speakers. This would allow me to use a different categorization of age that considers more than two age groups.

The quantitative and qualitative results obtained in the analysis of production data and the perception experiment indicate that factors related to style are very important in the variation under study. Thus, further exploring style-shifting in Asturian Spanish would greatly improve our understanding of the variation. This is especially important because a large number of speakers seem to be very aware of their own stylistic variation and because some of the conclusions drawn in this dissertation are based on the existence of those style shifts. In order to fully account for the effect that style-related variables have on the variation between Asturian and Spanish features, I plan on conducting a study in which speakers wear a recording device throughout the day, allowing me to capture interaction with different interlocutors, in different settings and addressing a wider range of conversational topics.

Due to time and space constraints, the research presented in this dissertation was limited to two of the differentiating features that characterize urban Asturian Spanish and that show variation in the speech of people from Gijón. As we saw in section 2.3.2, there are other features from Asturian that are often present in speech and that show alternation with their Spanish equivalents. The differential effect of social factors on the two linguistic variables explored in this dissertation and the differences observed in the social perceptions of those variables makes the future analysis of other features necessary to fully understand the social hierarchy in which they appear to be organized.
Finally, the importance of regional and urban identity in the alternation between Asturian and Spanish features makes the study of the variation in other geographical areas that are situated at different points along the rural/urban continuum an interesting continuation to this research. In general, we would expect to see an increase in the use of the Asturian variants as we move away from the urban centers. It is also possible that the same linguistic variants have different social meaning or that listeners use different linguistic cues to evaluate the identity of other speakers.
REFERENCES


Barajas, Jennifer. 2013. Vowel raising in Michoacán, Mexico. Paper presented at the OSU Congress on Hispanic and Lusophone Linguistics, Columbus, OH.

Barnes, Sonia and Scott Schwenter. 2013. An implicational hierarchy of contact features in Asturian Spanish. Paper presented at the 24th Conference on Spanish in the United States and 9th Conference on Spanish in Contact with other Languages, McAllen, TX.


Boersma, Paul and David Weenink. 2005. Praat: doing phonetics by computer (Version 5.3.32)


Croissant, Yves. 2011. mlogit: Multinomial logit model. R package version 0.2-3, URL: http://cran.r-project.org/web/packages/mlogit/


257


263


APPENDIX A: STIMULI USED IN THE PERCEPTION EXPERIMENT

Speaker 1 – male (under 40) o/u
a) hora y media si está bien, pero como en el trozo de ahí no hay autopista, pues el viaje acaba siendo más largo
b) hora y media si está bien, pero como en el trozo de ahí no hay autopista, pues el viaje acaba siendo más largo

Speaker 2 – female (over 40) o/u
a) Dijeron que se quedaba otro mes, así que a aguantar al encargao
b) Dijeron que se quedaba otru mes, así que a aguantar al encargau

Speaker 3 - female (under 40) o/u
a) Compraron un piso muy guapo en el dos mil ocho, pero está sin amueblar
b) Compraron un pisu muy guapu en el dos mil ocho, pero está sin amueblar

Speaker 4 - male (over 40) o/u
a) Hay un compañero, que ahora está de vacaciones, que perdió un brazo con la maquinaria
b) Hay un compañeru, que ahora está de vacaciones, que perdió un brazu con la maquinaria

Speaker 5 - male (under 40) as/es
a) Todos empezaron de aquella cuando salieron tantas plazas de policía
b) Todos empezaron de aquella cuando salieron tantes places de policía

Speaker 6 – male (over 40) as/es
a) Dediqué muchas horas a la empresa y a lo mejor dediqué menos a la familia
b) Dediqué muches hores a la empresa y a lo mejor dediqué menos a la familia

Speaker 7 - female (under 40) as/es
a) Nunca puedo abrir las ventanas porque como sacuden tanto por aquí
b) Nunca puedo abrir les ventanes porque como sacuden tanto por aquí.

Speaker 8 - female (over 40) as/es
a) Fue cuando construyeron escuelas en la zona esa donde están las tiendas del centro
b) Fue cuando construyeron escueles en la zona esa donde están les tiendes del centro
APPENDIX B: QUESTIONNAIRE USED IN THE PERCEPTION EXPERIMENT

En la siguiente grabación vas a escuchar a Pablo. Dale a play para escuchar la grabación. Puedes escucharla tantas veces como necesites. Después de escucharla, responde a las siguientes preguntas sobre Pablo.

Pablo suena:
- [ ] Nada masculino  [ ] Muy masculino
- [ ] Nada inteligente  [ ] Muy inteligente
- [ ] Informal  [ ] Formal
- [ ] De pueblo  [ ] De ciudad
- [ ] Pobre  [ ] Rico

¿Qué edad crees que tiene Pablo? (escoge solo uno)
- [ ] Menos de 30
- [ ] Entre 30 y 40
- [ ] Entre 40 y 50
- [ ] Entre 50 y 60
- [ ] Más de 60

¿Qué nivel de estudios crees que tiene Pablo?
- [ ] Nunca ha ido a la escuela
- [ ] EGB/Primaria
- [ ] Secundaria/Instituto/BUP
- [ ] Bachillerato/COU
- [ ] Diplomatura universitaria
- [ ] Licenciatura universitaria
- [ ] Modulo/FP

¿Cuánto crees que conoce Pablo a la persona con la que habla?
- [ ] Casi no se conocen
- [ ] Se conocen muy bien

¿De dónde crees que es Pablo? (selecciona solo uno)
- [ ] De Gijón
- [ ] De las Cuencas
- [ ] De un pueblo de Asturias
- [ ] De Oviedo
- [ ] De fuera de Asturias

¿A qué crees que se dedica Pablo?
- [ ] Trabaja en un bar/restaurante
- [ ] Trabaja en la construcción
- [ ] Trabaja en una tienda
- [ ] Es profesor

266
□ Es abogado/médico
□ Es agricultor/ganadero

¿Qué más piensas de Pablo?