ONZEminer
a digital tool for tracing the origins and development of New Zealand English

Heidi Quinn
University of Canterbury
heidi.quinn@canterbury.ac.nz

Transformations’07, 27 August 2007
“Language is a central feature of human identity. When we hear someone speak, we immediately make guesses about gender, educational level, age, profession, and place of origin. Beyond this individual matter, a language is a powerful symbol of national and ethnic identity.”

Spolsky (1999: 181)
Present-Day New Zealand English (NZE) is characterised by a unique combination of phonological and grammatical features that make it easy for us to distinguish New Zealanders from speakers of other varieties.
cc-fyn94-12a, born 1972
I mean the guys are a lot better [...] I think women are probably have to work at it a lot harder than y'know than the guys

cc-myn02-11a, born 1981
jumped out of the car and got the fire engine out and it turned out the fuel line fell off and put hot fuel on to the engine
The first generations of English-speaking New Zealanders sounded very different.

- lots of variation between speakers
- still strongly influenced by the varieties of English spoken by the early settlers, who mainly came from England, Scotland, and Ireland, often via Australia (cf. Gordon et al. 2004)
ia-98, born 1893, Otago, Irish parents
oh and when we were going to school we used to go to the pictures [...] I can remember the magic lantern shows and we were very delighted about those

ia-70b, born 1896, Central Otago
and I don't know why I did it but I got sugar and rubbed sugar into her hair to take the stuff out of course. she was a terrible mess
The ONZE Archives

• **Mobile Unit (mu)**
  95 speakers born 1851-1919, recorded by the New Zealand Broadcasting service after WWII

• **Intermediate Archive (ia)**
  about 90 speakers mostly born 1890-1935 and recorded by Rosemary Goodyear, Leslie Evans, and UC researchers in the early 1990s

• **Canterbury Corpus (cc)**
  currently 407 speakers born 1926-1985, recorded by students at the University of Canterbury since 1994
ONZEminer

• a browser-based research tool developed by Jen Hay and Robert Fromont

• freely available to download from: http://www.ling.canterbury.ac.nz/jen/onzeminer

• software currently integrated with ONZEminer Transcriber (http://trans.sourceforge.net/)
  Praat (http://www.praat.org/)
  CELEX (http://www.ru.nl/celex/)
ONZEminer

• ONZEminer is work in progress and its search and analysis facilities are constantly being refined by programmer Robert Fromont

• the program is already saving us hours of hand-searching and has allowed us to carry out linguistic comparisons on a much larger scale than was feasible before

• ONZEminer has also made it possible for undergraduate students to become involved in departmental research projects
projects using ONZEminer


- a project on Maori English by Margaret Maclagan and Jeanette King

- Katie Drager’s PhD research on language and social identity (cf. Drager 2006)

- Alex D’Arcy’s project on regional variation in Canterbury
projects using ONZEminer

• collaborative work on *give* with Stanford University (cf. Hay & Bresnan 2006; Bresnan & Hay in press)

• joint projects with Newcastle University (UK) on phrase-final /t/ (cf. Docherty et al. 2006) and on quotatives (current work by Alex D’Arcy and Isa Buchstaller)

• interactive corpus of German dialects (currently being created by Heidi Quinn)
onze miner

- install required components for sound to work
- filter speakers
- credits
<table>
<thead>
<tr>
<th>name</th>
<th>transcripts</th>
<th>corpus</th>
<th>gender</th>
<th>yob</th>
<th>class region</th>
<th>origin mother</th>
<th>origin father</th>
</tr>
</thead>
<tbody>
<tr>
<td>fyn94-12a</td>
<td>6</td>
<td>CC</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td>48</td>
<td>ME</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>MYNO2-11a</td>
<td>5</td>
<td>CC</td>
<td>M</td>
<td>1981</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>not transcribed</td>
<td>2</td>
<td>CC</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
</tbody>
</table>

Version 20061204.1602
Copyright © 2004-2006 ONZE Project University of Canterbury, NZ
This is free software distributed under the terms of the GNU General Public License
name     gender  class  year of birth  [notes]
fyn94-12a  F    N     1972  [notes]
MYN02-11a  M    N     1981  [notes]

Text/Regular Expression: why

Only search transcripts for which these are the main speakers.
Export results to Excel (CSV) [options]
[transcript types]

search

results

Found 8 results

1.   fyn94-12a-08.trs - fyn94-12a:
1.   I don't think oh it didn't rain I'm not sure why they didn't actually.

2.   fyn94-12a-09.trs - fyn94-12a:
2.   no results

3.   fyn94-12a-10.trs - fyn94-12a:
3.   you think like that that's why I li
<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Class</th>
<th>Year of Birth</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>fyn94-12a</td>
<td>F</td>
<td>N</td>
<td>1972</td>
<td>[notes]</td>
</tr>
<tr>
<td>MYN02-11a</td>
<td>M</td>
<td>N</td>
<td>1981</td>
<td>[notes]</td>
</tr>
</tbody>
</table>

**Text/Regular Expression:**

why

- **Only search transcripts for which these are the main speakers.**
- **Export results to Excel (CSV)**

**Microsoft Excel - results_why[1].csv**

<table>
<thead>
<tr>
<th>Number</th>
<th>Transcript</th>
<th>Speaker</th>
<th>Gender</th>
<th>Year of Birth</th>
<th>class</th>
<th>Sync</th>
<th>URL</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>fyn94-12a</td>
<td>fyn94-12a</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td>271.869</td>
<td><a href="http://local">http://local</a></td>
<td>I don't think oh it didn't rain I'm not sure why they didn't actually .</td>
</tr>
<tr>
<td>2</td>
<td>fyn94-12a</td>
<td>fyn94-12a</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td>285.174</td>
<td><a href="http://local">http://local</a></td>
<td>you think like that that's why I li</td>
</tr>
<tr>
<td>3</td>
<td>fyn94-12a</td>
<td>fyn94-12a</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td>119.929</td>
<td><a href="http://local">http://local</a></td>
<td>I don't know why . she was happy</td>
</tr>
<tr>
<td>4</td>
<td>fyn94-12a</td>
<td>fyn94-12a</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td>262.64</td>
<td><a href="http://local">http://local</a></td>
<td>I mean I had a awful birth that's probably why I think that</td>
</tr>
<tr>
<td>5</td>
<td>fyn94-12a</td>
<td>fyn94-12a</td>
<td>F</td>
<td>1972</td>
<td>N</td>
<td>4.254</td>
<td><a href="http://local">http://local</a></td>
<td>but it's why take that risk though it's you and a baby .</td>
</tr>
<tr>
<td>6</td>
<td>myn02-11a</td>
<td>MYN02-11M</td>
<td>M</td>
<td>1981</td>
<td>N</td>
<td>36.356</td>
<td><a href="http://local">http://local</a></td>
<td>quite understand why it stalled</td>
</tr>
<tr>
<td>7</td>
<td>myn02-11a</td>
<td>MYN02-11M</td>
<td>M</td>
<td>1981</td>
<td>N</td>
<td>175.69</td>
<td><a href="http://local">http://local</a></td>
<td>yeah . why ?</td>
</tr>
<tr>
<td>8</td>
<td>myn02-11a</td>
<td>MYN02-11M</td>
<td>M</td>
<td>1981</td>
<td>N</td>
<td>176.28</td>
<td><a href="http://local">http://local</a></td>
<td>why would you want a Holden ?</td>
</tr>
</tbody>
</table>
I don't know why she was happy.

Thanks to The Stanford Natural Language Processing Group

hoping that it would be a girl

cos I was thinking oh it could be a girl.
layered search

speakers

name          gender class year of birth [notes]
fyn94-12a     F     N      1972 [notes]
MYN02-11a     M     N      1981 [notes]

layers

☐ ab morphology
☐ ab syntax
☑ ab phonology
☐ orthography
☑ transcript

search across 1 words.

search regular expressions

☑ anything phonology [hint] followed by anything

☑ Only search transcripts for which these are the main speakers.
☐ Only show results from the first 5 transcripts.

[transcript types]

search
<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>check</th>
<th>type</th>
<th>dialect</th>
<th>accent</th>
<th>scope</th>
<th>layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>spk1</td>
<td>myn02-11a</td>
<td>no</td>
<td>native</td>
<td>local</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**transcript**

myn02-11a

**Wordlist**

hit  

1. hit  

hid
changes involving consonants

• loss of rhoticity and rise of linking & intrusive /r/ (cf. Hay & Sudbury 2005)

  ia-98 cc-myn02-11a
  magic lantern turned fire engine

• loss of /hw/ (cf. Schreier et al. 2003)

  why ia-70b cc-fyn94-12a
changes involving vowels

• closing diphthongs  
  (cf. Maclagan et al. 1999; Hay et al. in press)  
  now ia-98 cc-fyn94-12a

• the NEAR/SQUARE merger  
  (cf. Maclagan & Gordon 1996; Gordon & Maclagan 2001,  
  Langstrof 2004)  
  here hair ia-98 cc-fyn94-12a
changes involving vowels

• raising of TRAP and DRESS (Maclagan et al. 1999; Gordon et al. 2004; Maclagan & Hay 2007)

• centralisation of KIT (Gordon et al. 2004; Langstrof 2006)

• diphthongisation of FLEECE (Maclagan & Hay 2007)

The changes to TRAP, DRESS and KIT are illustrated in the vowel plot on the next slide.
Changes in NZE front vowels

note: The examples for the younger speaker come from a wordlist, whereas the tokens for the older speaker were produced in connected speech. This exaggerates the differences between their vowels.
Spectrograms for ‘hit’

ia-70b ‘hit’

cc-myn02-11a ‘hit’
grammatical changes

• subject-verb agreement with *be* (cf. Hay & Schreier 2004)

the volunteers *was* getting a little hard pressed (mu87a, male, born 1888)

there *was* just *heaps of shops* ’n you know both sides of the road (cc-fyn94-12a)
grammatical changes

• the use of (HAVE) *got* in present tense possessives (cf. Quinn 2004 & to appear)

now I *have* two young children (ia-70b)

I’ve *got* a soft spot for them (cc-myn02-11a)

he’s *got* the size for hockey (cc-fyn94-12a)
older CC = aged 45-60 when interviewed
younger CC = aged 20-30 when interviewed
the future of NZE

• in the 2006 census, 14.6% of New Zealanders identified themselves as Maori, and 7% as Pacific Islanders

• more than a third of the Maori and Pacific Island population were under the age of 15

(Statistics New Zealand 2007)
the future of NZE

• young New Zealanders with Maori and Pacific Island backgrounds are increasingly using English to express their ethnic identity (cf. Taumoeefolau et al. 2002; Starks & Reffell 2005 & 2006; Kennedy 2006; Benton 2007; Szakay 2006 & 2007; Hay et al. in press; Bell & Gibson forthcoming)

• you don’t have to be of Maori or Pacific Island descent to speak one of these new varieties of English
• **speaker 1**

  cos I mean in recent years yous have had a pretty s pretty solid score specially in prems like yous have done pretty good

• **speaker 2**

  eye for an eye tooth for a tooth [...] there’s a natural sort of connection with the church family

• **speaker 3**

  it sucks getting the sports results before you even see the game [...] always you can’t miss it when it’s on the front page

**note:** all three speakers come from Margaret Maclagan and Jeanette King’s Maori English corpus
Acknowledgements

Special thanks to:
• Robert Fromont, Jen Hay, and Katie Drager for help with ONZEminer, Praat, and vowel analysis
• Margaret Maclagan and Jeanette King for giving me access to their Maori English corpus
• Liz Gordon for drawing my attention to recent demographic data and for help with selecting speakers to demonstrate ONZEminer on
• Alex D’Arcy, Kate Kearns, and Andrew Carstairs-McCarthy for valuable comments on an earlier version of this talk
References


Bresnan, Joan and Jennifer Hay. in press. Gradient grammar: an effect of animacy on the syntax of 'give' in varieties of English. Lingua.


