A phonological variable in a textual medium: (ing) in online chat

Aaron J. Dinkin, University of Toronto
ajd@post.harvard.edu

“Instant messaging” (IM) is defined by the following properties (cf. Baron 2008):
• text-based, not spoken
• usually one-to-one communication
• Internet-mediated rather than e.g. transmitted by the telephone system
• expectation of real-time tempo of interaction

These make IM language a candidate for the written equivalent of vernacular speech—the natural, unmonitored style used with close acquaintances (cf. Labov 2006)—and therefore an ideal target for sociolinguistic analysis of written language.

Previous variationist analyses of IM language:
Squires (2012) on apostrophes: don’t vs. dont, etc.
• As expected, men favor the nonstandard variant of omitting apostrophes.
Tagliamonte & Denis (2008) on a variety of variables:
• Lexical/morphosyntactic variables: quotatives, deontic modality, intensifiers, &c:
  • IM has a higher rate of standard and conservative variants than speech
  • IM also has a greater diversity among how many variants are frequently used
• Orthographic variables: lowercase i for I; single-letter u for you
• Most users near-categorical one way or the other; little intra-user variation

Orthographic variables studied by Squires and T&D have no spoken equivalent.

The (ing) variable apparently exists in both spoken and written form—/ɪŋ/ vs. /ɪn/ in speech; -ing vs -in in writing
(with some additional minor variants in each).

But what is the relationship between the spoken and written versions of the variable?
Is the written use of -ing vs. -in controlled by the phonological process?
Or is it handled like other purely orthographic variables, independent of phonology?

Examining variation in (ing) in IM language thus may illuminate the nature of the relationship between speech and writing.

Consideration:

Data was collected by 22 first-year undergrads at University of Toronto as an assignment for a “Language and the Internet” seminar:
each was assigned to collect at least 1,000 words of one-on-one IM conversations
between themselves and similarly-aged peers, removing only identifying names.

Total size of corpus: about 22,000–23,000 words;
54 distinct IM chat participants:
• 30 male, 24 female
• mean age 18.5; age range 18–25; 34 aged 18
• variety of native languages: English (28), Mandarin (9), Hindi/Urdu (7);
  Cantonese, Gujarati, Spanish, Tagalog (2 each); Assyrian, Japanese (1 each)
• 34 grew up in Canada
other countries include: India, China, US, Taiwan, Pakistan, Philippines, et al.

Overall results:
(ing) is variable in IM, but -in variant is very infrequent:
• Only 17 tokens of -in per se, e.g.:
  H: I’M FREAKIN OUT
  VC: so we thinkin of reaching buffalo wild wings
• 4 tokens of alternate variants with no g:
  O: chillin like a villain, you?!
  O: haha fucken [–1]' fucken fuck i gotta get dressed msg me later have fun in class!
• 634 tokens of (ing) in corpus:
  • Tokens with obvious typos were included where interpretable (e.g., goinb for going)
  • Plurals of -ing nouns included (7 tokens, e.g. L: so do u have a lot of readings to do?)
  • Futural going to included (27 tokens); gonna not included (55 tokens)

Each participant is labeled by a 1- or 2-letter code.

634 tokens of (ing) in corpus:
• Tokens with obvious typos were included where interpretable (e.g., goinb for going)
• Plurals of -ing nouns included (7 tokens, e.g. L: so do u have a lot of readings to do?)
• Futural going to included (27 tokens); gonna not included (55 tokens)
• One seeming hypercorrection included (RB: its gonna be taking care of next week)

All participants but one produced at least one (ing) token.

This is really low! …isn’t it?

lowest rate of /ɪn/ use in any subgroup of Wagner’s data is over 30%.
So yes, 3.3% -in does seem really low compared to speech.

Tagliamonte & Denis (2008) find rate of nonstandard variants lower in IM than speech,
this data is consistent with that finding, but still seems unexpectedly low.
E.g., they find be like about one third as common in IM as in speech,
but it’s far from being as marginal in IM as -in seems to be.

1 Personal name redacted.
2 Except where otherwise noted, from here on “-in” will include -in-like variants as well.
Out of 53 IM participants,
- 42 never produced -\textit{in}
- 8 produced exactly one token of -\textit{in}
- 3 produced 4–5 tokens of -\textit{in} each (participants O, V, and ZB)

(V is the only one to use -\textit{in} more than half the time: 5 out of 7 tokens.)
This is consistent with Tagliamonte & Denis’s finding of more \textit{inter-user} variation than
\textit{intra-user} orthographic variation: the majority of participants are \textit{categorical}.
Their analysis: most individuals use orthographic features to establish a consistent \textit{personal style}, rather than as active sociolinguistic variables in their own usage.
That seems to be the case with (ing) as well, but -\textit{in} is so marginal that \textit{no one} uses it
categorically.

The per-user \textit{distribution of u vs. you} in this corpus is \textit{very similar} to T&D’s
(despite an overall \textit{higher total rate} of \textit{u}': 26% here vs. 9% in their data).
So the low rate of -\textit{in} seems more likely to be a \textit{fact about -in} in particular than about
nonstandard orthographic variants in this corpus overall.

-\textit{in} is found in \textit{multiple grammatical contexts}:
  - \textit{most frequent} in progressive verbs (e.g., H: \textit{I'M FREAKIN OUT}): 15/307 tokens (5%)
  - Anyone who used -\textit{in at all} used it for a progressive.
  - Other grammatical contexts: 6 -\textit{in} tokens out of 327 non-progressives (1.8%):
    - gerund (ZB: \textit{Solving for the sake of solvin})
    - monomorphic noun (ZB: \textit{I won’t be out till evenin then that ok?})
    - something (ZB: \textit{Probably by the end of the week or somethin})
    - misc.: \textit{fucking} (O: see above); TV show title (V: \textit{u finish breakin bad?})
Per $\chi^2$ test, difference between progressives and everything-else is significant ($p < 0.05$).

\textbf{Logistic regression} on writer gender, recipient gender, writer native language and
grammatical context finds all but writer gender significant\textsuperscript{4}.

<table>
<thead>
<tr>
<th></th>
<th>writer gender</th>
<th>native language</th>
<th>grammatical</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>.707</td>
<td>.935</td>
<td>.625</td>
</tr>
<tr>
<td>female</td>
<td>.278</td>
<td>.606</td>
<td>.382</td>
</tr>
<tr>
<td>Indic</td>
<td>.310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>.264</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chinese speakers disfavor -\textit{in}; Spanish speakers (in the “misc” category) favor -\textit{in}.
Walker (2013) found parallel results in \textit{speech} for heritage speakers of Chinese and
Romance languages, due to the substratal role of [\textit{\textasciitilde}] in those languages’ phonologies.

The fact that similar patterns are found for (ing) in speech and IM suggests that spoken
and written (ing) are the same variable after all.

Why then is the rate of -\textit{in} in IM so low?
Possible explanation: (ing) \textit{really is a phonological variable}, not an orthographic one.
I.e., perhaps in \textit{typing}, the phonological grammar can simply be (partly!) \textit{bypassed},
since no phonemic/phonological implementation is actually going to take place.
It’s possible to phonologically process things as you type, but not necessary.

In that case, the (ing) variable usually \textit{doesn’t even get activated} in IM messaging.

So morphosyntactic and lexical variables show robust variation in both IM and speech;
\textit{natively orthographic} variables are actively available for IM users to construct style;
but a \textit{phonological} variable like (ing) remains \textit{primarily a feature of speech},
even though it has a conventional orthographic representation.

\textbf{References}:
Baron, Naomi S. (2008). \textit{Always On: Language in an Online and Mobile World}. Oxford University
Press, New York.
Squires, Lauren (2012). “Whos punctuating what? Sociolinguistic variation in instant messaging”. In
Alexandra Jaffe, Jannis Androustopoulos, Mark Sebba, & Sally Johnson (eds.), \textit{Orthography as
language”. \textit{American Speech} 83:1–34.
Wagner, Suzanne Evans (2012). “Real-time evidence for age grad(ing) in late adolescence”.
Walker, James. (2013). “Ethnonoects at the intersection of phonological variables: Velar nasals in
Toronto English”. Paper presented at the LSA annual meeting, Boston.

\textsuperscript{4} Grammatical category does not remain significant when coded other than progressive vs. everything
else, however.

\textsuperscript{3} This includes \textit{u}, \textit{ur}, and \textit{arself} vs. \textit{you}, \textit{your}, \textit{you're}, and \textit{yourself};