Frequency Effects and the Lexical Split in the Use of [t], [s], [d], and [z] in Syrian Arabic

Rania Habib
Syracuse University
rhabib@syr.edu
LSA, January 9, 2009
1. Introduction

This study deals with the variable use of \([t]\) and \([s]\) in place of the Standard Arabic (SA) \([θ]\) and \([d]\) and \([z]\) in place of the SA \([ð]\) in the colloquial Arabic of rural migrant speakers to the city of Hims in Syria.
### Study variables

Table 1. Variants of the variables ($\Theta$) and ($\delta$) in Standard Arabic, the speech of non-migrant rural speakers, native Himsi speakers, and migrant rural speakers

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variants of Standard Arabic (SA)</th>
<th>Variants of non-migrant rural speakers (RCA)</th>
<th>Variants of native Himsi speakers (HCA)</th>
<th>Variants of migrant rural speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\Theta$)</td>
<td>[Θ]</td>
<td>[t] ~ [s]</td>
<td>[t] ~ [s]</td>
<td>[t] ~ [s]</td>
</tr>
<tr>
<td>($\delta$)</td>
<td>[ð]</td>
<td>[d] ~ [z]</td>
<td>[d] ~ [z]</td>
<td>[d] ~ [z]</td>
</tr>
</tbody>
</table>
1. Introduction

- According to Daher (1999) the use of [s] and [z] is more socially elevated than [t] and [d] in Damascene Arabic.

- Thus, highly educated speakers may produce [s] and [z] when aiming at producing the borrowed SA variants [Θ] and [ð] respectively.
1. Introduction

- In my original data, all speakers behave similarly regardless of their age, gender, social, class, education, occupation, and residential area.

- There is a puzzling *stable lexical split phenomenon* – two lists of words for each variable – some words are specifically used with [t] and others used with [s] in place of [θ], for example.

- Interchangeability does not exist; hence, *no variation*.
1. Introduction

• Such variability was traditionally explained in terms of two historical changes:

1. A complete merger between the SA interdental fricatives, [θ] and [ð], and the stops, [t] and [d] respectively, around the 14\textsuperscript{th} century (Schmidt 1974: 94; Daher 1998, 1999).

2. Sometime after the 14\textsuperscript{th} century, some [θ] and [ð] variants in borrowed words from SA were replaced in colloquial speech with [s] and [z] respectively (Birkeland 1952; Schmidt 1986: 57; Schulz 1981: 33; Daher 1999).
1. Introduction

- Markedness may have played a role historically in both changes:
  - The merger with the stops [t] and [d] involves reduction of the features [+Distributed] and [+Continuant].
  - The use of [s] and [z] involves reduction of the feature [+Distributed].
1. Introduction

- Today, markedness no longer plays a role in this variable use.

- The sounds /θ/ and /ð/ are absent from speakers’ phonemic inventory. Each of the four variants has its identical underlying form (e.g., /t/ is the underlying form for [t]).
1. Introduction: Linguistic Environment

- All four variants could occur in similar environments.

- The linguistic environment does not play a role in determining which variant should be used and when.

- Each of the variants is a separate phoneme; it is the underlying form for its surface form (i.e., the underlying form for [t] is /t/ and so on).
2. Research Questions

The study seeks to answer the questions:

1. How did the present puzzling lexical split phenomenon came to exist?

2. Why can’t markedness and the two mentioned historical changes provide complete explanation of this phenomenon?
3. Data

- My original corpus of words that contain the four variants [t] and [s] and [d] and [z] is taken from the naturally occurring speech of 52 rural migrant speakers to the city of Hims in Syria.

- The corpus consists of
  - 2767 words produced with [t]
  - 870 words produced with [s]
  - 1740 words produced with [d]
  - 1257 words produced with [z].
### 3.1. Examples of the stable use of [t] and [s]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ktiir</td>
<td>Much/a lot</td>
<td>1159</td>
<td>masalin/masalan</td>
<td>For example</td>
<td>530</td>
</tr>
<tr>
<td>2</td>
<td>mitl</td>
<td>Like/as</td>
<td>288</td>
<td>saaʔir</td>
<td>A proper noun</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>?aktar/?aktor</td>
<td>More</td>
<td>245</td>
<td>ḥadiis</td>
<td>Modern/talk</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>tlat</td>
<td>Three</td>
<td>153</td>
<td>ḥaadis</td>
<td>Accident</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>taanee/teenee</td>
<td>Second</td>
<td>149</td>
<td>saanawiyyii</td>
<td>High school</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>tlaati</td>
<td>Three</td>
<td>123</td>
<td>musaqqaf</td>
<td>Educated</td>
<td>12</td>
</tr>
</tbody>
</table>
3.2. Examples of the stable use of [d] and [z]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>haadaa</td>
<td>This (M)</td>
<td>607</td>
<td>?izaaz</td>
<td>If</td>
<td>567</td>
</tr>
<tr>
<td>hadool</td>
<td>These</td>
<td>135</td>
<td>kazaa</td>
<td>So/like this</td>
<td>231</td>
</tr>
<tr>
<td>haadii</td>
<td>This (F)</td>
<td>116</td>
<td>tzakkar</td>
<td>Remember (V)</td>
<td>35</td>
</tr>
<tr>
<td>?a kad</td>
<td>Took</td>
<td>114</td>
<td>lizaalik</td>
<td>For this reason</td>
<td>29</td>
</tr>
<tr>
<td>hadiik</td>
<td>That (F)</td>
<td>67</td>
<td>?staaz</td>
<td>Teacher/professor</td>
<td>27</td>
</tr>
<tr>
<td>?a kad</td>
<td>Take</td>
<td>59</td>
<td>kizb</td>
<td>Lying</td>
<td>26</td>
</tr>
</tbody>
</table>
4. Quantitative analysis: Frequency effects

- Since social factors and linguistic environment do not play a role in this stable phenomenon, I explored the role of frequency in the use of each variant.

- Similar words from all 52 speakers are grouped together.

- Highly frequent words (whose frequency is 10 or higher) are compared between corresponding variants, i.e., between [t] and [s] and between [d] and [z].
4.1. Frequency effects on the use of [t] and [s]

Table 2. Most frequent words produced with [t] and [s]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ktiir</td>
<td>Much/a lot</td>
<td>1159</td>
<td>masalin/masalan</td>
<td>For example</td>
<td>530</td>
</tr>
<tr>
<td>mitl</td>
<td>Like/as</td>
<td>288</td>
<td>saaʔir</td>
<td>A proper noun</td>
<td>19</td>
</tr>
<tr>
<td>?aktar/?aktor</td>
<td>More</td>
<td>245</td>
<td>ḥadiis</td>
<td>Modern/ talk</td>
<td>18</td>
</tr>
<tr>
<td>tlat</td>
<td>Three</td>
<td>153</td>
<td>ḥaadis</td>
<td>Accident</td>
<td>16</td>
</tr>
<tr>
<td>taanee/teenee</td>
<td>Second</td>
<td>149</td>
<td>saanawiiıyii</td>
<td>High school</td>
<td>13</td>
</tr>
<tr>
<td>tlaati</td>
<td>Three</td>
<td>123</td>
<td>musaqqaqaf</td>
<td>Educated</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 3. Contingency table showing the frequency use of [t] and [s] words and their conditional distribution in all speakers

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Freq &lt; 10 &amp; %</th>
<th>Freq &gt; 10 &amp; %</th>
<th>Total &amp; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]-words</td>
<td>96 (3%)</td>
<td>2671 (97%)</td>
<td>2767 (100%)</td>
</tr>
<tr>
<td>[s]-words</td>
<td>210 (24%)</td>
<td>660 (76%)</td>
<td>870 (100%)</td>
</tr>
<tr>
<td>Total &amp; %</td>
<td>306 (8%)</td>
<td>3331 (92%)</td>
<td>3637</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of [t]-words between high and low frequency

Figure 2. Distribution of [s]-words between high and low frequency
4.1. Frequency Effects on the use of [t] and [s]

- 23/50 words produced with [t] have very high frequency.

- Only 11/109 words produced with [s] have high frequency.

- Two-tailed one-sample t-tests for [t] vs. [s] showed that frequency is highly significant ([t]-words frequency:  \( t = 2.294, df = 60, p = 0.025 \); [s]-words frequency:  \( t = 2.334, df = 109, p = 0.021 \)).
### 4.2. Frequency effects on the use of [d] and [z]

Table 3. Most frequent words produced with [d] and [z]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>haadaa</td>
<td>This (M)</td>
<td>607</td>
<td>?izaa</td>
<td>If</td>
<td>567</td>
</tr>
<tr>
<td>hadool</td>
<td>These</td>
<td>135</td>
<td>kazaa</td>
<td>So/like this</td>
<td>231</td>
</tr>
<tr>
<td>haadii</td>
<td>This (F)</td>
<td>116</td>
<td>tzakkar</td>
<td>Remember (V)</td>
<td>35</td>
</tr>
<tr>
<td>?axad</td>
<td>Took</td>
<td>114</td>
<td>lizaalik</td>
<td>For this reason</td>
<td>29</td>
</tr>
<tr>
<td>hadiik</td>
<td>That (F)</td>
<td>67</td>
<td>?staaz</td>
<td>Teacher/professor</td>
<td>27</td>
</tr>
<tr>
<td>?axd</td>
<td>Take</td>
<td>59</td>
<td>kizb</td>
<td>Lying</td>
<td>26</td>
</tr>
</tbody>
</table>
Table 4. Contingency table showing the frequency use of [d] and [z] words and their conditional distribution in all speakers

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Freq &lt; 10 &amp; %</th>
<th>Freq &gt; 10 &amp; %</th>
<th>Total &amp; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>[d]-words</td>
<td>100 (6%)</td>
<td>1640 (94%)</td>
<td>1740 (100%)</td>
</tr>
<tr>
<td>[z]-words</td>
<td>255 (20%)</td>
<td>1002 (80%)</td>
<td>1257 (100%)</td>
</tr>
<tr>
<td>Total &amp; %</td>
<td>355 (12%)</td>
<td>2642 (88%)</td>
<td>2997</td>
</tr>
</tbody>
</table>
Figure 3. The distribution of [d]-words between high and low frequency

Figure 4. The distribution of [z]-words between high and low frequency
4.2. Frequency effects on the use of [d] and [z]

- 28/64 words produced with [d] have very high frequency.

- Only 12/138 words produced with [z] have high frequency.

- Two-tailed one-sample $t$-tests for [d] vs. [z] showed that frequency is highly significant ([d]-words frequency: $t = 2.762$, $df = 64$, $p = 0.007$; [z]-words frequency: $t = 2.063$, $df = 137$, $p = 0.041$).
5. Discussion

- The quantitative analysis shows that frequency plays a significant role in the current lexical split.

- It shows that highly frequent words are produced with [t] and [d] and infrequent words with [s] and [z].

- This analysis enables us to make predications about what may have happened diachronically.
5. Discussion

- It is puzzling that the words that are produced with the fricative sounds, [s] and [z], are higher in number than those produced with the stops, [t] and [d].

- If the former words are borrowed from SA, we would expect a smaller number of words to be borrowed.

- The data contradict that. The data give the impression that there has been some kind of influx of SA borrowings into the dialects, but this influx was stopped at a certain point in time by an opposing force.

- What is this opposing force?
5. Discussion

- If we take markedness alone into account, we would expect to have more words pronounced with [t] and [d] than with [s] and [z] because fricatives are more marked than stops cross-linguistically.

- One would expect more words produced with [t] and [d] not only based on markedness but also due to the complete merger that happened in the 14th century between the SA interdentals [θ] and [ð] and the stops [t] and [d] respectively.
5. Discussion

- We find more words produced with the fricatives [s] and [z] as a result of the second historical change that occurred sometime after the 14th century.

- However, this change did not seem to affect certain words that are produced with [t] and [d] (i.e., the most frequent words of all).
Discussion

- To explain the current stable lexical split phenomenon, we may have to follow one of two possibilities:

1. The first possibility will be to refute the theory of a complete merger in the 14\textsuperscript{th} century between stops and interdentals. That is, traces of interdentals continue to surface, although not as interdentals, rather as fricatives.

- Refuting the first theory leads to refuting the second theory of a second change after the 14\textsuperscript{th} century, which introduced the fricatives in place of the interdentals.
Discussion

- Refuting the second change suggests that the first change only affected highly frequent words on the basis of usage- or exemplar-based models (Bybee 2001; Pierrehumbert 2001).

- These models propose that words of higher frequency are more prone to reduction.

- In this sense, highly frequent words, such as *haadaa* and *ktiir*, underwent reduction of the SA features [+Continuant] and [+Distributed], but the least frequent words did not. However, there are no traces of interdentals in natural speech.
Discussion

- In addition, refuting the first historical change does not explain the two lists of words with fricatives, such as masalan and šizaa.

- These two lists also show reduction of the feature [+Distributed]. The few high frequency words with the fricatives do not behave like other highly frequent words that have also reduced the [+Continuant] feature.

- Why would some frequent words resist one reduction and not the other? In other words, why would some frequent words show reduction of two features and others show reduction of only one feature?
5. Discussion

- Furthermore, if there was not a complete merger, we would expect some kind of variation and probably a balanced number of words with the stops and the fricatives.

- We actually have no variation and continue to have a larger number of words produced with fricatives than with the stops.

- It seems this possibility is not borne out.
Discussion

2. The second possibility is that frequency has two main opposing effects (Bybee 2001: 11):

- In one effect of frequency “phonetic change often progresses more quickly in items with high token frequency”. Thus, high frequency words are more likely to undergo change, mainly reduction.

- For example, the deletion of [t] and [d] word-finally post-consonantally is more common in English in highly frequent words, such as *went, just, and and*, than in infrequent words.
Discussion

- This first frequency effect proposed by Bybee (2001) can explain the first historical change towards the use of [t] and [d] in place of [وث] and [دو] respectively.

- The SA interdentals were reduced to full stops; the features [+Distributed] and [+Continuant] were lost.

- This merger apparently happened gradually, affecting the most frequent words first and then the least frequent ones.
Discussion

- The second effect of token frequency contradicts the first effect, because “it makes items more resistant to change” (Bybee 2001: 12).

- In other words, high frequency words become resistant to a new swiping change because they have already developed an automatized production and become entrenched in their own phonemic representations and easier to access.
5. Discussion

- This second effect of frequency can account for the existence of [s]- and [z]-words alongside the [t]- and [d]-words.

- It seems that the second historical change did not affect the most frequent words in the dialects (i.e., words that are produced with the stops, [t] and [d]).

- The highly frequent words produced with the stops were not affected by the second historical change because they became entrenched in their own phonemic representation.
6. Conclusion

- Because the stable lexical split phenomenon presented in this study is not socially conditioned, frequency effects were explored against the two diachronic changes that led to the replacement of the SA [Θ] and [ð] with [t] and [d] first and later with [s] and [z] respectively.

- The findings of this study support Bybee’s (2001) usage-based theory, which provides explanation of the current stable phenomenon of the use of [t] and [s] and [d] and [z].
6. Conclusion

- The results showed that the current lexical split is due to two opposing frequency effects.

- The first frequency effect led to the first diachronic change and the merger with \([t]\) and \([d]\).

- The second frequency effect made highly frequent words resistant to the new swiping change: the use of \([s]\) and \([z]\) in place of \(\Theta\) and \(\delta\) in borrowed words from SA.

- That is why we observe the use of \([t]\) and \([d]\) in highly frequent words and \([s]\) and \([z]\) in less frequent words.
Thank You!

Rania Habib
Syracuse University
rhabib@syr.edu