The Rise and Fall of Homophones: A Window to Language Evolution

Jinyun Ke & Feng Wang
Language Engineering Laboratory
City University of Hong Kong
Hong Kong, China

Christophe Coupé
Laboratoire Dynamique de Langage
Institut des Sciences de l’Homme
Lyon, France

The rise of homophones:
- limitation of phonological resources

To avoid ambiguity, self-organization in the language system:
- disyllabification
- differentiation in grammatical classes
- differentiation in frequency levels

Computational modeling of homophone evolution:
- with the help of context, a high degree of homophony can be tolerated
Rise of homophones

The limitation of the phonological resources

- limited inventory of sounds (phonemes):
  - a large one: 15 vowels, 57 consonants, 8 tones (Yao language in China), possible 6840 distinctive CV syllables
  - however, systematic gaps and accidental gaps decrease the resources.

- limited number of categories in a space
  - eg. plain vowel space, no more than 5 categories in either high-low or back-front dimension.
  - categorical perception; the magical number 7. (G. Miller, 1956)
  - sound change is very likely to produce homophones. eg. vowel shift: the vowel often changes from one category to another, and two forms merge, eg. meat & meet
Hypothesis I:

The smaller the sound inventory, the larger the number of homophones.
Phonological inventory size vs degree of homophony in Chinese dialects

corr= -0.76

<table>
<thead>
<tr>
<th>taiyuan</th>
<th>wuhan</th>
<th>chengdu</th>
<th>yangzhou</th>
<th>hefei</th>
<th>changsha</th>
<th>suzhou</th>
<th>shuangfeng</th>
<th>wenzhou</th>
<th>jinan</th>
<th>xian</th>
<th>niang</th>
<th>beiing</th>
<th>jian'ou</th>
<th>meixian</th>
<th>yangjiang</th>
</tr>
</thead>
<tbody>
<tr>
<td>828</td>
<td>870</td>
<td>938</td>
<td>947</td>
<td>976</td>
<td>981</td>
<td>999</td>
<td>1001</td>
<td>1048</td>
<td>1063</td>
<td>1084</td>
<td>1111</td>
<td>1125</td>
<td>1241</td>
<td>1304</td>
<td>1319</td>
</tr>
<tr>
<td>15167</td>
<td>14072</td>
<td>12342</td>
<td>12208</td>
<td>11238</td>
<td>14203</td>
<td>12577</td>
<td>11307</td>
<td>14217</td>
<td>10387</td>
<td>9872</td>
<td>9252</td>
<td>11076</td>
<td>10641</td>
<td>7923</td>
<td>6894</td>
</tr>
</tbody>
</table>

C.C. Cheng: CCLANG; Hanyu Fangyan Zihui (Dictionary of Chinese Dialects)
Homophone existence

English:

Among the 5010 most frequent words in the Brown corpus, 998 words have homophones, eg. to, too, & two.


Chinese:

- In Modern Chinese Dictionary 現代漢語詞典 (1985), 80% of the monosyllables have homophones, and 55% of them are shared by five or more morphemes.

- An extreme case: syllable “yi4” has more than 90 homophones.
Fall of homophones: 
Self-organization to avoid ambiguity

- Di-syllabification: meanings represented by monosyllabic morphemes are expressed by words which combine several monosyllabic morphemes.

  eg. jian4  \(\rightarrow\) kan4 jian4

  见   看   见

  “see”  \(\rightarrow\) “look” “see”
Hypothesis II

more homophony

more disyllabification
Number of homophones vs degree of disyllabification

corr = 0.68
Differentiation in grammatical classes

wait (vi, vt, n) & weight (n, vt)

- Most of the words are of multiple grammatical classes
- 70.3% of homophones share at least one grammatical class
- When only considering the most frequent usage, the pairs of homophone words sharing the same grammatical class drops to 40%.

wait (vi) & weight (n)
Differentiation in frequency levels

<table>
<thead>
<tr>
<th>Distances of freq. ranking</th>
<th>Homophone pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10</td>
<td>36.91%</td>
</tr>
<tr>
<td>100</td>
<td>38.36%</td>
</tr>
<tr>
<td>200</td>
<td>8.55%</td>
</tr>
<tr>
<td>300</td>
<td>4.36%</td>
</tr>
<tr>
<td>400</td>
<td>4.36%</td>
</tr>
<tr>
<td>500</td>
<td>3.64%</td>
</tr>
<tr>
<td>&gt;500</td>
<td>3.82%</td>
</tr>
</tbody>
</table>

The two pairs closest in frequency ranking

- **their & there**: rank 42 & 40;
- **weight & wait**: rank 455 & 462;

(The lowest rank in the whole word list is 555)
<table>
<thead>
<tr>
<th>Distances of freq. ranking</th>
<th>Homophone pairs</th>
<th>Randomly selected pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10</td>
<td>36.91%</td>
<td>70.00%</td>
</tr>
<tr>
<td>100</td>
<td>38.36%</td>
<td>26.18%</td>
</tr>
<tr>
<td>200</td>
<td>8.55%</td>
<td>2.00%</td>
</tr>
<tr>
<td>300</td>
<td>4.36%</td>
<td>1.09%</td>
</tr>
<tr>
<td>400</td>
<td>4.36%</td>
<td>0.18%</td>
</tr>
<tr>
<td>500</td>
<td>3.64%</td>
<td>0.00%</td>
</tr>
<tr>
<td>&gt; 500</td>
<td>3.82%</td>
<td>0.55%</td>
</tr>
</tbody>
</table>

The frequency distance distribution of the homophone pairs is not a spurious phenomenon.
Replacing words homophonous to taboo words by synonyms.

eg. Bloomfield’s examples:

- In American English, rooster and donkey are replacing cock and ass, as the latter two words are homophonous with words of body parts. “In such cases there is little real ambiguity, but some hearers react nevertheless to the powerful stimulus of the taboo-word, having called forth ridicule or embarrassment, the speaker avoids the innocent homonym.”

(--Language 1933)
Homophony arises due to the limitation of phonological resources, to satisfy a large lexical need.

One reason for a lexicon to tolerate a high degree of homophony is the help of context during communication.
Naming Games (Steels, 1996)

Simulate the interaction between agents communicating meanings with utterances.

A number of meanings
A number of distinctive utterances
Creating new words
learn new words

empty lexicon $\rightarrow$ a set of shared associations: (meaning, utterance)
The effect of meaning/utterance ratio
Success of interaction
Percentage of homophones
Percentage of synonyms
Coherence

meaning/utterance = 3
Two-word communication:

eg: “flower” “vase”

[fa] [pin]

“computer” “screen”

[kom] [pin]
The effect of context

two-word communication
two-word communication

meaning/utterance = 3
Homophony is an unavoidable phenomenon due to the limitation of phonological resources.

The language system self-organizes in various ways to decrease the possibility of confusion implied by homophones:

- disyllabification
- differentiation in grammatical classes
- differentiation in frequency levels
Implication to the study of language evolution

Language evolves in a self-organizing way. Individuals only focus on their own communication need (effectiveness, efficiency, learnability etc.). Global structure emerges due to the local interactions.

Any change or emergence occurs and spreads to the whole population through individual interactions. eg. the loss of words homophonous to taboo words.


Computer models provide a viable paradigm to embody the investigation of various assumptions, conditions and factors in the study of language evolution.
Acknowledgement

- Professor William S-Y Wang, Language Engineering Laboratory, research project 901000, City University of Hong Kong
- Professor Jean-Marie Hombert, Laboratoire Dynamique du Langage, Lyon