Word Doubling in New Englishes

Manfred Sailer

Seminar für Englische Philologie
Universität Göttingen

ISLE 2011, Boston
21. 6. 2011
Outline

1 Introduction

2 Doubling in British, Singapore, and Hong Kong English

3 Corpus study
   - Study 1: Raw data
   - Study 2: Reduplication vs. repetition

4 Conclusion
Introduction: Word doubling

- Doubling: uttering the same word twice in a row
- Competence doubling: Total reduplication
  - semantically or pragmatically meaningful
  - competence phenomenon

(1) Go outside walk-walk. [Singapore Engl, Wee (2008)]

- Performance doubling: repetition
  - Pause-filling doubling
  - not part of the message
  - depends on speech situation/ speaker
  - performance phenomenon

(2) Say aye what what has what has happen [ICE-SG, S2A-040]
Outline

1 Introduction

2 Doubling in British, Singapore, and Hong Kong English

3 Corpus study
   - Study 1: Raw data
   - Study 2: Reduplication vs. repetition

4 Conclusion
Total reduplication from a general-linguistic perspective


- Typologically highly frequent (Graz reduplication database: 62 of 80 languages [reduplication.uni-graz.at/redup/])
- Prominent in L1 acquisition, but irrelevant in L2 learning
- Prominent in creoles, but barely present in pidgins (Bakker and Parkvall, 2005)

⇒ Total reduplication is an indication of a high degree of nativeness of a grammatical system
Fluency

- Levelt (1989): high fluency in native language:
  - high speech rate: 2–3 words per second
  - low error rate: 1/1,000 words
  - short pauses, few repairs

- Schmidt (1992): fluency ≠ proficiency
  ...we identify fluency with the processing of language in real time, rather than with language as the object of knowledge

- Lennon (1990): speech-pause relation; frequency of filled pauses and repetitions

- L1 speakers: fewer/shorter pauses than L2 speakers
  - L1 speakers: fewer filled pauses/repetitions than L2 speakers
Schneider’s Dynamic Model

- Schneider (2003, 2007)
- 5 phases in dialect genesis:
  - Phase 1: Foundation
  - Phase 2: Exonormative stabilization
  - Phase 3: Nativization
  - Phase 4: Endonormative stabilization
  - Phase 5: Differentiation
- Predictions:
  - Reduplication: grammatical innovation; only from phase 3/4 on.
  - Repetition: Up to phase 2/3: foreign language varieties in the indigenous strand
Doubling in New Englishes

- Data: British English (GB), Singapore English (SG), Hong Kong English (HK)
- Components of the International Corpus of English (ICE)

Background:
- research on total reduplication (Stolz, 2008)
- research on second language fluency (Schmidt, 1992)
- Dynamic Model (Schneider, 2003, 2007)

Hypotheses:
- GB: native variety with little reduplication
- SG: near-native variety with grammaticalized reduplication
- HK: second language variety with little reduplication
- Reduplication: more in SG than in GB and HK
- Repetition: more in HK than in SG and GB
1 Introduction

2 Doubling in British, Singapore, and Hong Kong English

3 Corpus study
   - Study 1: Raw data
   - Study 2: Reduplication vs. repetition

4 Conclusion
Total reduplication in GB

- Quirk et al. (1985): doubling of intensifying adverbs
  (3) very very good

- Ghomeshi et al. (2004): Contrastive fokusreduplication
  (4) I’ll make the tuna salad, and you make the SALAD-salad.

- Similar constructions:
  Quirk et al. (1985): coordinative structure \((X \text{ and } X)\)
  (5) she kept getting thinner and thinner

  Jackendoff (2008): N-P-N-construction
  (6) a. day by day b. picture after picture
Repetition in GB

- GB: first language variety
- Expect: Fewer marks of dysfluency than in foreign language varieties
Singapore

Schneider (2007): Phase 4
- Official languages: Englisch, Malay, Chinese (Mandarin), Tamil
- English-based bilingualism
- barely monolingual native speakers of English
- Excellent knowledge of English for Singaporeans born after 1970
- Many special patterns

Total reduplication in Malay, Mandarin (Wee, 2004), and Tamil (Wiltshire and Marantz, 2000)
Reduplication in SG

Several patterns of reduplication:

- Prenominal elements: intensification (Wee, 2004; Wong, 2004)
  
  (7) Don’t always eat sweet-sweet [= very sweet] things.

- Verbs: attenuation, continuity (Wee, 2004)
  
  (8) Don’t always stay in the house.
  Go outside walk-walk [= stroll].

- Nouns: affectionate marking (Wee, 2004)
  
  (9) Where is your boy-boy [= boyfriend/son]?
Repetition in SG

- Advanced second language variety
- Expect: Fewer marker of dysfluency than in foreign language varieties
- Deterding (2007): repetition more frequent than reduplication
Hong Kong

Schneider (2007): Phase 2–3

- Official languages: English, Cantonese, Mandarin
- Population: primarily Cantonese speaking
- Barely native speakers of English
- Knowledge of English: ca. 43% of the population in 2001
- Orientation towards British English; some special patterns
- Complaint tradition

Total reduplication in Cantonese (Matthews and Yip, 2002)
Doubling in HK

- Total reduplication:
  - Not documented in descriptions of the variety, such as Setter et al. (2010)
  - Ansaldo (2010): no systematic reduplication in HK.

- Repetition
  - If phase 2/3: learner/foreign language variety
  - Expect: more repetition than in native/ more native-like varieties
  - Setter et al. (2010): many cases of repetition in the data
Hypotheses

- H1: There is more reduplication in SG than in GB or HK.
- H2: There is more repetition in HK than in GB and SG.
Outline

1. Introduction

2. Doubling in British, Singapore, and Hong Kong English

3. Corpus study
   - Study 1: Raw data
   - Study 2: Reduplication vs. repetition

4. Conclusion
International Corpus of English (ICE)

- ice-corpora.net/ice/
- Written and spoken English after 1989
- Parallel corpus components for various varieties: ICE-GB, ICE-SG, ICE-HK
- Each component: 1 mio words, 500 files à 2,000 words
  - 300 files: spoken English
  - 200 files: written English
Working with corpora

- Corpus-based (qualitative):
  - Manual retrieval of the data
  - Does a particular phenomenon occur in the corpus?

- Corpus-driven (quantitative)
  - Automatic retrieval of the data
  - How common is a phenomenon?

- No distinction between performance and competence data
- Can we disentangle reduplication from repetition?
Corpus-driven: Frequency of a pattern

- Distribution of doubling in a corpus

Stefanowitsch (2007): Word doubling in English and German
  - Brown Corpus (1 mio words, written GB, 1960s)
  - Automatic extraction of all occurrences of the patterns: “X X” and “X and X”
  - Results: “X X”: 37 hits
    “X and X”: 102 hits
  - $\chi^2$-test: difference between the patterns is significant
    $\chi^2 = 51.65$, $p < 0.001$
Corpus-driven: Productivity of a pattern

- Productivity:
  - Pattern occurs with many different words.
  - Pattern occurs with new words.


- Number of . . .
  - Tokens: How often does the pattern occur?
  - Types: different words that participate in the pattern?
  - hapax legomena: words that occur exactly once in the pattern.

- Probability of finding a new word with a given pattern:

  strict produktivity: \[ P = \frac{\# \text{hapax legomena}}{\# \text{Token}} \]

  global productivity: \( P^* \): 2-dimensional, relates \( P \) and the number of Types
Doubling data from the ICE segments

- Extract pattern “X X” from ICE-GB, ICE-SG, ICE-HK, using the script from Stefanowitsch (2007)
- Compute productivity
- Hypotheses:
  - H1: More reduplication in ICE-SG than in ICE-GB and ICE-HK
  - H2: More repetition in ICE-HK than in ICE-SG and ICE-GB
- Can productivity distinguish between reduplication and repetition?
Results

<table>
<thead>
<tr>
<th>corpus</th>
<th>Word doubling (tokens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>6.651</td>
</tr>
<tr>
<td>SG</td>
<td>6.094</td>
</tr>
<tr>
<td>HK</td>
<td>17.039</td>
</tr>
</tbody>
</table>

- about the same amount of doubling in GB and SG
- more than 2.5 times more doubling in HK than in GB or SG
Productivity

Productivity: $P = \text{hapax/tokens}$

<table>
<thead>
<tr>
<th></th>
<th>tokens</th>
<th>types</th>
<th>hapax</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>6.651</td>
<td>744</td>
<td>452</td>
<td>0.07</td>
</tr>
<tr>
<td>SG</td>
<td>6.094</td>
<td>924</td>
<td>586</td>
<td>0.10</td>
</tr>
<tr>
<td>HK</td>
<td>17.039</td>
<td>1.305</td>
<td>834</td>
<td>0.05</td>
</tr>
</tbody>
</table>

- Highest productivity for SG
- Productivity for GB and HK similar

Global productivity $P^*$: 2-dimensional measure: $P$ and types
Global productivity reflects reduplication: Doubling more productive in SG than in GB and HK.

While productivity for GB and HK similar, global productivity shows distinct patterns. Reflex of repetition?
Effect of repetition

- Token frequency of doubling: SG = GB, SG < HK, HK > GB
- High number of doubling in HK reflects fluency difference due to higher processing load in foreign language.
- Expectation: Fluency
  - GB: native language
  - ≥ SG: second language
  - > HK: foreign language
- Case studies:
  - filled pauses (*uhm*-doubling)
  - detailed sample analysis
Case study I: filled pauses

- Indicator of dysfluency: frequency of filled pauses: \( uh(m) \)
- Hypothesis: more filled pauses in HK than in GB and SG
  Equal number of filled pauses in GB and SG
- Results:

<table>
<thead>
<tr>
<th>corpus</th>
<th>( uh(m) )</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>12,642</td>
<td>1.19</td>
</tr>
<tr>
<td>SG</td>
<td>12,068</td>
<td>1.09</td>
</tr>
<tr>
<td>HK</td>
<td>42,088</td>
<td>2.86</td>
</tr>
</tbody>
</table>

- Pause-filling by \( uh(m) \) supports the fluency hypotheses.
Case study II: Sample

- For each corpus: random sample of 300 hits from the doubling data
- Type of doubling: (GB)
  - (potential) reduplication/ “potentially intended doubling”:
    I had a really really good supper last night
  - repetition: we used a a slightly different uh r rhythmic quality to it
  - unclear: Yes Yes I ’ll tell Jane that you ’ve done
  - other (wrong annotation, names, . . . ): Building Regulations 9 9
- Hypotheses:
  - reduplication: most reduplication in SG
  - repetition: most repetition in HK
Case study II: Results

- Hypotheses:
  - reduplication: most reduplication in SG
  - repetition: most repetition in HK

- Results: Type of doubling

<table>
<thead>
<tr>
<th>corpus</th>
<th>reduplication</th>
<th>repetition</th>
<th>unclear</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>12 (4.78%)</td>
<td>239 (95.22%)</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>SG</td>
<td>19 (8.15%)</td>
<td>214 (91.85%)</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>HK</td>
<td>7 (2.95%)</td>
<td>230 (97.05%)</td>
<td>39</td>
<td>24</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 6.5504, \, df = 2, \, p < 0.05 \]

- Most important cells: reduplication in SG and HK
Reduplication data in the GB sample

(10) a. which is quite quite a nice advantage
    b. I ’ve been applying quite quite regularly since I ’ve been really really sort of working . . .
    c. I had a really really good supper last night
    d. Very bad Very minor Very Very minor
    e. you ’re already very very good at your job

(11) a. Have you ever ever seen anybody who was addicted to exercise?

(12) a. affecting our everyday everyday lives
    b. . . . any any any sort of questions of clarification about what you ’ve actually been presented with . . .

(13) I love the way they they refer to everything as all all our relations in including the stars

(14) It makes such a difference having you. Love Love Love Love and More Love
Reduplication data in the HK sample

(15) GB-like pattern:
   a. I’m very very very good at school
   b. Noel say that you are very very very busy
   c. the education system here makes children feel very very afraid of their studies
   d. Good good good barbecue pork

(16) Other function words:
   Actually, I learnt how to play guitar. Yeah, but only only the simple songs.

(17) V-doubling:
   a. but their mother uh dislike dislike uh secondary school talk [taught?] by Chinese
   b. I just like love love the atmosphere yah
Reduplication data in the SG sample

(18) a. That’s why you have to be very very fast
   b. that museum is totally supported by a very very wealthy individual . . .

(19) a. No never never
   b. it has stopped manufacturing Ewok Village for a long long time

(20) a. she is living on her own now now that her auntie is not there anymore
   b. So they just go in they look look whatever looks interesting and then they just take it out

(21) a. Come come come come ask ask
   b. can you just split so they come over here
   c. That’s all my report Uh wait wait wait
Summary: Corpus findings

- Overall most doubling in HK, equal doubling in GB and SG
- Productivity: most productive in SG
- Fluency:
  - Filled pauses: independent evidence for lowest fluency in HK; same rate in GB and SG.
  - Repetition: most repetition and least reduplication in HK
- Reduplication: SG reduplication strongest effect in the samples
Outline

1 Introduction

2 Doubling in British, Singapore, and Hong Kong English

3 Corpus study
   - Study 1: Raw data
   - Study 2: Reduplication vs. repetition

4 Conclusion
Summary

- Doubling: competence and performance phenomenon
- Hypothesis:
  - New Englishes above phase 3: more reduplication than GB; equal repetition as GB.
  - New Englishes below phase 3: not more reduplication than GB; more repetition than GB.
- Corpus data doesn’t discriminate between reduplication and repetition
- Productivity: identify grammaticalized pattern
- Repetition: correlates with filled pauses (uhm)
Conclusion

- GB
  - native variety
  - few reduplication patterns
  - mild repetition ratio

- Singapore:
  - non-native variety
  - English as a second language
  - but: variety of its own with natural grammar
  - productive total reduplication
  - mild repetition ratio

- Hong Kong:
  - non-native variety
  - English as a foreign language.
  - high degree of repetition.
  - no independent reduplication patterns.
Future directions

- Different corpora? blogs (Deterding, 2007)
- More recent development in Hong Kong
- More ICE components
- Integration of audio data
- Reduplication – repetition – pragmatic doubling?
...thank you, thank thank thank you

(www, Singapore)


