

# **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

# Outline

- 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

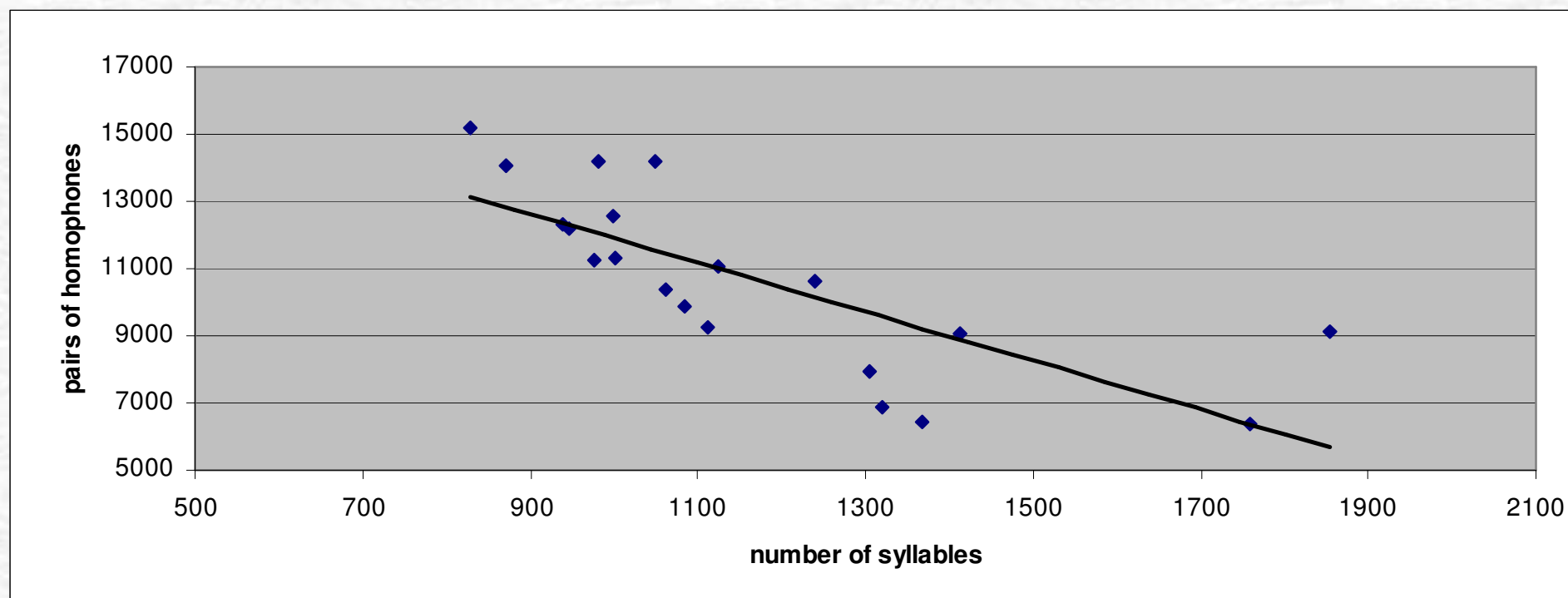
# Hypothesis I:

- The smaller the sound inventory, the larger the number of homophones.

# Phonological inventory size vs degree of homophony in Chinese dialects

corr= -0.76

tai yuan	wuhan	cheng du	yang zhou	hefei	chang sha	su zhou	shuang feng	wen zhou	ji'nan	xi'an	nan chang	beijing	jian'ou	mei xian	yang jiang
828	870	938	947	976	981	999	1001	1048	1063	1084	1111	1125	1241	1304	1319
15167	14072	12342	12208	11238	14203	12577	11307	14217	10387	9872	9252	11076	10641	7923	6894



# Homophone existence

## English:

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

- (Brown Corpus: the Standard Corpus of Present-Day American English. Nelson and Kučera, 1989)

## 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 → kan4 jian4

见 看 见

“see” → “look” “see”

# Hypothesis II

more homophony

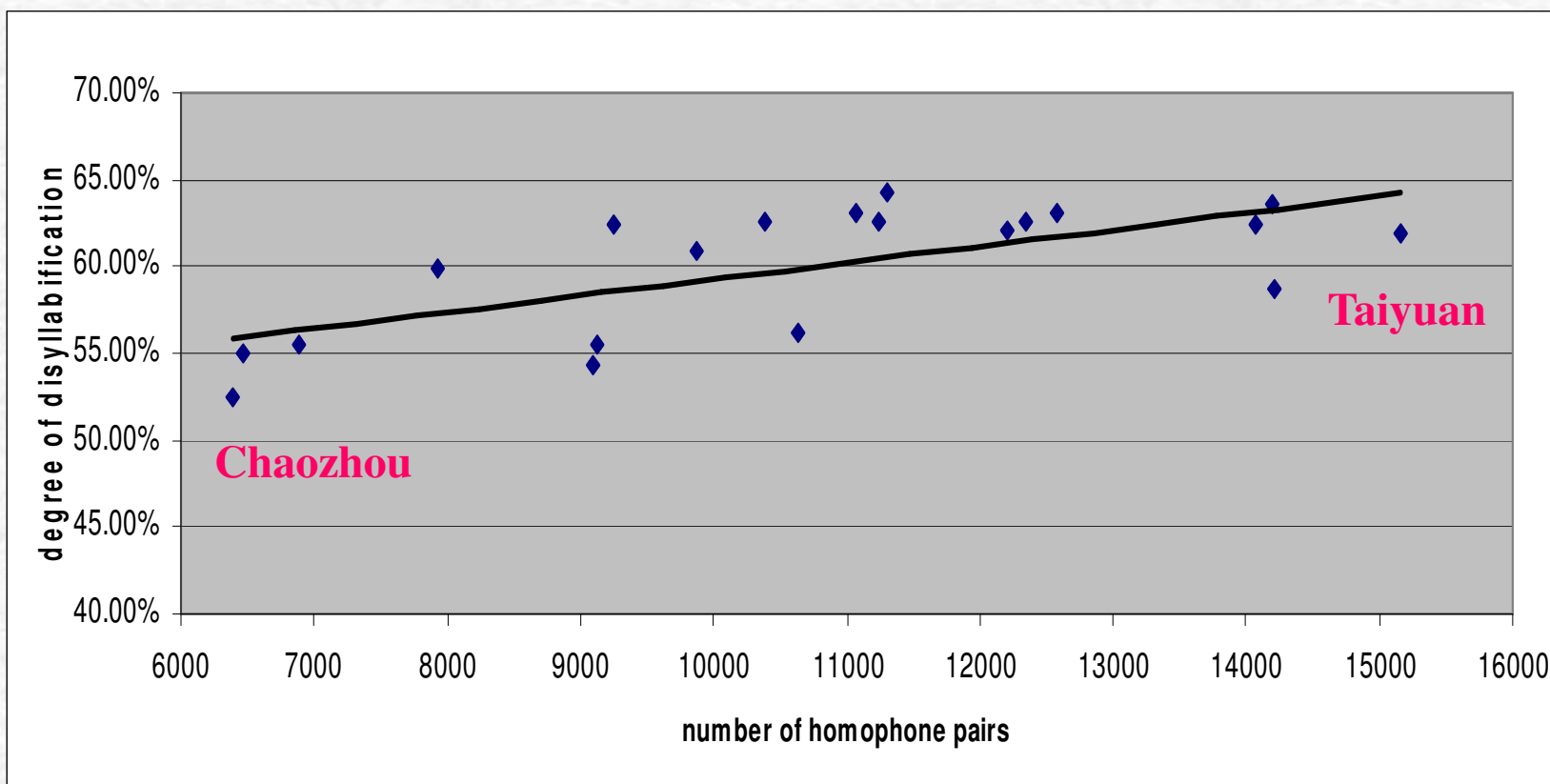


more disyllabification



# 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

Distances of freq. ranking	Homophone pairs
<10	36.91%
100	38.36%
200	8.55%
300	4.36%
400	4.36%
500	3.64%
>500	3.82%

**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)

Distances of freq. ranking	Homophone pairs	Randomly selected pairs
< 10	36.91%	70.00%
100	38.36%	26.18%
200	8.55%	2.00%
300	4.36%	1.09%
400	4.36%	0.18%
500	3.64%	0.00%
> 500	3.82%	0.55%

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)

# Computational modeling

- 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

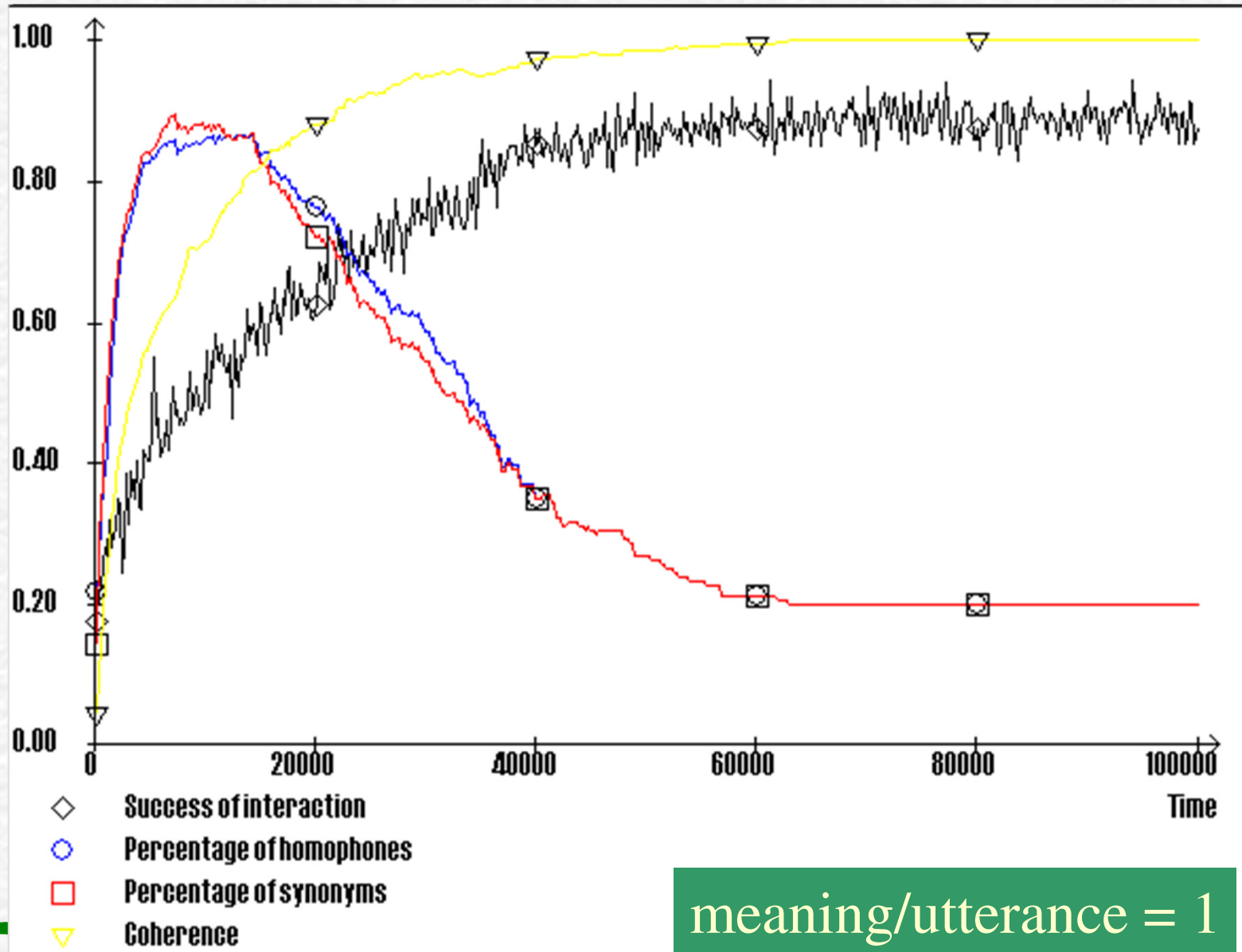
Creating new words

learn new words

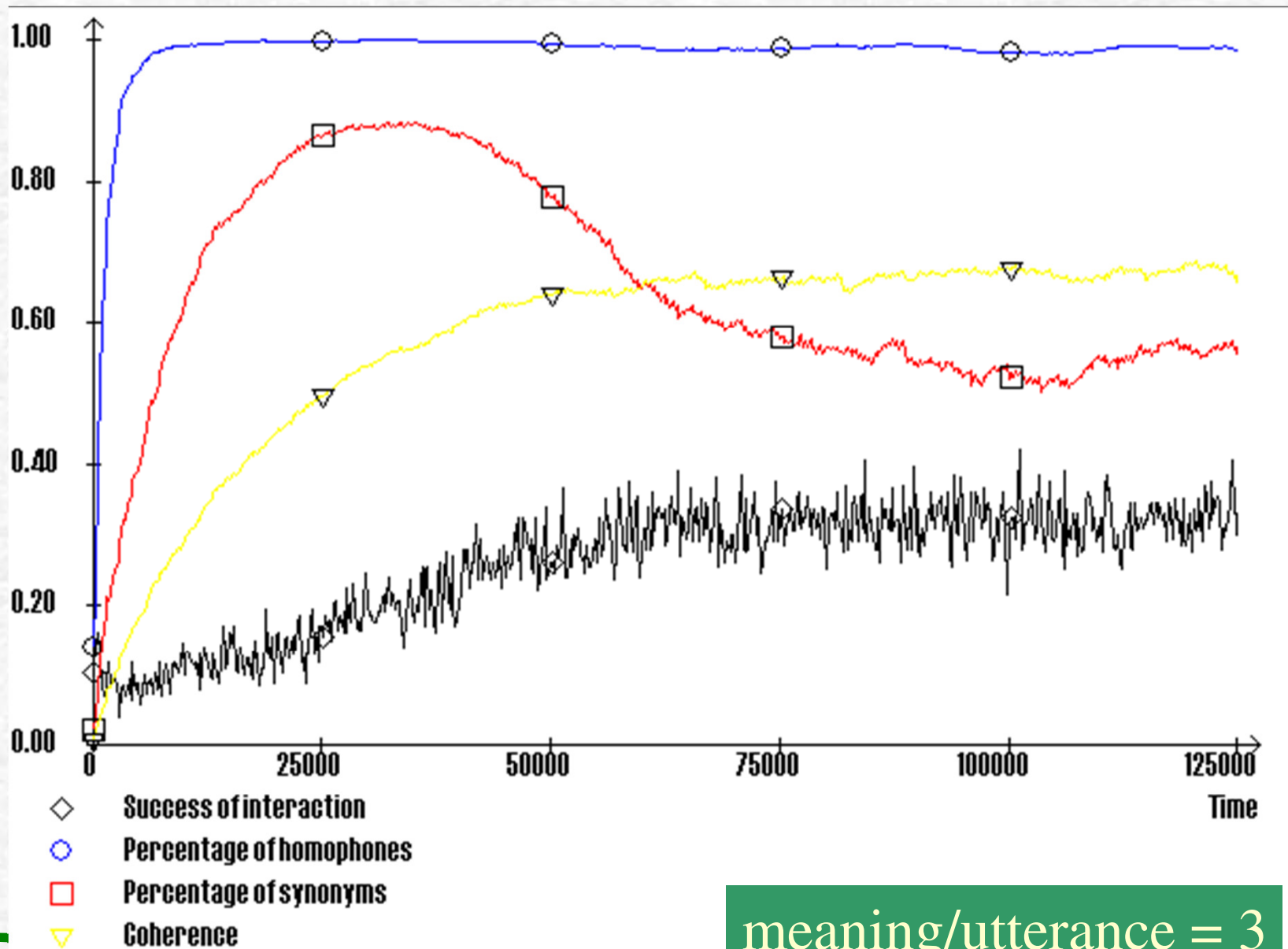
Speaker Listener

empty lexicon  $\rightarrow$  a set of shared associations:(meaning, utterance)

# The effect of meaning/utterance ratio







meaning/utterance = 3

Two-word communication:

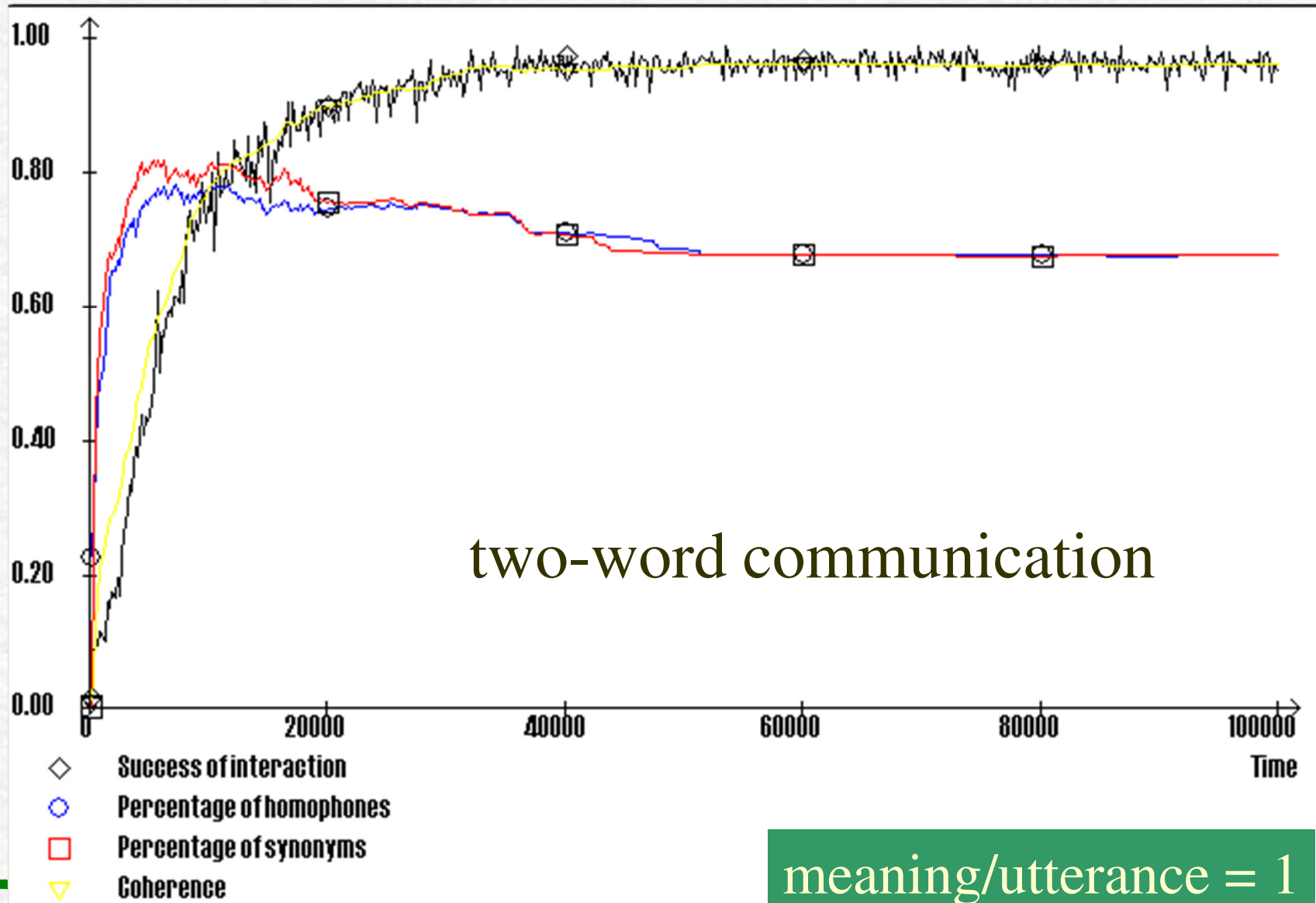
eg: "flower" "vase"

[fa] [pin]

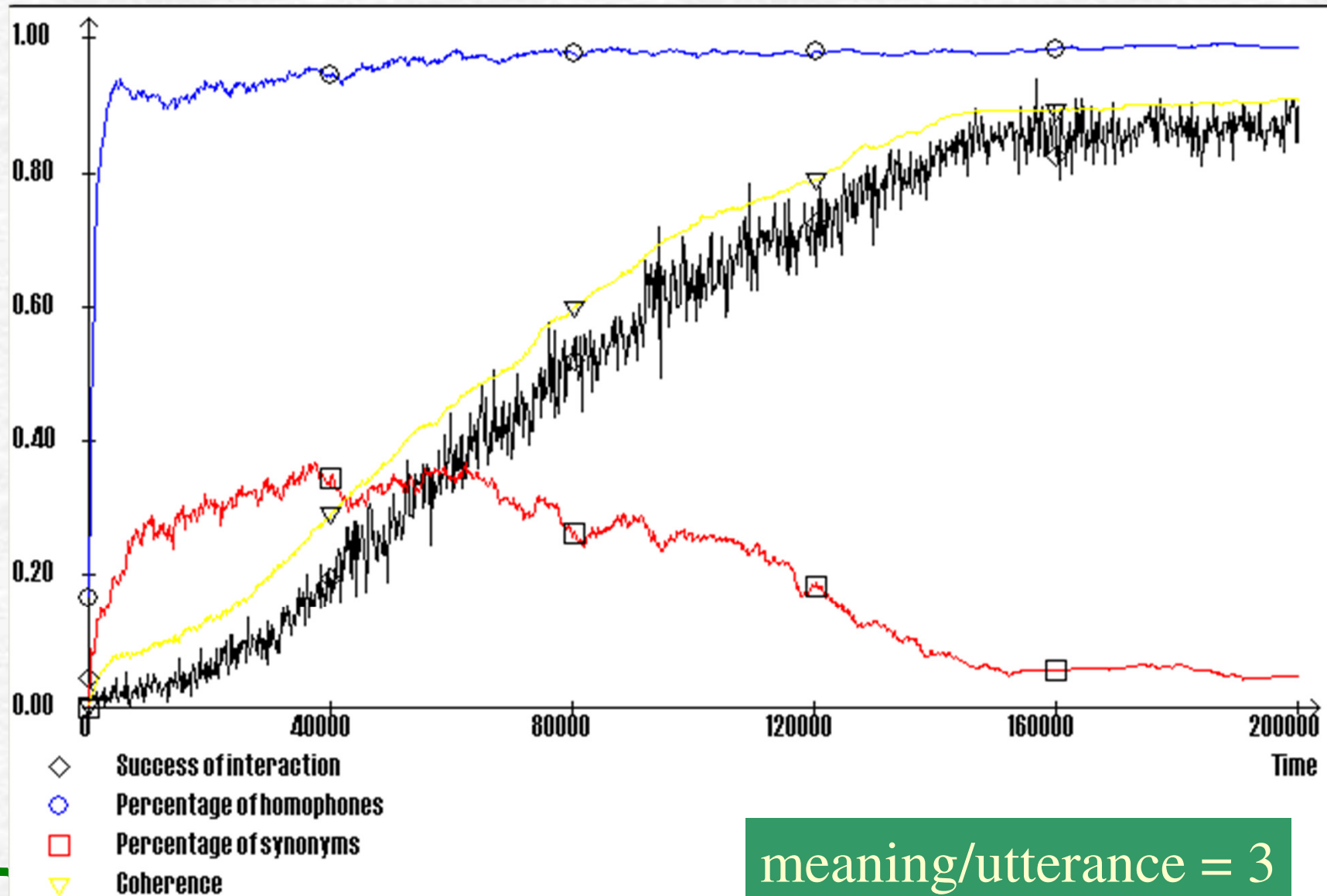
"computer" "screen"

[kom] [pin]

# The effect of context



## two-word communication



meaning/utterance = 3

# Conclusions

- 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.
  - “The invisible hand hypothesis” (R. Keller 1994); L. Steels, et al.
- 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