Robustness Analysis of 入力 Adaptive 中文

Mike Tian-Jian Jiang, Cheng-Wei Lee, Chad Liu, Yung-Chun Chang,
Wen-Lian Hsu
Institute of Information Science, Academia Sinica

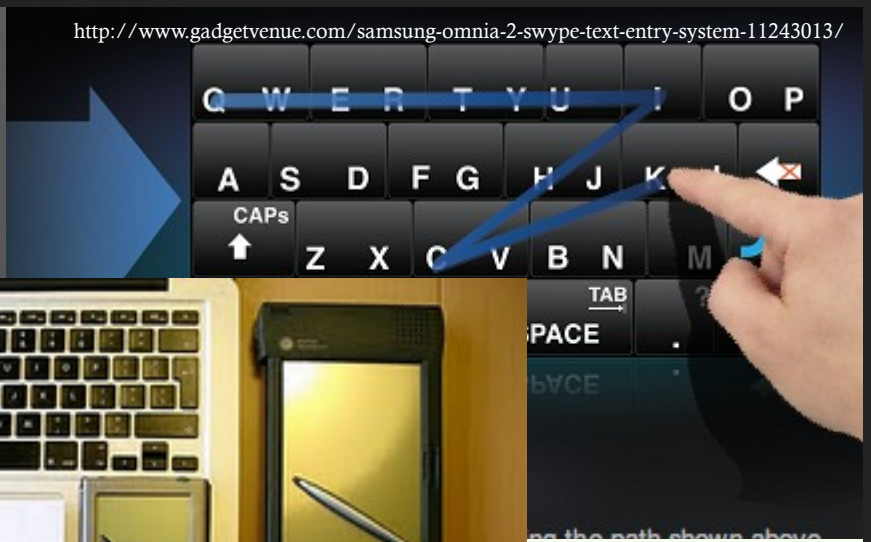
CC BY

入力, Input Method (IM)

Text Entry

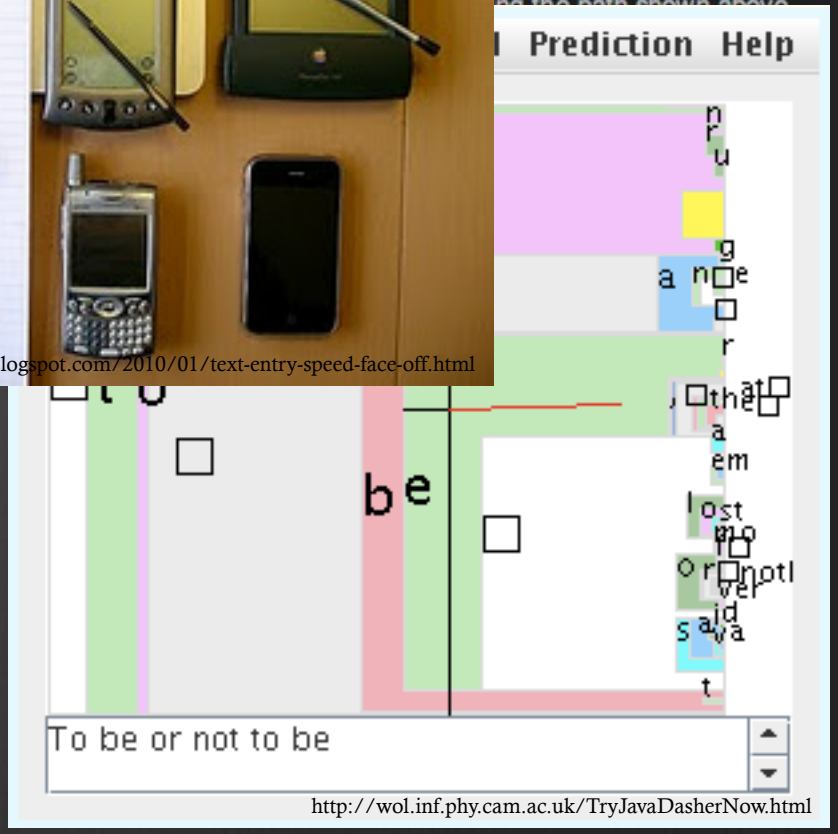

Normal alphanumeric gestures
Punctuation mode - tap to enter, backsp
space return
tab
er mode - prefix with
Accents
http://en.wikipedia.org/wiki/File:Palm_Graffiti_gestures.png

http://www.inquisitr.com/97719/circboard-brings-easier-text-entry-to-xbox-360-just-needs-investors/


http://www.gadgetvenue.com/samsung-omnia-2-swype-text-entry-system-11243013/
CAPs
TAB


http://myapplenewton.blogspot.com/2010/01/text-entry-speed-face-off.html

http://research.microsoft.com/en-us/um/redmond/groups/cue/mobileinteraction/


Prediction
Help
To be or not to be
http://wol.inf.phy.cam.ac.uk/TryJavaDasherNow.html

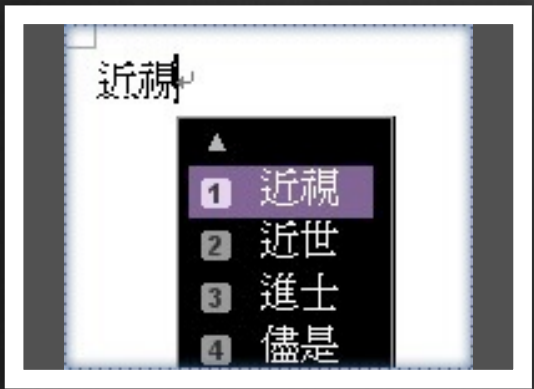
Radical vs. Phonetic

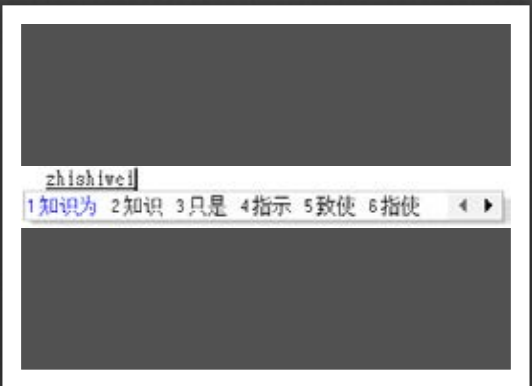
Homographs vs. Homophones

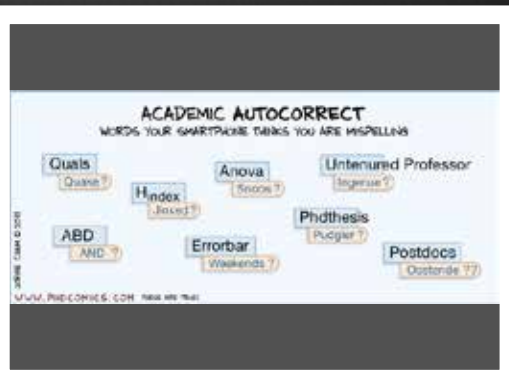
http://thereality.nl/potd/1552/donderdag-31-december-2009.html

http://www.ehow.com/how_6788906_use-multi_tap.html

http://en.wikipedia.org/wiki/File:ITap_on_Motorola_C350.jpg

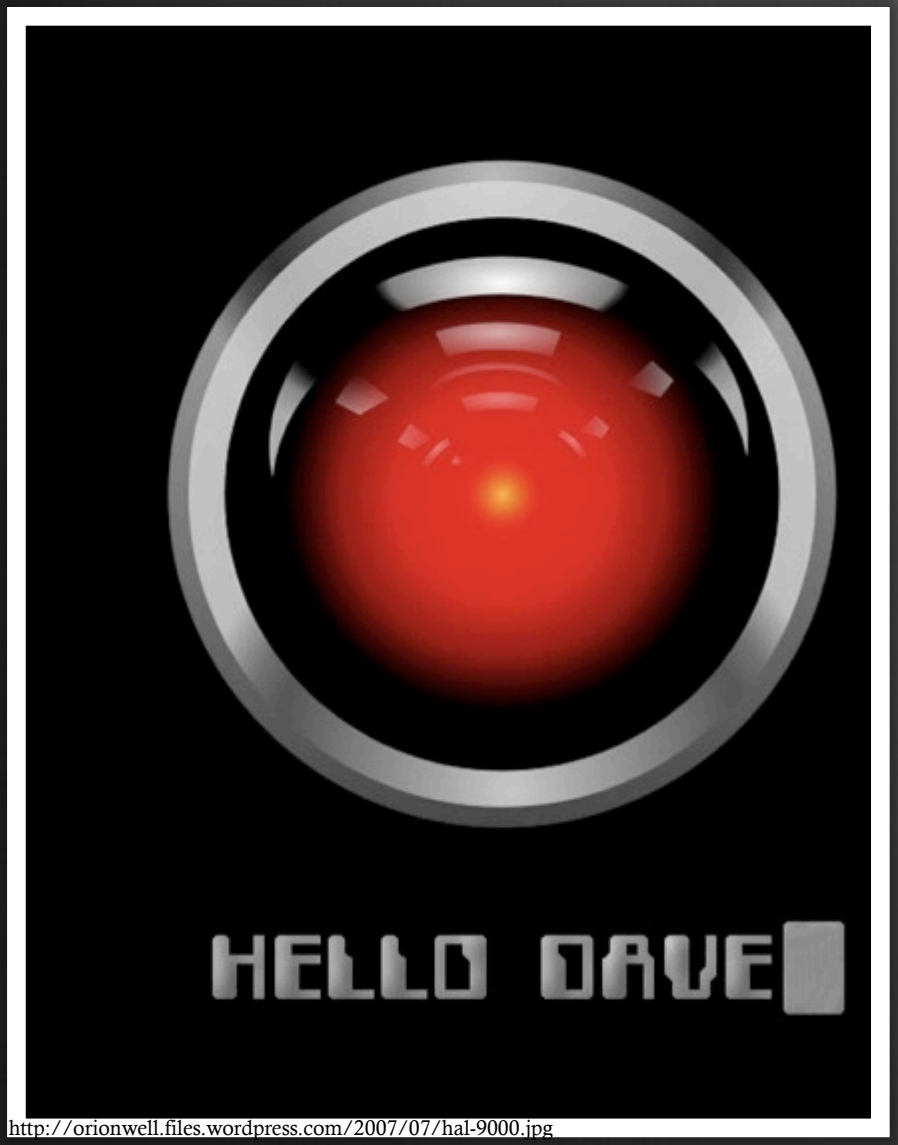






Disambiguation

To predict or not



http://orionwell.files.wordpress.com/2007/07/hal-9000.jpg

HCI, NLP (, SE)

- Unified error metrics (Soukoreff and MacKenzie, 2001)
- Error correction (Arif and Stuerzlinger, 2010)
- Reused vocabulary (Tanaka-Ishii et al., 2003)
- Backward compatibility (Suzuki and Gao, 2005)

"Prediction and spell correction can be very annoying if they are not smart enough. For many applications, user input can be very noisy (imagine voice recognition or typing on a small screen), so the input methods must be robust against such noise. Finally, there is no standard data set or evaluation metric, which is necessary for quantitative analysis of user input experience."

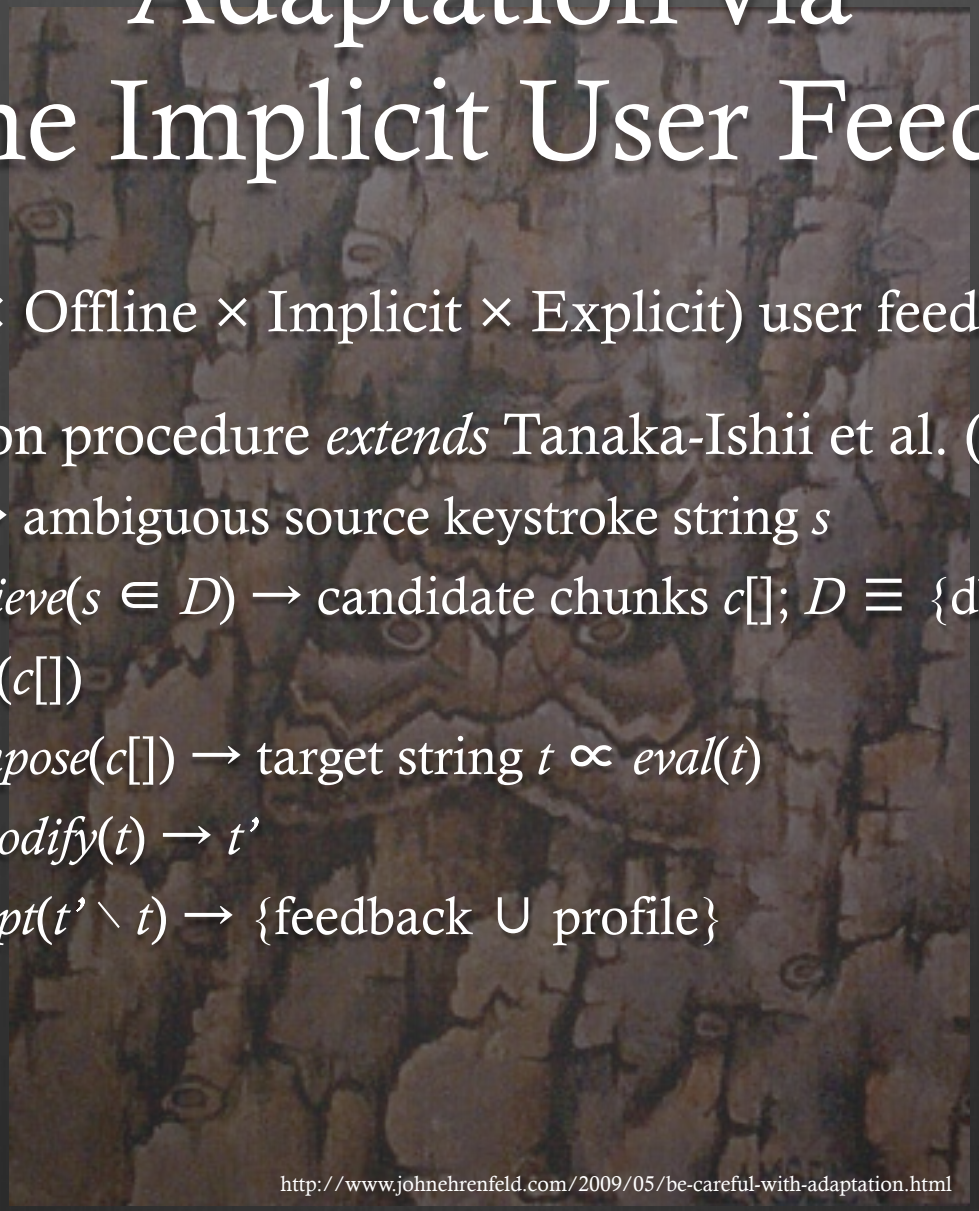
– WTIM 2011 statements of call for papers

Prediction & Adaptation

Properties of Chinese Phonetic IM

Adaptation via Online Implicit User Feedback

- (Online × Offline × Implicit × Explicit) user feedback
- Adaptation procedure *extends* Tanaka-Ishii et al. (2003)
  - User → ambiguous source keystroke string s
  - IM *retrieve*($s \in D$) → candidate chunks $c[]$; $D \equiv \{\text{db} \cup \text{profile}\}$
  - IM *sort*($c[]$)
  - IM *compose*($c[]$) → target string $t \propto eval(t)$
  - User *modify*(t) → t'
  - IM *adapt*($t' \setminus t$) → {feedback ∪ profile}



http://www.johnehrenfeld.com/2009/05/be-careful-with-adaptation.html

http://www.personal.psu.edu/afr3/blogs/SIOW/2011/10/live-long-and-prosper.html

Dilemma

Type long and (either right or wrong things) prosper

Amortized Cost

Trade-off between benefic and cost of error correction

…based on Unified Error Metrics

Related to
minimum string distance (MSD) error and
key-stroke per character (KSPC)

With Fitts' law and Hick's law

Notations

- *P*: presented text
- *T*: transcribed text
- *IS*: input stream
- *C*: number of correct characters in *T*
- *F*: keystrokes for fixing in *IS* like editing, modifier, or navigation.
- *INF*: number of incorrect yet not fixed errors in *T*
- *IF*: number of incorrect but fixed errors (keystrokes in *IS* that are not *F* and not in *T*)

MSD Error Rate

$$\frac{MSD(P,T)}{\overline{S_A}} \times 100\%$$

- *P*: the quick brown fox
- *T*: the quixck brwn fox

- MSD(*P*, *T*) = 2 here
- Only for *T* without editing process

KSPC

|IS| / |T|

Unified Error Metrics

$$Total\ Error\ Rate = \frac{INF + IF}{C + INF + INF} \times 100\%$$

$$MSD\ Error\ Rate = \frac{INF}{C + INF} \times 100\%$$

$$KSPC \approx \frac{C + INF + IF + F}{C + INF}$$

- *T*: the quixck brwn fox
- *T'*: the quixck brown fox

- *Total Error Rate*(*T'*) = (2/18)%
- *MSD Error Rate*(*T'*) = (1/17)%
- *KSPC*(*T'*) = 19/17

Fitts' and Hick's

$$t_F = a + b\log_2(\frac{d}{w} + 1)$$

$$t_H = b\log_2(n + 1)$$

- t: time
- a, b: empirical constants
- d: distance to target
- w: width of target
- n: number of equal possible choices
- Index of difficulty (ID): $\log_2((d/w) + 1)$

Error Correction Conditions

- None, Forced, or Recommended conditions
  - No relations between typing speed and correction attempt
- Spectrum of Recommended condition

Situation	**Fixed characters**	***INF***	***IF***	***F***
S_0	none	INF_0	0	0
S_i	some	INF_i	IF_i	F_i
S_{all}	all	0	IF_{all}	F_{all}

Maximally Amortized Cost

$$AC = \frac{Wasted\ Bandwidth}{Utilized\ Bandwidth} = \frac{\frac{INF + IF + F}{C + INF + IF + F}}{\frac{C}{C + INF + IF + F}} = \frac{INF + IF + F}{C}$$

$$\frac{INF_0}{C} \le AC_i = \frac{INF_i + IF_i + F_i}{C} \le \frac{IF_{all}}{C} + \frac{F_{all}}{C} = \frac{INF_0}{C} + \frac{F_{all}}{C}$$

$$penalty = \frac{t_H \times INF_0 + t_F \times \max(D)}{C + INF_0} \qquad reward = \frac{C}{C + INF_0}$$

$$AC_{modification} = \frac{penalty}{reward} = \frac{t_H \times INF_0 + t_F \times \max(D)}{C}$$

$$MAC = \frac{INF_0}{C} + AC_{modification}$$

Vocabulary Reuse

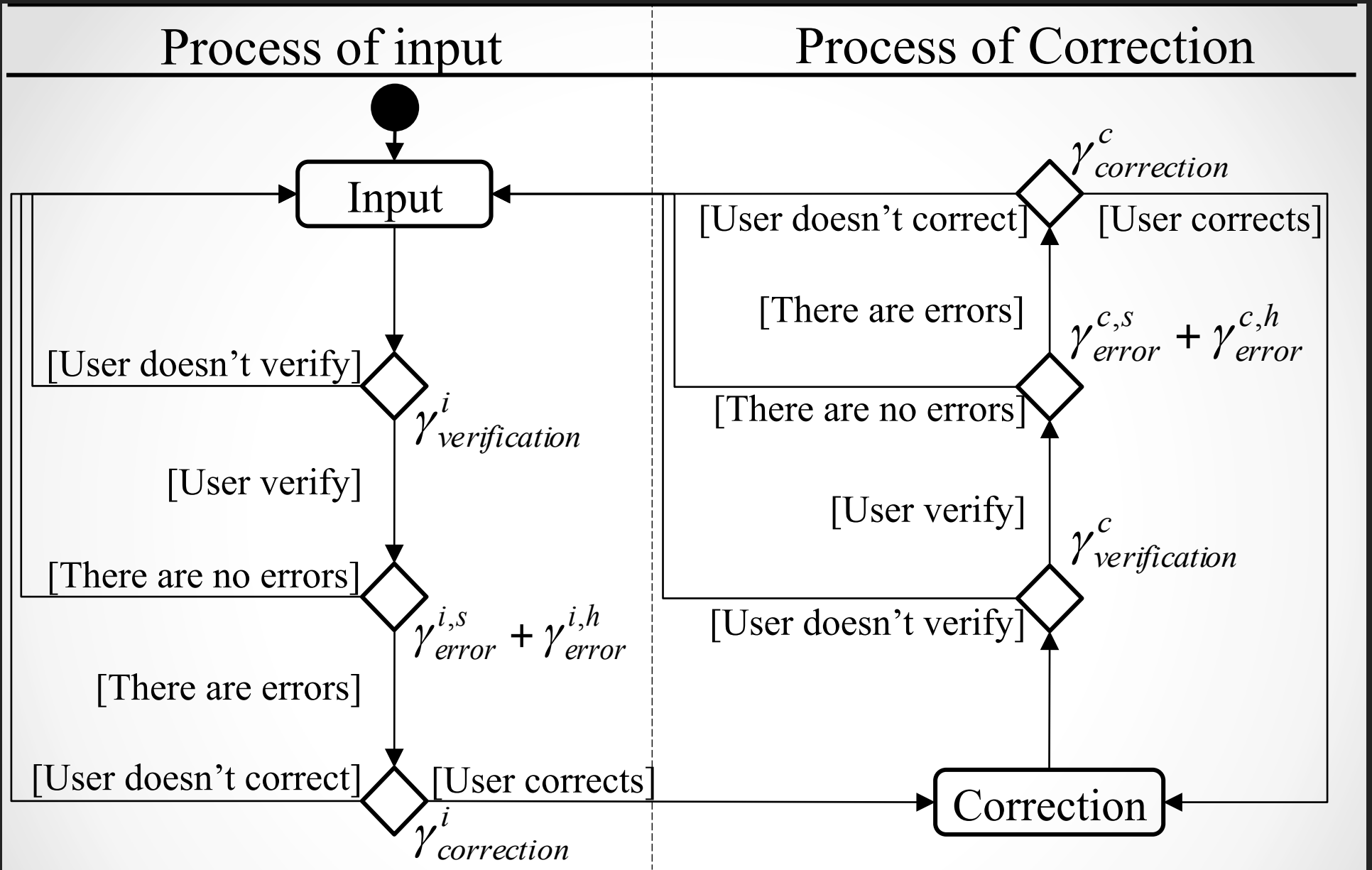
70 – 80 % vocabulary reused only after a small offset window in KB
(such that simulations of typing repeatedly are representative enough)

Backward Compatibility

Error Ratio = |errors by adaptation| / |corrections by adaptation|

… however,

Error correction can be complicated.

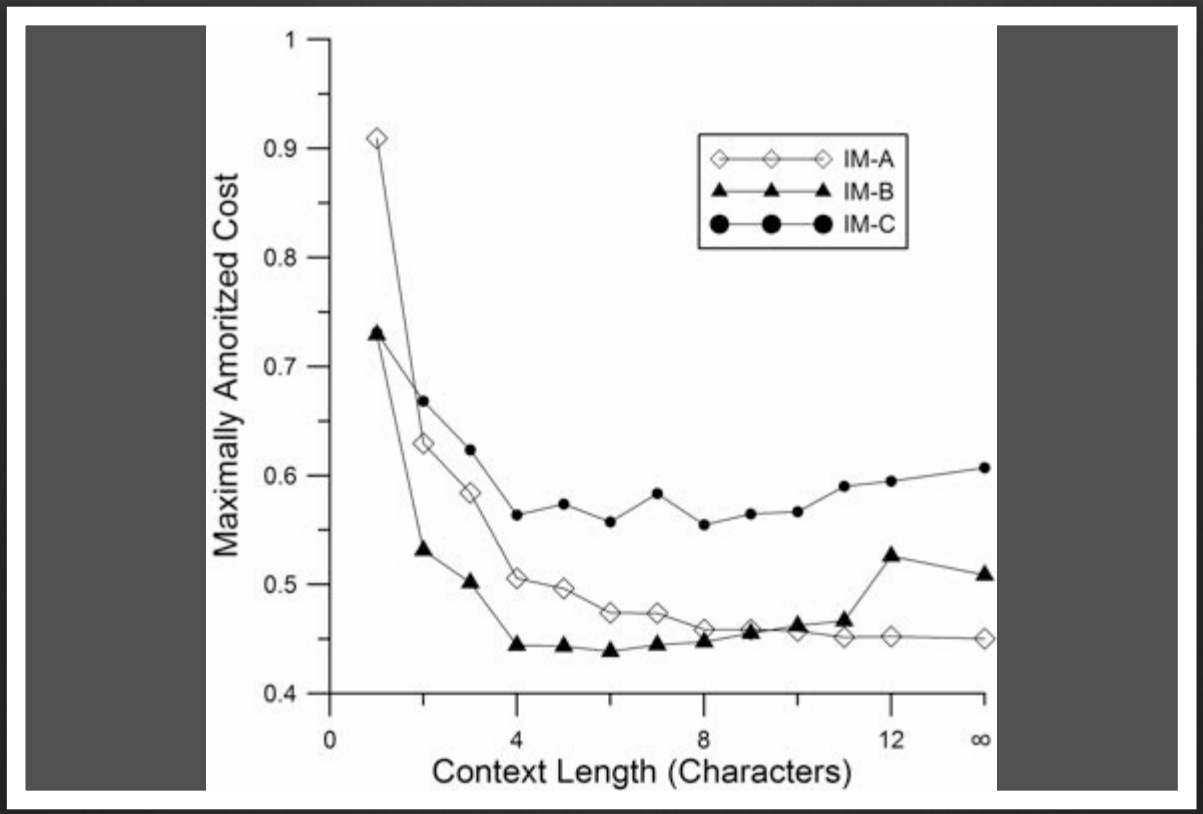


$$\rho^{h}_{correction} = \gamma^{i}_{verification} + \gamma^{i}_{correction} + \gamma^{c}_{verification} + \gamma^{c}_{correction}$$

$$\rho^{s}_{error} + \rho^{h}_{error} = \gamma^{i,s}_{error} + \gamma^{i,h}_{error} + \gamma^{c,s}_{error} + \gamma^{c,h}_{error}$$

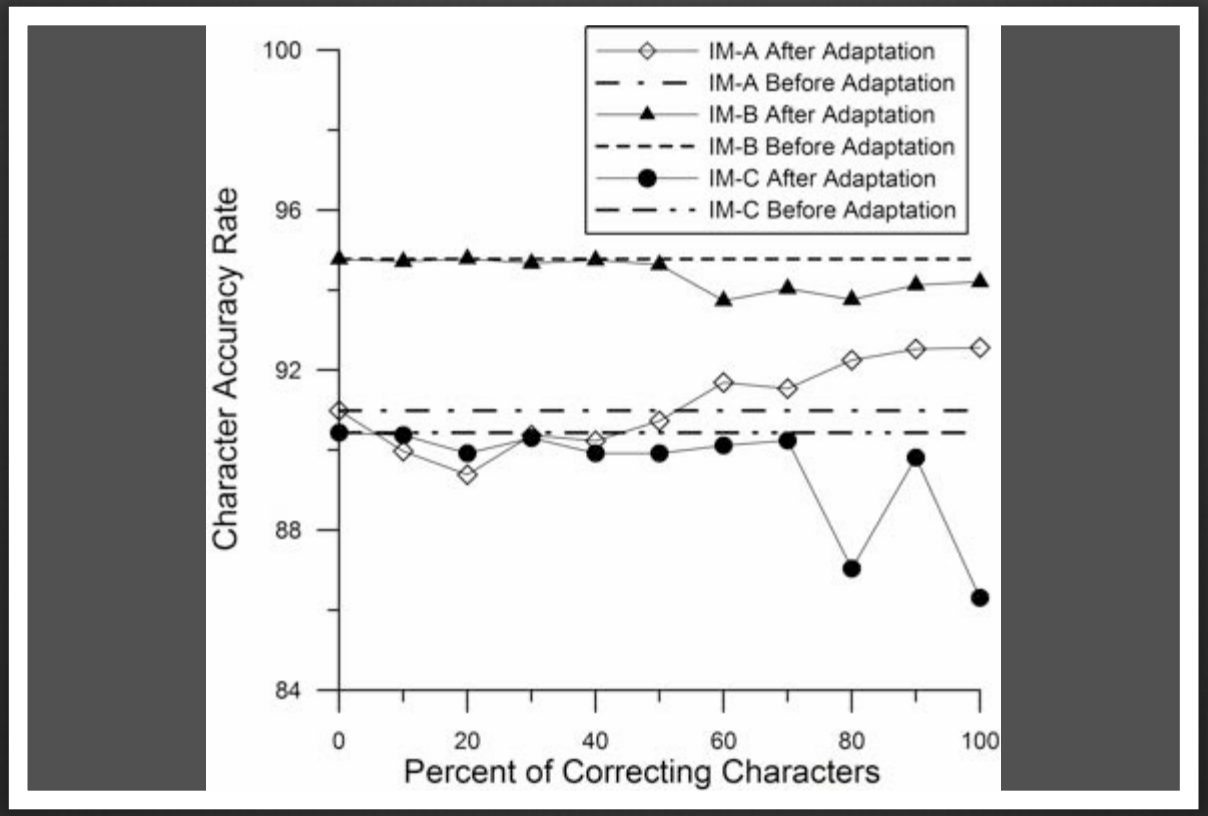
Simulation

- 3 IMs A, B, and C
  - Phonetic method
    - Bopomofo (Zhuyin)
      - Daqian keyboard layout
- Data
  - Academia Sinica Balanced Corpus
    - 4,000 sentences
      - 39,469 words
- Independent variables
  - Context length k
  - $\rho^{h}_{correction}$



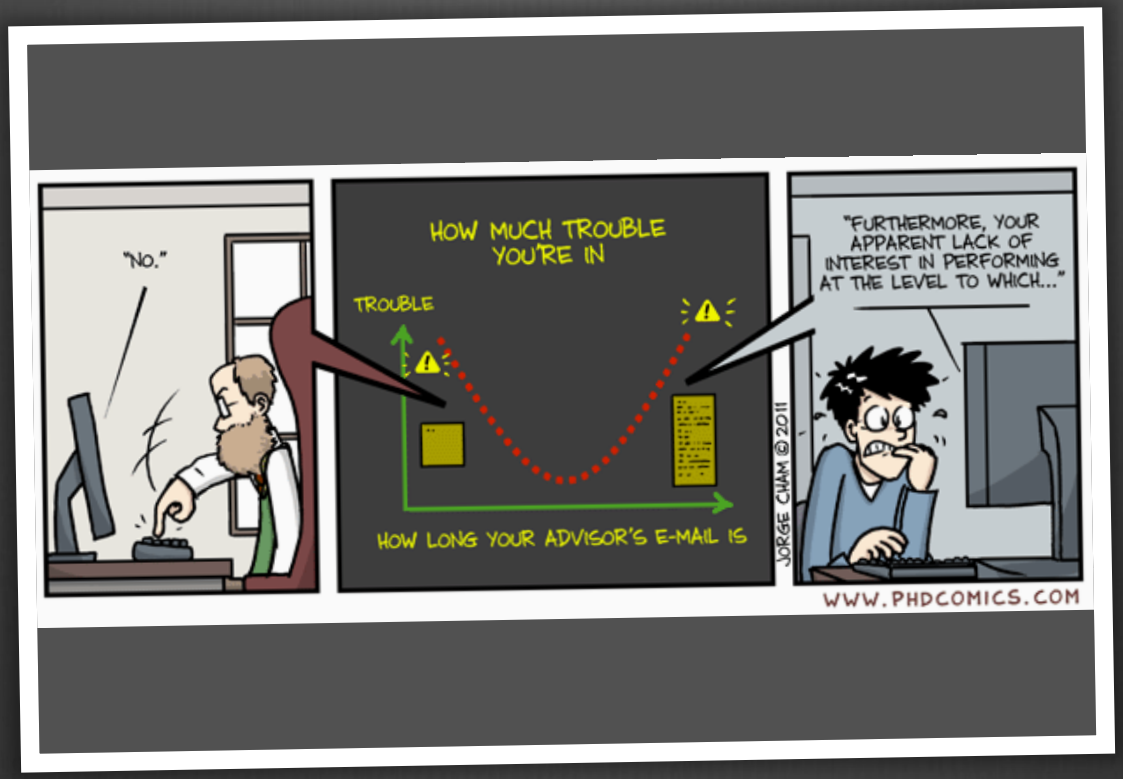
Comparison of *MAC*

IM-A seems to be different to others?



GBC at Context Length 6

Again, IM-A is segregated……



…more aspects wanted

Than this V curve…

$\rho^h_{correction}$	E^f_{max}	E^f_{avg}	E^b_{max}	E^b_{avg}	U_{max}	U_{avg}
10%	0	0.00	1	0.00	30	5.73
20%	2	2.00	1	1.00	22	8.30
30%	0	0.00	1	1.00	31	13.00
40%	4	2.40	3	1.45	51	12.05
50%	3	2.20	2	1.20	111	23.25
60%	2	2.00	6	2.50	57	20.85
70%	2	2.00	8	2.60	56	22.55
80%	5	2.35	9	3.00	35	18.75
90%	5	2.40	10	2.90	33	18.00
100%	5	2.25	18	3.55	29	16.50

Table 2. Error-tolerance level of IM-A

$\rho^h_{correction}$	E^f_{max}	E^f_{avg}	E^b_{max}	E^b_{avg}	U_{max}	U_{avg}
10%	0	0.00	1	1.00	33	8.00
20%	0	0.00	1	1.00	10	2.75
30%	2	0.00	2	1.05	33	9.95
40%	0	0.00	2	1.05	37	13.80
50%	2	2.00	2	1.20	31	10.45
60%	2	2.00	2	1.20	19	14.45
70%	3	2.13	4	1.70	28	11.65
80%	2	2.00	4	2.20	21	10.15
90%	5	2.45	3	2.25	24	12.10
100%	3	2.25	4	2.55	25	13.45

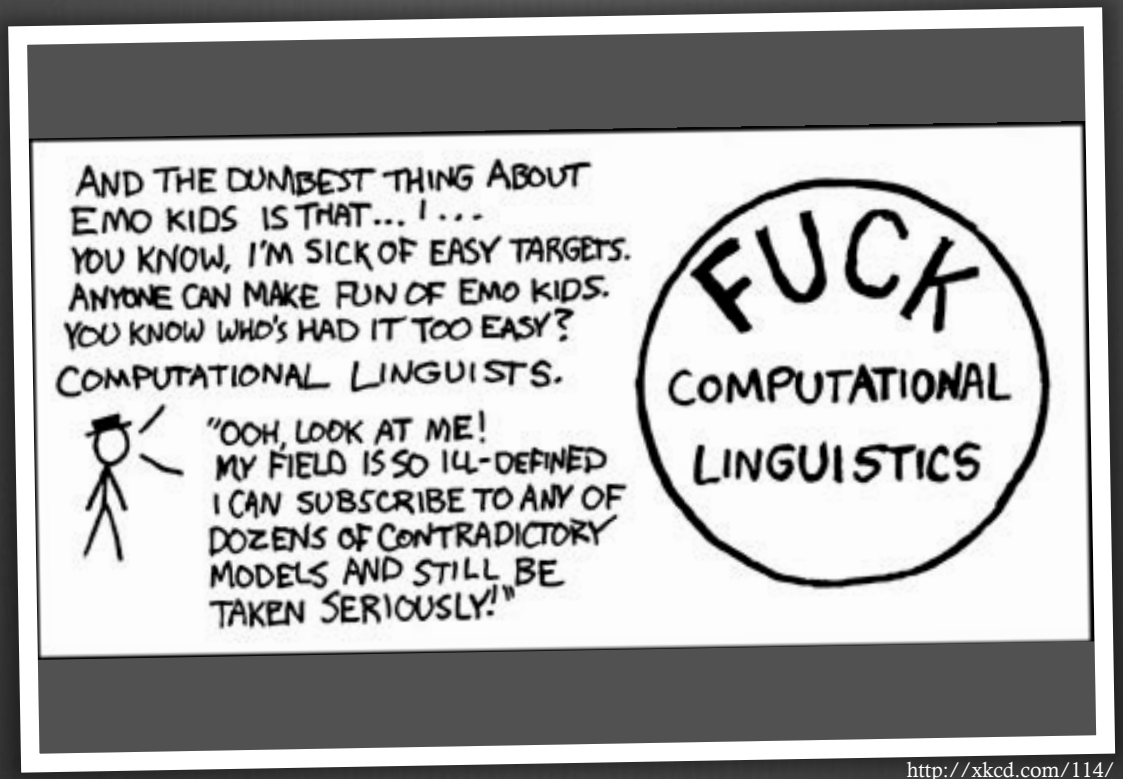
Table 3. Error-tolerance level of IM-B

Error Tolerance Level

- Futile Effort (E^f): |never adapted chunks|
- Beneficial Effort (E^b): |adapted chunks|
- Utility (U): |before forgotten|
- E^b(IM-A) vs. E^b(IM-B)
  - "his" or "her" or "its"?
    - 他 or 她 or 它
      - /ta1/

Correlation Coefficient to CAR

	E^f_{avg}	E^b_{avg}	U_{avg}
IM-A	0.49	0.92	0.66
IM-B	-0.78	-0.62	-0.51



http://xkcd.com/114/

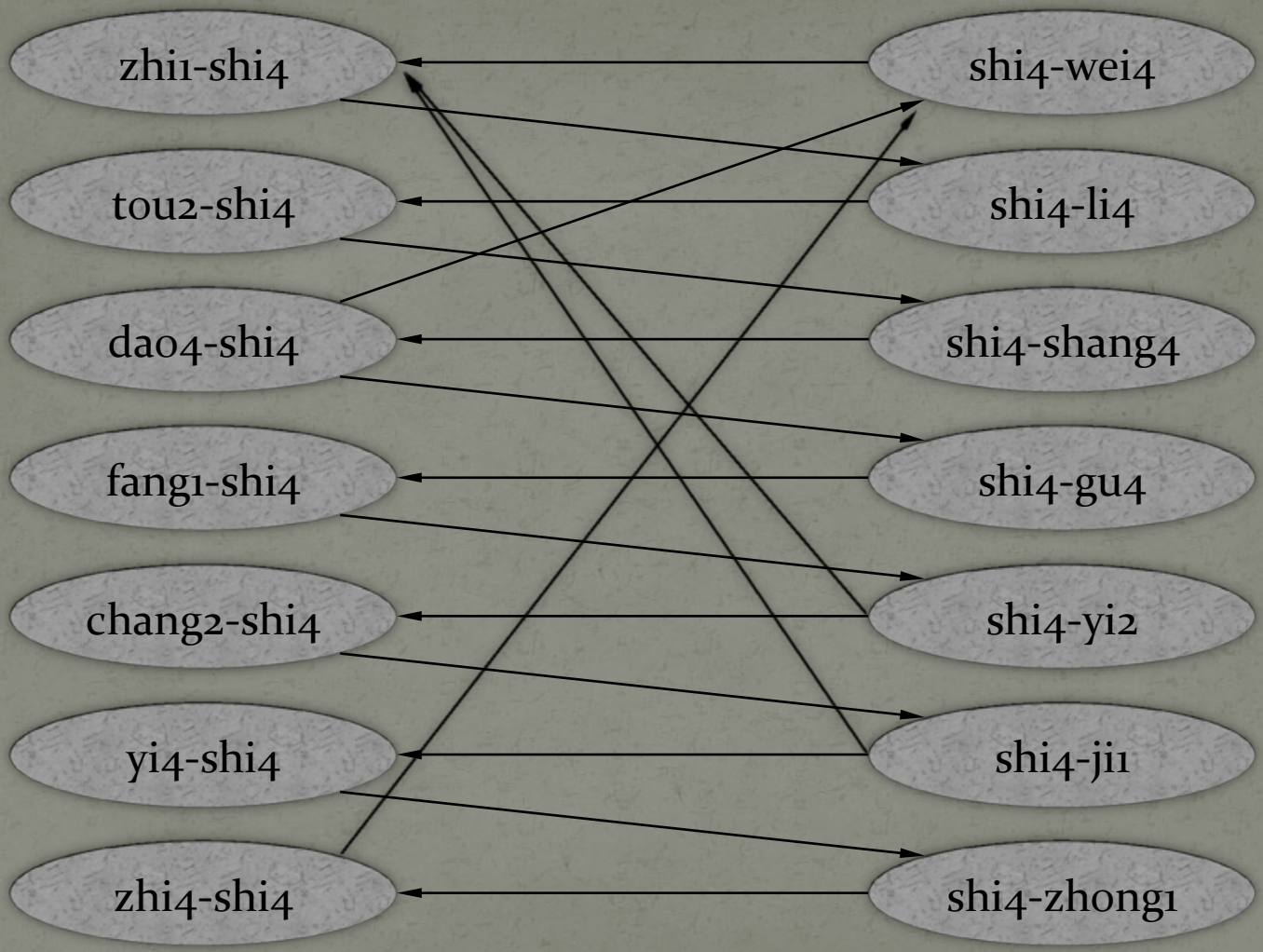
How about a shared task?

Just my humble suggestion :p

Thank YOU…?

Or do we have time for…

Many MORE Things…


zhi1-shi4
tou2-shi4
dao4-shi4
fang1-shi4
chang2-shi4
yi4-shi4
zhi4-shi4
shi4-wei4
shi4-li4
shi4-shang4
shi4-gu4
shi4-yi2
shi4-ji1
shi4-zhong1

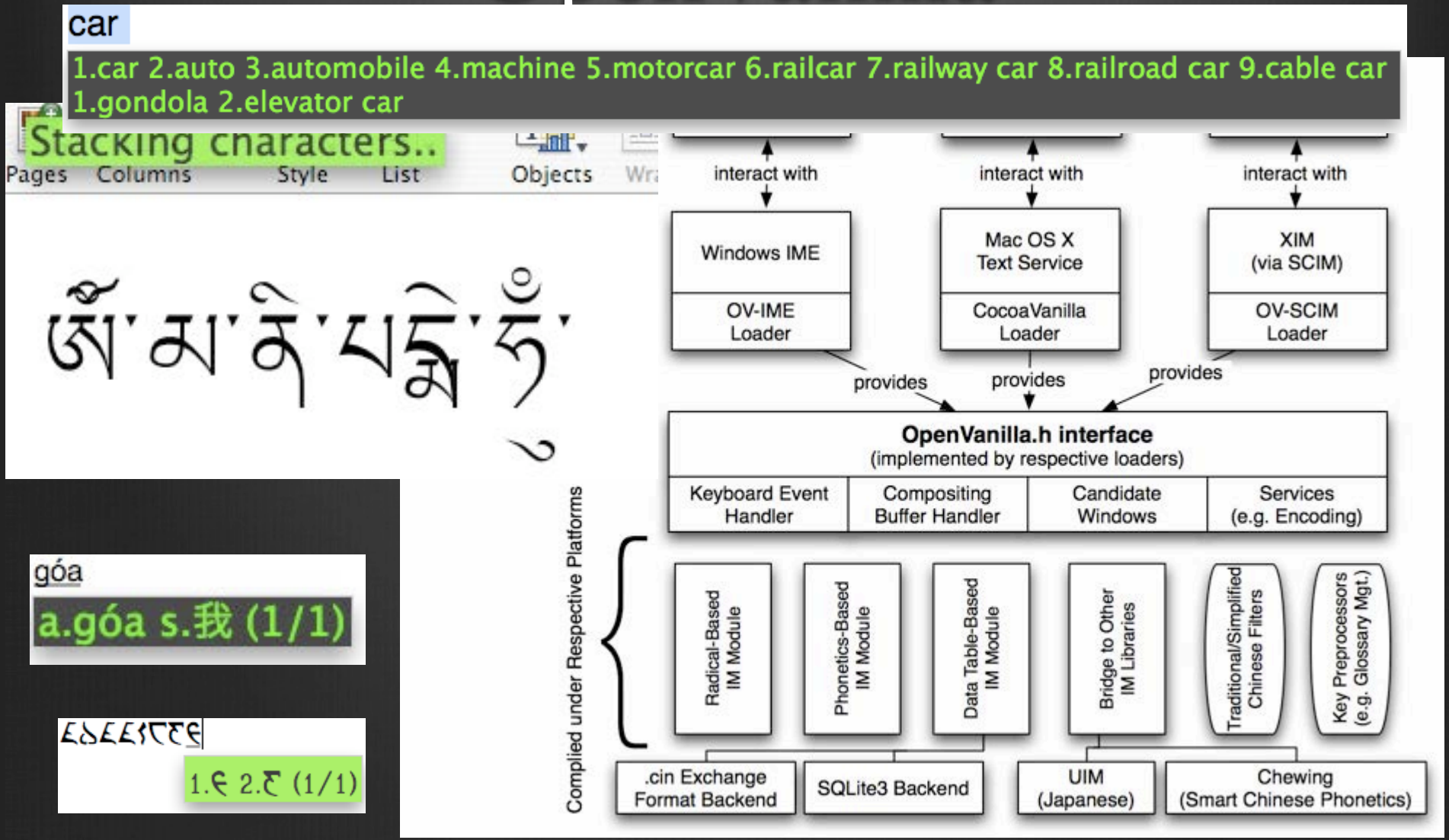
Reduced *n*-gram

British Rail Enquiries

P(Enquiries | British, Rail)

P(Enquiries | Rail)

OpenVanilla



SearchTyping

http://www.youtube.com/watch?v=jgl23E6wzVA

http://www.youtube.com/watch?v=xJsXaPe_f8c

Thank YOU!

Any question or comment?