Cross-linguistic investigations of oral and silent reading

Christophe Coupé[1,2], Yoon Mi Oh[1,2], François Pellegrino[1,2] & Egidio Marsico[1,2]

[1]Laboratoire Dynamique du Langage, CNRS - Université Lyon 2, Lyon, France [2]Institut Rhône-Alpin des systèmes complexes, Lyon, France

Background

- Relationship between oral reading rate and i) linguistic features & ii) general cognitive processes (Chetail, 2014; Ferrand, 2000; Naveh-Benjamin & Ayres, 1986; New et al., 2006)
- Investigation of oral and silent reading rates with respect to the acquisition of reading (Fuchs et al., 2009; O'Brien et al., 2013; Wright, 2011)
- Non-linear positive correlation between self-paced reading time and word expectation (Smith & Levy, 2013)
- Compensation between oral reading rate and average amount of information carried by syllables (Pellegrino et al., 2011)

Research goals

- Address an under-researched question: *how do silent and oral reading rates vary cross-linguistically?*
- Better understand the cognitive and articulatory processes underlying reading: *what is the impact of syllabic complexity on oral and silent reading rates?*
- Study the relationship between text length and reading duration: *what are the effects of increasing word predictability and cognitive load?*

Main findings

- Silent and oral reading rates are strongly correlated across languages

→ **Cross-linguistic differences in word structure complexity influence phonological processing in both reading modes**

- Results from (Pellegrino et al., 2011) are confirmed and extended

→ **Information density and both silent and oral reading rates are negatively correlated at language-level**

- A logarithmic relationship exists between text lengths and reading durations, for both silent and oral reading

→ **Word predictability seems to increase with longer texts.**

- Sex is a significant predictor of oral but not silent reading rate

→ **A sociolinguistic effect of sex when it comes to orality?** (Jacewicz et al., 2009)

- Languages with different writing systems have similar reading rates

→ **The writing system does not seem to impact reading speed**

Perspectives

- Evaluate participants' reading skills and text comprehension (e.g. with self-paced reading) to better assess inter-individual variation
- Record silent and oral rates in a more symmetrical fashion

Data and methodology

Written material

- 15 short English texts from (Campione & Veronis, 1998) translated into 8 languages: *Cantonese* (YUE), *Finnish* (FIN), *French* (FRA), *Japanese* (JPN), *Korean* (KOR), *Mandarin Chinese* (CMN), *Serbian* (SRP) and *Thai* (THA)

Collection of reading times

- Recordings with Rocme! (Ferragne et al., 2013)
- 10 native speakers - 5 men & 5 women - per language; no strict control on age or social diversity
- Two steps: 1) each text read silently, duration recorded 2) each text read aloud three times, speech and duration recorded only the third time
- 15 texts x 10 speakers x 8 languages = 1,200 (subject, text) pairs

Methodology

- Pauses longer than 150ms in the oral recordings discarded with *Praat*
- Computations of information density, syllabic rate and duration;
- Use of Vietnamese as a reference language to normalize computations and avoid quantifying semantic content

Average quantity of information per syllable for each text T_k in language L, composed of $\sigma_k(L)$ syllables:

$$I_L^k = \frac{S_k}{\sigma_k(L)}$$

Syllabic Information Density:

$$ID_L^k = \frac{I_L^k}{I_{VI}^k} = \frac{S_k}{\sigma_k(L)} \times \frac{\sigma_k(VI)}{S_k} = \frac{\sigma_k(VI)}{\sigma_k(L)}$$

Silent and oral reading rates:

$$SilSR_L^{k,sp} = \frac{\sigma_k(L)}{SilD_L^{k,sp}}, OrSR_L^{k,sp} = \frac{\sigma_k(L)}{OrD_L^{k,sp}}$$

$SilD_L^{k,sp}$ and $OrD_L^{k,sp}$: durations of silent and oral readings of text T_k by speaker sp in language L

- 39 (subject, text) pairs removed as outliers - 1161 pairs for analysis
- Correlation coefficients and mixed-effects (M-E) regression models

References

1. Campione, E. & Veronis, J. "A multilingual prosodic database", *Proc. of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, 3163-3166. 1998.
2. Chetail, F. "Effect of number of syllables in visual word recognition: new insights from the lexical decision task", *Language, Cognition and Neuroscience*. DOI: 10.1080/23273798.2013.876504. 2014.
3. Ferragne, E., Flavier, S. & Fressard, C. "ROCme! Software for the recording and management of speech corpora", *Proc. of Interspeech 2013*, Lyon, France, 25-29 august. 2013.
4. Ferrand, L. "Reading aloud polysyllabic words and nonwords: The syllabic-length effect re-examined", *Psychonomic Bulletin & Review*, 7: 142-148, 2000.
5. Fuchs, L. S., Fuchs, D., Hosp, M. K. & Jenkins, J. R. "Scientific Studies of Reading Oral Reading Fluency as an Indicator of Reading Competence : A Theoretical, Empirical, and Historical Analysis", *Sci. Stud. Read.*, 5(3): 239–256, 2009.
6. Jacewicz, E., Fox, R.A., O'Neill, C. & Salmons, J. "Articulation rate across dialect, age, and gender", *Language Variation and Change*, 21(2): 233-256, 2009.
7. Naveh-Benjamin, M. & Ayres, T. J. "Digit Span, Reading Rate, and Linguistic Reactivity", Q. J. Exp. Psychol., 38A: 739–751, 1986.
8. New, B., Ferrand, L., Pallier, C. & Brysbaert, M. "Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project", *Psychonomic Bulletin & Review*, 13: 45-52. 2006.
9. O'Brien, B. A., Wallot, S., Haussmann, A. & Kloos, H. "Using Complexity Metrics to Assess Silent Reading Fluency: A Cross-Sectional Study Comparing Oral and Silent Reading", *Sci. Stud. Read.*: 1-20, 2013.
10. Pellegrino, F., Coupé, C. & Marsico, E. "Across-language perspective on speech information rate", *Language* 87: 539–558. 2011.
11. Smith, N. J. & Levy, R. "The effect of word predictability on reading time is logarithmic", *Cognition*, 128(3): 302–19, 2013.
12. Wright, J. A. "The Impact of Oral Fluency and Silent Fluency on the Comprehension of Fourth Graders", *Diss. Louisiana State University*, 2011.

Results

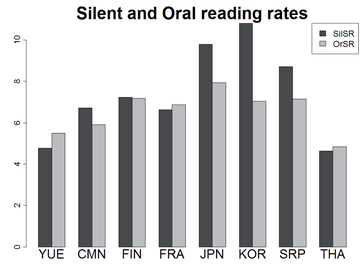

Figure 1: Silent and oral reading rates (in #syl/s)

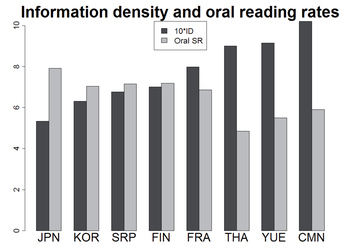

Figure 2: ID (unitless) and ORSr (#syl/s)

Data set	Correlation coef.
All data (N = 1161)	Pearson's R: .60***
Averaged by speaker (N = 80)	Pearson's R: .67***
Averaged by language (N = 8)	Spearman's Rho: .81**

Table 1: Correlation between silent and oral SR

Comparison of oral and silent syllabic rates

- Noticeable differences of SilSRs and OrSRs between languages (fig. 1)
- Strong positive correlations between silent and oral reading rates (table 1)
- Result confirmed by M-E model: significant effects on *SilSR* of *OrSR*, *Language* and *Sex* as fixed effects, and of *Text* and *Subject* as random effects ($p < .001$*** for all effects). No effect of *Sex*

Balance between Information Density and Syllabic Rate

- Strong negative correlation between *ID* and both *OrSR* and *SilSR* at language level (Spearman's Rho = -.81, p=.021*) (fig. 2)
- Result confimed by M-E models: significant effects of *ID*, *Language*, *Text* and *Subject* ($p < .001$***) on both *SilSR* and *OrSR*. Significant effect of *Sex* only on *OrSR* (p=.019*)

Relation between duration and text length

- Weak correlation between *SilD* and the number of syllables (σ) (Pearson's $R = .11$, $p < .001$***), stronger correlation between *OrD* and σ (Pearson's $R = .71$, $p < .001$***)
- Comparison of 3 different M-E models with *OrD* as dependent variable, *Sex* and either i) σ, ii) $\log(\sigma)$ or iii) $\exp(\sigma)$ as fixed predictors, and *Language*, *Text*, and *Subject* as random predictors:

→ Significant effects for *Text*, *Subject*, *Language* and *Sex* in all three models

→ Best prediction obtained with $\log(\sigma)$. Significant improvement over the two other models ($p < .001$***)

- Similar results with *SilD*, but weaker prediction and no effect of *Sex*

Financial support: «Investissements d'Avenir» Program, operated by the French National Research Agency (ANR-11-IDEX-0007) - ASLAN Laboratory of Excellence (ANR-10-LABX-0081), Université of Lyon

Contact: Christophe COUPE, Christophe.Coupe@cnrs.fr