Within-Word Text Segmentation on Lexical Processing and Reading

Yu-Chi Tai
Shun-nan Yang
Andrew Reder
John Hayes
James Sheedy
Why part a word into segments?

• Frst, let’s look at some txt msgz...

  – Hey can u do me a favour, take a pic of urself n send me it, i'm playin cards n i'm missin the joker!!

  – i want u 2 know that our friendship means a lot 2 me. U cry i cry. U lauf i lauf. U jump out’f a window... I look down & then... i lauf again.

(visual ➔ sound ➔ semantic transformation)
The Question

• Will a word be more readily for process if presented in a way that is consistent with how we mentally represent or processing the word?

Specifically...
• Will visually segmenting a word into units based on certain hypothesized processes differentially affect lexical processing?

Some hypothesized lexical processes:
  – Familiar words are quickly identified as a whole... through...
    • Whole word identification?
    • Word shape?
    • ...
  – Alternatives may be needed for longer or unfamiliar words...
    • Sound processing...
    • Morpheme processing...
    • Other sub-word unit activation...
Experiment 1. Effect of within-word segmentation on threshold word recognition

• Subjects:
  – 46 native-English speakers (34F, 12M; age 18-60, mean=23.8)
  – 20/20+ binocular vision

• Stimulus:
  – 5 words a row, 7-13 letters of similar frequency (10-15 time/10⁶ words)
  – in 12-pt Consolas (8 x 9 pixels per character, 0.234° x 0.264° from 50 cm)
  – 3 segmentation conditions: default, morpheme-seg., syllable-seg.
  – Segmentation by inserting 2 extra white pixels (0.06°)

• Task: Step-back threshold visual acuity test
(a) Testing view
  dependent  --  default (no segmentation)
  dependent  --  morpheme segmentation
  dependent  --  syllable segmentation

(b) Enlarged view (600%)
Result:
Better **Threshold legibility** with segmentation

*Figure 2.* Effect of within-word segmentation on threshold size of word recognition. The threshold size represents the minimum angular size where a word can still be recognized, at the logMAR unit. Smaller logMAR indicates better legibility or being able to be recognized from a smaller angular size. The boxes show the 84% confidence interval of the estimated threshold size. The middle line inside a box represents the estimated mean. Non-overlapped boxes indicate significant difference between the corresponding conditions at $\alpha=0.05$. This rule applies to the following figures.
Experiment 2: Effect of prime segmentation on single word naming and lexical decision

- Subjects:
  - 49 native-English speakers (27F, 22M; Age 18-60, mean=25.7)
  - 20/25+ binocular near vision (33 optometry students)

- Materials
  - Priming effect on single word naming and lexical decision
    - Word naming (read aloud the word)
    - Lexical decision (Can the word be used as a noun)
  - 4 Segmentation types: syllable, morpheme, default, default length matched to syllable
  - Word length (7-14 letters): 7-9 vs. 10-14 letters
  - Word frequency: High frequency, Low frequency, Pseudowords
    - verb-like endings: -aim, -erge, -ieve, -olve
    - noun-like endings: -int, -ior, -ium, -oon
Fixation cross (until button press)

SOA (500 ~ 1500 milliseconds)

button press

Prime: in one of 3 formats
- conducive - unsegmented
- conducive - syllable segmented
- conducive - morpheme segmented (80 milliseconds)

Mask (500 milliseconds)

Target word: in one of 3 groups
- High frequency word
- Low frequency word
- Pseudoword (stay until oral response)

Next trial...
Results: Depend on the task...
Lexical decision: morpheme; Naming: syllable

Accuracy:
• Lexical decision
  – Segmentation effect: **Morpheme** is better
  – Frequency effect: High-freq > Low-freq > pseudowords
  – Length effect: 7-9 letters > 10-14 letters

• Word naming
  – Segmentation effect: **Syllable** is better
  – Frequency effect: High-freq > Low-freq > Pseudowords
  – Length effect: 7-9 letters > 10-14 letters
Accuracy:
• Significant improvement with morpheme segmentation for lexical decision
• No significant effect for read aloud
RT

• Lexical decision
  – Segmentation effect: no sig. effect
  – Frequency effect: High < Low < pseudowords
  – Length effect: 7-9 > 10-14

• Word naming
  – Segmentation effect: no sig. effect
  – Frequency effect: High < Low < pseudowords
  – Length effect: 7-9 > 10-14
Experiment 3: Segmentation effect on parafoveal preview and sentence comprehension

• Task:
  • Read 60 sentences under the text disappearing paradigm.
  • Each sentence was segmented entirely in one of 3 conditions (default, morpheme, syllable).
  • A different statement was presented after subject finished reading a sentence. Subject was asked to respond whether this statement was congruent with the previous sentence. A true/false response was required for each statement.
The honest garbageman returned the money to the owner.

A garbageman found money and returned to the owner. (expected answer: YES)

A professor returned the money to the owner. (Expected answer: No)

Drift correction (until button press)

Sentence to read (until button press)

Button press

Next trial...
Example of sentence display

Reading progress...

1. Fixation before sentence display

2. Beginning of sentence reading

3. Beginning of fixation on the word N-1 before the target word N

4. 60~80 ms after the onset of fixation on word N-1 till button press

Sentence display and eye position...

The honest garbageman returned the money to the owner.

The honest garbageman returned the money to the owner.

The honest returned the money to the owner.
Result:

- Sentence comprehension accuracy was overall low (< 50%) for the task → A brief (80 ms) preview of the target word from periphery is not enough to catch it.
- Syllable segmentation made it even worse (more chucks... harder to see?)

![Graph showing estimated sentence comprehension accuracy in Exp 2]
Experiment 4: Segmentation effect on whole passage reading

• Read “The Naval Treaty” by Arthur Conan Doyle for 20 minutes
• Each page was randomly selected among the 3 segmentation conditions.
• Comprehension test:
  – 1 multiple-choice question for every page;
  – The questions are considered page-independent, which means each question based on the text in the page just-read without linking to earlier text.
Results of Exp. 4

• Effect of within-word segmentation on a complicated task like reading is subtle and is likely modulated by individual’s reading strategy and word decoding skills.

• Using word decoding skills (from Exp 2) and reading speed (from Exp 4) as covariates, we found ...
  – Reading comprehension was improved ...
    • 10% for good word decoders, whether they read slow or fast.
    • 4 ~ 5% for poor word decoders, if they read slowly.
  – Reading speed was slowed down with segmentation.
Results of Exp. 4

– Fixation duration was slightly increased for faster readers but decreased for poorer readers → enhance their word-decoding efficiency;

– Saccade amplitude was longer for faster readers with both segmentations, and longer for slower readers only with syllable conditions.
  • The increment in saccade amplitude was about .05°, slightly larger than the space of the 2 inserted pixels.

– Words per fixation was slightly more with segmentation for Slower word decoders → Larger portion of words was processed w/ segmentation for slower decoders.
Result of Exp. 4

– Slightly higher blink rate with both segmentation conditions than with default.

Overall, these results suggest that...

Within-word segmentation seemed to help reading comprehension and word decoding, but at the cost of reading speed, possibly from slightly longer fixation duration and higher blink rate (which can help relieve some discomfort from cornea surface).
Comprehension:
• For good word decoders, reading comprehension was 10% better with segmentation;
• For poor word decoders, comprehension improved with syllable segmentation when reading slowing down.
Reading speed:
• 10 ~ 15 words/min slower with segmentation.
• Faster word decoders and good comprehension were related to better reading speed.
Fixation duration
• Faster readers dwelled longer with segmentation conditions;
• Slower readers stayed shorter in syllable segmentation condition.

Saccade amplitude
• No gain for saccade amplitude.
Words per fixation
• Slower word decoders processed slightly more words in each fixation (or “use fewer fixations to process each word”) with segmentation than in the default condition.
• No difference on EM efficiency among segmentation for faster decoders, though tended to be better without.
Blink rate increased with segmentation

Estimated blink rate (number of blinks per minute) in Exp. 3
Conclusion

The effect of within-word segmentation depends on the task:

– With unlimited access (exp1), it improves the accuracy of word recognition.

– With short (e.g., 80ms of priming) exposure, segmentation effect varies by tasks: morpheme seg. helps lexical decision and syllable seg. helps read-aloud.

– With brief (80 ms) peripheral preview, it does not help.

– With regular sentence reading, it interacts with reader’s strategy and capability.

– In general, syllable segmentation helps poor word decoders and morpheme segmentation helps good decoders more.

Within-word segmentation can be a tool to assist readers of different levels...
Acknowledgement

This study was supported by
Microsoft Advanced Reading Technology Group.