The Meta-Science of Adults Statistical Word Segmentation: Part 1

Lauren Skorb
Joshua K. Hartshorne
Department of Psychology, Boston College
Poster available at iSatbc.org

BACKGROUND & MOTIVATION

Statistical word segmentation:
- Identification of words based solely on the statistical properties of speech stream (e.g., Romburg & Saffran, 2010)
- Theoretically important: learning language with little explicit instruction
- Extensively studied phenomenon: 100+ published experiments

Meta-Science:
- More information about our literature in order to interpret results and improve methodologies
- Replicability varies by discipline, but little research on the robustness of psycholinguistics (Aarts & Laats, 2016; Open Science Collaboration, 2015)
- Help determine power for specific literatures (Miehswardt et al., 2018)
- Better understand & utilize methods (e.g., Belward et al., 2011)

Research questions:
Q1 What are the most reliable findings? Where is more investigation most useful?
Q2 What factors influence replicability?
Q3 How reproducible is the literature?
Q4 Are studies sufficiently powered?

OVERVIEW OF METHODOLOGY

Data and analyses available at: https://osf.io/ehu7q/

1. Online replications: (Q1, Q2 & Q3)
   Part 1: 6 of 130 experiments replicated
     [2-3] Exp. 1 and 2 from Saffran et al. (1999)
     [4-5] Exp. 1 and 3 from Finn & Hudson Kam (2008)
     Ongoing: Replicate adult word segmentation exps (100+)

2. Three Replications: Online & In Lab: (Q1 & Q2)
   Part 1: Comparison of in-lab and online replications of [1]

3. Meta-analysis: (Q1 & Q4)
   Part 1: Preliminary p-curve analyses of main effects and modulators (30 of 148 studies coded)
   Ongoing: Complete p-curve analyses and meta-analytic regression analyses

** Help us confirm that we have collected all of the relevant papers by reviewing our list here: https://goo.gl/forms/Klq0xAC12kzixex1

THE THREE REPLICATIONS: ONLINE & IN LAB

<table>
<thead>
<tr>
<th>Exp. 1 of Saffran et al. 1996</th>
<th>Attention screen?</th>
<th>Main effect</th>
<th>Modulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original in lab</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online rep. 1</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online rep. 2</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In lab rep.</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A1 & A2 Main effects are robust—the pattern of significance remains the same with changing venues, samples, and attention screen

A1 & A2 Modulating effects are not robust—we fail to replicate pattern of significance with changing venues, samples, and attention screen

CONCLUSIONS

A1: Main effect (ME) of statistical word segmentation is most reliable finding, while the modulators do not appear to be reliable.

1. Meta analysis: P-curves show strong evidence for both main effects & modulators (significant right skew)
2. Online replications: We replicate ME but not modulators
3. In-lab replication: ME replicates w/o attention screen

A2: Preliminary observation: The most reliable findings have larger sample sizes, more test items, & open materials/data.

A3: We ran into issues in trying to reproduce some of the original studies including errors in the papers & obsolete technology.

A4: The p-curve analyses show that the literature is sufficiently powered to detect the main effect and modulators; however, our failure to replicate the modulating effects suggests low power in the selected studies.

P-curve definition:
A tool that plots the reported p-values in a literature. True effects show a right skew, and a left skew suggests selective reporting or p-hacking (Simonsohn et al., 2014)

Our analyses:
P-curve analysis of 62 main effects (top) and 19 modulators (bottom)

Results:
A1: P-curve analyses find evidential value for both main effects & modulators (significant right skew)
A4: Both main effects and modulators show significant power (99% and 97%)