Word Shift: A General Method for Visualizing and Explaining Pairwise Comparisons Between Texts

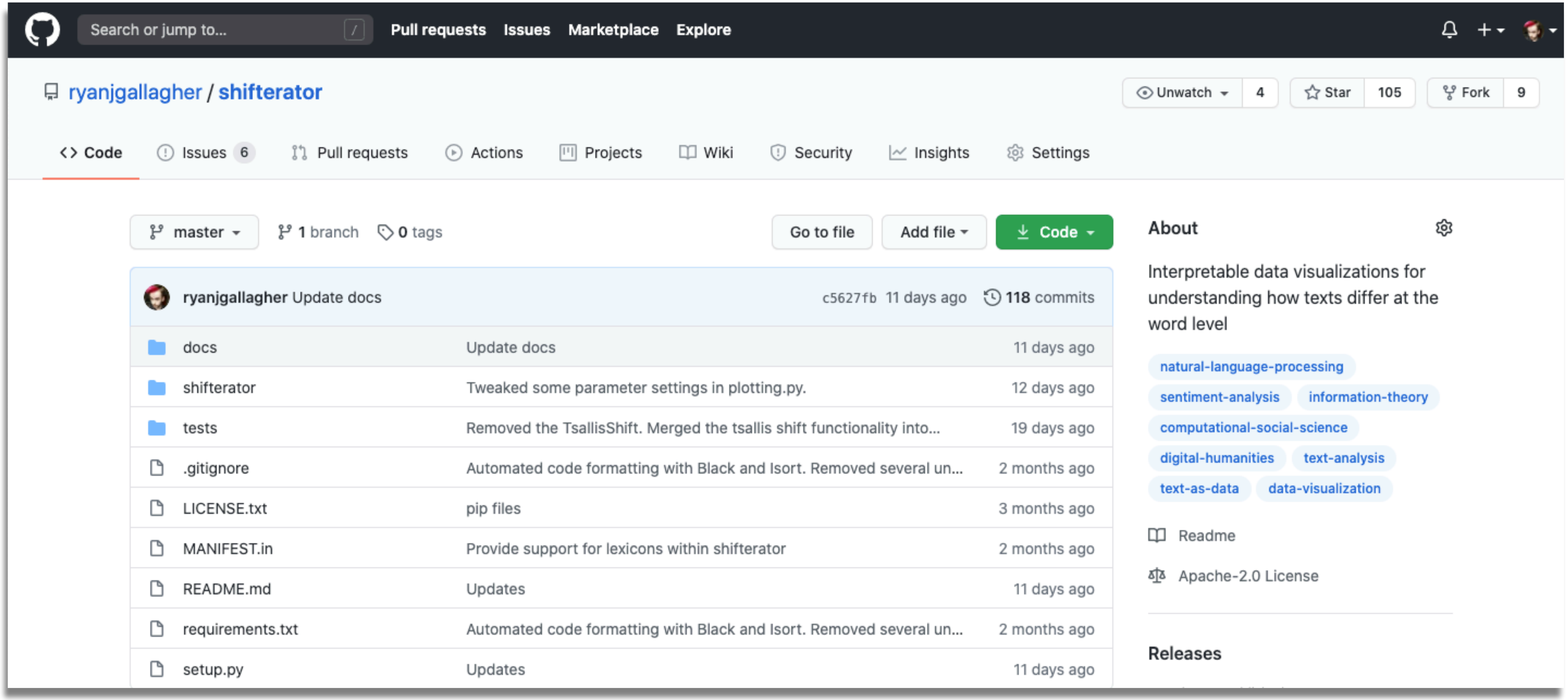
Ryan J. Gallagher
@ryanjgallag

netsi
Northeastern University
Network Science Institute

CoMM Lab

Talk Outline

1. Review common text comparison measures, including dictionary measures
2. Show how differences between texts can be visualized at the word level
3. Review the basic form of the word shift graphs
4. Introduce generalized word shift graphs for weighted averages
5. Discuss a case study about Twitter and 280 character tweets



https://github.com/ryanjgallagher/shifterator

https://shifterator.readthedocs.io

```
pip install shifterator
```

How do we compare two texts?

Measures for Comparing Texts: Proportions

One of the simplest ways of comparing two texts is by comparing how often a word appears in each of them

Measures for Comparing Texts: Proportions

One of the simplest ways of comparing two texts is by comparing how often a word appears in each of them

If τ is a word in our vocabulary, then we compare its relative frequency in each text

$$\delta p_\tau = p_\tau^{(2)} - p_\tau^{(1)}$$

Measures for Comparing Texts: Proportions

One of the simplest ways of comparing two texts is by comparing how often a word appears in each of them

If τ is a word in our vocabulary, then we compare its relative frequency in each text

$$\delta p_\tau = p_\tau^{(2)} - p_\tau^{(1)}$$

We can rank words by this difference!

$p_2 - p_1 > 0$ word is more common in second text

$p_2 - p_1 < 0$ word is more common in first text

Proportion Shift

Case study: presidential speeches by Lyndon B. Johnson and George W. Bush

Proportion Shift

Case study: presidential speeches by Lyndon B. Johnson and George W. Bush

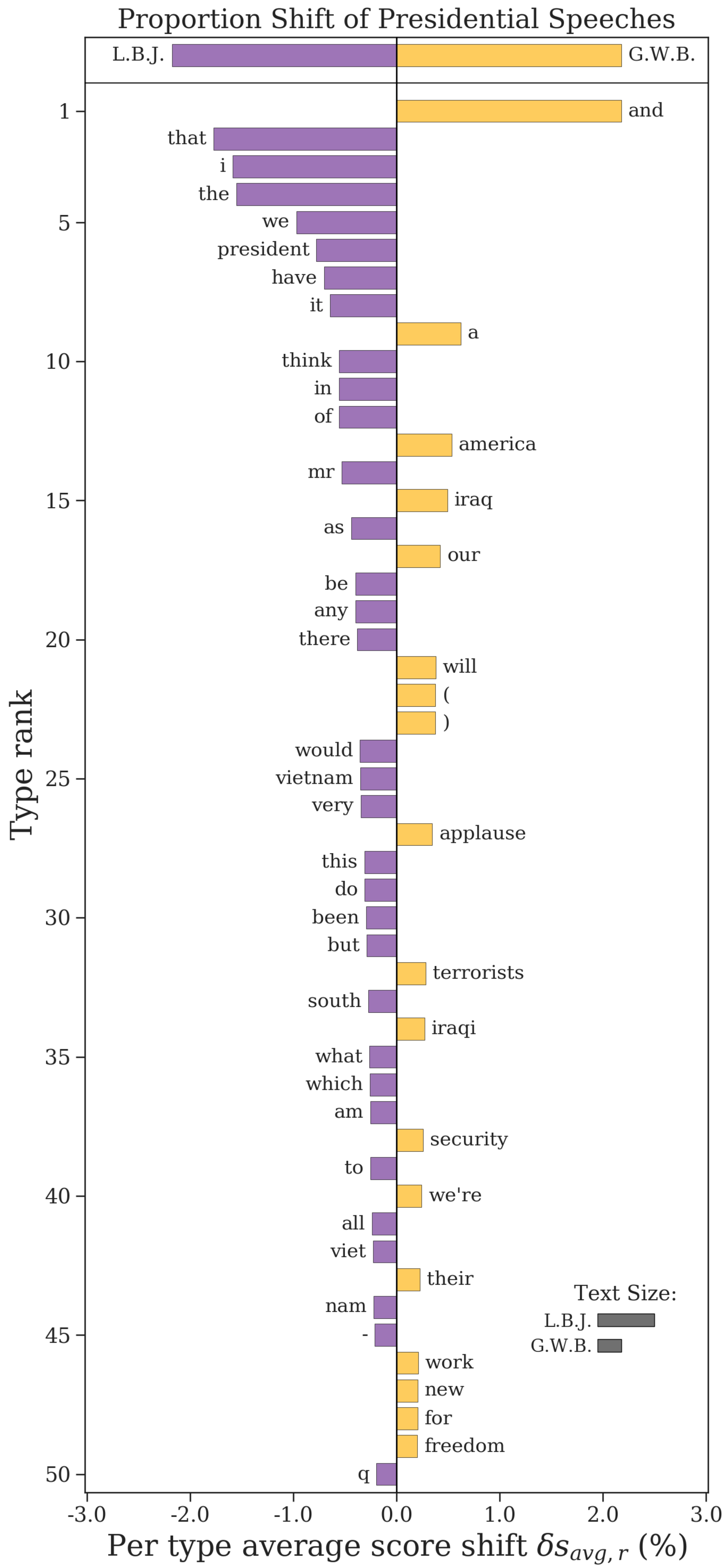
$$\delta p_\tau = p_\tau^{(G.W.B.)} - p_\tau^{(L.B.J.)}$$

Proportion Shift

Case study: presidential speeches by Lyndon B. Johnson and George W. Bush

$$\delta p_\tau = p_\tau^{(G.W.B.)} - p_\tau^{(L.B.J.)}$$

Used relatively more by L.B.J



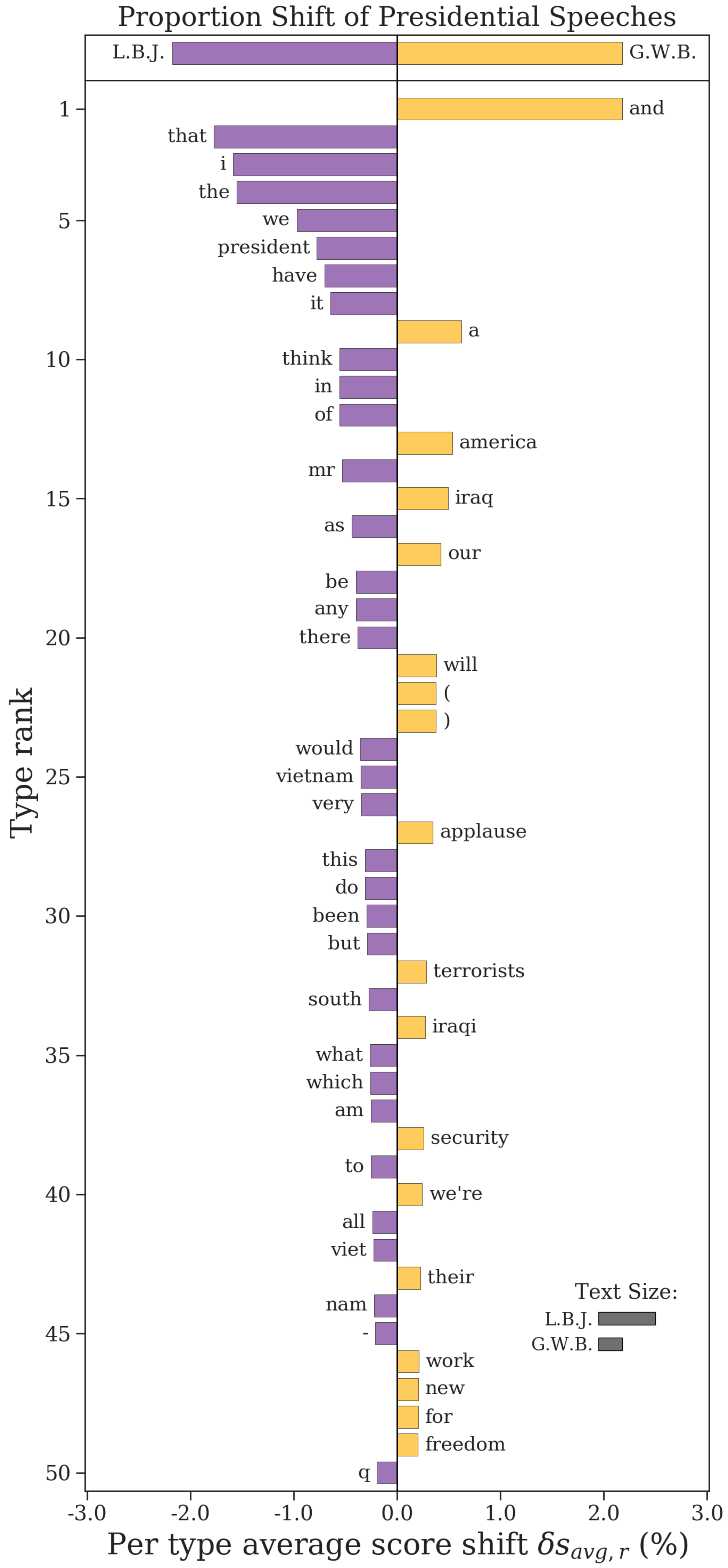
Used relatively more by G.W.B.

Proportion Shift

Case study: presidential speeches by Lyndon B. Johnson and George W. Bush

$$\delta p_\tau = p_\tau^{(G.W.B.)} - p_\tau^{(L.B.J.)}$$

Used relatively more by L.B.J



Used relatively more by G.W.B.

Relative text size comparison

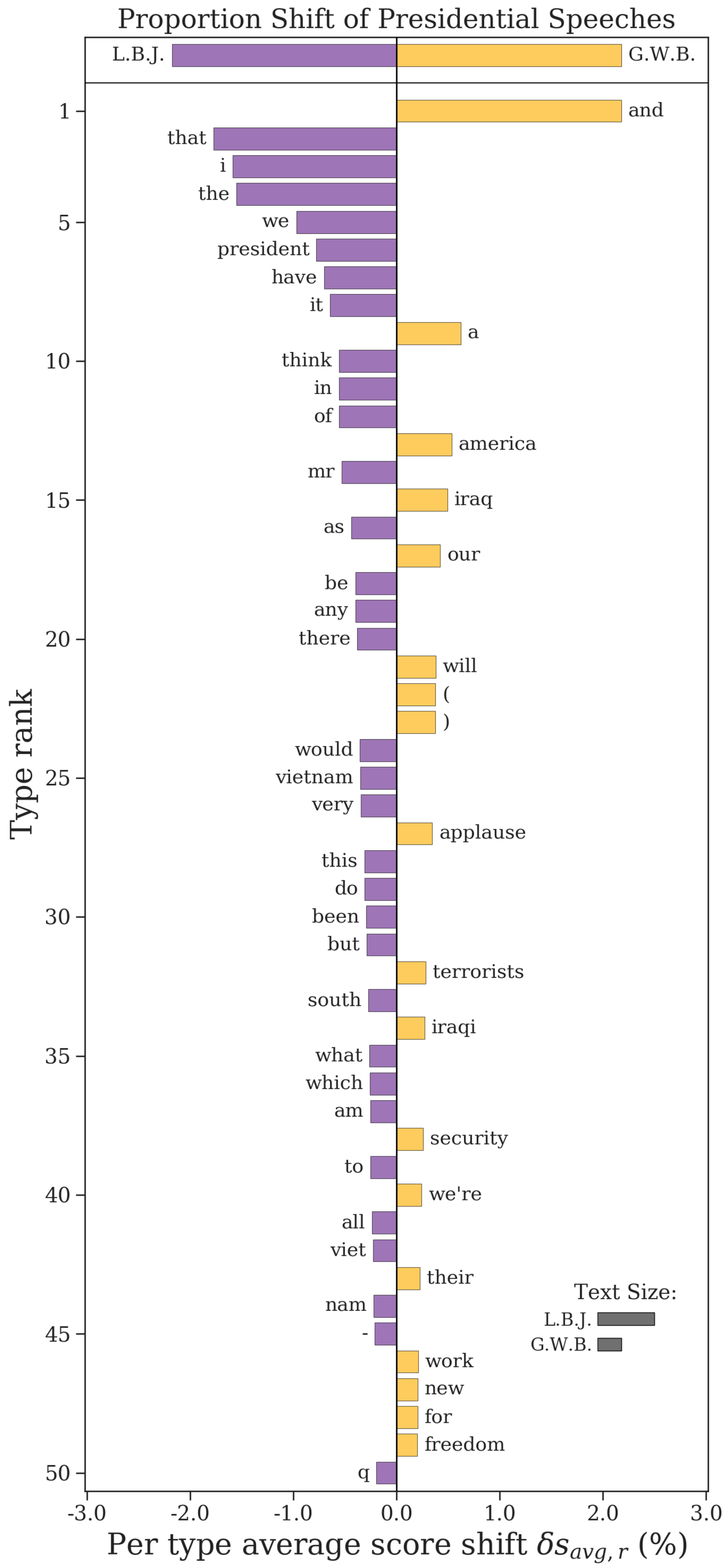
Over 2x as much text in L.B.J's speeches compared to G.W.B.

Proportion Shift

Case study: presidential speeches by Lyndon B. Johnson and George W. Bush

$$\delta p_\tau = p_\tau^{(G.W.B.)} - p_\tau^{(L.B.J.)}$$

Used relatively more by L.B.J



Used relatively more by G.W.B.

Relative text size comparison

Over 2x as much text in L.B.J's speeches compared to G.W.B.

```
import shifterator as sh

p_shift = sh.ProportionShift(type2freq_1=type2freq_1,
                             type2freq_2=type2freq_2)
```

Measures for Comparing Texts: Shannon Entropy

Entropy attempts to account for both how frequent and how “surprising” each word is

$$H(P) = \sum_{\tau} p_\tau \log \frac{1}{p_\tau}$$

Measures for Comparing Texts: Shannon Entropy

Entropy attempts to account for both how frequent and how “surprising” each word is

$$H(P) = \sum_{\tau} p_\tau \log \frac{1}{p_\tau}$$

surprisal
of word τ

Measures for Comparing Texts: Shannon Entropy

Entropy attempts to account for both how frequent and how “surprising” each word is

$$H(P) = \sum_{\tau} p_{\tau} \log \frac{1}{p_{\tau}}$$

average
surprisal

Measures for Comparing Texts: Shannon Entropy

Entropy attempts to account for both how frequent and how "surprising" each word is

$$H(P) = \sum_{\tau} p_\tau \log \frac{1}{p_\tau}$$

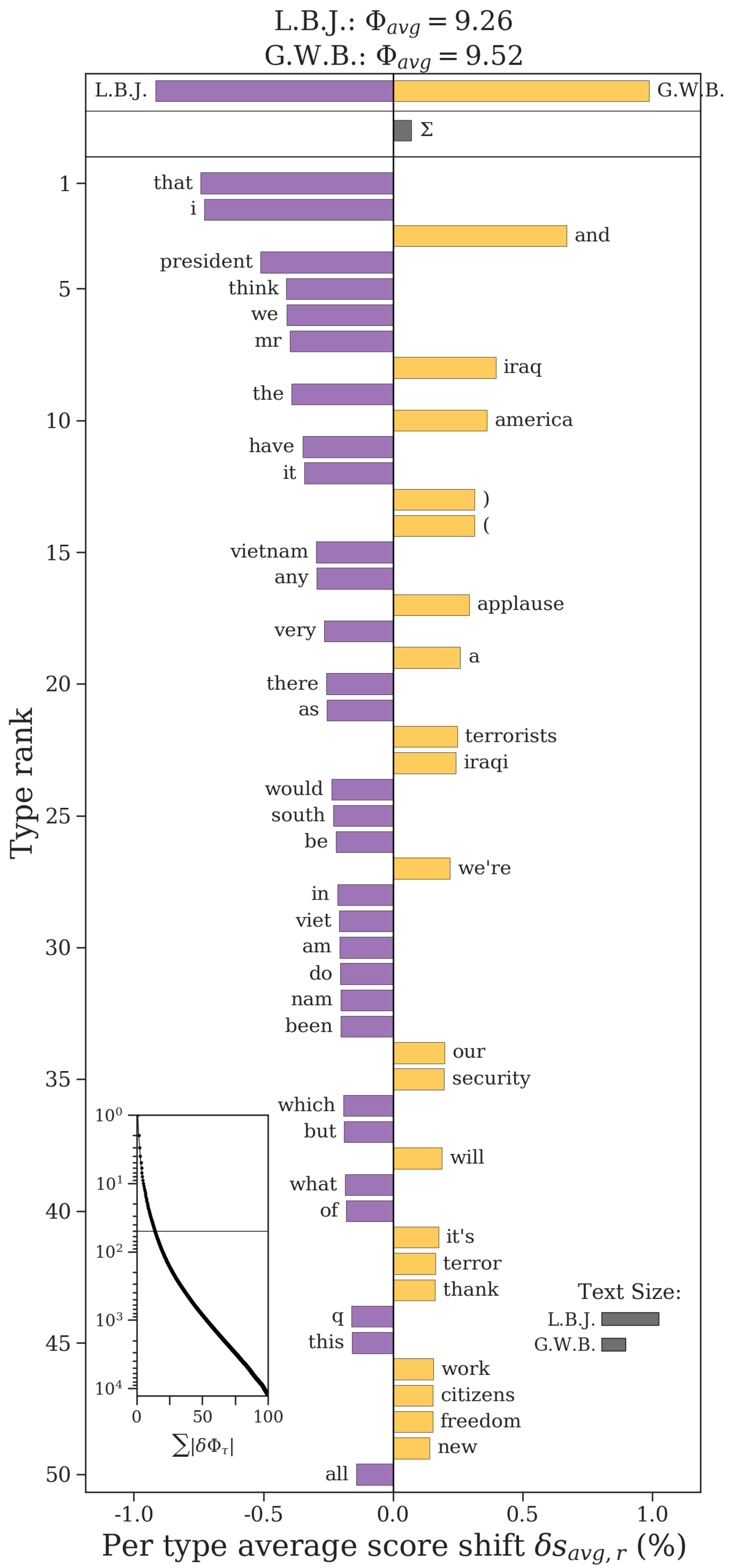
We can compare two texts by comparing contributions to the entropy of each text

$$\delta H = H\left(P^{(2)}\right) - H\left(P^{(1)}\right) = \sum_{\tau} p_\tau^{(2)} \log \frac{1}{p_\tau^{(2)}} - p_\tau^{(1)} \log \frac{1}{p_\tau^{(1)}}$$

Shannon Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

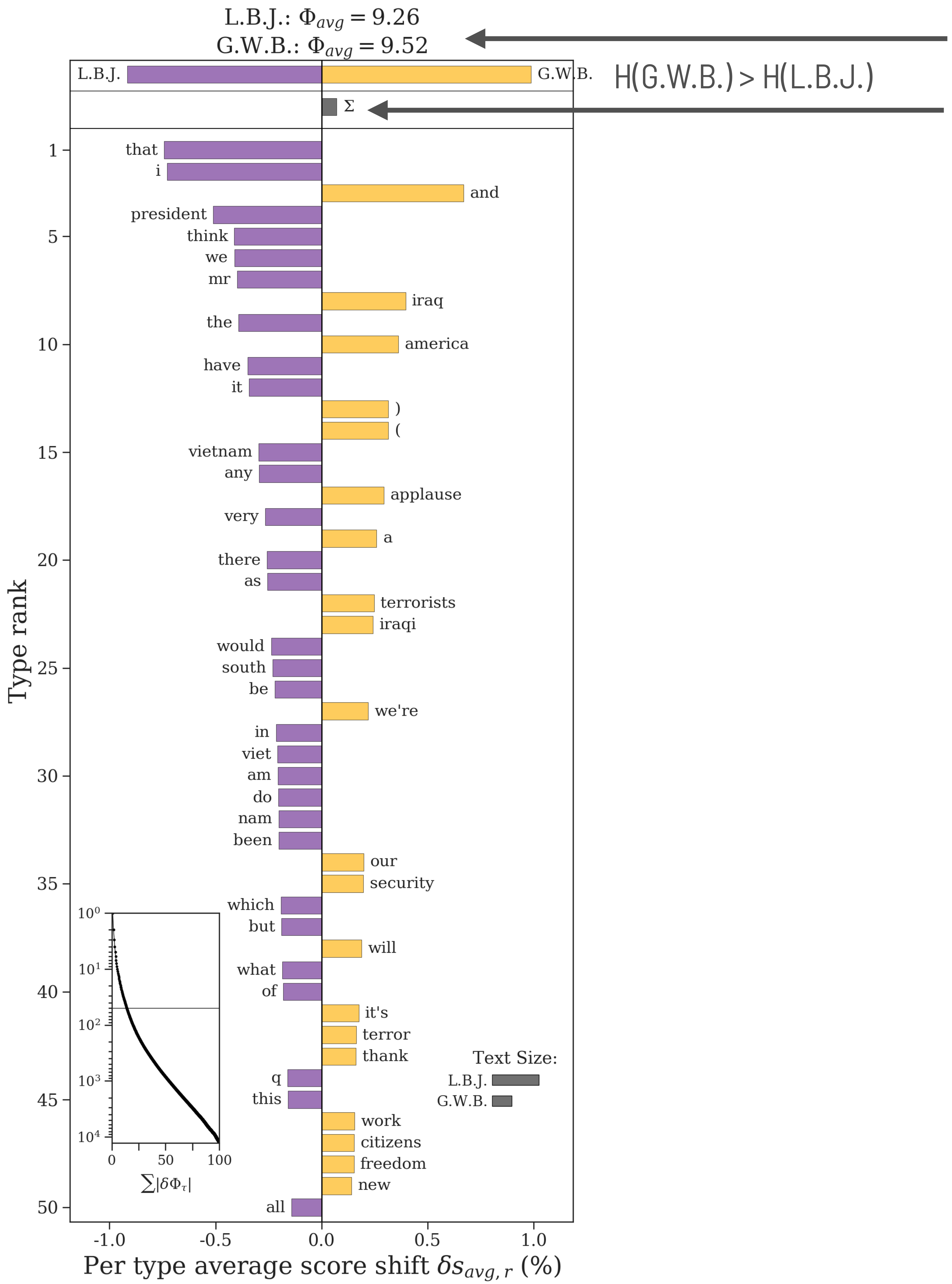
Important for interpreting word direction



Shannon Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

Important for interpreting word direction



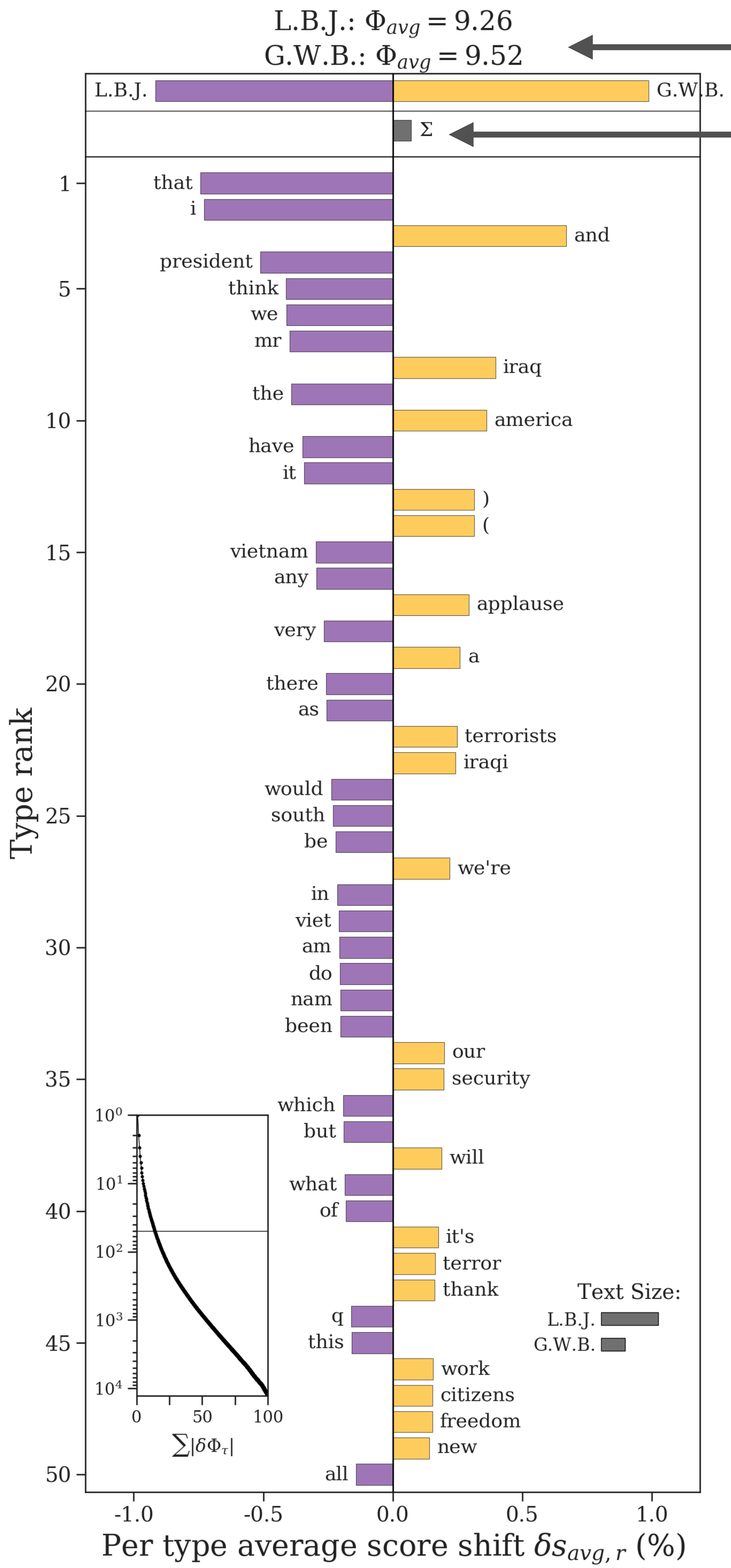
Shannon Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

Important for interpreting word direction

Counteract H(GWB) > H(LBJ)

Entropy difference would be even greater otherwise



H(G.W.B.) > H(L.B.J.)

Directly make H(GWB) > H(LBJ)

Shannon Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

Important for interpreting word direction

H(G.W.B.) > H(L.B.J.)

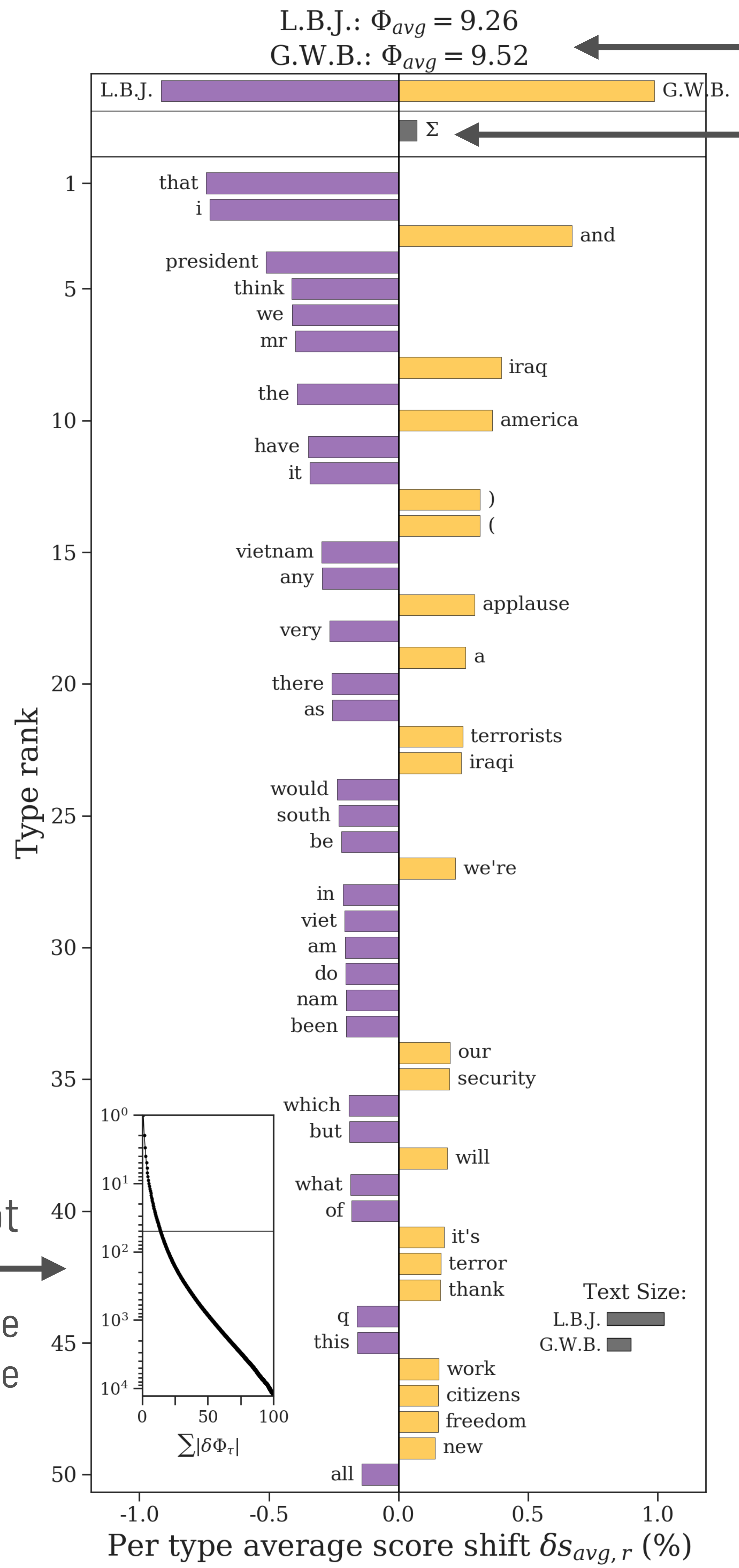
Counteract H(GWB) > H(LBJ)

Entropy difference would be even greater otherwise

Directly make H(GWB) > H(LBJ)

Cumulative contribution plot

Only a small fraction of the total entropy difference is explained by the top 50 words visualized here



Shannon Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

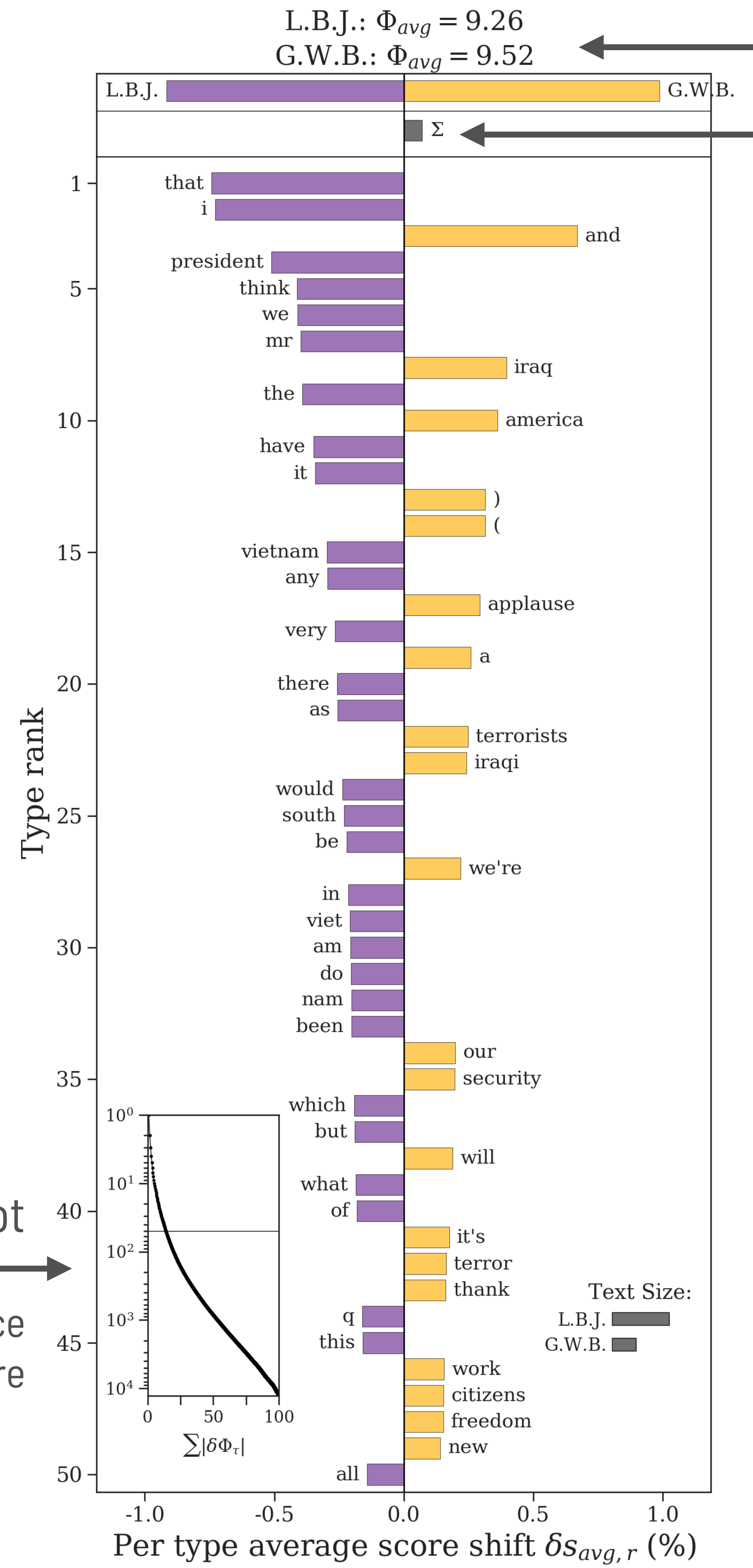
Important for interpreting word direction

Counteract H(GWB) > H(LBJ)

Entropy difference would be even greater otherwise

Cumulative contribution plot

Only a small fraction of the total entropy difference is explained by the top 50 words visualized here



H(G.W.B.) > H(L.B.J.)

Directly make H(GWB) > H(LBJ)

```
e_shift = sh.EntrophShift(type2freq_1=type2freq_1,
                          type2freq_2=type2freq_2,
                          base=2)
```

Measures for Comparing Texts: Tsallis Entropy

We can generalize entropy to emphasize either common or uncommon words

$$H_\alpha(P) = \frac{1}{1-\alpha}\left(\sum_\tau p_\tau^\alpha - 1\right)$$

$\alpha < 1$ emphasizes rare words

$\alpha = 1$ balances between rare and frequent words, equivalent to Shannon entropy

$\alpha > 1$ emphasizes common words

Measures for Comparing Texts: Tsallis Entropy

We can generalize entropy to emphasize either common or uncommon words

$$H_\alpha(P) = \frac{1}{1-\alpha}\left(\sum_\tau p_\tau^\alpha - 1\right)$$

Like the Shannon entropy, we can difference between the Tsallis entropies of two texts

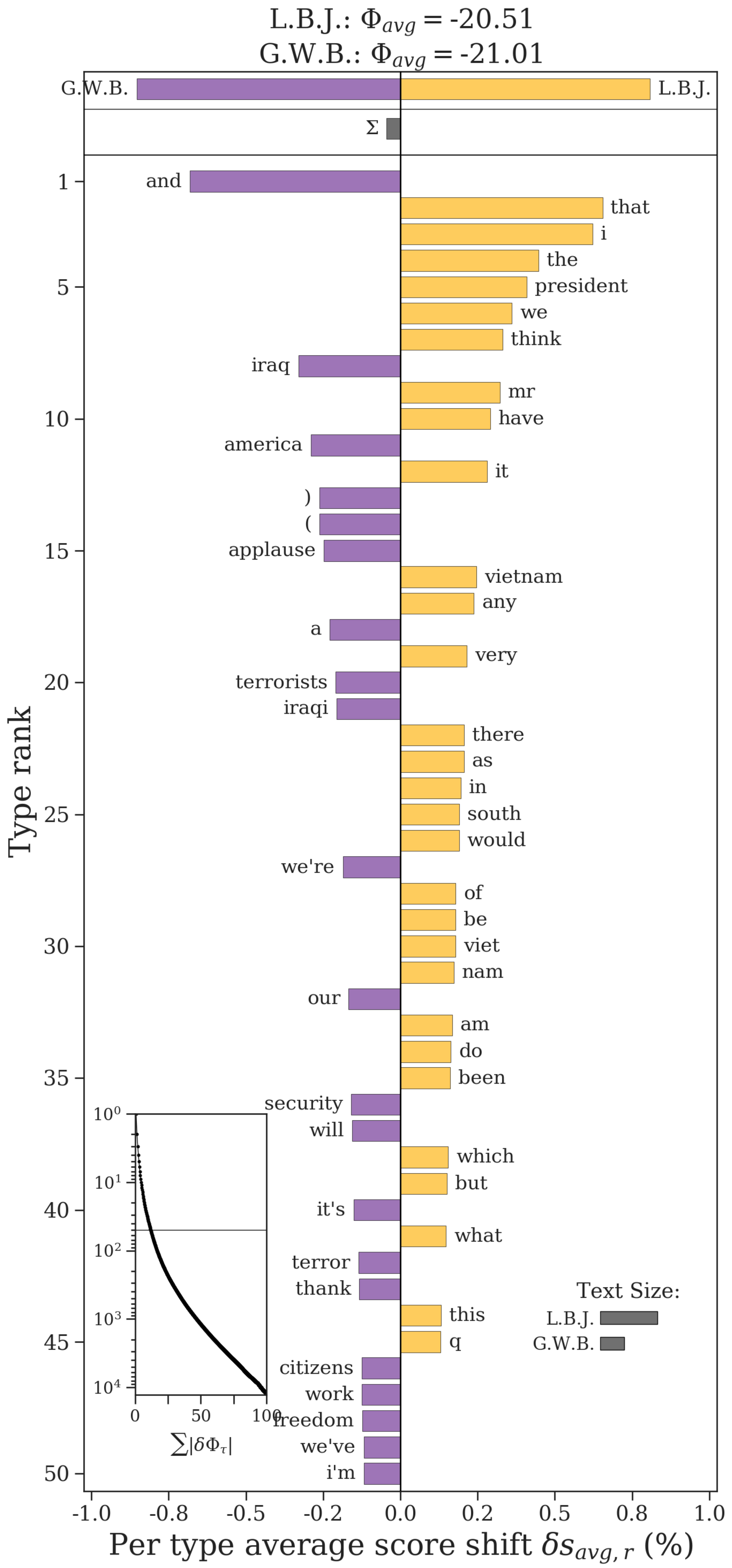
$$\delta H_\alpha = H_\alpha\left(P^{(2)}\right) - H_\alpha\left(P^{(1)}\right) = -\,p_\tau^{(2)}\left[\frac{\left(p_\tau^{(2)}\right)^{\alpha-1}}{\alpha-1}\right] + p_\tau^{(1)}\left[\frac{\left(p_\tau^{(1)}\right)^{\alpha-1}}{\alpha-1}\right]$$

Tsallis Entropy Shift

Note: We're calculating H(G.W.B) - H(L.B.J)

Here, $\alpha = 0.8$

```
e_shift = sh.EntrophShift(type2freq_1=type2freq_1,
                          type2freq_2=type2freq_2,
                          base=2,
                          alpha=0.8)
```



Measures for Comparing Texts: Kullback-Leibler Divergence

Sometimes we want to compare one text to a *reference text*

Measures for Comparing Texts: Kullback-Leibler Divergence

Sometimes we want to compare one text to a *reference text*

Say $P^{(1)}$ is the reference, and $P^{(2)}$ is the comparison. The *Kullback-Leibler divergence* (KLD) is

$$D^{(KL)}\left(P^{(2)}\,||\,P^{(1)}\right) = \sum_{\tau} p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(1)}} - p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(2)}}$$

Measures for Comparing Texts: Kullback-Leibler Divergence

Sometimes we want to compare one text to a *reference text*

Say $P^{(1)}$ is the reference, and $P^{(2)}$ is the comparison. The *Kullback-Leibler divergence* (KLD) is

$$D^{(KL)}\left(P^{(2)} \,||\, P^{(1)}\right) = \sum_{\tau} p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(1)}} - p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(2)}}$$

surprisal of $p_{\tau}^{(1)}$

surprisal of $p_{\tau}^{(2)}$

Measures for Comparing Texts: Kullback-Leibler Divergence

Sometimes we want to compare one text to a *reference text*

Say $P^{(1)}$ is the reference, and $P^{(2)}$ is the comparison. The *Kullback-Leibler divergence* (KLD) is

$$D^{(KL)}\left(P^{(2)} \,||\, P^{(1)}\right) = \sum_{\tau} p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(1)}} - p_{\tau}^{(2)} \log \frac{1}{p_{\tau}^{(2)}}$$

weighted by $p_{\tau}^{(2)}$

Measures for Comparing Texts: Kullback-Leibler Divergence

Sometimes we want to compare one text to a *reference text*

Say $P^{(1)}$ is the reference, and $P^{(2)}$ is the comparison. The *Kullback-Leibler divergence* (KLD) is

$$D^{(KL)}\left(P^{(2)} \,||\, P^{(1)}\right) = \sum_{\tau} p_\tau^{(2)} \log \frac{1}{p_\tau^{(1)}} - p_\tau^{(2)} \log \frac{1}{p_\tau^{(2)}}$$

Drawback: only well-defined if ***all*** the words in the reference text are also in the comparison text

Measures for Comparing Texts: Jensen-Shannon Divergence

The *Jensen-Shannon divergence* (JSD) attempts to account for the shortcomings of the KLD

Measures for Comparing Texts: Jensen-Shannon Divergence

The *Jensen-Shannon divergence* (JSD) attempts to account for the shortcomings of the KLD

We first define a mixture text M

$$M = \pi_1 P^{(1)} + \pi_2 P^{(2)}$$

Measures for Comparing Texts: Jensen-Shannon Divergence

The *Jensen-Shannon divergence* (JSD) attempts to account for the shortcomings of the KLD

We first define a mixture text M

$$M = \pi_1 P^{(1)} + \pi_2 P^{(2)}$$

Then, the JSD is the average KLD of each text from the mixture text

$$D^{(JS)}\left(P^{(1)} || P^{(2)}\right) = \pi_1 D^{(KL)}\left(P^{(1)} || M\right) + \pi_2 D^{(KL)}\left(P^{(2)} || M\right)$$

Measures for Comparing Texts: Jensen-Shannon Divergence

The *Jensen-Shannon divergence* (JSD) attempts to account for the shortcomings of the KLD

We first define a mixture text M

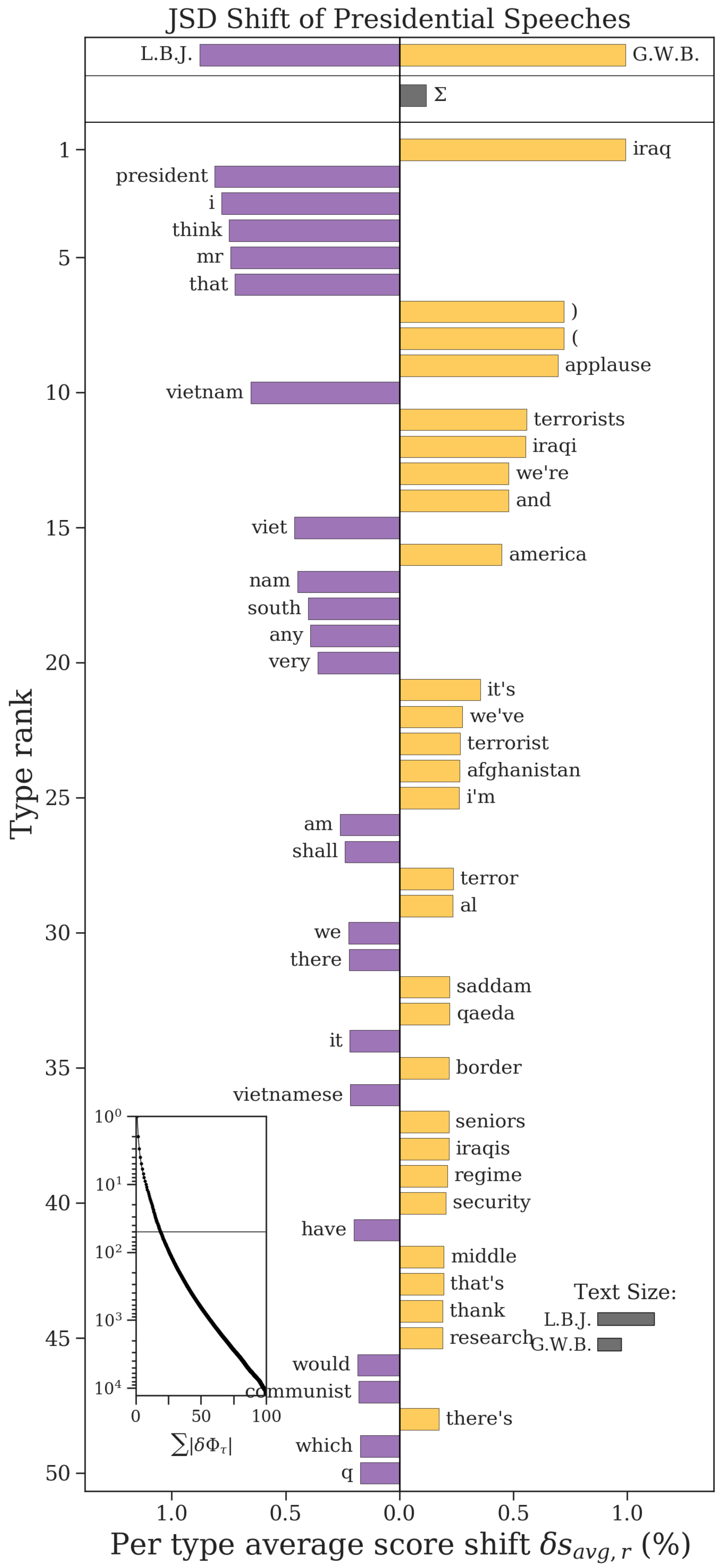
$$M = \pi_1 P^{(1)} + \pi_2 P^{(2)}$$

Then, the JSD is the average KLD of each text from the mixture text

$$D^{(JS)}\left(P^{(1)} || P^{(2)}\right) = \pi_1 D^{(KL)}\left(P^{(1)} || M\right) + \pi_2 D^{(KL)}\left(P^{(2)} || M\right)$$

$$= \sum_{\tau} m_\tau \log \frac{1}{m_\tau} - \left(\pi_1 p_\tau^{(1)} \log \frac{1}{p_\tau^{(1)}} + \pi_2 p_\tau^{(2)} \log \frac{1}{p_\tau^{(2)}} \right)$$

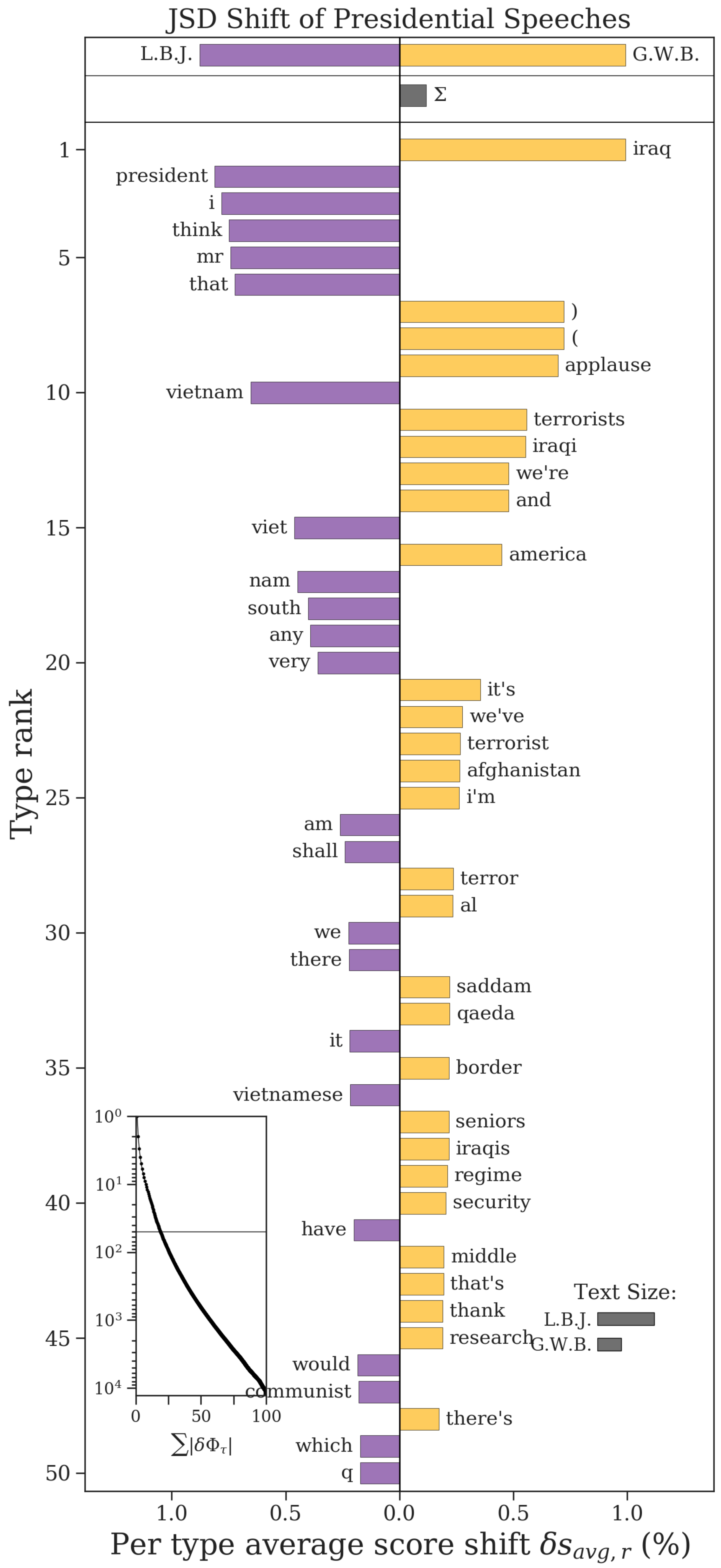
JSD Shift



All positive contributions

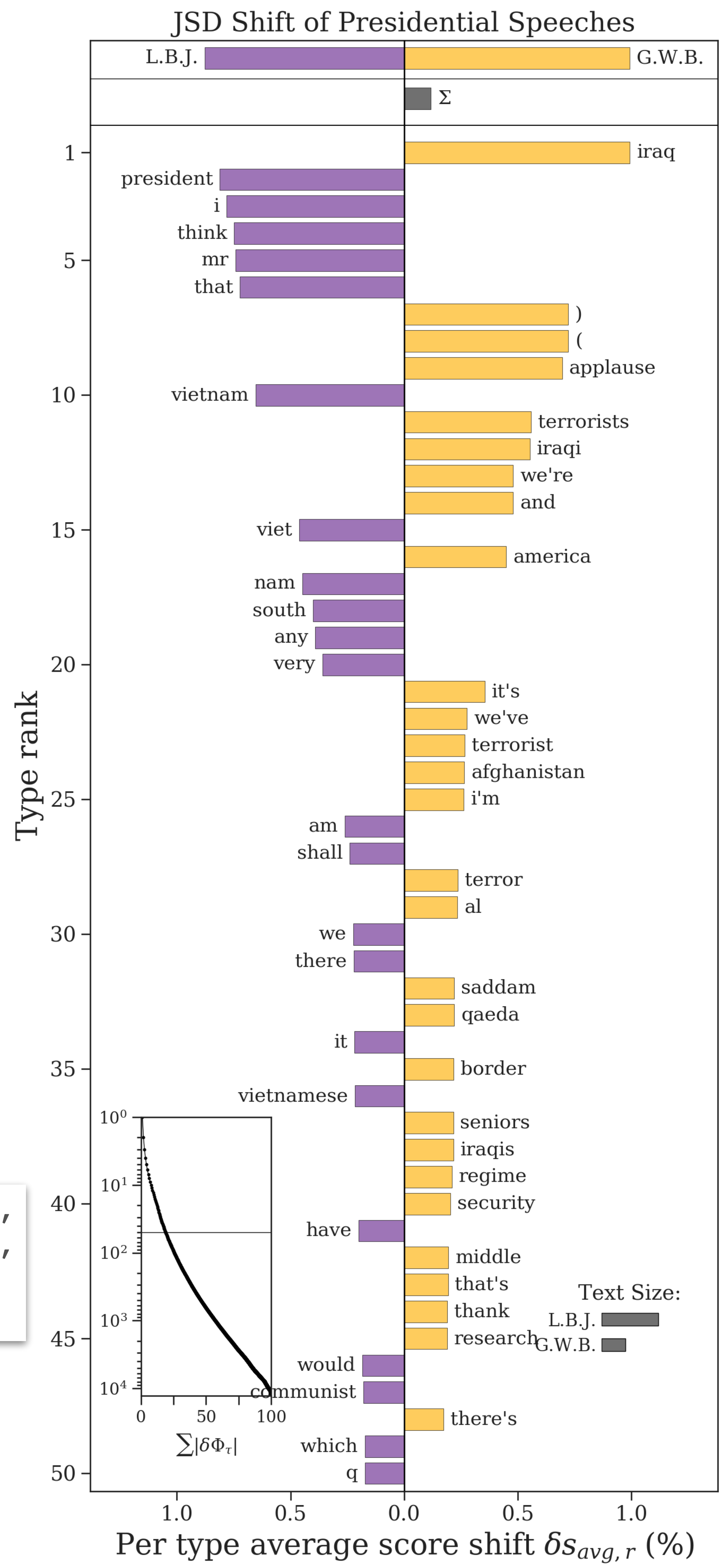
JSD Shift

Used relatively more by L.B.J



Used relatively more by G.W.B.

JSD Shift



Used relatively more by L.B.J

Used relatively more by G.W.B.

```
jsd_shift = sh.JSDivergenceShift(type2freq_1=type2freq_1,
                                  type2freq_2=type2freq_2,
                                  base=2,
                                  alpha=1.0)
```

Measures for Comparing Texts: Dictionary Scores

Dictionary methods assign a weight, or score, to each word in the vocabulary. If done carefully, scores can "measure" sentiment, hatefulness, respect, morality, or any number of other theoretical constructs

Measures for Comparing Texts: Dictionary Scores

Dictionary methods assign a weight, or score, to each word in the vocabulary. If done carefully, scores can "measure" sentiment, hatefulness, respect, morality, or any number of other theoretical constructs

We calculate the average score by taking a weighted average over all words

$$\Phi = \sum_{\tau} \phi_\tau p_\tau$$

Measures for Comparing Texts: Dictionary Scores

Dictionary methods assign a weight, or score, to each word in the vocabulary. If done carefully, scores can "measure" sentiment, hatefulness, respect, morality, or any number of other theoretical constructs

We calculate the average score by taking a weighted average over all words

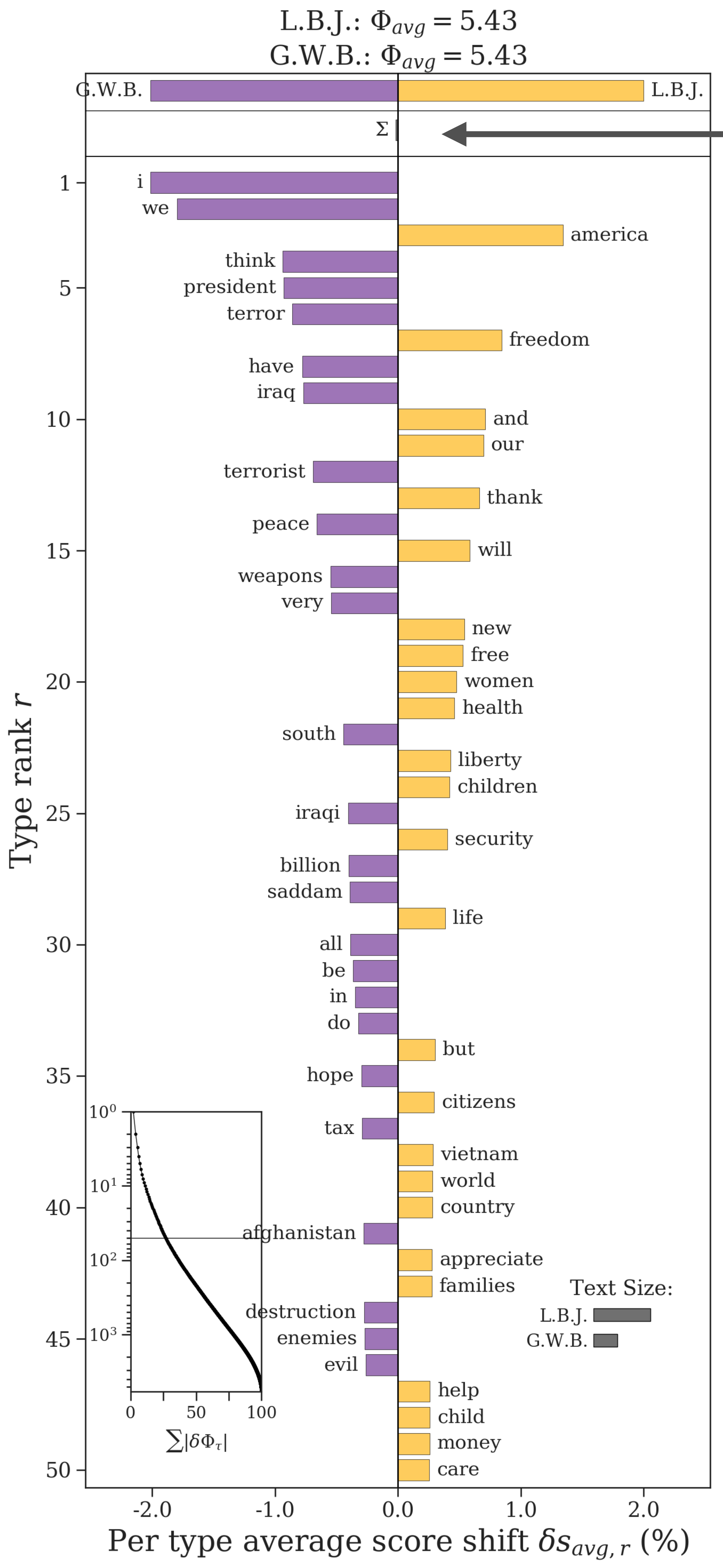
$$\Phi = \sum_{\tau} \phi_\tau p_\tau$$

We can get an individual word's contribution to the difference between two average scores

$$\delta\Phi = \sum_{\tau} \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

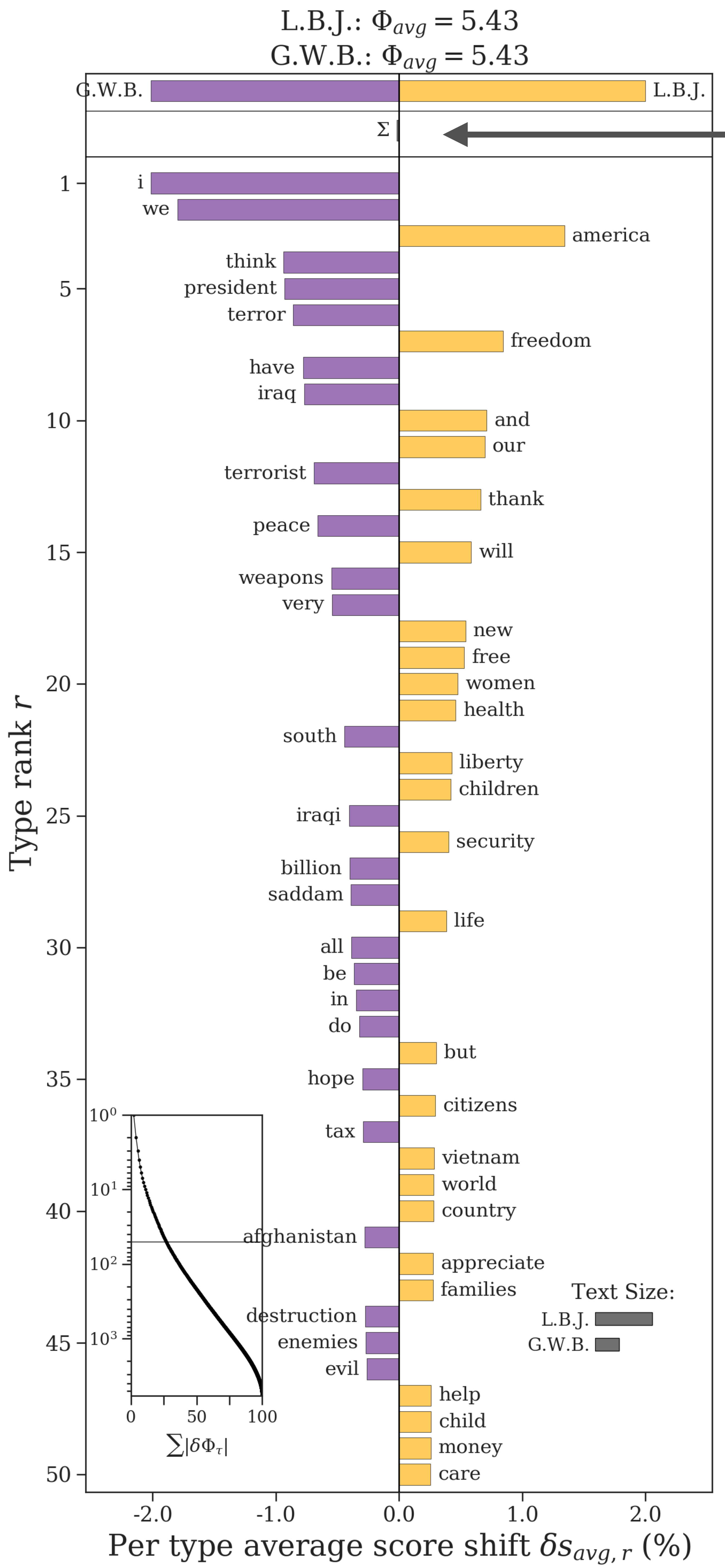


G.W.B. < L.B.J. (very slightly)

Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

G.W.B. < L.B.J. (very slightly)



Directly contribute to G.W.B. < L.B.J

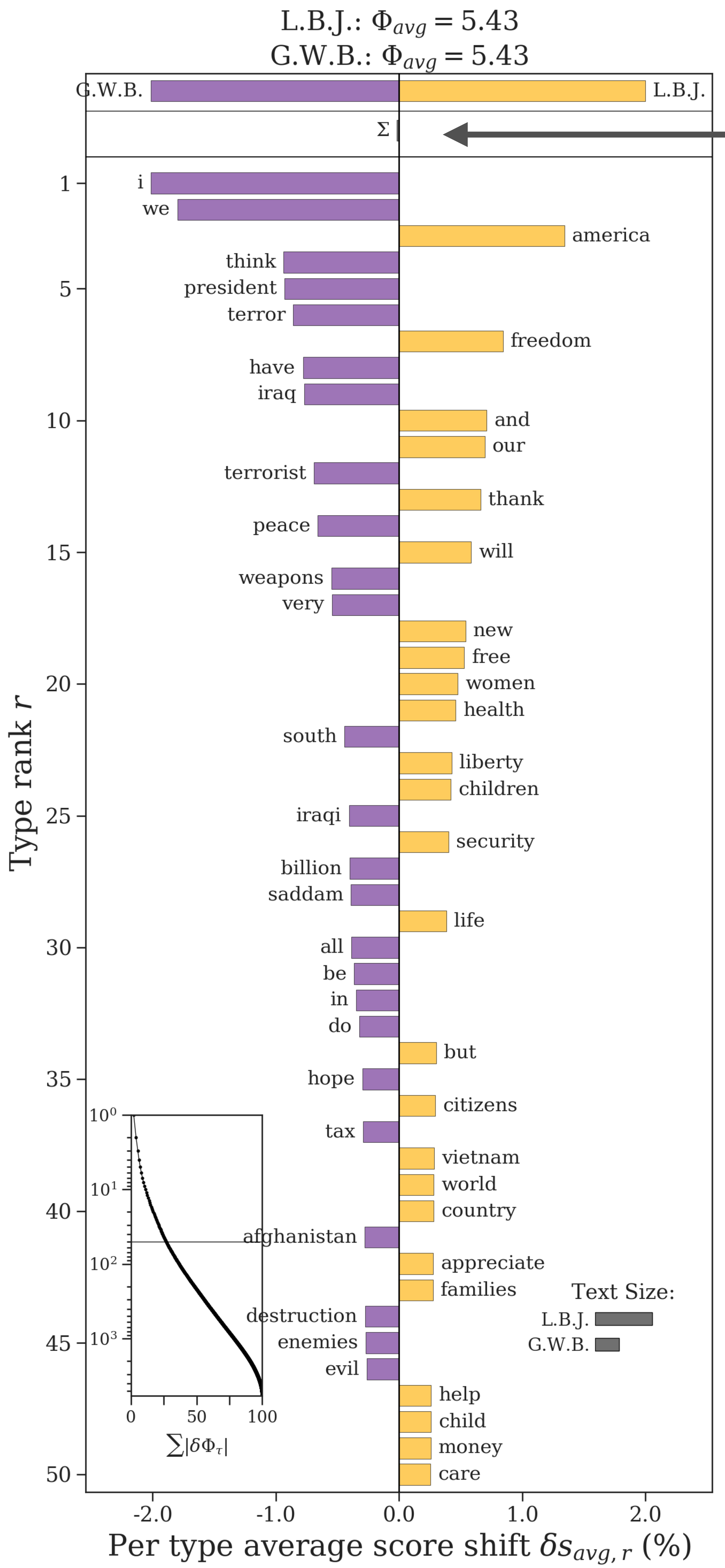
Counteract G.W.B. < L.B.J.

Sentiment difference would be even greater otherwise

Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

Directly contribute to G.W.B. < L.B.J



G.W.B. < L.B.J. (very slightly)

Counteract G.W.B. < L.B.J.

Sentiment difference would be even greater otherwise

```
w_shift = sh.WeightedAvgShift(type2freq_1=type2freq_1,
                              type2freq_2=type2freq_2,
                              type2score_1='labMT_English')
```

Measures for Comparing Texts

Measure	Advantages	Drawbacks

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words
Shannon entropy	Accounts for how “surprising” a word is	Surprisal weighting can’t always offset common words

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words
Shannon entropy	Accounts for how “surprising” a word is	Surprisal weighting can’t always offset common words
Tsallis entropy	Tunability between rare and common words	Requires ad hoc choice of parameter

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words
Shannon entropy	Accounts for how "surprising" a word is	Surprisal weighting can't always offset common words
Tsallis entropy	Tunability between rare and common words	Requires ad hoc choice of parameter
Kullback-Leibler divergence	Measures divergence from reference text	Only well-defined when texts have all the same words

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words
Shannon entropy	Accounts for how "surprising" a word is	Surprisal weighting can't always offset common words
Tsallis entropy	Tunability between rare and common words	Requires ad hoc choice of parameter
Kullback-Leibler divergence	Measures divergence from reference text	Only well-defined when texts have all the same words
Jensen-Shannon divergence	Effective at drawing out differences across the word distribution	Difficult to interpret word-level contributions

Measures for Comparing Texts

Measure	Advantages	Drawbacks
Proportions	Simple, interpretable	Emphasizes small differences between common words
Shannon entropy	Accounts for how “surprising” a word is	Surprisal weighting can’t always offset common words
Tsallis entropy	Tunability between rare and common words	Requires ad hoc choice of parameter
Kullback-Leibler divergence	Measures divergence from reference text	Only well-defined when texts have all the same words
Jensen-Shannon divergence	Effective at drawing out differences across the word distribution	Difficult to interpret word-level contributions
Dictionary scores	Theoretical concepts can be encoded through user-defined weights	Potential serious concerns about measurement validity

For any measure where we can get individual word contributions, we should always plot a simple word shift plot

For any measure where we can get individual word contributions, we should always plot a simple word shift plot

For any measure that we can write as a weighted average or difference in weighted averages, we can go further

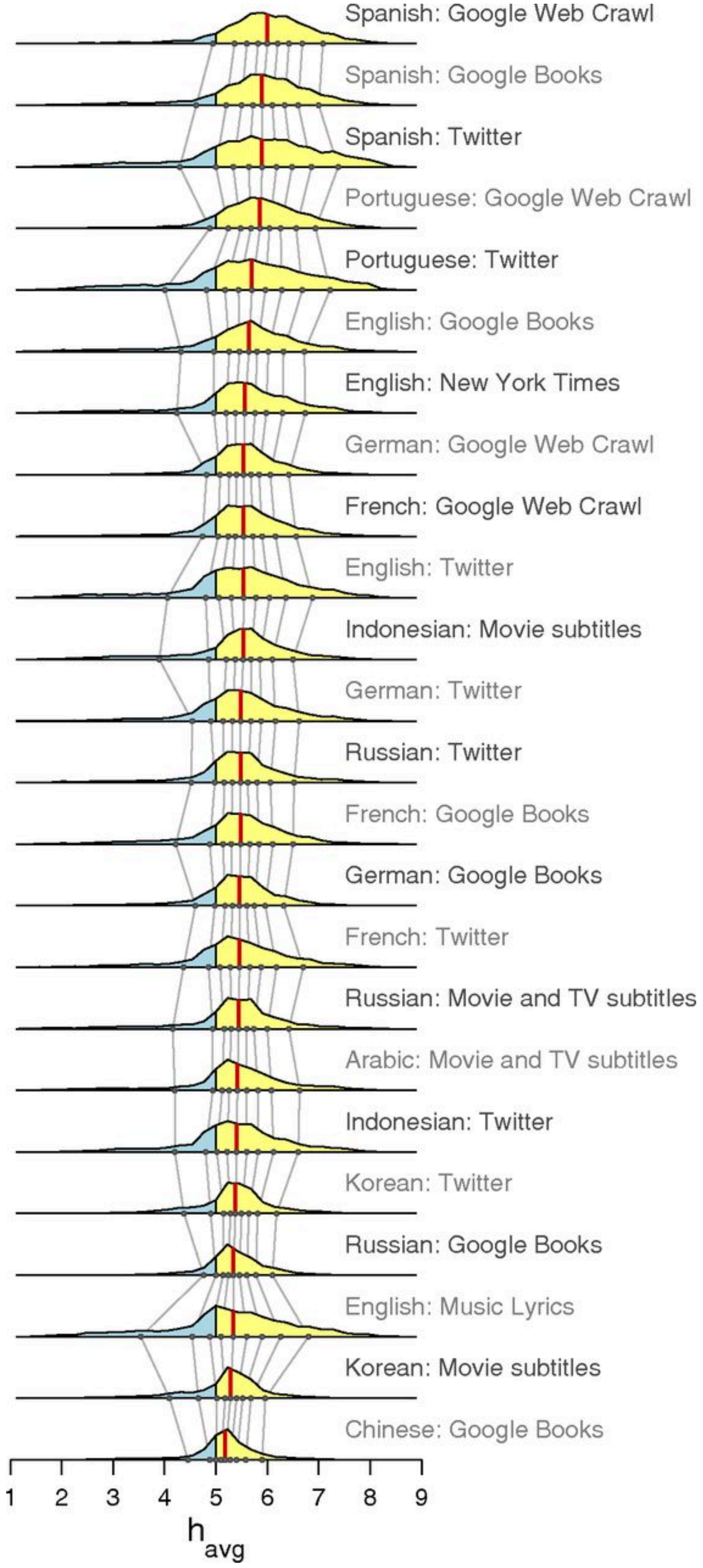
Reference Scores

Consider sentiment analysis as an example

Reference Scores

Consider sentiment analysis as an example

The Story Lab found that there is a universal positivity bias in human language



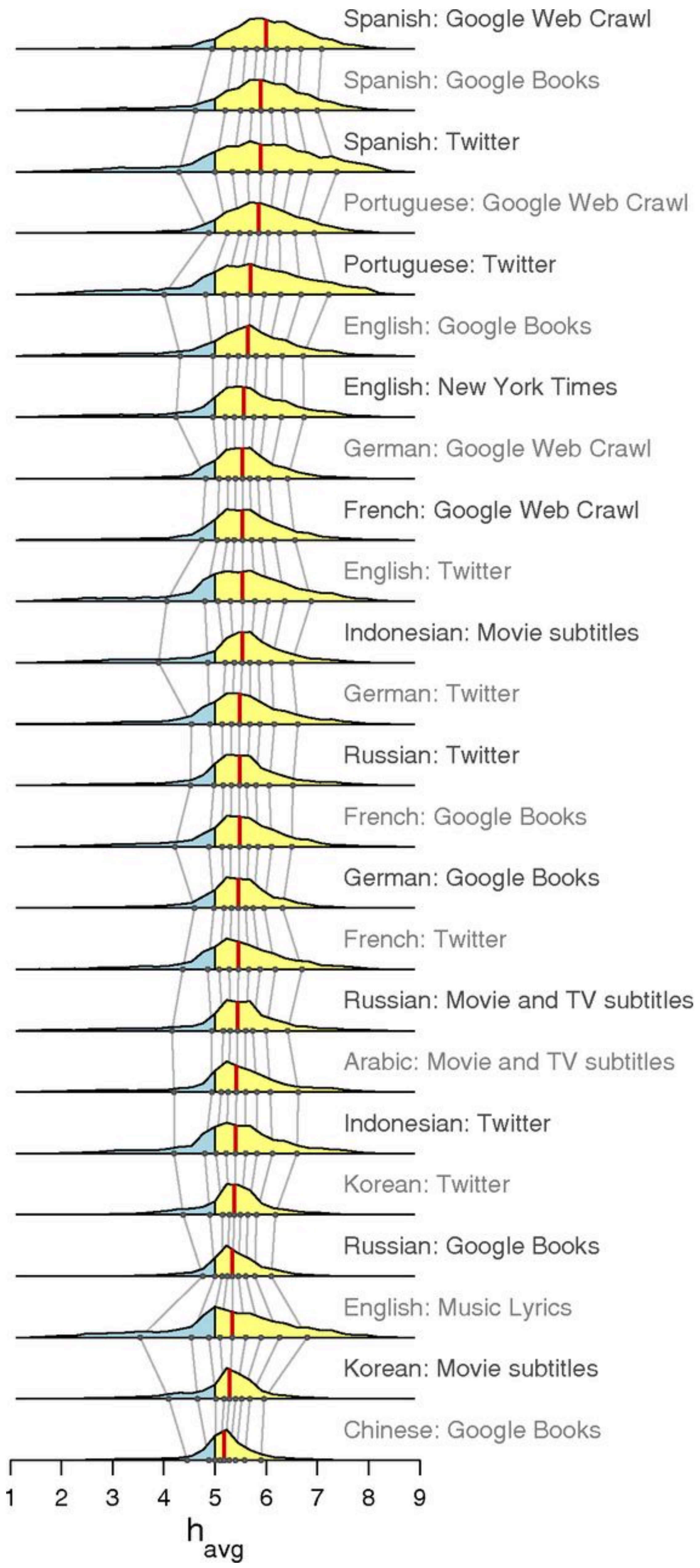
Reference Scores

Consider sentiment analysis as an example

The Story Lab found that there is a universal positivity bias in human language

The bias is *with respect to a* ***reference***

Qualitatively, we know that labMT words with scores > 5 are *positive* and those with scores < 5 are *negative*



Reference Scores

We can encode qualitatively different regimes of scores in our word shifts by applying a *reference* score

Reference Scores

We can encode qualitatively different regimes of scores in our word shifts by applying a *reference* score

We can rewrite any difference of weighted averages to incorporate a reference score

$$\delta\Phi = \sum_{\tau} \phi_\tau p_\tau^{(2)} - \phi_\tau p_\tau^{(1)}$$
$$= \sum_{\tau} \left(\phi_\tau - \Phi^{(ref)}\right)\left(p_\tau^{(2)} - p_\tau^{(1)}\right)$$

Reference Scores

We can encode qualitatively different regimes of scores in our word shifts by applying a *reference* score

We can rewrite any difference of weighted averages to incorporate a reference score

$$\delta\Phi = \sum_{\tau} \phi_\tau p_\tau^{(2)} - \phi_\tau p_\tau^{(1)}$$

$$= \sum_{\tau} \left(\phi_\tau - \Phi^{(ref)}\right) \left(p_\tau^{(2)} - p_\tau^{(1)}\right)$$

word score
with respect
to reference

Reference Scores

We can encode qualitatively different regimes of scores in our word shifts by applying a *reference* score

We can rewrite any difference of weighted averages to incorporate a reference score

$$\delta\Phi = \sum_{\tau} \phi_\tau p_\tau^{(2)} - \phi_\tau p_\tau^{(1)}$$

$$= \sum_{\tau} \left(\phi_\tau - \Phi^{(ref)}\right) \left(p_\tau^{(2)} - p_\tau^{(1)}\right)$$

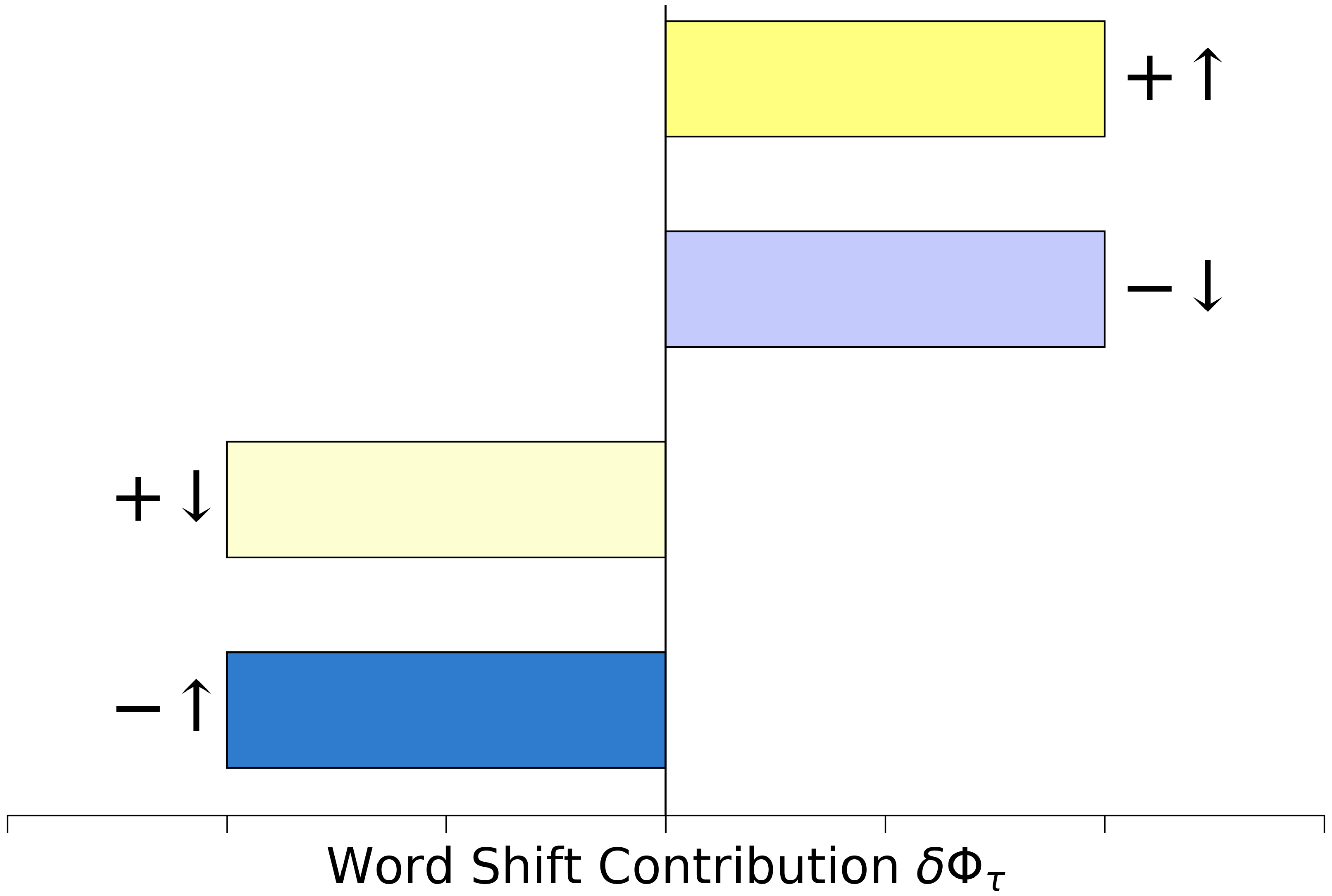
difference in frequency

Word Contributions

$$\delta\Phi_\tau = \underbrace{\left(\phi_\tau - \Phi^{(ref)}\right)}_{+/-} \underbrace{\left(p_\tau^{(2)} - p_\tau^{(1)}\right)}_{\uparrow/\downarrow}$$

Word Contributions

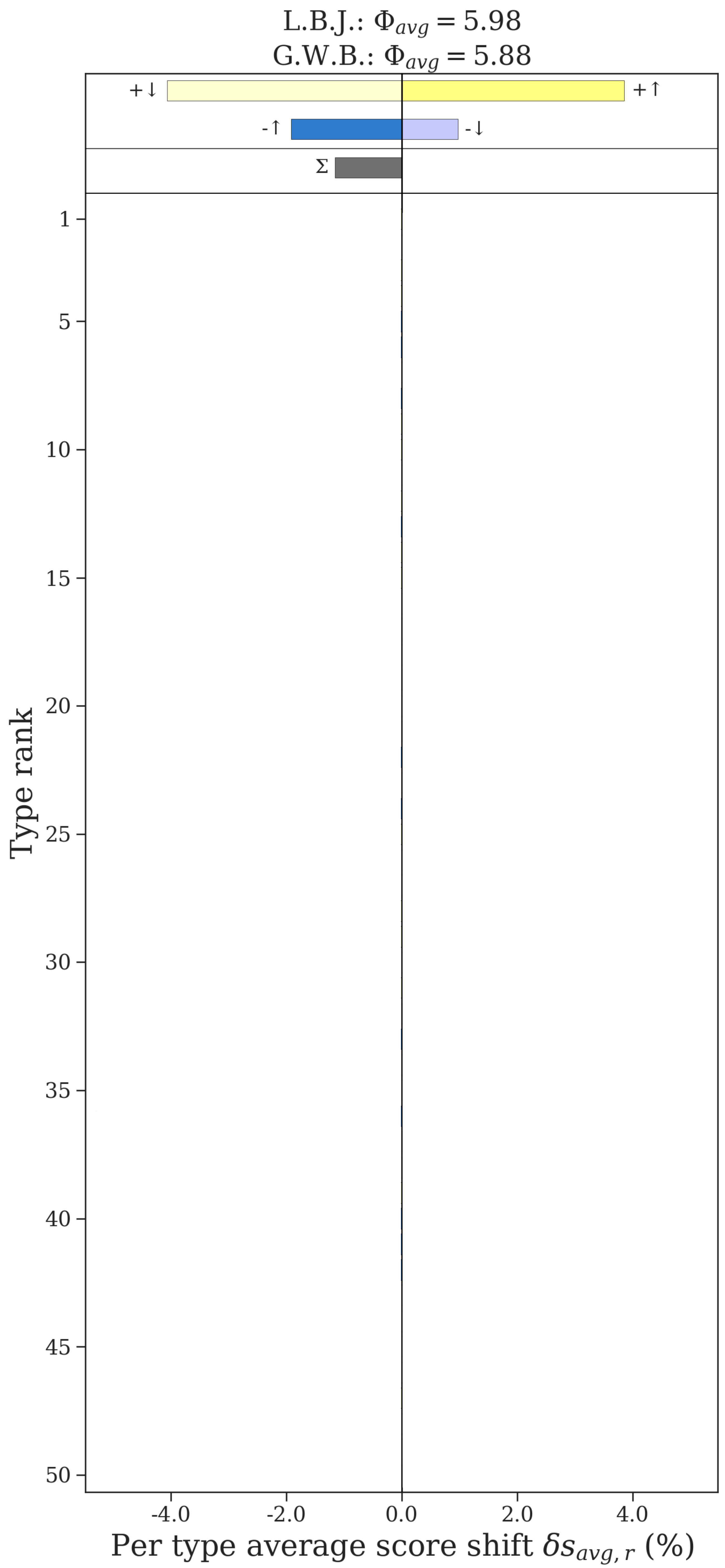
$$\delta\Phi_\tau = \underbrace{\left(\phi_\tau - \Phi^{(ref)}\right)}_{+/-} \underbrace{\left(p_\tau^{(2)} - p_\tau^{(1)}\right)}_{\uparrow/\downarrow}$$



Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

$$\Phi^{(ref)} = 5$$



+ ↑ Relatively positive word used more often

− ↓ Relatively negative word used less often

+ ↓ Relatively positive word used less often

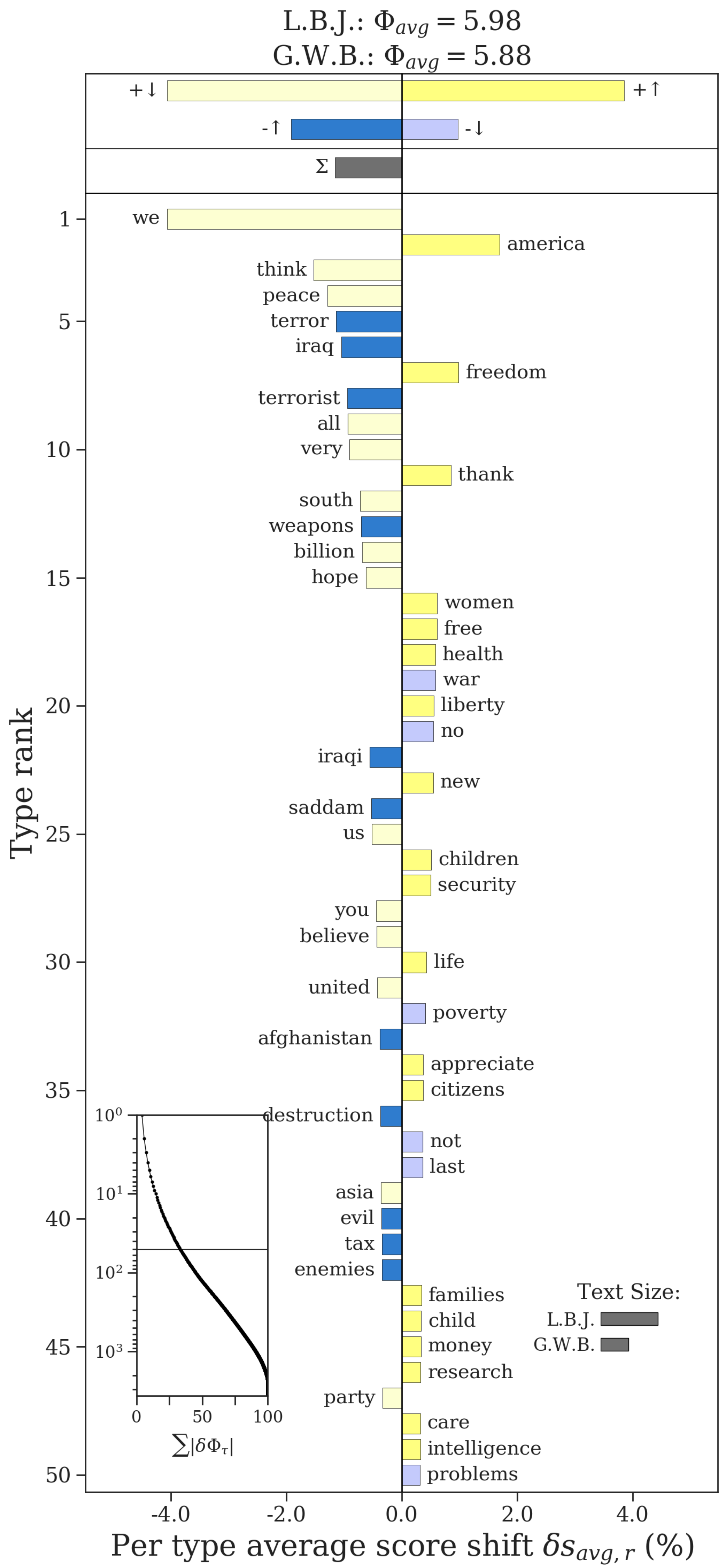
− ↑ Relatively negative word used more often

Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

$$\Phi^{(ref)} = 5$$

Directly contribute to G.W.B. < L.B.J



+ ↑ Relatively positive word used more often

− ↓ Relatively negative word used less often

+ ↓ Relatively positive word used less often

− ↑ Relatively negative word used more often

Counteract G.W.B. < L.B.J.

Sentiment difference would be even greater otherwise

Generalized Word Shifts

Before, we assumed that a word's score is the same across both texts

This limits our ability to use the full word shift framework for any of the entropy-based measures, or for dictionary-based analyses using domain-adapted dictionaries

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_{\tau} \phi_{\tau}^{(2)} p_{\tau}^{(2)} - \phi_{\tau}^{(1)} p_{\tau}^{(1)}$$

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_{\tau} \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_{\tau} \left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]\left(p_\tau^{(2)} - p_\tau^{(1)}\right) + \frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)$$

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_\tau \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_\tau \left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)} \right] \left(p_\tau^{(2)} - p_\tau^{(1)} \right) + \frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)} \right) \left(\phi_\tau^{(2)} - \phi_\tau^{(1)} \right)$$

average
score

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_\tau \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_\tau \left[\frac{1}{2} \left(\phi_\tau^{(1)} + \phi_\tau^{(2)} \right) - \Phi^{(ref)} \right] \left(p_\tau^{(2)} - p_\tau^{(1)} \right) + \frac{1}{2} \left(p_\tau^{(1)} + p_\tau^{(2)} \right) \left(\phi_\tau^{(2)} - \phi_\tau^{(1)} \right)$$

difference between average score and reference

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_{\tau} \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_{\tau} \left[\frac{1}{2} \left(\phi_\tau^{(1)} + \phi_\tau^{(2)} \right) - \Phi^{(ref)} \right] \left(p_\tau^{(2)} - p_\tau^{(1)} \right) + \frac{1}{2} \left(p_\tau^{(1)} + p_\tau^{(2)} \right) \left(\phi_\tau^{(2)} - \phi_\tau^{(1)} \right)$$

difference in frequency

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_{\tau} \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_{\tau} \left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]\left(p_\tau^{(2)} - p_\tau^{(1)}\right) + \frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)$$

average frequency

Generalized Word Shifts

We can generalize word shifts to account for changes in scores

$$\delta\Phi = \sum_{\tau} \phi_\tau^{(2)} p_\tau^{(2)} - \phi_\tau^{(1)} p_\tau^{(1)}$$

$$= \sum_{\tau} \left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]\left(p_\tau^{(2)} - p_\tau^{(1)}\right) + \frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)$$

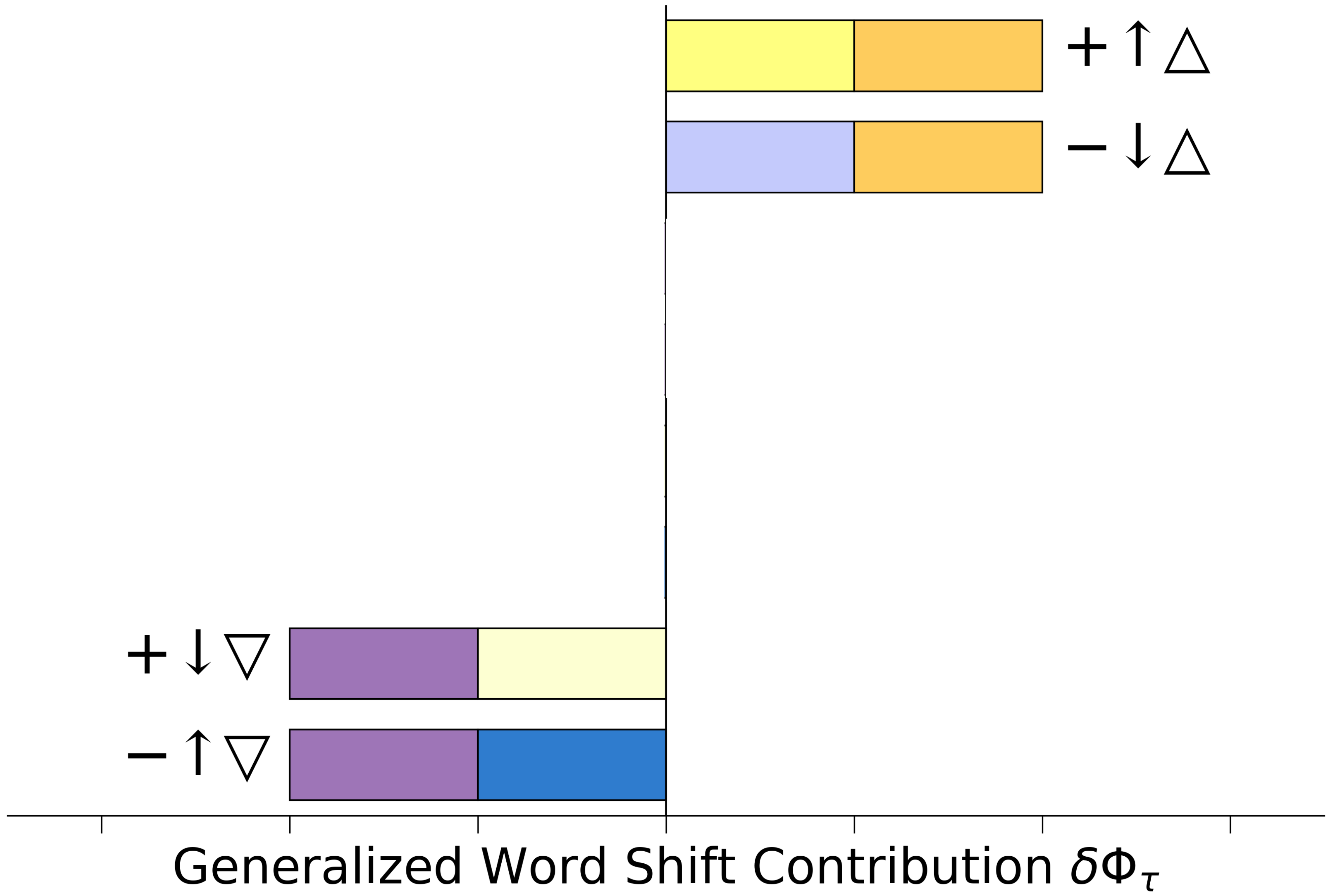
difference in scores

Word Contributions

$$\delta\Phi_\tau = \underbrace{\left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]}_{+/-} \underbrace{\left(p_\tau^{(2)} - p_\tau^{(1)}\right)}_{\uparrow/\downarrow} + \underbrace{\frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)}_{\triangledown/\triangle}$$

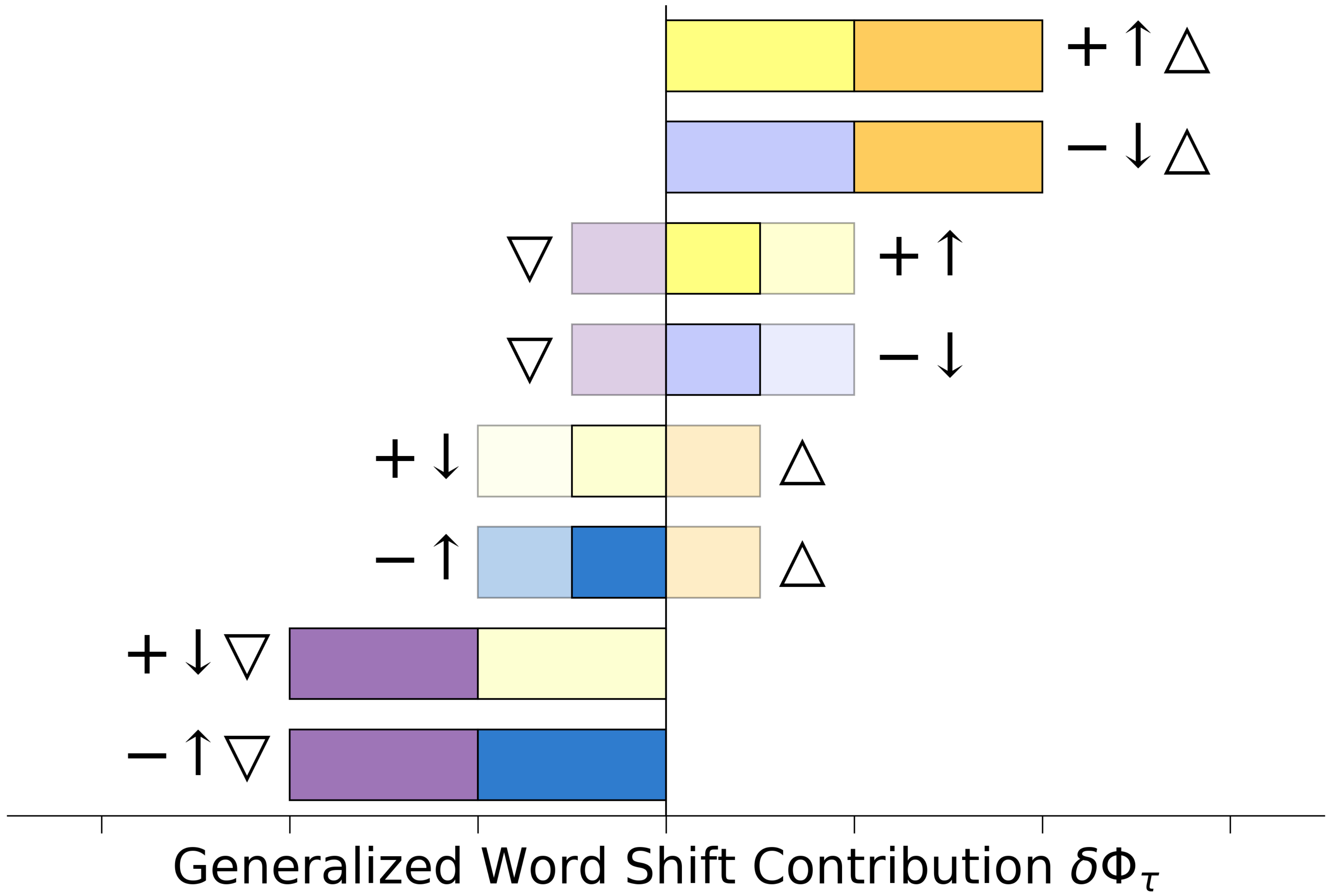
Word Contributions

$$\delta\Phi_\tau = \underbrace{\left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]}_{+/-} \underbrace{\left(p_\tau^{(2)} - p_\tau^{(1)}\right)}_{\uparrow/\downarrow} + \underbrace{\frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)}_{\bigtriangledown/\bigtriangleup}$$



Word Contributions

$$\delta\Phi_\tau = \underbrace{\left[\frac{1}{2}\left(\phi_\tau^{(1)} + \phi_\tau^{(2)}\right) - \Phi^{(ref)}\right]}_{+/-} \underbrace{\left(p_\tau^{(2)} - p_\tau^{(1)}\right)}_{\uparrow/\downarrow} + \underbrace{\frac{1}{2}\left(p_\tau^{(1)} + p_\tau^{(2)}\right)\left(\phi_\tau^{(2)} - \phi_\tau^{(1)}\right)}_{\bigtriangledown/\bigtriangleup}$$



Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

$$\Phi^{(ref)} = 5$$

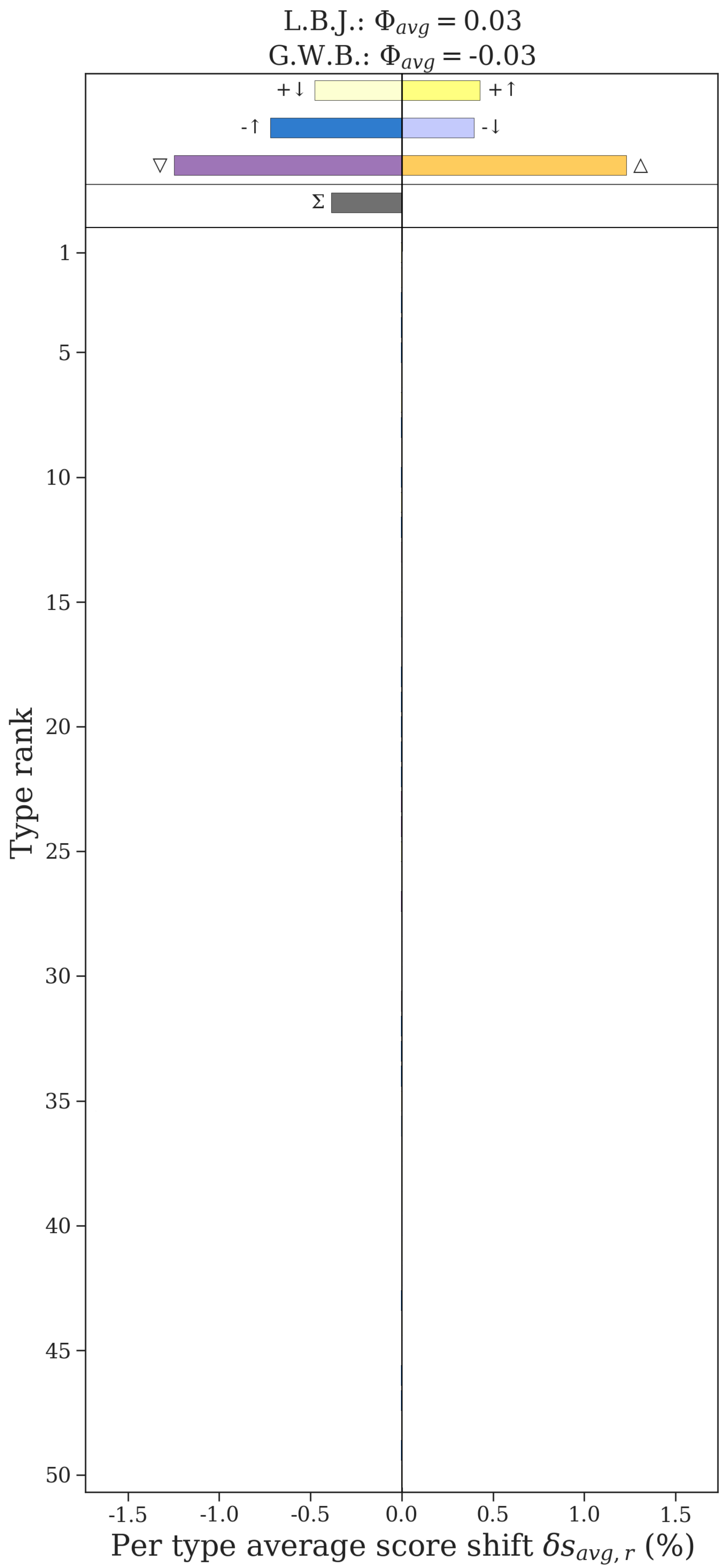
Using domain-adapted dictionaries
for the 1960s and 2000s

Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

$$\Phi^{(ref)} = 5$$

Using domain-adapted dictionaries for the 1960s and 2000s



+ ↑ Relatively positive word used more often

− ↓ Relatively negative word used less often

+ ↓ Relatively positive word used less often

− ↑ Relatively negative word used more often

△ Higher word positivity than before

▽ Lower word positivity than before

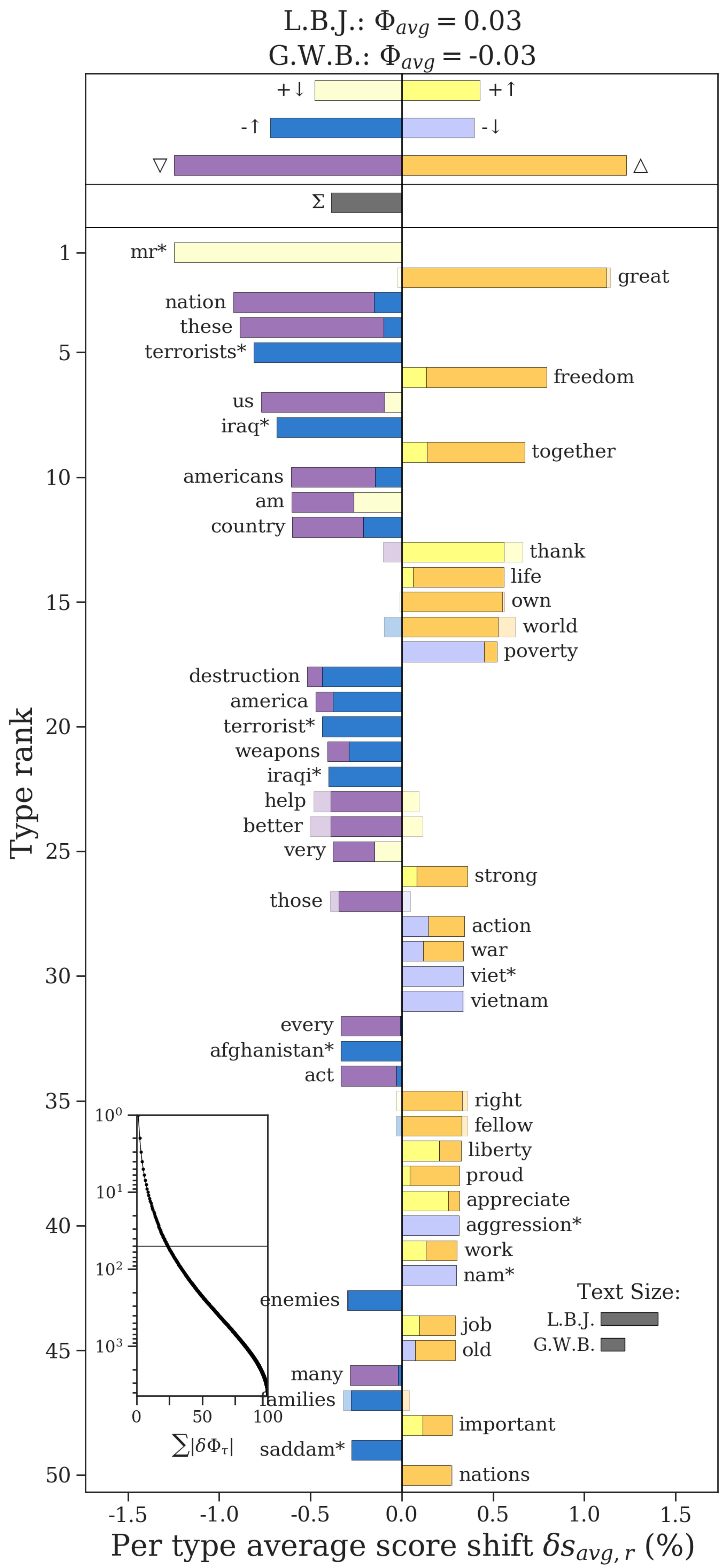
Sentiment Shift

$$\delta\Phi = \Phi^{(G.W.B.)} - \Phi^{(L.B.J.)}$$

$$\Phi^{(ref)} = 5$$

Using domain-adapted dictionaries for the 1960s and 2000s

Directly contribute to G.W.B. < L.B.J



+↑ Relatively positive word used more often

−↓ Relatively negative word used less often

+↓ Relatively positive word used less often

−↑ Relatively negative word used more often

△ Higher word positivity than before

▽ Lower word positivity than before

Counteract G.W.B. < L.B.J.

Sentiment difference would be even greater otherwise

Comparison Measures as Weighted Averages

Measure	Word Contribution $\delta\Phi_\tau$
Proportions	
Shannon entropy	
Tsallis entropy	
Kullback-Leibler divergence	
Jensen-Shannon divergence	
Generalized JSD	

Comparison Measures as Weighted Averages

Measure	Word Contribution $\delta\Phi_\tau$
Proportions	$p_\tau^{(2)} - p_\tau^{(1)}$
Shannon entropy	$-p_\tau^{(2)} \log p_\tau^{(2)} + p_\tau^{(1)} \log p_\tau^{(1)}$
Tsallis entropy	$-p_\tau^{(2)} \left[\frac{\left(p_\tau^{(2)}\right)^{\alpha-1}}{\alpha - 1} \right] + p_\tau^{(1)} \left[\frac{\left(p_\tau^{(1)}\right)^{\alpha-1}}{\alpha - 1} \right]$
Kullback-Leibler divergence	$-p_\tau^{(2)} \log p_\tau^{(1)} + p_\tau^{(1)} \log p_\tau^{(1)}$
Jensen-Shannon divergence	$p_\tau^{(2)} \pi_2 \left(\log p_\tau^{(2)} - \log m_\tau \right) - p_\tau^{(1)} \pi_1 \left(\log m_\tau - \log p_\tau^{(1)} \right)$
Generalized JSD	$-p_\tau^{(2)} \pi_2 \left[\frac{\left(p_\tau^{(2)}\right)^{\alpha-1} - m_\tau^{\alpha-1}}{\alpha - 1} \right] - p_\tau^{(1)} \pi_1 \left[\frac{m_\tau^{\alpha-1} - \left(p_\tau^{(1)}\right)^{\alpha-1}}{\alpha - 1} \right]$

Comparison Measures as Weighted Averages

Measure	Word Contribution $\delta\Phi_\tau$
Proportions	$p_\tau^{(2)} - p_\tau^{(1)}$
Shannon entropy	$-p_\tau^{(2)} \log p_\tau^{(2)} + p_\tau^{(1)} \log p_\tau^{(1)}$
Tsallis entropy	$-p_\tau^{(2)} \left[\frac{\left(p_\tau^{(2)}\right)^{\alpha-1}}{\alpha-1} \right] + p_\tau^{(1)} \left[\frac{\left(p_\tau^{(1)}\right)^{\alpha-1}}{\alpha-1} \right]$
Kullback-Leibler divergence	$-p_\tau^{(2)} \log p_\tau^{(1)} + p_\tau^{(2)} \log p_\tau^{(2)}$
Jensen-Shannon divergence	$p_\tau^{(2)} \pi_2 \left(\log p_\tau^{(2)} - \log m_\tau \right) - p_\tau^{(1)} \pi_1 \left(\log m_\tau - \log p_\tau^{(1)} \right)$
Generalized JSD	$-p_\tau^{(2)} \pi_2 \left[\frac{\left(p_\tau^{(2)}\right)^{\alpha-1} - m_\tau^{\alpha-1}}{\alpha-1} \right] - p_\tau^{(1)} \pi_1 \left[\frac{m_\tau^{\alpha-1} - \left(p_\tau^{(1)}\right)^{\alpha-1}}{\alpha-1} \right]$

Case Study: 280 Character Tweets

In early November 2017, Twitter began rolling out a new 280 character limit for tweets (up from 140 characters)

Case Study: 280 Character Tweets

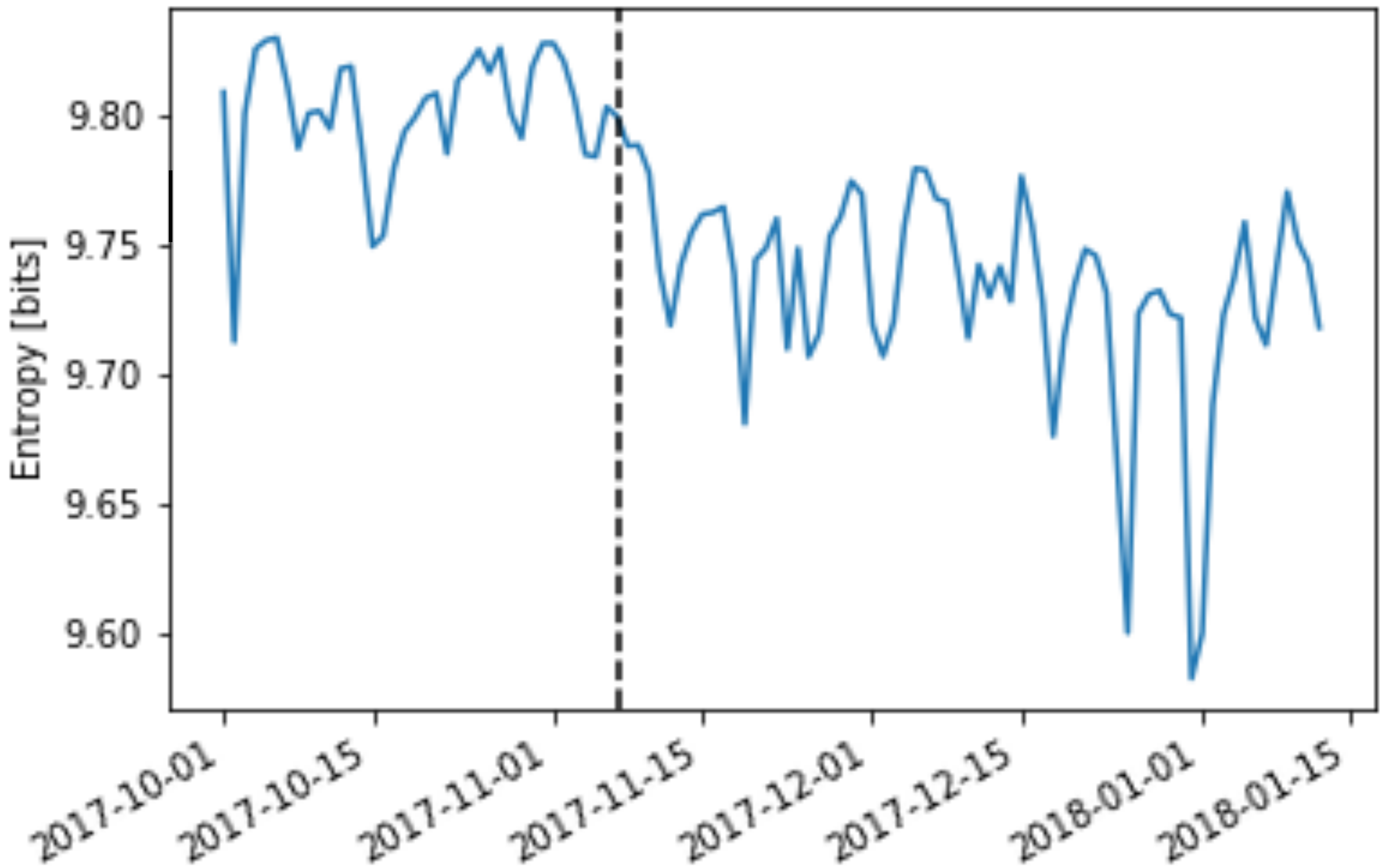
In early November 2017, Twitter began rolling out a new 280 character limit for tweets (up from 140 characters)

Question: How did that change the information content of tweets?

Case Study: 280 Character Tweets

In early November 2017, Twitter began rolling out a new 280 character limit for tweets (up from 140 characters)

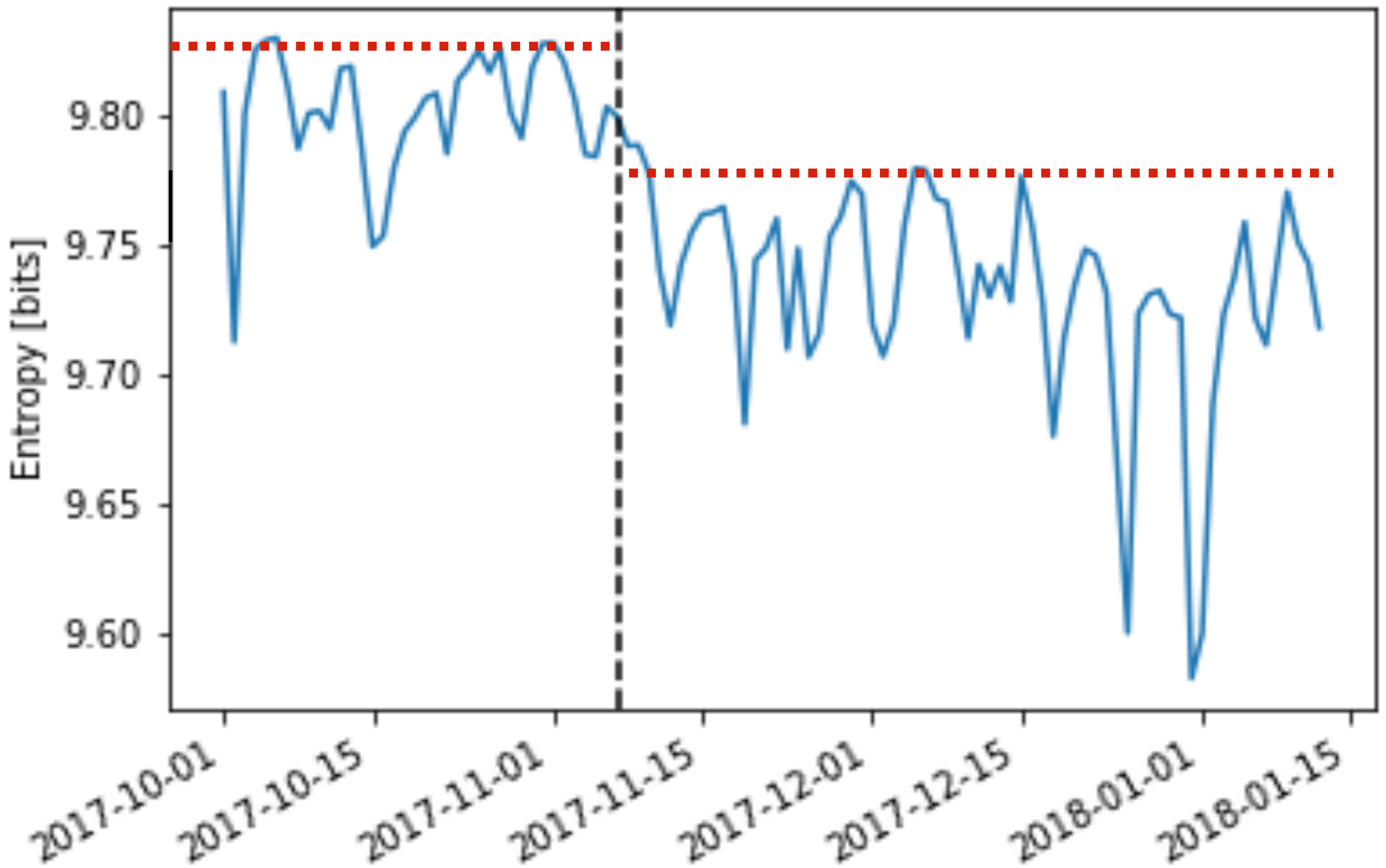
Question: How did that change the information content of tweets?



Case Study: 280 Character Tweets

In early November 2017, Twitter began rolling out a new 280 character limit for tweets (up from 140 characters)

Question: How did that change the information content of tweets?



Entropy over entire before and after periods

Twitter Entropy Shift

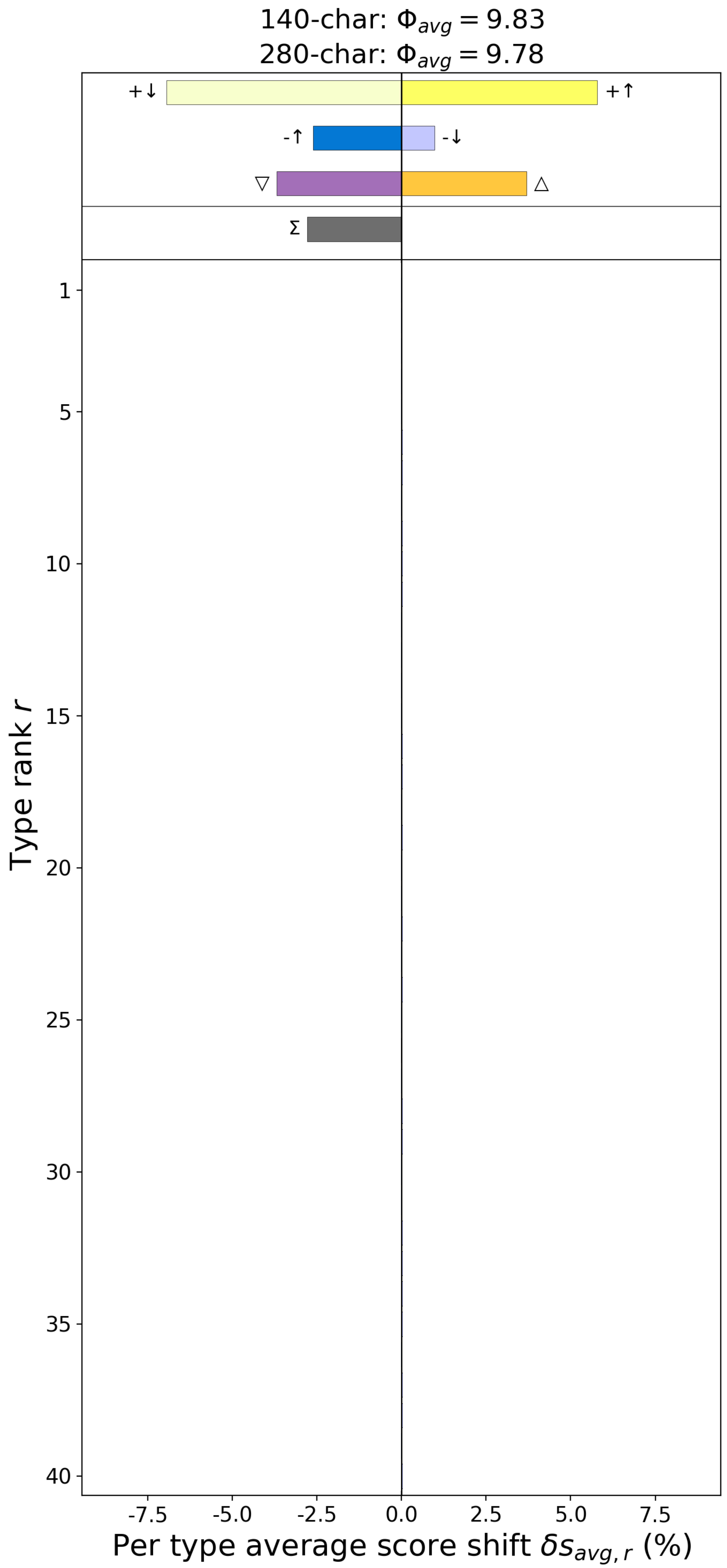
$$\delta H = H^{(280)} - H^{(140)}$$

$$\Phi^{(ref)} = H^{(140)}$$

Twitter Entropy Shift

$$\delta H = H^{(280)} - H^{(140)}$$

$$\Phi^{(ref)} = H^{(140)}$$



+ ↑ Relatively surprising word used more often

− ↓ Relatively unsurprising word used less often

+ ↓ Relatively surprising word used less often

− ↑ Relatively unsurprising word used more often

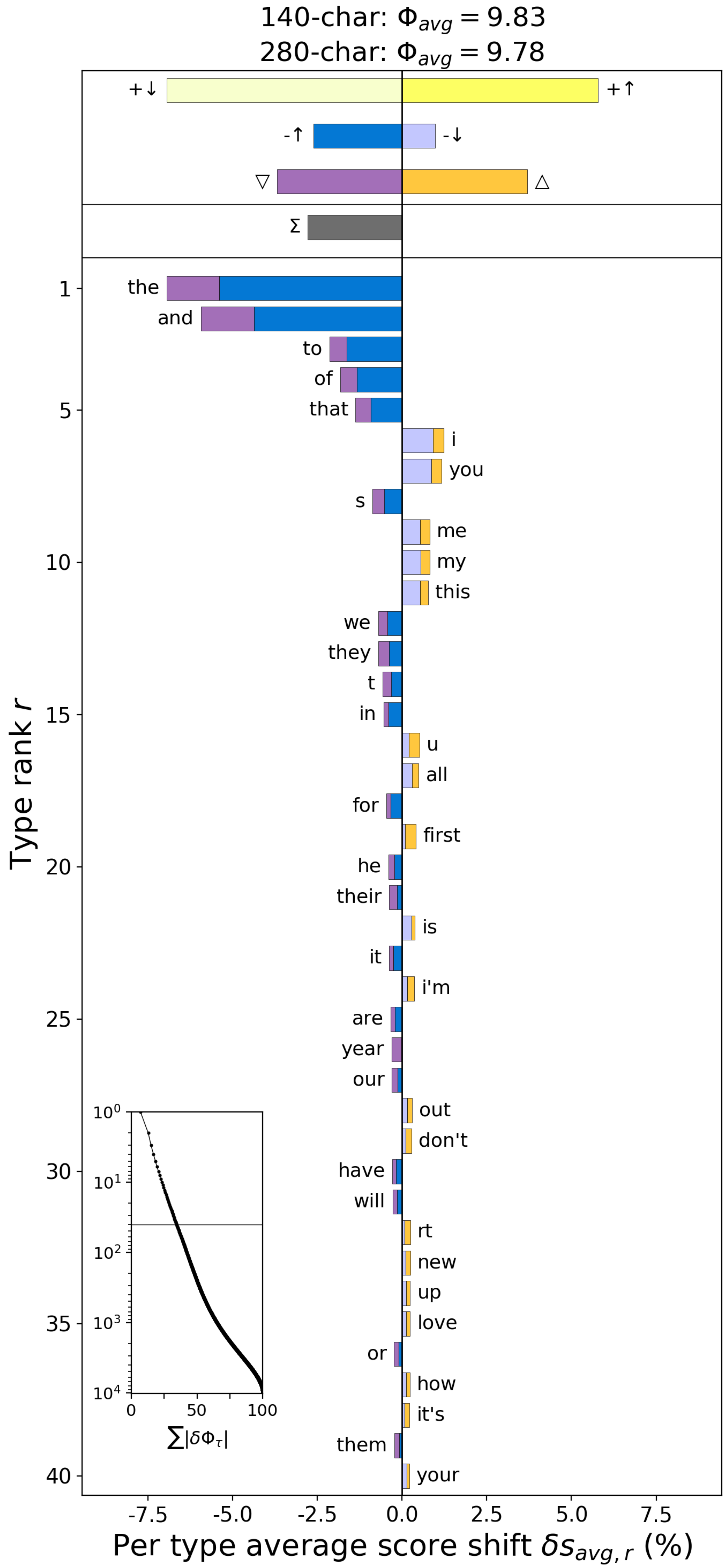
△ Higher surprisal than before

▽ Lower surprisal than before

Twitter Entropy Shift

$$\delta H = H^{(280)} - H^{(140)}$$

$$\Phi^{(ref)} = H^{(140)}$$



Directly contribute to H(280) < H(140)

+ ↑ Relatively surprising word used more often

− ↓ Relatively unsurprising word used less often

+ ↓ Relatively surprising word used less often

− ↑ Relatively unsurprising word used more often

△ Higher surprisal than before

▽ Lower surprisal than before

Counteract H(280) < H(140)

Entropy difference would be even greater otherwise

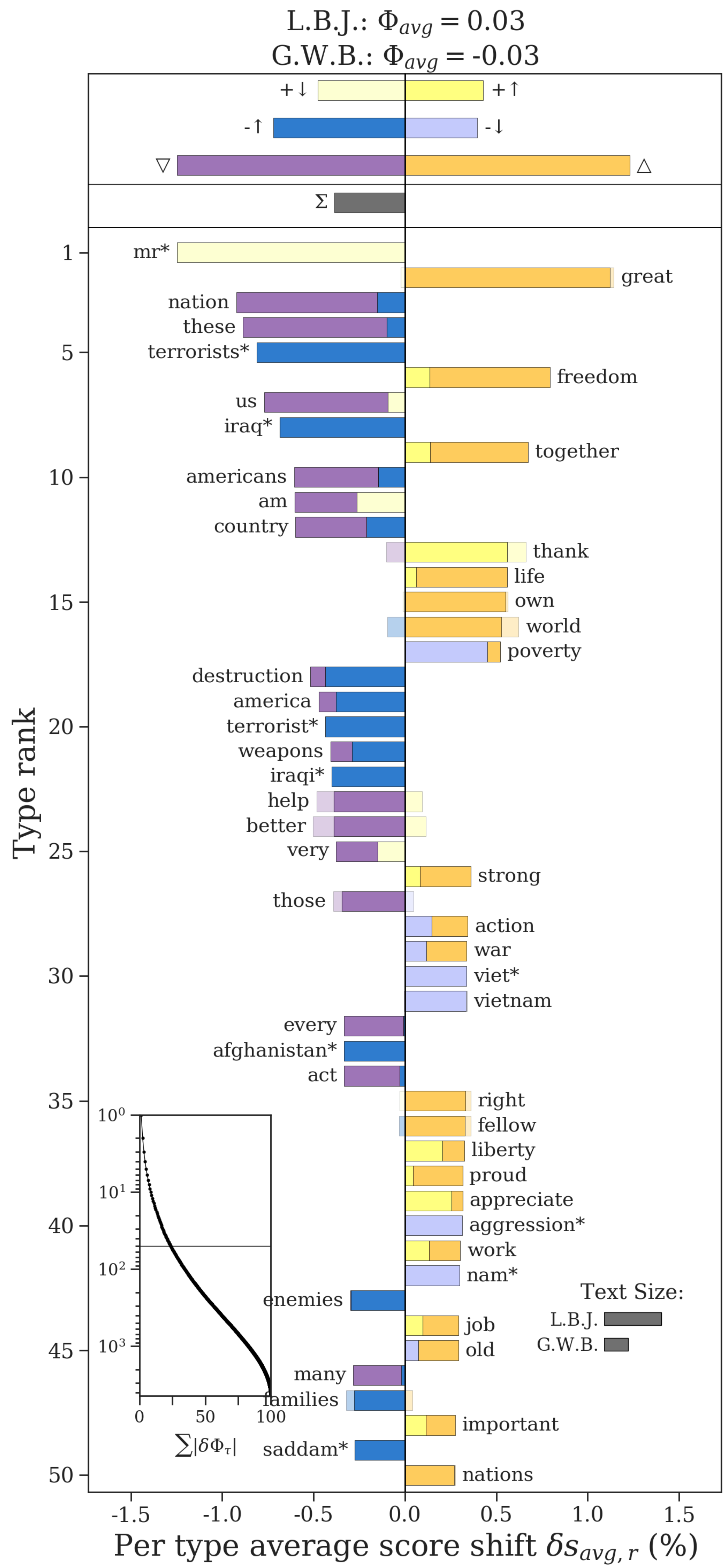
Conclusion

1. Look at the words!
2. We can visualize any measure where individual word contributions can be extracted
3. We can use a detailed word shift decomposition to visualize any weighted average
4. Many common measures can be reformulated as weighted averages

All visualizations were made using the Shifterator Python package

https://github.com/ryanjgallagher/shifterator

```
pip install shifterator
```



Collaborators

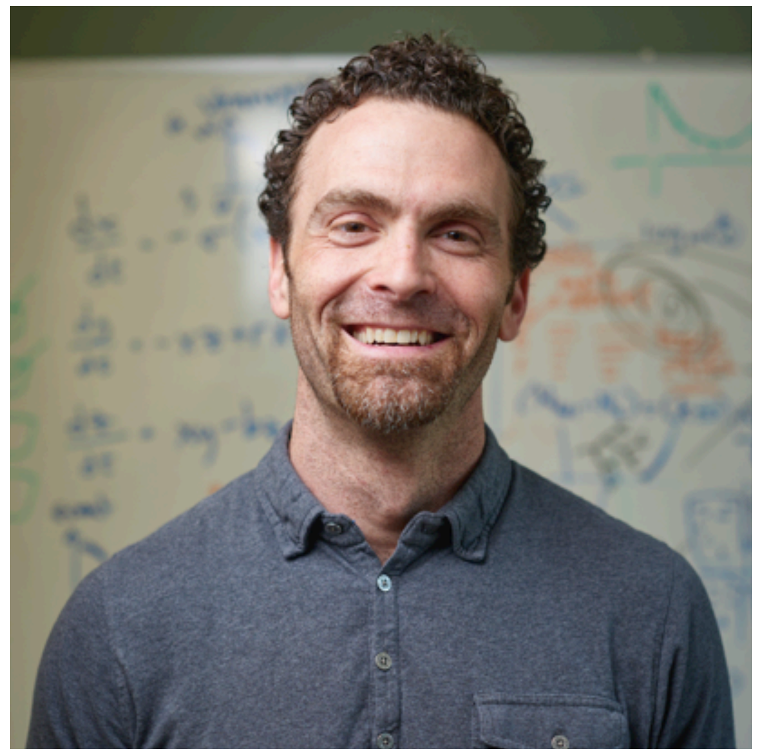
Morgan Frank
MIT

Colin Van Oort
University of Vermont

Lewis Mitchell
University of Adelaide

Aaron Schwartz
University of Vermont

Andy Reagan
MassMutual



Chris Danforth
University of Vermont

Peter Dodds
University of Vermont

Thank you for your time!

Ryan J. Gallagher

@ryanjgallag

ryanjgallag@gmail.com

netsi

Northeastern University

Network Science Institute

CoMM Lab