Grammar checking and readability
Readability

• Given a text, how readable is it?
• Factors:
  – Vocabulary
  – Grammatical complexity
  – Domain (subject area)
• Relies mostly on corpus statistics
• Several different formulas
Readability

• Important initial facts to remember:
• Distinct from legibility (typeface and font)
• Completed grade level doesn’t necessarily correlate with reading level
Why is the Topic of Readability Important?

- Many Americans, even those who have graduated from high school, read at about an 8th grade reading level. But much of the writing they encounter and are expected to be able to read is written on a higher grade reading level (10th grade reading level is common).
Why is the Topic of Readability Important?

As with literacy, it relates to the following:

– Safety (for example, technical writing and instructional brochures)
– Education (teaching materials)
– Sales and marketing (assembly instructions, owners’ manuals, policy manuals, how-to books, etc.)
– Participation in society (forms, applications, employment policy manuals, voting, etc.)
– Law (contracts, jury duty instructions, etc.)
– National Defense

Research into Readability

- During the last century, many readability studies have been done and various formulas have been proposed.
- Two variables that are commonly integrated into readability formulas:
  - Semantic difficulty (commonly measured in the vocabulary)
  - Syntactic difficulty (commonly measured in sentence length)
- No firm consensus exists on which approach to readability formulas is best
## Flesch’s Approach

<table>
<thead>
<tr>
<th>Reading Ease Score</th>
<th>Style Description</th>
<th>Estimated Reading Grade</th>
<th>Estimated Percent of U.S. Adults (1949)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 30:</td>
<td>Very Difficult</td>
<td>College graduate</td>
<td>4.5</td>
</tr>
<tr>
<td>30 to 40:</td>
<td>Difficult</td>
<td>13&lt;sup&gt;th&lt;/sup&gt; to 16&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>33</td>
</tr>
<tr>
<td>50 to 60:</td>
<td>Fairly Difficult</td>
<td>10&lt;sup&gt;th&lt;/sup&gt; to 12&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>54</td>
</tr>
<tr>
<td>60 to 70:</td>
<td>Standard</td>
<td>8&lt;sup&gt;th&lt;/sup&gt; and 9&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>83</td>
</tr>
<tr>
<td>70 to 80:</td>
<td>Fairly Easy</td>
<td>7&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>88</td>
</tr>
<tr>
<td>80 to 90:</td>
<td>Easy</td>
<td>6&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>91</td>
</tr>
<tr>
<td>90 to 100:</td>
<td>Very Easy</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>93</td>
</tr>
</tbody>
</table>
HOW TO USE THIS CHART
Draw a straight line from your measured "Words per Sentence" to your measured "Syllables per Word." The intersection of this line with the center column shows your readability score. The minimum score for Plain English is 60.
### Flesch’s Tabulated Scores for Different Types of Readings

<table>
<thead>
<tr>
<th>Publication</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comics</td>
<td>92</td>
</tr>
<tr>
<td>Consumer ads in magazines</td>
<td>82</td>
</tr>
<tr>
<td>Movie Screen</td>
<td>75</td>
</tr>
<tr>
<td>Seventeen</td>
<td>67</td>
</tr>
<tr>
<td>Reader's Digest</td>
<td>65</td>
</tr>
<tr>
<td>Sports Illustrated</td>
<td>63</td>
</tr>
<tr>
<td>New York Daily News</td>
<td>60</td>
</tr>
<tr>
<td>Atlantic Monthly</td>
<td>57</td>
</tr>
<tr>
<td>Time</td>
<td>52</td>
</tr>
<tr>
<td>Newsweek</td>
<td>50</td>
</tr>
<tr>
<td>Wall Street Journal</td>
<td>43</td>
</tr>
<tr>
<td>Harvard Business Review</td>
<td>43</td>
</tr>
<tr>
<td>New York Times</td>
<td>39</td>
</tr>
<tr>
<td>New York Review of Books</td>
<td>35</td>
</tr>
<tr>
<td>Harvard Law Review</td>
<td>32</td>
</tr>
<tr>
<td>Standard auto insurance policy</td>
<td>10</td>
</tr>
<tr>
<td>Internal Revenue Code</td>
<td>minus 6</td>
</tr>
</tbody>
</table>
Dale–Chall formula

To apply the formula:

1. Select several 100-word samples throughout the text.
2. Compute the average sentence length in words (divide the number of words by the number of sentences).
3. Compute the percentage of words NOT on the Dale–Chall word list of 3,000 easy words.
4. Compute this equation:

\[ \text{Raw Score} = 0.1579 \times (\text{PDW}) + 0.0496 \times (\text{ASL}) + 3.6365 \]

where:
Raw Score = uncorrected reading grade of a student who can answer one-half of the test questions on a passage.
PDW = Percentage of Difficult Words not on the Dale–Chall word list.
ASL = Average Sentence Length
Other factors

- Image words, abstraction, direct and indirect statements, types of narration and sentences, phrases, and clauses.
- Difficult concepts.
- Idea density.
- Human interest.
- Nominalization.
- Active and passive voice.
- Embeddedness.
- Structural cues.
- The use of images.
- Diagrams and line graphs.
- Highlighting.
- Fonts and layout.
- Document age.
Grammar checking

- Prescriptive vs. descriptive grammar
- Some institutions have their own writing style
- Writing tools can enforce these
- Approaches:
  - Grammar engineering
  - Statistical analysis
  - N-grams (language models)
Controlled languages

• Form of language usage restricted by grammar and vocabulary rules
• Sublanguage: specialized jargon used for communication between experts in certain fields of knowledge
• Goals
  – Consistent authoring, eliminate ambiguity, reduce misinterpretation, enable high-quality translation
Church Basic English

- Can we define a core LDS vocabulary?
- Can we use it to increase readability?
- ...


Senses of “serve (v)”:  

stative verb meanings:  
  1. function  
  3. effectuate, bring about  
  8. promote, benefit  
  9. serve, do, spend, pass  
  12. suffice, do, answer, serve, satisfy, fulfill  

social verb meanings:  
  2. do duty in a specific function  
  4. be used by, as a utility  
  7. serve  
  10. attend to, assist  
  13. do military service  

contact verb meanings:  
  11. serve, process, swear out  
  14. breed  

consumption verb meanings:  
  5. help with food or drink  
  6. serve up, dish out, dish up  

competition verb meaning:  
  15. put the ball into play
Senses of “serve (v)”: usages from General Conference

stative verb meanings:
1. function: “...I think it will serve a purpose...”
3. effectuate, bring about: “...they had only served to confirm my faith...”
8. promote, benefit: “…the stickiness of the berry served to attach it to the tree...”
9. serve, do, spend, pass: “…I served a mission...”
12. suffice, do, answer, serve, satisfy, fulfill: “…it may serve other purposes...”

social verb meanings:
2. do duty in a specific function: “... serve as a visiting teacher...”
4. be used by, as a utility: “…this tabernacle has served us well...”
7. serve: “…the opportunity we have had of serving the Church...”
10. attend to, assist: “God help us all to serve one another...”
13. do military service: “We have young men serving in the military...”

contact verb meanings:
11. serve, process, swear out: “People go around serving notice...”
14. breed (N/A)

consumption verb meanings:
5. help with food or drink: “…we had been served our breakfast...”
6. serve up, dish out, dish up: “…the wild music that is served up...”

competition verb meaning:
15. put the ball into play: “…in tennis if we didn’t serve well...”
WS complexity by text type

<table>
<thead>
<tr>
<th>Text Type</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GosFun</td>
<td>59</td>
</tr>
<tr>
<td>GConf</td>
<td>223</td>
</tr>
<tr>
<td>BoM</td>
<td>115</td>
</tr>
<tr>
<td>WSJ</td>
<td>105</td>
</tr>
<tr>
<td>Xerox</td>
<td>60</td>
</tr>
<tr>
<td>EUmtg</td>
<td>294</td>
</tr>
<tr>
<td>GosPrn</td>
<td>104</td>
</tr>
<tr>
<td>JSPam</td>
<td>103</td>
</tr>
</tbody>
</table>
Global English

“The widespread use of [Basic English] would be a gain to us far more desirable and fruitful than the annexation of great provinces.”

Winston Churchill
“Children need to learn English to succeed in the world today”

<table>
<thead>
<tr>
<th>Agree (Percent %)</th>
<th>South Korea</th>
<th>Indonesia</th>
<th>India</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98</td>
<td>96</td>
<td>93</td>
<td>92</td>
</tr>
</tbody>
</table>
Web Content

English 68%

Japanese 6%
German 6%
Chinese 4%
French 3%
Spanish 2%
Russian 2%
Italian 2%
Portuguese 1%
Korean 1%
Other 5%

Source: Vilaweb.com, as quoted by eMarketer
Worldwide Convergence

Global English

Church Global English

Church Linguistic Correlation

- Identify essential Church vocabulary
- Church Vocabulary Lists
- 30 Years of General Conference talks
Church Linguistic Correlation
Essential Church Vocabulary

Foundation Principles and Ordinances Document

A Testimony of God the Father and His Son Jesus Christ and the Gospel
The Witness and Inspiration of the Holy Ghost
The Keys, Offices, Quorums, and Authority of the Priesthood

Church Vocabulary Lists

Aaronic Priesthood
Apostle
area (administrative area of the Church)
Area Authority
Area Presidency
Articles of Faith
Articles of Faith (text of)
atonement
auxiliary (organization)
baptism
baptism prayer (text of)
bishop
Bishopric
Book of Mormon
Essential Church Vocabulary
Church Global English Coverage:

93.2%  General Conference
99.6%  Gospel Fundamentals
99.7%  Joseph Smith pamphlet

<table>
<thead>
<tr>
<th># word families</th>
<th>Joseph Smith pamphlet</th>
<th>Gospel Fundamentals</th>
<th>Elder Oaks (Apr 03)</th>
<th>Elder Maxwell (Apr 03)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.98</td>
<td>0.95</td>
<td>0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>1000</td>
<td>0.99</td>
<td>0.99</td>
<td>0.88</td>
<td>0.82</td>
</tr>
<tr>
<td>1500</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>2000</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Worldwide Convergence

Global English

Essential Church Vocabulary

Church Global English

1830 word families

- Church Vocabulary Lists
- 30 Years of General Conference talks
- Global English
My *beloved* brethren and sisters throughout the world, we send *greetings* in the name of our Redeemer. We send our love and our blessing. I *commend* you most warmly on what you are doing to move forward the work of the Lord.

Now and again I quietly *reflect* on the growth and *impact* of this work. I *reflect* on that meeting with a few present in the Peter Whitmer *farmhouse* on the 6th of April 1830. Here the Church was organized, and here began the long march which has brought it to its present *stature*.

My *dear* brethren and sisters throughout the world, we *welcome* you in the name of our Redeemer. We send our love and our blessing. I *praise* you most warmly on what you are doing to move forward the work of the Lord.

Now and again I quietly *think about* the growth and *influence* of this work. I *think about* that meeting with a few present in the Peter Whitmer *home* on the 6th of April 1830. Here the Church was organized, and here began the long march which has brought it to its present *position*. 
On the evening before he was *martyred* in Carthage, the Prophet Joseph Smith bore testimony to his guards. He testified of the *divine* *authenticity* of the Book of Mormon. He bore witness of the administration of angels and that the kingdom of God was again established upon the earth.
Words Appearing for First Time
(General Conferences Oct 2001 - Apr 2003)

anchorman
cumulatively
daredevils
dvd
freebies
generic
olympians
pagers
pleat
portfolio
shooed
skunk
stenosis
tricuspid
venues
beseechingly
bullhorn
chug
clowns
evacuation
fascinatingly
gullibility
headstrong
heft
impish
kennels
lioness
lordly
metaphorically
paddies
affectations
antivirus
avocational
chafing
cubicle
emulative
etymological
fridge
megabytes
hyperventilating
languished
locomotives
perfectionism
pigsty
portaging
amok
appeased
bos'n's
cyclopaedia
derides
effacing
ferreted
footholds
gill
huff
lecherous
napkins
nonjudgmentalism
postmodernism
sitcom
Global English by Church Department

<table>
<thead>
<tr>
<th></th>
<th>Distinct Words</th>
<th>Words Not on Base List</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTC (&quot;Preach My Gospel&quot;)</td>
<td>7,256</td>
<td>1,020</td>
<td>14%</td>
</tr>
<tr>
<td>Curriculum</td>
<td>2,753</td>
<td>354</td>
<td>13%</td>
</tr>
<tr>
<td>CES</td>
<td>4,038</td>
<td>665</td>
<td>17%</td>
</tr>
<tr>
<td>Family History</td>
<td>7,754</td>
<td>2,497</td>
<td>32%</td>
</tr>
<tr>
<td>Mormon.org</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>