**Appendix A: Subcorpora Composition and Simplicity Analyses**

In addition to what we have written in the main article text about our analysis, the following technical points regarding the composition of our subcorpora and the analysis may be of interest:

* Since sentence length impacts readability scores, punctuation was left up to the judgement of the transcribers. When they thought that the sentence had come to an end, they introduced an end-of-sentence mark (full stop, question mark etc.). As noted in the text, leaders were each transcribed by multiple students to avoid individual transcription preferences/idiosyncrasies biasing results. The authors (who are fluent in English, French and Italian) checked a random selection of each (20,000 words per leader).
* Words were counted using the *Taltac2* (www.taltac.com) software following a light tokenization procedure, i.e. capital letters determined by the syntactic context (e.g. at the beginning of sentences) were decapitalized (to avoid, for example, the software considering “And” and “and” as two distinct forms) and running words (or “tokens”) were identified by a set of boundaries such as blanks or apostrophes, but not in terms of multi-words such as “credit card” or “tug of war”. To take an example: “l’amico” (“the friend” in Italian) counts as two tokens because the apostrophe is considered as a word boundary, unlike hyphens (e.g. “by-election” counts as one token). Fixed expressions like “credit card” are counted as two tokens. It is worth recalling here the difference between “token” and “type”, i.e. a word form that may occur repeatedly in a corpus. For example, the sentence “I know that you know that I know” contains eight tokens but only four types, because “I”, “know” and “that” are repeated. All our subcorpora can be considered statistically reliable since their type-token ratios are always smaller than 1:5 and the percentage of hapax legomena (word-types occurring only once in a corpus) is below 50 per cent.[[1]](#footnote-1)
* Lexical richnesswas calculated as the type-token ratio (TTR)[[2]](#footnote-2) and hapax percentage (i.e. the percentage of words occurring only once out of the total amount of types in a corpus; henceforth: hapax) of our subcorpora. We calculated these using *Taltac2*. To calculate the Lemma-Token Ratio (LTR) and Lexical density, the identification of lemmas and word classes (i.e. adjectives, nouns, verbs etc.), known as Part-of-Speech or POS-Tagging, was carried out using *Treetagger*, a software developed for several languages available from: http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.
* To assess lexical difficulty, we calculated the incidence of words considered difficult in the English and Italian languages. The Dale-Chall formula calculates at which US grade level an English text may be deemed accessible based on its length, along with the incidence of a list of 3,000 common words considered familiar to most fourth-grade students (Dale and Chall, 1948; Chall and Dale 1995). Scores were calculated using the software freely available from: https://github.com/joaopalotti/readability\_calculator. In the case of Italian, we used the www.corrige.it website to calculate the incidence of the Vocabolario di Base (VdB, i.e. the basic vocabulary). This is a list of approximately 7,000 lemmas that are among the most frequent in a wide range of texts and/or used to describe everyday concepts familiar to most Italian speakers (De Mauro 1997). In short, for both English and Italian, the assumption is that the percentage of common words in a corpus is an indicator of linguistic simplicity. Consequently, the higher the percentage of words in a corpus that are not on the list of words considered common, the more difficult it is for the average person to understand. No such tool is available for French. While the “échelle Dubois Buyse” (Mayer et al. 1995) provides a list of 3,787 words said to be known by 80 per cent of the French population, it focuses on the orthography of written text rather than comprehension proper, and we therefore decided to discard it.

The remainder of Appendix A provides additional details about the composition of each subcorpus for our 10 leaders and the simplicity analyses conducted. As in the article text, we list these results below country-by-country.

**UNITED STATES**

**Table A.1: Donald Trump speeches**

|  |  |  |
| --- | --- | --- |
|  | **Place** | **Date** |
| 1 | Monessen, Pennsylvania | 06.28.2016 |
| 2 | New York City, New York | 07.16.2016 |
| 3 | Cleveland, Ohio | 07.21.2016 |
| 4 | Green Bay, Wisconsin | 08.05.2016 |
| 5 | Detroit, Michigan | 08.08.2016 |
| 6 | Fayetteville, North Carolina | 08.09.2016 |
| 7 | Youngstown, Ohio | 08.15.2016 |
| 8 | Dimondale, Michigan | 08.19.2016 |
| 9 | Akron, Ohio | 08.22.2016 |
| 10 | Jackson, Mississippi | 08.24.2016 |
| 11 | Washington DC | 08.30.2016 |
| 12 | Phoenix, Arizona | 08.31.2016 |
| 13 | Detroit, Michigan | 09.03.2016 |
| 14 | Philadelphia, Pennsylvania | 09.07.2016 |
| 15 | Washington DC | 09.09.2016 |
| 16 | Miami, Florida | 09.16.2016 |
| 17 | West Palm Beach, Florida | 10.13.2016 |
| 18 | Gettysburg, Pennsylvania | 10.22.2016 |
| 19 | Valley Forge, Pennsylvania | 11.01.2016 |
| 20 | Eau Claire, Wisconsin | 11.01.2016 |
| 21 | Pensacola, Florida | 11.02.2016 |
| 22 | Raleigh, North Carolina | 11.07.2016 |

**Table A.2: Hillary Clinton speeches**

|  |  |  |
| --- | --- | --- |
|  | **Place** | **Date** |
| 1 | New York City, New York | 06.11.2015 |
| 2 | Charlotte, North Carolina | 07.05.2016 |
| 3 | Hampton, New Hampshire | 09.05.2016 |
| 4 | Washington DC | 06.14.2016 |
| 5 | Raleigh, North Carolina | 06.23.2016 |
| 6 | Atlantic City, New Jersey | 07.06.2016 |
| 7 | Philadelphia, Pennsylvania | 07.11.2016 |
| 8 | Portsmouth, New Hampshire | 07.12.2016 |
| 9 | Springfield, Illinois | 07.13.2016 |
| 10 | Cincinnati, Ohio | 07.18.2016 |
| 11 | Washington DC | 07.26.2016 |
| 12 | Johnstown, Pennsylvania | 07.30.2016 |
| 13 | Pittsburgh, Pennsylvania | 07.30.2016 |
| 14 | Youngstown, Ohio | 07.30.2016 |
| 15 | Columbus, Ohio | 07.31.2016 |
| 16 | Tampa, Florida | 07.22.2016 |
| 17 | Philadelphia, Pennsylvania | 07.28.2016 |
| 18 | Omaha, Nebraska | 08.01.2016 |
| 19 | Commerce City, Colorado | 08.03.2016 |
| 20 | Warren, Michigan | 08.11.2016 |
| 21 | Reno, Nevada | 08.25.2016 |
| 22 | Greensboro, North Carolina | 09.15.2016 |
| 23 | Washington DC | 09.16.2016 |
| 24 | Washington DC | 09.18.2016 |
| 25 | Philadelphia, Pennsylvania | 09.19.2016 |
| 26 | Orlando, Florida | 09.21.2016 |
| 27 | Charlotte, North Carolina | 10.02.2016 |
| 28 | Akron, Ohio | 10.03.2016 |
| 29 | Columbus, Ohio | 10.10.2016 |
| 30 | Miami, Florida | 10.11.2016 |

**Readability**

All indices were calculated using the following website: https://www.webfx.com/tools/read-able/.

**Table A3: Flesch-Kinkaid Reading Ease Ranking System**

|  |  |
| --- | --- |
| **Score** | **Comprehensibility** |
| 90-100 | Very easy |
| 80-89 | Easy |
| 70-79 | Fairly easy |
| 60-69 | Standard |
| 50-59 | Fairly difficult |
| 30-49 | Difficult |
| 0-29 | Very Difficult |

**Table A4: Flesch-Kincaid Grade Levels**

|  |  |
| --- | --- |
| **US Grade Level** | **Student Age** |
| 12 | 17 |
| 11 | 16 |
| 10 | 15 |
| 9 | 14 |
| 8 | 13 |
| 7 | 12 |
| 6 | 11 |
| 5 | 10 |
| 4 | 9 |
| 3 | 8 |
| 2 | 7 |

**Table A5: Dale-Chall Grade Level Scale**

|  |  |
| --- | --- |
| **Score** | **Grade Level** |
| 4.9 and below | Grade 4 |
| 5.0-5.9 | Grades 5-6 |
| 6.0-6.9 | Grades 7-8 |
| 7.0-7.9 | Grades 9-10 |
| 8.0-8.9 | Grades 11-12 |
| 9.0-9.9 | College |
| 10 and above | College graduate |

**Table A.6: Readability Scores (Trump and Clinton)**

|  |  |  |
| --- | --- | --- |
|  | **Trump** | **Clinton** |
| Flesch-Kincaid Reading Ease | 72.4 | 71 |
| Flesch-Kincaid Grade Level | 6.3 | 7.1 |
| Gunning Fog Score | 8.4 | 9.4 |
| SMOG Index | 7 | 7.5 |
| Coleman Liau Index | 9.2 | 9.6 |
| Automated Readability Index | 4.9 | 6.4 |
| Complex words % | 10.97% | 10.75% |
| Average words per sentence | 12.55 | 15.02 |
| Average syllables per word | 1.44 | 1.43 |

**Table A.7: Lexical Richness (Trump and Clinton)**

|  |  |  |
| --- | --- | --- |
|  | **Trump** | **Clinton** |
| Tokens (N) | 102,976 | 102,016 |
| Types (V) | 6,121 | 6,614 |
| (V/N)% | 5.944 | 6.483 |
| Hapax% | 40.042 | 43.045 |
| (Lemmas/V)% | 4.086 | 4.524 |

**Table A.8: Lexical Density and Difficult Words (Trump and Clinton)**

|  |  |  |
| --- | --- | --- |
|  | **Trump** | **Clinton** |
| Content words | 58.5775 | 57.0733 |
| Grammar words | 41.4225 | 42.9267 |
| Dale-Chall score | 4.205 | 4.618 |

**Note:** Grammar and content word scores are percentages, Dale-Chall score is the grade level.

**FRANCE**

**Table A.9: Marine Le Pen speeches**

|  |  |  |
| --- | --- | --- |
|  | **Place** | **Date** |
| 1 | Fréjus | 09.18.2016 |
| 2 | Paris | 12.09.2016 |
| 3 | Lyon | 02.05.2017 |
| 4 | Pierrelatte | 02.25.2017 |
| 5 | Nantes | 02.26.2017 |
| 6 | Mont Saint-Michel | 02.27.2017 |
| 7 | Chateauroux | 03.11.2017 |
| 8 | Mirande | 03.09.2017 |
| 9 | Paris | 03.13.2017 |
| 10 | Metz | 03.18.2017 |
| 11 | Saint Raphael | 03.21.2017 |
| 12 | N’Djaména | 03.23.2017 |
| 13 | Lille | 03.27.2017 |
| 14 | Bordeaux | 04.02.2017 |
| 15 | Monswiller | 04.05.2017 |
| 16 | Ajaccio | 04.09.2017 |
| 17 | Arcis-sur-Aube | 04.11.2017 |
| 18 | Pageas | 04.13.2017 |
| 19 | Henin-Beaumont | 04.23.2017 |
| 20 | Paris | 05.07.2017 |

**Table A.10: Emmanuel Macron speeches**

|  |  |  |
| --- | --- | --- |
|  | **Place** | **Date** |
| 1 | Bobigny | 11.16.2016 |
| 2 | Angers | 02.28.2017 |
| 3 | Caen | 03.04.2017 |
| 4 | Talence | 03.09.2017 |
| 5 | Reims | 03.17.2017 |
| 6 | Dijon | 03.25.2017 |
| 7 | Mayotte | 03.26.2017 |
| 8 | Marseille | 04.01.2017 |
| 9 | Furiani | 04.07.2017 |
| 10 | Besançon | 04.11.2017 |
| 11 | Pau | 04.12.2017 |
| 12 | Chatellerault | 04.18.2017 |
| 13 | Nantes | 04.19.2017 |
| 14 | Victoire | 04.23.2017 |
| 15 | Arras | 04.26.2017 |
| 16 | Albi | 05.04.2017 |

**Readability**

The Kandel-Moles score was calculated using the following website: https://visual-seo.com/it/SEO-Software-Features/Readability-Analysis. The higher the score, the more readable the text.

The LIX index was calculated with the software available from https://kwichmann.github.io/my\_a2z/Week02/lix/. The lower the score, the more readable the text.

The average word length (in letters) and sentence length (in words) were calculated using www.corrige.it.

**Table A.11: Kandel-Moles Reading Ease Ranking System**

|  |  |
| --- | --- |
| **Index** | **Difficulty** |
| 90-100 | Very easy |
| 80-89 | Easy |
| 70-78 | Fairly easy |
| 60-69 | Standard |
| 50-59 | Fairly difficult |
| 30-49 | Difficult |
| 0-29 | Very difficult |

**Table A.12: Readability Scores (Le Pen and Macron)**

|  |  |  |
| --- | --- | --- |
|  | **Le Pen** | **Macron** |
| Kandel-Moles Index | 57 | 65 |
| LIX Index | 48 | 44 |
| Average words per sentence | 22.27 | 21.51 |
| Average letters per word | 4.76 | 4.48 |

**Table A.13: Lexical Richness (Le Pen and Macron)**

|  |  |  |
| --- | --- | --- |
|  | **Le Pen** | **Macron** |
| Tokens (N) | 105,774 | 104,074 |
| Types (V) | 11,086 | 7,463 |
| (V/N)% | 10.481 | 7.171 |
| Hapax% | 51.515 | 48.841 |
| (Lemmas/V)% | 5.676 | 3.841 |

**Table A14: Lexical Density (Le Pen and Macron)**

|  |  |  |
| --- | --- | --- |
|  | **Le Pen** | **Macron** |
| Content words | 53.1586 | 51.272 |
| Grammar words | 46.8414 | 48.728 |

**Note:** Grammar and content word scores are percentages.

**UNITED KINGDOM**

**Table A.15: Nigel Farage speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Sedgley | 07.17.2012 | UKIP public meeting |
| 2 | Belfast | 07.09.2013 | UKIP Campaign speech |
| 3 | Canada | 07.NA.2013 | UKIP public meeting |
| 4 | Northwich | 01.24.2014 | UKIP public meeting |
| 5 | Chalfont Saint Peter | 01.25.2014 | UKIP public meeting |
| 6 | Torquay | 02.28.2014 | UKIP Spring Conference |
| 7 | Basingstoke | 04.05.2014 | UKIP public meeting |
| 8 | Gateshead | 04.23.2014 | UKIP public meeting |
| 9 | Swansea | 04.30.2014 | UKIP public meeting |
| 10 | Derby | 05.01.2014 | UKIP public meeting |
| 11 | Cambridgeshire | 05.02.2014 | UKIP public meeting |
| 12 | London | 05.17.2014 | UKIP public meeting |
| 13 | Margate | 05.18.2014 | UKIP public meeting |
| 14 | Eastbourne | 06.07.2014 | UKIP Annual Conference |
| 15 | South Thanet | 08.26.2014 | Candidature speech |
| 16 | Switzerland | 10.09.2014 | Lecture in support of Sons of Libertas |
| 17 | Rochester & Strood | 11.04.2014 | UKIP public meeting |
| 18 | Maryland | 02.27.2015 | CPAC 2015 |
| 19 | Dudley | 04.07.2015 | UKIP Campaign Speech |
| 20 | Grimsby | 04.08.2015 | UKIP public meeting |
| 21 | Westminster | 09.04.2015 | NO campaign launch |
| 22 | Doncaster | 09.25.2015 | UKIP Conference |
| 23 | Gateshead | 10.12.2015 | No to the EU |
| 24 | Chester | 11.06.2015 | No to the EU |
| 25 | Basingstoke | 11.16.2015 | Speech after terrorist attacks in Paris |
| 26 | London | 11.20.2015 | No to the EU |
| 27 | Leeds | 11.30.2015 | No to the EU |
| 28 | Bolton | 02.08.2016 | Brexit Speech |
| 29 | Llandudno | 02.27.2016 | UKIP Spring Conference |
| 30 | Thornbury | 03.05.2016 | UKIP South West Conference |
| 31 | Peterborough | 03.16.2016 | No to the EU |
| 32 | London | 04.29.2016 | Speech on immigration and Brexit |
| 33 | Worcester | 04.30.2016 | UKIP event, Worcestershire County Cricket Club |
| 34 | Bolton | 05.25.2016 | UKIP & Labour Leave Conference |
| 35 | London | 06.22.2016 | Last speech before the Referendum |

**Table A.16: Ed Miliband speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Manchester | 09.28.2010 | 2010 Labour Conference |
| 2 | Llandudno | 02.19.2011 | 2011 Welsh Labour Conference |
| 3 | London | 06.13.2011 | Speech at the Coin Street Neighbourhood Centre |
| 4 | Liverpool | 09.27.2011 | 2011 Labour Conference |
| 5 | Dundee | 03.03.2012 | 2012 Scottish Labour Conference |
| 6 | London | 06.22.2012 | Speech to the Institute for Public Policy Research |
| 7 | London | 07.02.2012 | Speech to Fabian Society |
| 8 | Manchester | 10.02.2012 | 2012 Labour Conference |
| 9 | Inverness | 04.20.2013 | 2013 Scottish Labour Conference |
| 10 | Ipswich | 04.08.2013 | Local election speech |
| 11 | London | 07.09.2013 | Speech at the St. Bride’s Foundation |
| 12 | Bournemouth | 09.10.2013 | Trades Union Congress |
| 13 | Brighton | 09.24.2013 | 2013 Labour Conference |
| 14 | Birmingham | 04.08.2014 | Rally speech to supporters |
| 15 | London | 03.01.2014 | 2014 Labour Special Conference |
| 16 | Perth | 03.21.2014 | 2014 Scottish Labour Conference |
| 17 | Glasgow | 05.01.2014 | Speech on Scottish independence at Emirates Arena |
| 18 | Redbridge | 05.01.2014 | Campaign Launch, Redbridge Leisure Centre |
| 19 | Manchester | 05.12.2014 | Speech on the NHS |
| 20 | Thurrock | 05.27.2014 | Local Heroes Reception |
| 21 | Nottingham | 06.12.2014 | GMB Congress |
| 22 | London | 06.17.2014 | Speech to Labour Friends of Israel |
| 23 | London | 07.03.2014 | Policy Network Inclusive Prosperity Conference |
| 24 | London | 07.19.2014 | National Policy Forum |
| 25 | Cumbernauld | 09.10.2014 | Speech at Forge Community Centre |
| 26 | Manchester | 09.23.2014 | 2014 Labour Annual Conference |
| 27 | Great Yarmouth | 12.15.2014 | Launch of Labour's 2nd election pledge |
| 28 | Salford | 01.05.2015 | 2015 General election campaign |
| 29 | Swansea | 02.14.2015 | 2015 Welsh Labour conference |
| 30 | Stratford | 03.27.2015 | 2015 Campaign Launch |
| 31 | Manchester | 04.13.2015 | 2015 Manifesto Launch |

**Table A.17: David Cameron speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Birmingham | 10.01.2008 | 2008 Conservative Conference |
| 2 | Manchester | 10.08.2009 | 2009 Conservative Conference |
| 3 | London | 04.07.2010 | Speech on the 2010 General Election |
| 4 | Brecon | 04.10.2010 | 2010 Welsh Manifesto launch |
| 5 | Belfast | 05.04.2010 | 2010 Campaign speech |
| 6 | London | 05.11.2010 | 2010 General election victory speech |
| 7 | London | 07.08.2010 | Civil Service Live event |
| 8 | Birmingham | 10.06.2010 | 2010 Conservative Conference |
| 9 | Liverpool | 02.14.2011 | Speech at Liverpool Hope University |
| 10 | Manchester | 10.05.2011 | 2011 Conservative Conference |
| 11 | Bluewater | 06.25.2012 | Speech at Bluewater shopping centre in Kent |
| 12 | Birmingham | 10.10.2012 | 2012 Conservative Conference |
| 13 | London | 01.23.2013 | Speech about the EU’s future |
| 14 | London | 05.09.2013 | 2013 Global Investment Conference |
| 15 | Manchester | 10.02.2013 | 2013 Conservative Conference |
| 16 | London | 04.05.2014 | 2014 National Conservative Conference speech |
| 17 | Aberdeen | 09.15.2014 | Scottish Independence Speech |
| 18 | Birmingham | 09.29.2014 | 2014 Conservative Conference |
| 19 | Manchester | 03.28.2015 | 2015 Conservative Spring forum |
| 20 | Corsham | 03.30.2015 | 2015 Campaign launch speech |
| 21 | Wiltshire | 04.13.2015 | 2015 Conservative Party manifesto launch |
| 22 | London | 05.05.2015 | 2015 Campaign Speech |
| 23 | Manchester | 10.07.2015 | 2015 Conservative Conference |
| 24 | Llangollen | 03.03.2016 | 2016 Welsh Conservative conference |
| 25 | London | 04.10.2016 | 2016 Conservative Spring forum |
| 26 | London | 05.09.2016 | Speech on the UK's strength and security in the EU |

**Table A.18: Flesch-Kinkaid Reading Ease Ranking System**

|  |  |
| --- | --- |
| **Score** | **Comprehensibility** |
| 90-100 | Very easy |
| 80-89 | Easy |
| 70-79 | Fairly easy |
| 60-69 | Standard |
| 50-59 | Fairly difficult |
| 30-49 | Difficult |
| 0-29 | Very Difficult |

**Table A.19: Flesch-Kincaid Grade Levels**

|  |  |
| --- | --- |
| **US Grade Level** | **Student Age** |
| 12 | 17 |
| 11 | 16 |
| 10 | 15 |
| 9 | 14 |
| 8 | 13 |
| 7 | 12 |
| 6 | 11 |
| 5 | 10 |
| 4 | 9 |
| 3 | 8 |
| 2 | 7 |

**Table A.20: Dale-Chall Grade Level Scale**

|  |  |
| --- | --- |
| **Score** | **Grade Level** |
| 4.9 and below | Grade 4 |
| 5.0-5.9 | Grades 5-6 |
| 6.0-6.9 | Grades 7-8 |
| 7.0-7.9 | Grades 9-10 |
| 8.0-8.9 | Grades 11-12 |
| 9.0-9.9 | College |
| 10 and above | College graduate |

**Readability**

All indices were calculated by the following website: https://www.webfx.com/tools/read-able/.

**Table A.21: Readability Scores (Farage, Miliband and Cameron)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Farage** | **Miliband** | **Cameron** |
| Flesch-Kincaid Reading Ease | 63.2 | 72.9 | 71.8 |
| Flesch-Kincaid Grade Level | 10.6 | 6.6 | 7.1 |
| Gunning Fog Score | 13 | 9 | 9.6 |
| SMOG Index | 8.9 | 6.9 | 7.2 |
| Coleman Liau Index | 8.8 | 9.6 | 9.6 |
| Automated Readability Index | 10.6 | 5.9 | 6.6 |
| Complex words % | 9.54% | 9.66% | 9.72% |
| Average words per sentence | 24.61 | 13.99 | 15.49 |
| Average syllables per word | 1.4 | 1.42 | 1.41 |

**Table A.22: Lexical Richness (Farage, Miliband and Cameron)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Farage** | **Miliband** | **Cameron** |
| Tokens (N) | 103,275 | 102,858 | 103,618 |
| Types (V) | 6,208 | 5,965 | 7,391 |
| (V/N)% | 6.011 | 5.799 | 7.133 |
| Hapax% | 43.847 | 42.129 | 47.179 |
| (Lemmas/V)% | 4.262 | 4.019 | 4.842 |

**Table A.23: Lexical Density and Difficult Words (Farage, Miliband and Cameron)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Farage** | **Miliband** | **Cameron** |
| Content words | 55,026 | 55,9148 | 57,5175 |
| Grammar words | 44,974 | 44,0852 | 42,4825 |
| Dale-Chall score | 4.654 | 4.267 | 4.326 |

**Note:** Grammar and content word scores are percentages, Dale-Chall score is the grade level.

**ITALY**

**Table A.24: Matteo Salvini speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Milan | 02.22.2014 | Basta Euro (‘Enough of the Euro’) tour |
| 2 | Buja | 04.05.2014 | Basta Euro tour |
| 3 | Verona | 04.06.2014 | Rally for arrested independence activists |
| 4 | Bergamo | 04.13.2014 | Basta Euro tour |
| 5 | Treviso | 04.24.2014 | Basta Euro tour |
| 6 | Monza | 04.30.2014 | Basta Euro tour |
| 7 | Cagliari | 05.02.2014 | Basta Euro tour |
| 8 | Genoa | 05.03.2014 | Basta Euro tour |
| 9 | Pontida | 04.04.2014 | Lega Nord annual Pontida rally |
| 10 | Lamezia Terme | 05.05.2014 | Basta Euro tour |
| 11 | Rome | 05.11.2014 | Basta Euro tour |
| 12 | Padua | 07.20.2014 | Lega Nord Federal Congress |
| 13 | Milan | 10.18.2014 | ‘Stop the Invasion’ rally. |
| 14 | Rome | 12.19.2014 | Presentation of ‘Noi con Salvini’ |
| 15 | Bergamo | 12.28.2014 | Berghem Frecc |
| 16 | Rome | 02.28.2015 | ‘Renzi a casa!’ (Send Renzi home!) rally |
| 17 | Montesilvano | 03.21.2015 | Lega Nord rally |
| 18 | Terni | 04.14.2015 | Regional elections rally |
| 19 | Reggio Emilia | 04.25.2015 | Lega Nord rally |
| 20 | Catania | 05.27.2015 | Lega Nord elections campaign rally |
| 21 | Martinengo | 06.02.2015 | Lega Nord rally |
| 22 | Pontida | 06.21.2015 | Lega Nord annual Pontida rally |
| 23 | Milan | 07.04.2015 | ‘Euro o Libertà’ (Euro or freedom) rally |
| 24 | Pontida | 09.18.2016 | Lega Nord annual rally |
| 25 | Florence | 11.12.2016 | #iovotoNO (I vote ‘no’) |

**Table A.25: Matteo Renzi speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Rome | 12.08.2013 | PD primaries victory rally |
| 2 | Milan | 12.15.2013 | PD national assembly |
| 3 | Rome | 02.13.2014 | PD meeting |
| 4 | Palermo | 05.14.2014 | European elections rally |
| 5 | Cesena | 05.16.2014 | European elections rally |
| 6 | Bergamo | 05.20.2014 | Local and European elections rally |
| 7 | Rome | 05.22.2014 | European elections rally |
| 8 | Rome | 06.14.2014 | PD national assembly |
| 9 | Rome | 12.08.2014 | Speech at #factory365 (Young PD meeting) |
| 10 | Rome | 12.14.2014 | PD national assembly |
| 12 | Florence | 12.13.2015 | Closing speech at Leopolda 5 |
| 13 | Rome | 02.07.2016 | Speech at PD training school |
| 14 | Rome | 03.13.2016 | PD national assembly |
| 15 | Rome | 04.04.2016 | PD meeting |
| 16 | Milan | 05.31.2016 | Local elections rally |
| 17 | Rome | 07.04.2016 | PD meeting |
| 18 | Rome | 07.23.2016 | PD national assembly |
| 19 | Rome | 10.29.2016 | Referendum rally in Piazza del Popolo |
| 20 | Florence | 11.06.2016 | Closing speech at ‘Leopolda 6’ meeting |

**Table A.26: Angelino Alfano speeches**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Place** | **Date** | **Event** |
| 1 | Rome | 12.05.2013 | Presentation of the NCD logo |
| 2 | Rome | 12.07.2013 | NCD Convention |
| 3 | Rome | 02.05.2014 | The centre-right in the Third Republic conference |
| 4 | Reggio Calabria | 02.08.2014 | NCD rally |
| 5 | Rome | 02.17.2014 | NCD elected representatives convention |
| 6 | Milan | 03.01.2014 | NCD Congress |
| 7 | Rivisondoli | 03.14.2014 | Meeting: Responsibly…European! |
| 8 | Bologna | 03.16.2014 | NCD conference on the family |
| 9 | Catania | 03.22.2014 | NCD rally |
| 10 | Palermo | 03.23.2014 | NCD Convention |
| 11 | Agrigento | 03.29.2014 | NCD Convention |
| 12 | Messina | 03.30.2014 | #Insieme ‘Together’ rally |
| 13 | Rome | 05.10.2014 | NCD European election rally |
| 14 | Rome | 09.03.2014 | NCD assembly |
| 15 | Frascati | 09.11.2014 | Speech at the Magna Carta Summer School |
| 16 | Chianciano Terme | 09.13.2014 | National meeting of the UDC |
| 17 | Ceglie Messapica | 10.12.2014 | National meeting of the NCD |
| 18 | Rome | 12.03.2014 | National NCD executive meeting |
| 19 | Rome | 03.30.2015 | Presentation of ‘Area Popolare’ |
| 20 | Rome | 09.10.2015 | Speech on the plan for the South |
| 21 | S. Giovanni Rotondo | 09.11.2015 | Speech at the opening of UDC national meeting |
| 22 | Sorrento | 09.27.2015 | Speech at Sorrento Summer School |
| 23 | Sorrento | 09.29.2015 | Speech at Sorrento Summer School |
| 24 | Teramo | 10.16.2016 | NCD referendum rally |

**Readability**

The LIX index was calculated with the software available from https://kwichmann.github.io/my\_a2z/Week02/lix/. The lower the score, the more readable the text.

The Gulpease score was calculated using the following website: www.corrige.it. The higher the score, the more readable the text. Gulpease scores are matched with the three main education levels of the Italian population: licenza elementare (grade 5), licenza media (grade 8), diploma superiore (grade 13).

**Table A.27: Gulpease Reading Ease Ranking System**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Very easy** | **Easy** | **Difficult** | **Very difficult** | **Almost incomprehensible** |
| Grade 5 | 100-96 | 95-81 | 80-71 | 70-56 | 55-0 |
| Grade 8 | 100-81 | 80-61 | 60-51 | 50-36 | 0-35 |
| Grade 13 | 100-71 | 70-41 | 40-31 | 30-11 | 0-10 |

The average word length (in letters) and sentence length (in words) were provided by www.corrige.it.

**Table A.28: Readability Scores (Salvini, Renzi and Alfano)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Salvini** | **Renzi** | **Alfano** |
| Gulpease Index | 62 | 56 | 50 |
| LIX Index | 49 | 56 | 65 |
| Average words per sentence | 17.22 | 22.52 | 33.78 |
| Average letters per word | 4.51 | 4.7 | 4.8 |

**Table A.29: Lexical Richness (Salvini, Renzi and Alfano)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Salvini** | **Renzi** | **Alfano** |
| Tokens (N) | 102,232 | 103,009 | 101,005 |
| Types (V) | 9,384 | 9,294 | 8,199 |
| (V/N)% | 9.179 | 9.023 | 8.117 |
| Hapax% | 48.21 | 48.838 | 45.262 |
| (Lemmas/V)% | 4.762 | 4.658 | 4.365 |

**Lexical density and difficulty**

The incidence of the basic vocabulary (Vocabolario di base=VdB) was calculated using the website www.corrige.it. The VdB is subdivided into three layers:

Fundamental: comprising the 1,991 most frequent lemmas in Italian;

Highly used: the next 2,750 most frequent lemmas in Italian;

Highly available: an additional 2,337 lemmas selected because, regardless of their frequency, they are considered known to the vast majority of speakers.

This version of the VdB was first published in 1980 (De Mauro, 1980) based on corpora collected in the previous decades. It was updated in 2016 (available from: https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana), but since no widely used software has implemented it, as a double-check we include calculations made with the *Bran* software available from https://github.com/zorbaproject/Bran.

**Table A.30: Lexical Density and Difficult Words (Salvini, Renzi and Alfano)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Salvini** | **Renzi** | **Alfano** |
| Content words% | 58.7352 | 59.0851 | 57.3324 |
| Grammar words% | 41.2648 | 40.9149 | 42.6676 |
| VdB1980% | 91.3 | 92.42 | 93.15 |
| Non-VdB1980% | 8.7 | 7.58 | 6.85 |
| Fundamental% | 85.71 | 87.01 | 87.45 |
| Highly used% | 4.46 | 4.55 | 5.27 |
| Highly available% | 1.13 | 0.87 | 0.85 |
| VdB2016% | 83.95 | 83,54 | 84,53 |
| Non-VdB2016% | 16.05 | 16.46 | 15.47 |

**BIBLIOGRAPHY**

Bérubé, Nicolas, Maxime Sainte-Marie, Philippe Mongeon and Vincent Larivière. 2018. “Words by the tail: Assessing lexical diversity in scholarly titles using frequency-rank distribution tail fits.” *PLoSONE* 13(7): e0197775. Available at: https://doi.org/10.1371/journal.pone.0197775.

Bolasco, Sergio. 1999. *Analisi multidimensionale dei dati*. Rome: Carocci.

Chall, Jeanne Sternlicht and Edgar Dale. 1995. *Readability Revisited: The New Dale-Chall Readability Formula*. Cambridge (MA): Brookline Books.

Dale, Edgar and Jeanne S. Chall. 1948. “A Formula for Predicting Readability.” *Educational Research Bulletin*, 27: 11–20.

De Mauro Tullio. 1997. *Guida all’uso delle parole*. Rome: Editori riuniti.

Lebart, Lucien and André Salem. 1994. *Statistique textuelle*. Paris: Dunot.

Mayer, Georges, François Ters and Daniel Reichenbach. 1995. *L’Echelle Dubois Buyse D’Orthographe Usuelle Francaise*. Paris: Pedagogie Mdi.

1. See Lebart and Salem (1994) and Bolasco (1999). With 51.5% hapaxes, Marine Le Pen is an exception but, since the deviation is limited, it does not affect the overall reliability of the subcorpus. [↑](#footnote-ref-1)
2. TTR has been termed “the most used and intuitive way to measure lexical diversity on the basis of word repetition patterns” (Bérubé et al. 2018: 4). [↑](#footnote-ref-2)