



Spotlight Session: Using MAXDictio

Dr. Daniel Rasch daniel.rasch@rub.de

Assistant Professor
German University of Administrative Sciences Speyer
Chair of Political Science

MAXDictio is accessed as a plug-in within MAXQDA. If you use MAXQDA Plus or MAXQDA Analytics Pro, MAXQDA's main ribbon menu contains an additional tab called "MAXDictio".

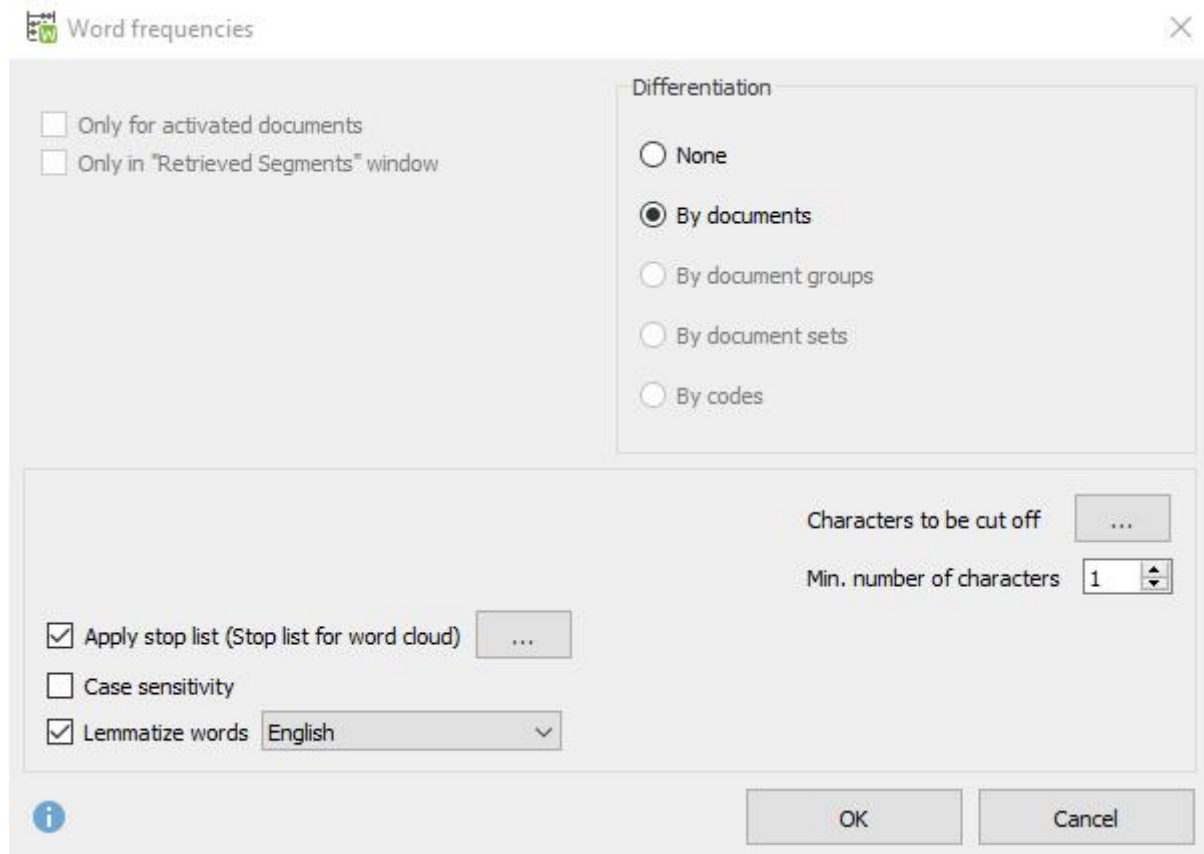
Literature

- Benoit, Kenneth & Laver, Michael (2003): Estimating Irish party policy positions using computer wordscoring: the 2002 election – a research note. *Irish Political Studies*, 18:1, 97-107.
- Boräng, F., et al. (2014). "Identifying frames: A comparison of research methods." *Interest Groups & Advocacy* 3(2): 188-201.
- Klüver, Heike (2009): Measuring Interest Group Influence Using Quantitative Text Analysis. *European Union Politics*, 10(4), 535–549.
- Laver, Michael, Benoit, Kenneth, & GARRY, John (2003): Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review*, 97(2), 311-331.
- Lowe, Will: Software for Content Analysis – A Review, <http://dl.conjugateprior.org/preprints/content-review.pdf>
- Proksch, Sven-Oliver and Slapin, Jonathan B. (2009): How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany'. *German Politics*, 18:3, pp. 323 – 344.
- Schonhardt-Bailey, Cheryl (2005): Measuring ideas more effectively: an analysis of Bush and Kerry's national security speeches. *PS - Political Science and Politics*, 38 (4). pp. 701-711.



Word frequencies

- Open the MAXQDA-project „MAXDictio_Rasch“.
- Let's start with the Word frequencies function and differentiate between the two documents. Go to **MAXDictio > Word Frequencies** in the ribbon-menu.



- As you can see, you can apply a stop list or even lemmatize the words.
- 'Stop lists' are lists of words or other data items which, for a specific reason, should be ignored or bypassed by a particular data processing operation.
- 'Lemmatizing' the identified words: they will be simplified to their word stems (*lemmas*) by using a lemma lexicon in the chosen language. For example, if a text contains the words "gave", "given", and "gives", MAXDictio will list the base form "give" in the results table only.
- In addition, you can differentiate the results by documents, document groups, document sets and even by codes.



- The results table shows the words, their length, their frequency, percentage of all words, the rank, in how many documents it appeared and how often it appeared in each text.
- We found 3667 words in total, but only 1032 different words, which makes a type-token-relation (TTR) of 0.28. The higher the TTR becomes, the more complex a text.

Word Cloud: Word frequencies

In 2 documents (3667 words total) 1032 Words (TTR = 0,2814)

Display top ranks

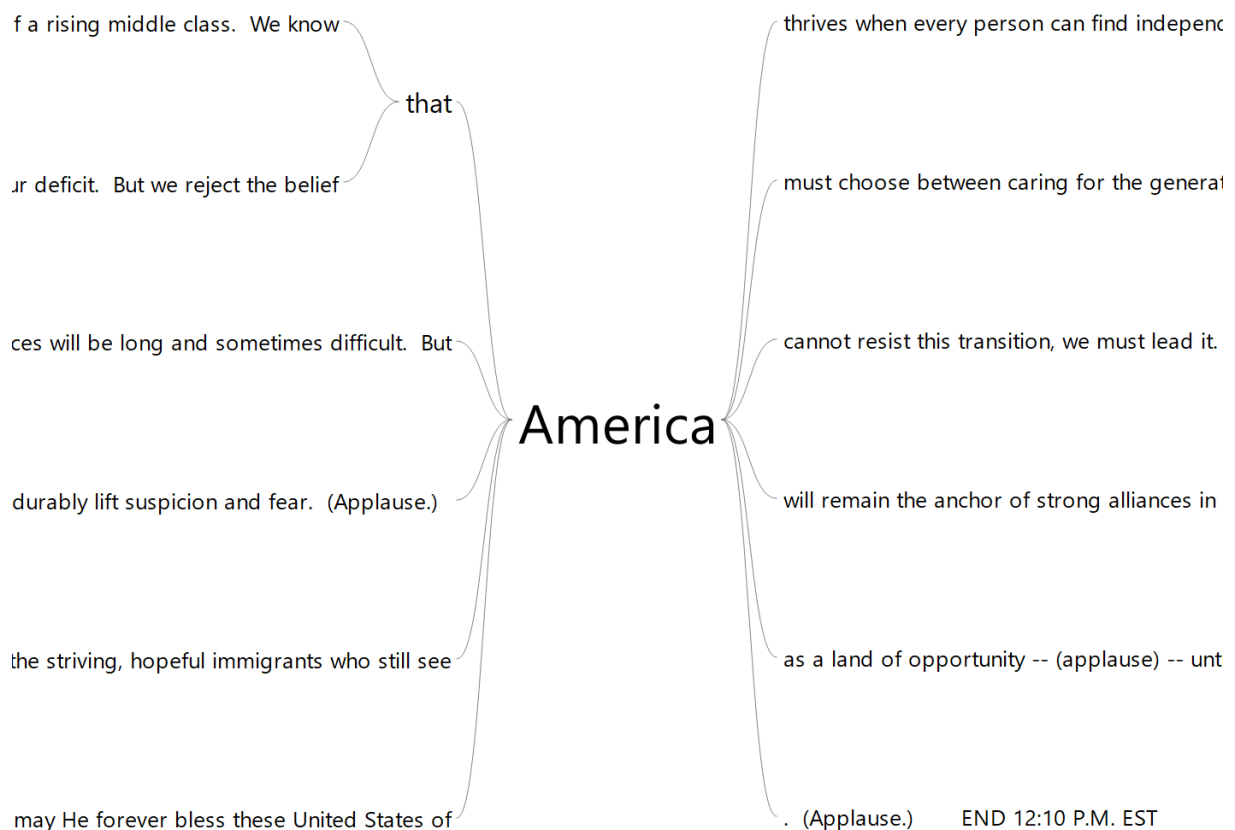
| Word | Word length | Frequency | % | Rank | Documents | Documents % | Obama | Trump |
|----------|-------------|-----------|------|------|-----------|-------------|-------|-------|
| the | 3 | 168 | 4,58 | 1 | 2 | 100,00 | 103 | 65 |
| and | 3 | 158 | 4,31 | 2 | 2 | 100,00 | 88 | 70 |
| our | 3 | 124 | 3,38 | 3 | 2 | 100,00 | 76 | 48 |
| of | 2 | 118 | 3,22 | 4 | 2 | 100,00 | 69 | 49 |
| be | 2 | 115 | 3,14 | 5 | 2 | 100,00 | 63 | 52 |
| to | 2 | 103 | 2,81 | 6 | 2 | 100,00 | 66 | 37 |
| we | 2 | 86 | 2,35 | 7 | 2 | 100,00 | 60 | 26 |
| a | 1 | 79 | 2,15 | 8 | 2 | 100,00 | 57 | 22 |
| that | 4 | 72 | 1,96 | 9 | 2 | 100,00 | 65 | 7 |
| will | 4 | 60 | 1,64 | 10 | 2 | 100,00 | 20 | 40 |
| for | 3 | 41 | 1,12 | 11 | 2 | 100,00 | 26 | 15 |
| in | 2 | 34 | 0,93 | 12 | 2 | 100,00 | 21 | 13 |
| | 1 | 33 | 0,90 | 13 | 2 | 100,00 | 32 | 1 |
| not | 3 | 33 | 0,90 | 13 | 2 | 100,00 | 23 | 10 |
| this | 4 | 32 | 0,87 | 15 | 2 | 100,00 | 22 | 10 |
| we | 3 | 27 | 0,74 | 16 | 2 | 100,00 | 7 | 20 |
| have | 4 | 27 | 0,74 | 16 | 2 | 100,00 | 14 | 13 |
| american | 8 | 25 | 0,68 | 18 | 2 | 100,00 | 10 | 15 |
| applause | 8 | 25 | 0,68 | 18 | 1 | 50,00 | 25 | 0 |
| all | 3 | 24 | 0,65 | 20 | 2 | 100,00 | 12 | 12 |
| but | 3 | 23 | 0,63 | 21 | 2 | 100,00 | 12 | 11 |
| us | 2 | 23 | 0,63 | 21 | 2 | 100,00 | 21 | 2 |
| america | 7 | 21 | 0,57 | 23 | 2 | 100,00 | 5 | 16 |
| with | 4 | 21 | 0,57 | 23 | 2 | 100,00 | 13 | 8 |
| make | 4 | 20 | 0,55 | 25 | 2 | 100,00 | 12 | 8 |

- You can now sort the results by each column.
- You can also combine words by dragging one word and drop it on the other, for example with the words "America" and "American".
- And, if you now want to add certain words to the stop list, just double click the very first column of this word.



Interactive word tree

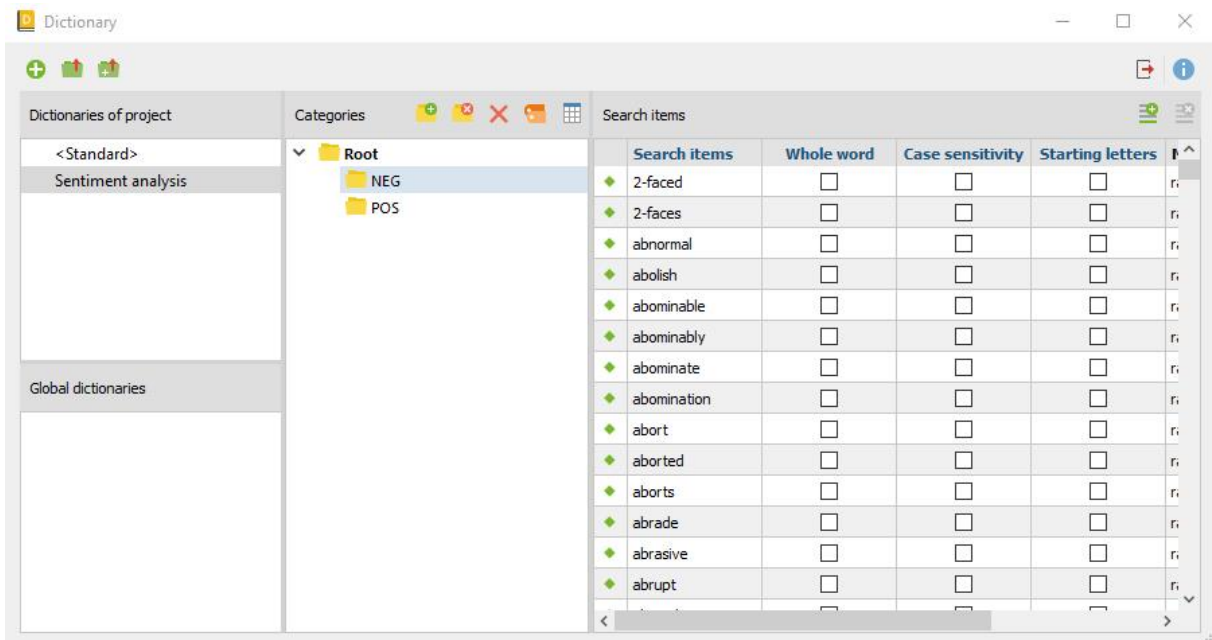
- The 'Keyword-in-context' function can be used to compare specific words within its context. The more frequently a word or word combination occurs, the more distinctly it appears in the tree. The 'Word Tree' function offers two-way interactivity:
 - First, it can be navigated via its individual "branches" in order to visualize words in their contexts.
 - Second, the data is interactively linked to the original texts, so you can view words and phrases in MAXQDA's "Document Browser".
- One of the main functions of the Word Tree is to search for interesting words within a text and to explore them in their respective contexts. To this end, a search field is located at the top of the window. Hit the Enter key to begin the search process. If there is a hit, it will go directly to the new root of the tree.
- Let's look at the word "America" in Obama's speech and in its context.





Dictionary-Based Content Analysis

- Dictionary-based content analysis allows you to count the category frequencies for texts (or parts of text) based on the dictionary.
- Independent from the texts, you will have to create a dictionary or encyclopedia first before you can work with it. This dictionary consists, on the one hand, of an unlimited number of categories. On the other hand, it includes the words that are attached to these categories.
- Let's discover the main differences between Obama's and Trump's speeches using a classical sentiment analysis. You can download the dictionaries at different linguistic institutes.
- Here, two categories are created by importing the linguistic dictionaries via Excel showing a collection of negative and positive words.



- Run a quantitative content analysis using the dictionary.

| Document group | Document | Words | NEG | POS |
|----------------|----------|-------|------|------|
| ◆ | Trump | 1.478 | 47,8 | 52,2 |
| ◆ | Obama | 2.189 | 56,1 | 43,9 |



- You can sort by the different columns and you can see that Obama’s speech is longer than Trump’s and contains more negative words (56.1 %) in comparison to Trump (47.8 %). It is contrary to the positive words, where Trump used more positive words (52.2 %) than Obama (43.9) in his speech.
- You can now code the results, look at the context or words or transform the results into a document variable.
- If you code the words, you can later qualitatively analyze the coded segments or present them in certain ways, for example with the help of the Code-Matrix-Browser or using the export function for Excel.

| Code System | Obama | Trump |
|-------------|-------|-------|
| MAXDictio | | |
| POS | ● | ● |
| NEG | ● | ● |

