

November 2021

# How to determine patent quality based on qualitative claims-analysis?

Dierk-Oliver Kiehne (InTraCoM GmbH)

---

## Summary

Patent valuation has become a central role when analysing companies or doing deep analysis of patent portfolios. Here, indicators like citations or family sizes are often used and mentioned in literature. However, these are measuring patent quality indirectly: A big patent family for example, does reflect the willingness of an owner to lock out as much competitors in as much markets as possible. Thereof a strong ambition of utilisation is derived from and therefore a high quality. However, the most important part of a patent representing the legal core of the invention are the claims: they reflect what the owner claims to be his monopoly with his invention for the lifetime of the patent. Is this easy to bypass, e.g. because the claim range is so small or so highly specific, the blocking effect and thereby the quality is suffering. Some statistical and textual approaches investigate in this analysis what makes a claim valuable in terms of a big claims coverage. The direct quality indication is derived from above mentioned secondary indicators like foreign forward citations and oppositions being a value-driving blocking effect indicator. The analysis shows that there is a specific scheme that obviously leads to a higher citation frequency as well as a higher opposition frequency and thereby a higher quality.

## Introduction

Patent quality and patent values are often used for more sophisticated analysis of patent landscapes, patenting trends, deep company due diligences, for example. When measuring patent quality many easy to determine indicators are often used. However, these are measuring an indirect footprint of patents (e.g. forward citations), not the patent itself. The evaluation of a patent quality is impacted by many factors, different studies have already shown<sup>1</sup> however, the most important quality driver are supposed to be the patent claims<sup>2</sup>, especially the independent ones respectively the wide coverage of those<sup>3</sup>: The claims of a patent represent the core of an invention. The question therefore arises as to why no intensive quality analysis has been carried out here so far. The answer is: Because it is very difficult. A patent attorney may explain a valuable claim with the "2 finger rule" which means that a good claim is so short that it can be covered with only 2 fingers. In fact, it is not that simple since many chemical patents don't even contain words but simply a chemical formula, others contain a short but highly specified claim making it easy to bypass by simply adding, removing or modifying a feature.

In literature, some claims related indicators are mentioned like the total amount of dependent or independent claims<sup>2,3</sup>, the total amount of words<sup>4</sup> or the number of nouns in the claims<sup>5</sup>.

---

<sup>1</sup> Squicciarini, Mariagrazia, Hélène Dernis, and Chiara Criscuolo. "Measuring patent quality: Indicators of technological and economic value." (2013).

<sup>2</sup> Narin, Patent bibliometrics, *Scientometrics* 30 (1) (1994) 147–155

<sup>3</sup> Trappey, Amy JC, et al. "A patent quality analysis for innovative technology and product development." *Advanced Engineering Informatics* 26.1 (2012): 26-34.

<sup>4</sup> Malackowski, James E., and Jonathan A. Barney. "What Is Patent Quality--A Merchant Banc's Perspective." *Nouvelles-Journal of the Licensing Executives Society* 43.2 (2008): 123

<sup>5</sup> Dang, Jianwei, and Kazuyuki Motohashi. "Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality." *China Economic Review* 35 (2015): 137-155.

The following analysis is investigating 2 different features and tries to find out their quality impact tendencies.

## Method

2 different approaches based on the independent claims are suggested here. In order to determine whether these are indicating quality and to what extent, correlation analysis were performed: the indicators were primary correlated with the total number of oppositions. Oppositions are done by any third party after a patent was granted. Typically, an opposition is filed either when a patent is close to the state of the art and the opponent believes that a prior art can be proved or when a competitor is feeling bothered. Often, both is the case. However, the competitor must be bothered by a specific patent, otherwise the cost and effort for an opposition proceeding were not justified. Hence, the analysis here uses oppositions as a strong indicator of blocking effect: The greater the blocking effect, the more likely it is that a competitor will be compelled to restrict the blocking effect of a patent in opposition proceedings. In any case oppositions proceedings are always claims-related in contrast to e.g. forward citations: A patent may be cited by a foreign source also because there is relevant text mentioned in the description text, not necessarily in the (independent) claims. For the analysis this means that oppositions are a better correlation indicator than e.g. foreign forward citations.

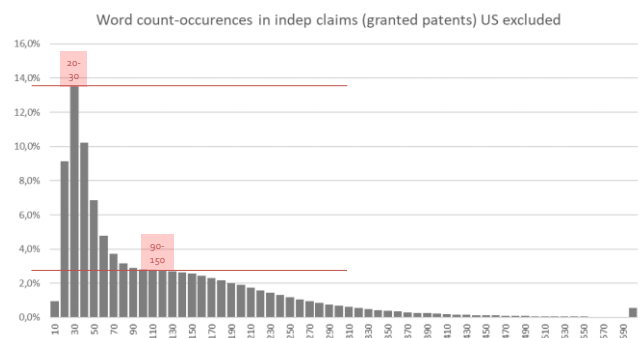
The first analysis that is introduced is based on the method presented by Malackowski and Barney. However here, only the independent claims were examined and only the part of the text in which the invention-related part is formulated, e.g. the part following "characterized by", was focused, if available. The number of features described in an invention has further been subject to analysis: To infringe a patent, all features according to the claimed invention must be infringed. Accordingly, the fewer these are, the more difficult it is to circumvent. Accordingly, the analysis examined how many features linked by "and" are present in the independent claims.

Further, only the claims of granted patents were used for the analysis. Here, an automated text analysis must also consider certain stop words or meta-texts that are often found like e.g. "claims 1-4 cancelled".

## Results

In the first part of the analysis only the total amount of words were counted according to the method description before.

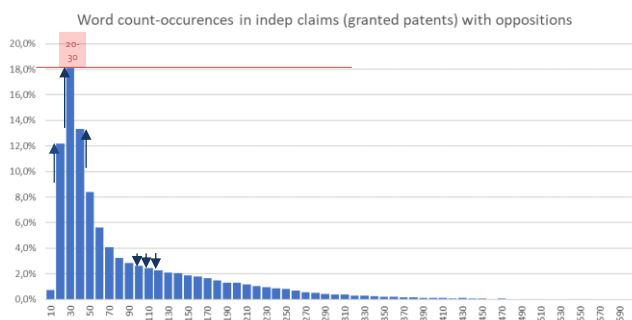
The distribution can be seen in the following graph:



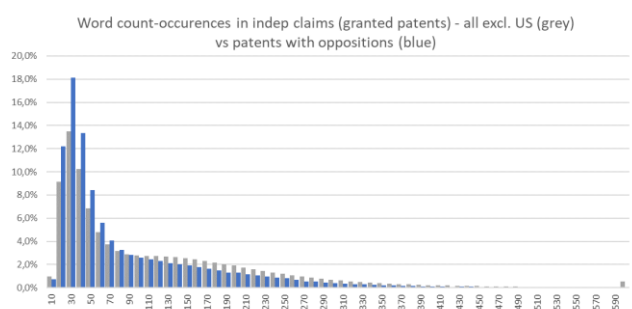
Graph 1: histogram of word-count-occurrences in independent claims of granted patents

The graph shows an interesting distribution with a quite clear top in the region of 20-30 words – 13.5% of all investigated patents have that number of words. A strong "belly" is also seen in the region of 90-130 words where the distribution is almost constant: 19% of all patents are in this region. US patents do not have the same opposition procedure as e.g. European patents, therefore US patents were excluded from the analysis. In total 3,728,061 granted patents were analysed.

For the second part of the analysis, only those patents were filtered that had at least one opposition during the oppositions phase. Anyway, the picture of word count distribution is a significant different one and it becomes even more significant when the graphs are directly put in one chart (graph 3).



Graph 2: histogram of word-count-occurrences in independent claims of granted patents that underwent at least one opposition proceeding.



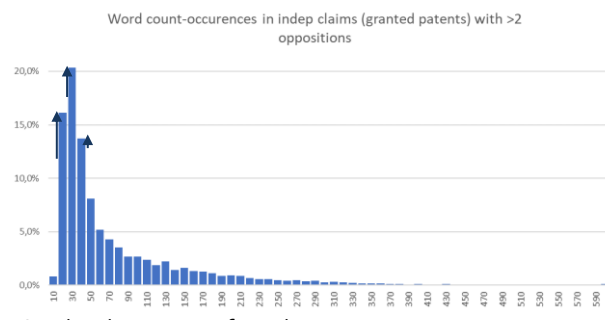
Graph 3: the histogram in direct comparison to Graph 1.

Especially the chart in Graph 3 perfectly shows the significant difference in word-count-occurrences: Patents that underwent an opposition proceeding typically have way shorter independent claims.

Interestingly, the “belly” at 90-130 words as seen in graph 1 has completely disappeared for those patents having oppositions.

For the analysis 118,244 granted patents with at least 1 opposition have been analysed.

This effect becomes even more significant, when patents having multiple oppositions of different opponents were analysed: the trend as seen in graph 2 has significantly increased again towards shorter claims: the top at 20-30 words is now represented by more than 20% of the patents with at least 3 oppositions. The groups 10-40 words are here represented by more than 50% of the patents and have significantly grown compared to patents with only one opposition (43,6%). A very clear picture.

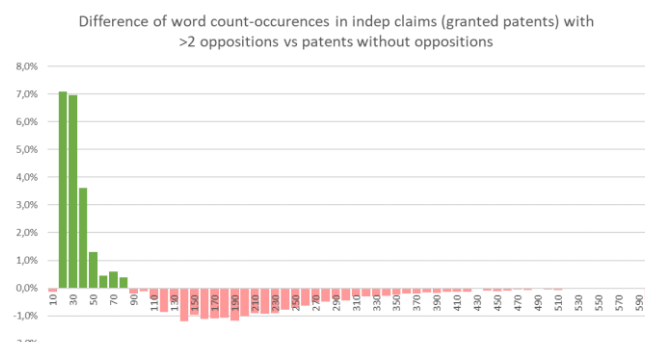


Graph 4: histogram of word-count occurrences in independent claims of granted patents with at least 3 oppositions.

For the analysis 9,690 granted patents with at least 3 oppositions have been analysed.

It can be observed that those patents obviously bothering multiple competitors have typically the shortest claims, the centre of gravity is shifted to the left in the distribution (graph 4) and became even more dominant compared to graph 2. A very clear indication towards shorter claims: The shorter the (independent) claims were, the higher the probability of an opposition was.

In order to illustrate this trend in more detail, the difference between the frequencies was calculated. Here, claim lengths of granted patents with at least 3 oppositions were compared to those that had no oppositions.

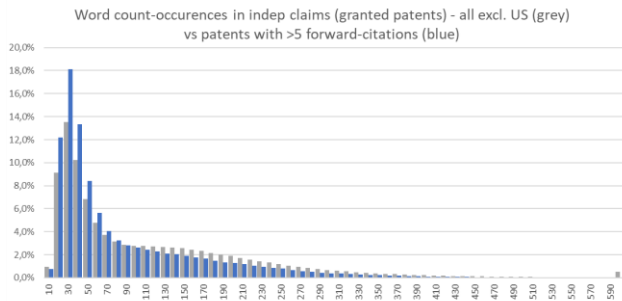


Graph 5: word-count occurrences in independent claims of patents with a high amount of oppositions versus those having no oppositions

The chart in graph 5 shows most significantly the trends of this analysis- the shorter the claims were the more likely an opposition was. However, there seems to be a minimum of at least 10 words that are required to be a “good” claim.

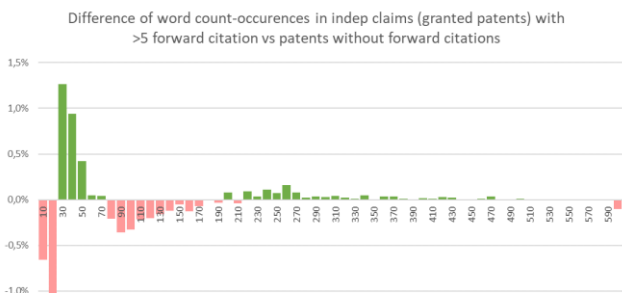
Just for comparison also the forward citations were correlated with the number of words. As described already, forward citation do not necessarily only refer to the independent claims

we are focusing here. Patents may be cited due to a dependent claim or referring to the description text or even a figure that is used. Thereby it is less independent claim-related than oppositions are.



Graph 6: word-count occurrences in independent claims comparing often cited patents with more than 5 forward citations all patents (grey)

However, the analysis is showing a similar picture than the analysis with the oppositions. For a better comparability, the US Patents were also excluded here (graph 6). 77,365 granted patents had more than 5 forward citations here. According to graph 5 we calculated the changes in word-number-occurrences of high frequently cited versus not cited patents.



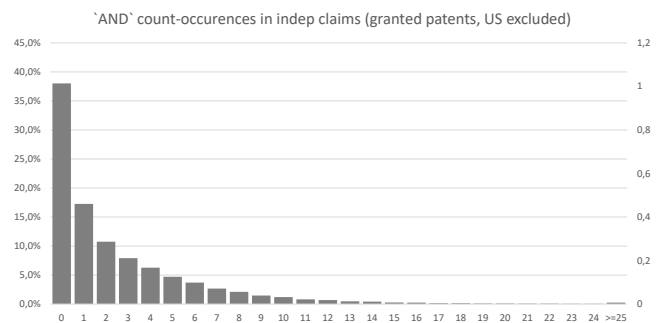
Graph 7: word-count occurrences in independent claims of patents with a high amount of oppositions versus those having no oppositions

Generally, the short independent claims tend to be more often cited (as expected to be the more valuable patents) as seen in the highest green bars in graph 6 and confirms the analysis that was made using oppositions even though with less significance. However, the biggest surprise are the very short independent claims represented in the biggest red bar in the graph (graph 7). Here, the oppositions-analysis showed the highest positive difference, the cited-by-analysis shows the biggest negative difference. That means that the shortest independent claims with 10-20 words are less likely to be cited but more likely to bother

competitors resulting in oppositions. For the word-count occurrences 30-50 both quality indicators – oppositions and forward citations – show equally a positive correlation.

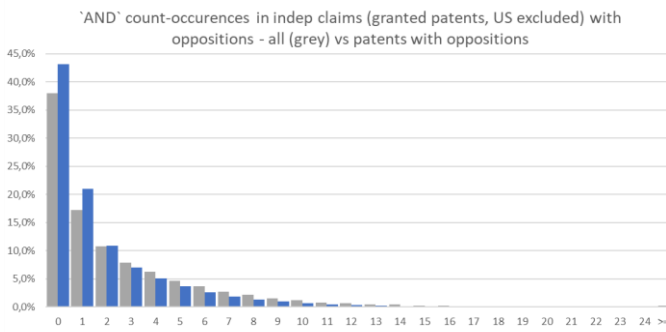
A second indicator based on claims that has been investigated here is the amount of features in an independent claim. In order to infringe a patent, at least all the mentioned features must be infringed. So, it can be assumed that the more features are mentioned in the claims, the less likely they might be infringed and the easier it is to bypass. Hence, it can be assumed vice versa, the less features are mentioned, the broader the claims coverage and the more difficult it is to bypass. The total amount of different features was measured here by counting the words “and” as an indicator for the connection of different features.

The same methodology as before was used here: histograms were created by measuring how often the word “and” was used in the independent claims. In the following graph 8, the distribution of occurrences shows all granted patents (US excluded as already explained before, in order to have a reference for the later correlation with oppositions):



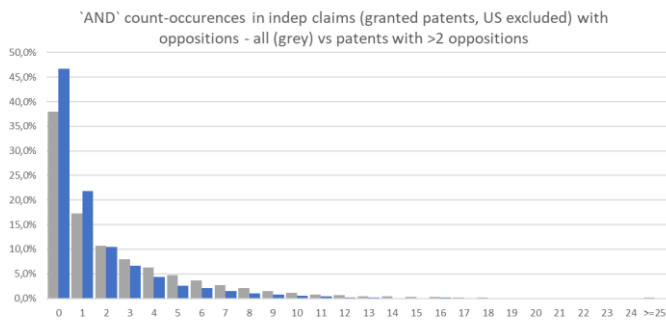
Graph 8: feature occurrences in independent claims by counting the word “and”

The distribution shows a very clear pattern of a Weibull ( $k=0.5$ ) distribution: the strongest area is the one where only one feature is given where no “and” were counted. According to the analysis done before counting words in general, the same histogram was performed with patents where at least one opposition was filed. Graph 9 shows impressively a pattern which is very similar the former analysis: Those patents having at least one opposition tend to have less features: the group of “0” is bigger.



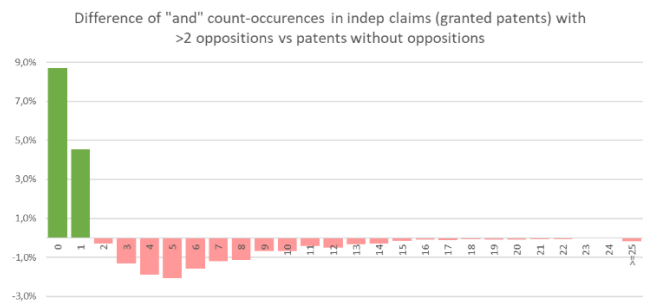
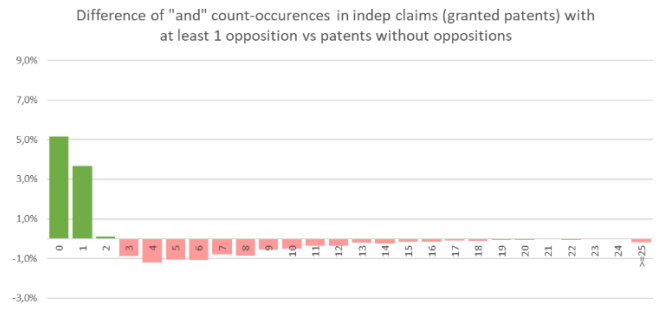
Graph 9: feature occurrences in independent claims by counting the word "and" with oppositions (blue) compared to all granted patents (grey). Obvious: the group with 0 "and" was rising from 38% (all granted patents) to 43% (granted patents with oppositions)

In order to confirm this trend, the patents with more than 2 oppositions were also compared. As mentioned before, multiple (unsuccessful) oppositions indicate a big blocking effect and potential usability in terms of license revenues.



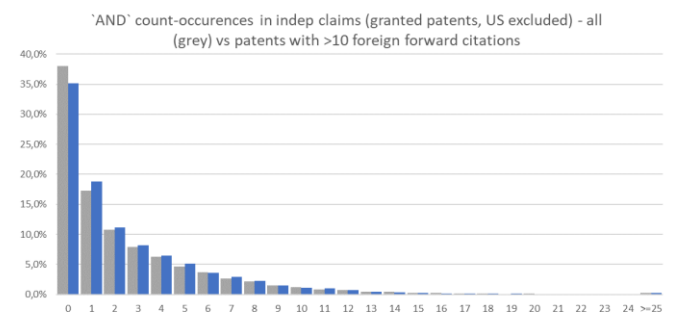
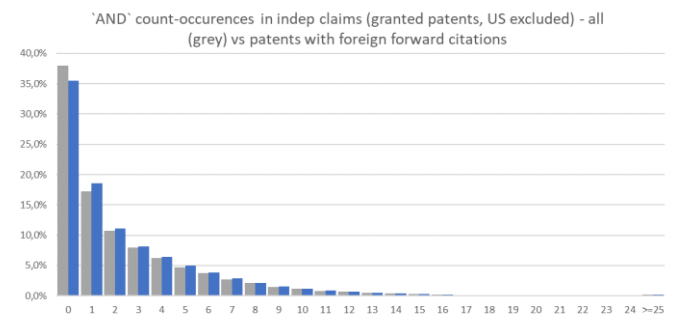
Graph 10: same analysis as in graph 9 but with more than 2 oppositions (blue) compared to all granted patents (grey). The group with 0 "and" increased again to 47% (granted patents with more than 2 oppositions)

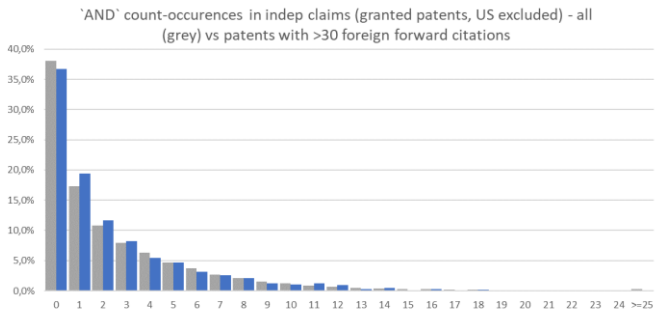
An even better and more obvious picture can be achieved by directly comparing occurrences of patents with oppositions to patents without. The following graphs indicate the same trend as already seen in Graph 9 and 10. The pattern that can be seen here is surprisingly very clear: Again, the differences of occurrence were calculated and put into one graph. In order to make the graphs for all patent with oppositions and those with a high opposition rate more easy to compare and to see the very obvious trend, the used scaling on the y axis is identical.



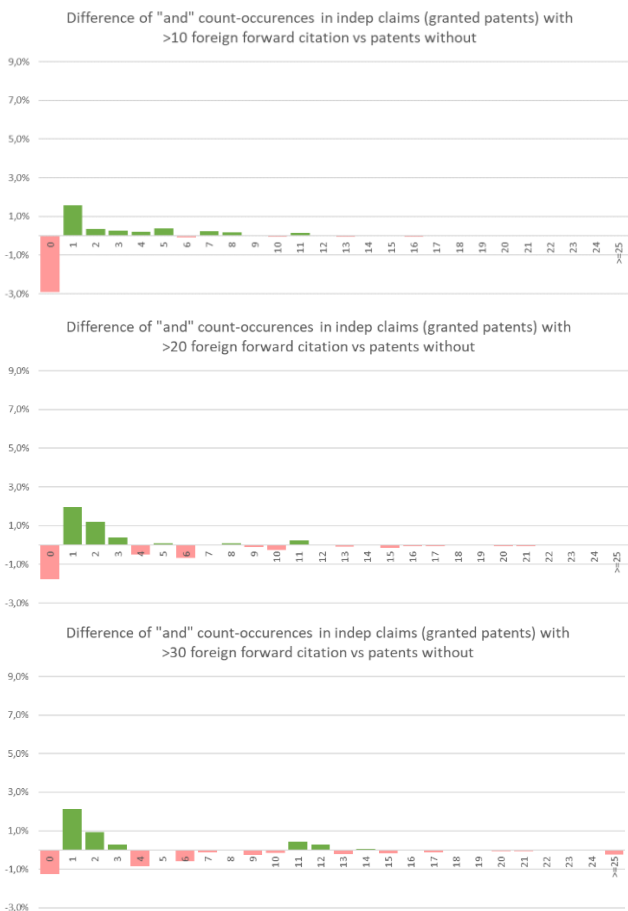
Graph 11: occurrence differences of the word "and" in independent claim of granted patents with at least one (above) respectively more than 2 oppositions (below) compared to granted patents with no oppositions

The analysis has been repeated using the foreign forward citations as correlation indicator. The results were similar, however different to a certain extent.





Graph 12: occurrence of the word "and" in independent claim of granted patents with at least one (above) respectively more than 10 (middle) and more than 30 (below) foreign forward citations compared to all granted patents



Graph 13: occurrence differences of the word "and" in independent claims of granted patents with more than 10 (above), 20 (middle) and 30 (below) foreign forward citations compared to granted patents with no forward citations

For the forward citations the picture doesn't have that statistically significant appearance. However, the picture seems to be more manifest, the more patent with forward citations are filtered. Most interestingly however, the group counting no "and" occurrences has here the strongest negative difference, means those

patents having only one feature are obviously comparably less often cited. This is a completely different picture than for the oppositions. For the group of 2 occurrences the patterns are congruent, for the group of 2-3 occurrences, the foreign forward citations show a positive difference (patents with 2-3 "and" occurrences are relatively more often cited) however they are less often involved in opposition proceedings.

**Conclusion**

The number of words in the independent claims are a quality indicator: the shorter the better, that could be proved with two different correlations, the forward-citations and the number of oppositions. The ultra-short independent claims had in comparison the highest probability for an opposition but were less likely cited. However, the oppositions are a better claim-related indicator. Both correlations show the highest quality correlation in a word-frequency range of 20-50 words. Here seems to be highest probability for a high valuable patent. The analysis did only consider granted patents: Quite often patents are applied with short claims having a big coverage but suffering big changes during examination stage: either complete claims are deleted or require adjustments that limit the breadth by adding additional features - and words thereby. This means that the word-count-based indicator of the independent claims has a better quality-prognosis when a patent is already granted.

For the "and" occurrences-count the same analysis comes to somehow contradicting results. Patents don't have an "and" occurrence are most often opposed but less often cited. Patents having 1 "and" occurrence have also comparably often oppositions and are most often cited. Patents with an "and" occurrence of 2-3 times are still comparably higher citations but are less often involved in oppositions. Both correlation-models agree only in the one "and" occurrence, however the oppositions-model shows a higher significance: the differences shown (graph 11) are higher (compared to graph 13).