



Universiteit Utrecht

Words count

A linguistic analysis of the possible impact of the European Court of Auditors' special reports

Thomas Schillemans and Marija Aleksovka



Table of Contents

- Introduction 3
- About LIWC (= method)..... 4
- The analysis 6
- The results 8
- Patterns in the special reports..... 9
- Patterns in formal responses..... 11
- Two archetypal reports 13
- Conclusions and reflections..... 15

Introduction

The business of governing is in important ways captured in words. Governments and other public organizations are working on a steady and unyielding stream of assessment reports, including many cases of policy advice and evaluation. Audit offices, advisory bodies, evaluation committees, and research bodies are active at all levels of government. If an incident happens or if there is a political impasse, a temporary committee is usually set up to investigate what went wrong and to suggest ways forward. All those different bodies eventually produce reports. Those reports are full of *words*. And those words matter.

Politico-administrative actors intuitively know that words count: *c'est le ton qui fait la musique*. Many successful politico-administrative actors are wordsmiths who are always able to find the right words and the right intonation at the right moment. The chosen wording and the nature of the analysis has an impact on how a message comes across and therefore also on the question of what the recipient of the message can and may do with it. For that reason, especially in recent years, quite some attention has been paid to the language of policy-making (Van Twist 2018) and to the framing of issues by policy actors (De Bruijn et al 2012). At the same time, those analyses are often still highly intuitive and dependent on specific cases. We know *that* the words matter and we have a great deal of insight into how analyses can be framed and understood; at the same time, it is not very clear *how* they matter and *what words* matter.

In this report we report on a more systematic exploration of the linguistic characteristics of reports and their possible effects on the impact of those reports. We use the automated text analysis of LIWC (Linguistic Inquiry & Word Count). This program has been developed for the analysis of texts of individuals but can also be used for the analysis of reports. The software analyses texts on linguistic characteristics, for example on the question of how positively or negatively a report is voiced or to what extent a more or less analytical reasoning emerges from it. These types of attributes can be important for the impact of reports in governmental settings.

The advantage of automated text analysis via LIWC is that texts can be analyzed for a large number of linguistically relevant features that can be important in a wide variety of contexts. By exposing textual patterns in reports (and other policy texts), by comparing reports with each other, and also by relating them to other characteristics, we aim to further our understanding of how words in audit reports matter. As a result, more specific knowledge of the impact of the wording of (audit) report is gained.

The analysis in this report is based on 60 special reports by the European Court of Auditors, which formed a sample of a larger research project focusing 233 reports in total. Our analysis will focus mostly on the results for the special reports yet at some points we will also refer to results from the wider sample.

All those reports are understood as performance, evaluation, or accountability reports. They have in common that they assess a certain policy against certain standards, but then also make suggestions and recommendations for improvement. They also have in common that the assessed organization gives a formal response which is publicly available and in which it is indicated how the recipient will deal with the (sometimes

mandatory) recommendations. At times the formal response is hesitant and the wording suggests that the recipient body has difficulty endorsing the analysis and appreciating the recommendations. This may well be the harbinger of a restrained, if not reluctant, handling of recommendations. Sometimes the response is warm and generous, and the assessed organization expresses a lot of support and appreciation for the analysis and recommendations. In that case it is conceivable that a more enthusiastic and motivated approach to the recommendations lies ahead. This raises the question: do reports that are received more positively actually have special linguistic characteristics?

In this report we explore the applicability of the LIWC software for what we call performance accountability reports in three steps. First we analyse the special linguistic characteristics of these reports. We then analyse the special features of the formal responses to those reports. And finally, we answer the question whether successful reports – that is, reports that are received more positively – have special linguistic characteristics. The analysis will show that there are indeed two archetypal types of reports - we will call them *analytical* and *directing* reports - one of which is indeed received much more positively than the other. This raises one tip of the veil about the question of when a performance accountability report effectively affects the policy of the assessed organization. We will reflect briefly on the implications of this at the end of this report, but before we get to our analysis, an explanation is first given about the LIWC-research software and how we applied it.

About LIWC (= method)

The Linguistic Inquiry and Word Count (LIWC) program is an automated quantitative text analysis program. The program scans entered texts and then indicates what percentage of the total number of words in the text refers to a large number of linguistic characteristics, such as emotions – such as fear, horror, and amusing – and cognitive processes – such as analysing, conceiving, and realizing – and more specific information about the text, for example the percentage of pronouns or the percentage of prepositions.

There is a long academic tradition of systematic text analysis. For example, it is used in psychology because the words we use may reveal something about ourselves. This is already implicit in the well-known expression of a *Freudian slip*: the words from the slip "betray" what the speaker actually thinks, wants, or feels. Similarly, texts can also give us insight into people's desires and needs (Parker 2014) and a systematic analysis of texts can help to expose liars (Newman et al 2003). The application of this type of analysis is very diverse. The LIWC was originally developed to analyse expression of emotions but can also be used to say something about the use of words in a text in a general sense (Zijlstra et al., 2004: 272). An example makes the application of the LIWC clear. Cohn, Mehl and Pennebaker (2004) investigated discussion groups on the internet after the 9/11 attacks on the World Trade Centre in New York. With the help of the LIWC program, they signalled an increase in negative emotions immediately after the attacks. In addition, they signalled more social involvement and saw that there

was more talk of "we" instead of "me" than in the weeks before the attacks. After two weeks the word use seemed to normalize again. This finding is relevant for subsequent trauma and crisis research.

There is also a flourishing tradition of text analysis in international relations. It is relevant there, because there is often very little knowledge available about foreign powers – think of the Soviet Union earlier or presently IS. It is then unclear what the other intends and how the other interprets events. The official messages are often also not reliable because they serve a propagandistic purpose. But text analysts can look for the hidden messages in the texts, for the proverbial Freudian slips and implicit language characteristics, which uncover what the leaders of foreign powers really think. A modern translation of this is, for example, the research by Van Esch (2014) and others into the views and ideas that political leaders hint at in their texts about the euro crisis. There is also research showing that the language of political leaders changes prior to the escalation of political violence, both nationally and internationally (Suedfeld & Tetlock, 2014: 598).

Automated text analysis has, like all things in life, advantages and disadvantages. In the past, texts were always analysed manually by teams of researchers who first had to be trained and partly had to do double work in order to arrive at a reliable coding of texts (Tetlock et al 2014). Text analysis was the work of people and that is time-consuming and error-ridden. Exactly on those points automated text analysis offers a solution: it still takes time because texts must be prepared for analysis, but of course much less time than manually going through page after page. The reliability is also much better. Computers can crash, get hacked or get lost, but in principle they do not make mistakes. Entering the same text invariably leads to the same linguistic analysis. When the text analysis is done by people, there is always some noise between two coders. Automated text analysis therefore has the advantage that it is much more reliable and much less time-consuming.

On the other hand, people understand texts better than computers and can therefore analyse more precisely and subtly than computers. People recognize irony and sarcasm, for example, and understand the difference between "all patients live, it goes really well" and "all patients died, it goes really well". For the computer, the meaning of both identical clauses is fairly similar as the second part expresses a positive emotion, the human assessor of course recognises the enormous differences. In a more scientific sense: the external validity of human coders is better (see also Conway et al., 2014), they are much better able to understand what a text really means.

This issue of external validity means that the results of the analysis in this report should not be interpreted *at face value* and that it is important to check whether statements are also recognizable to those involved. A limitation of the LIWC program is that it does not provide insight into the context in which words are used, which can lead to some distortion in the results (Zijlstra et al., 2004: 280). A statement like 'the organization does not have to be afraid of a bad result' will be (rightly!) interpreted by the reader as positive (posemo) but is understood by LIWC as negative (negemo) because of the word choices 'bad' and 'afraid'. It therefore remains important to carefully weigh the results of the text analysis and, where necessary, to supplement it with other research and direct reading of texts. At the same time, the gravity of this problem also varies per analytical category. For example,

the software counts the number of words and it can safely be assumed that this is done reliably (much more reliably than by people). On the other hand, it also analyses more complex constructions such as the analytical content of reasoning; the risks of (slight) misinterpretation is greater there.

The analysis

How does the LIWC program work? The user opens a text in the program, the program scans the text and compares each word in the text with words in a special dictionary. The words in the dictionary are divided into different word categories. The total number of words is then calculated and what percentage of this total belongs to the various categories. The program thus calculates the percentage of verbs, nouns, etc., but also what percentage of words has a positive or negative emotion. All words in the text are placed in one or more categories. For example, the word "crying" is placed in the "verbs" category, but also in the "emotional processes" main category and in a subcategory thereof, "negative emotions".

Once a text has been entered into the program, it is automatically analysed for all available categories which results in numerical scores that are of little significance in themselves. They acquire meaning through comparison with other similar texts. We have limited our analysis to a small proportion of the word categories that are available within the LIWC, because they are not all relevant to the performance accountability reports.

We present them in four groups: form characteristics, tone, cognitive processes, and power relations. We always explain what the characteristic is, why it can be relevant for the impact of an audit report, and, if it does not speak for itself, we provide an example. The examples are taken from the total sample of 233 reports that we analysed.

1. Form characteristics

First of all, LIWC distinguishes a number of formal characteristics where evaluating organizations can make specific style considerations.

- **WC.** Word count. Displays the length of a report.
- **WPS.** Words per sentence. Indicative level of difficulty of the text.
- **SixLtr.** Number of six letter words; so relatively longer words. This too is indicative of the level of difficulty of the text.
- **Number.** Use of numbers.

In general it could be expected that a longer report, with more difficult words and numbers, and more complex sentences, could be experienced as "heavier" by the reader. It is therefore potentially more authoritative, which

could lead to a stronger impact. Conversely, it can also be assumed that a shorter report, with more simple sentence constructions and use of words, could be more accessible and, for that very reason, could have a stronger impact on the reader.

2. *Tone*

The LIWC analyses texts for, what it calls, positive and negative emotions and therefore essentially on the amount of positive and negative qualifications in texts.

- **Posemo** indicates "positive emotions". An example would be the sentence "It's good to see". For example, reports speak of a "brave precursor role" or an organization "may count itself as lucky".
- **Negemo** points to negative emotions, for example in the sentence "It's a shame to see". In this category we find for example sentences such as 'this is regretted by parties', 'it seems that there is a certain fear', 'it is very unfortunate that little or no progress has been made', and 'the fact that the [organization] currently has little investment room is seen as very unfortunate'. Reports that score high on "negative emotions" thus have a negative tone.
- **Tone** links the positive and negative emotions (see 'posemo' and 'negemo') with each other and weighs them against each other. For example, it is conceivable for a report to make both a striking number of positive and striking numbers of negative statements. In that case, this category indicates which of these prevails.

3. *Cognitive mechanisms*

Cognitive mechanisms include a series of analytical categories that deal with the reasoning in a text. The text analysis can then serve as a tool to gain insight into the psychological processes – emotions and thoughts – that lie behind word choices (Zijlstra et al., 2005: 56). This is more difficult to interpret but also potentially more interesting. We discuss them one by one.

- **Cause** indicates causal reasoning along the lines of "A is a result of B". An example from one of the reports is: "it takes a long time for the maintenance service to come by and that is because they are not allowed to bundle complaints".
- **Insight** points to insightful reasoning such as "It appears that...".
- **Discrep** indicates the mentioning of discrepancies such as "on the one hand there is appreciation for one's own view, on the other hand an earlier involvement of stakeholders is desired".
- **Tentat** is about tentative formulations such as "although maybe not entirely genuine, this has probably also helped to determine the opinion of many stakeholders".

- **Certain** deals with strong claims such as “the position of the [organisation] is undeniably important in the region”.
- **Inhib** is about limiting, with words like ‘block’, ‘limit’, and ‘stop’.
- **Inclusive** is about words that are inclusive. It is about words such as ‘and’, ‘with’, and ‘including’.
- **Exclusive** is about words that are exclusive. It is about words such as ‘but’, ‘without’, and ‘excluding’.
- **Cogmech** is then the *sum* of the above categories.

In a general sense, cognitive mechanisms say a lot about how a text is written. This applies to the overarching concept of ‘cogmech’, but perhaps even more so to underlying concepts such as, for example, whether a text is written tentatively (‘it could be that ...’) or is very firmly written (‘it is that ...’). This will have effects on the reader of the text.

4. **Power relations**

Finally, a fourth category is about the balance of power between the sender of a message (here the ECA) and the entity being addressed. There are two important subcategories.

- **Achieve** indicates performance orientation. Words such as ‘achieving’, ‘succeeding’, ‘competing’, or ‘initiating’ indicate performance orientation, for example.
- **Clout**, lastly, refers to the self-assured taking of a higher social position. A sentence such as “Come here for a moment”, for example, has the same content but comes across very differently – has much more ‘clout’ – than the sentence “Would you please come here?”. This variable is complicated but also of acute importance for assessors. After all, they speak against the organization with the aim of helping in accountability and learning. In that case, is it better to speak from a more authoritative position – that is, with more clout – or is it better to communicate on a more socially equivalent basis? There are from a theoretical point of view both advantages and disadvantages for both forms of social positioning for evaluators. ‘Clout’ is a compound variable that is calculated according to the following formula: Clout = we + you + social - i - swear - negate - excl (Kacewicz et al 2014).

The results

In the results section we discuss the sub questions that were asked for this report. We first look at the characteristics of the reports and then we analyse the formal responses to those reports. Further analysis shows that there are essentially two archetypal reports, we dub them analytical and directive. In conclusion, it becomes clear that one of these two types is generally met with a much more favourable responses than the other.

Patterns in the special reports

Characteristics reports

What are the linguistic characteristics of the special reports, compared to other performance evaluation reports?

Below we report the means for the 60 special reports, compared to the means of the 233 reports in total.

Table 1: means special reports ECA and responses compared to means full sample of reports

		ECA	Mean all
Statistic		Mean	Mean
	Attributes reports		
WC	nr of words	11,118.120	11,865.000
WPS	Long sentences	28.624	22.893
Sixltr	Long words	30.677	32.643
Cause	Causal reasoning	2.175	1.089
Insight	Giving insights	2.089	1.562
Discrep	Pinpointing discrepancy	0.806	0.684
tentat	Tentative phrasing	1.285	0.821
certain	Certain phrasing	0.878	0.538
achieve	Focus on achieving	2.952	1.100
posemo	Positive emotions	1.884	1.197
negemo	Negative emotions	0.755	0.447
number	Use of numbers	6.792	5.872
Tone	Tone: net positive – negative emotions	1.129	0.751
Clout	Taking position of authority	0.103	0.333
Analytic		50.622	55.364

The comparison of LIWC-outcomes is not ideal, as the other reports were written in Dutch. LIWC has been translated, yet, as languages differ in width and complexity of reasoning, so will the numbers generated by LIWC do as well. Nevertheless, some features stand out.

Compared to the other performance evaluation reports, the special reports exhibit some special features.

First of all, the special reports are somewhat shorter than other comparable types of reports, with longer sentences yet more short words.

Second, the tone is overall somewhat more 'emotional' in the use of both positive and negative emotions. This could be the product of language (English vs Dutch) or a different 'emotional strategy' in writing.

Third, there is a very striking focus on words relating to 'achievement'. This would probably, also looking at the reports substantively, suggest a relatively strong focus on outcomes of specific procedures or policies.

Finally, the use of numbers is above average, which seems fitting for an *audit* institution.

Relations between characteristics reports

A second question to look at is: how are those characteristics of reports related. We thus look at relationships *between* characteristics within reports. The figure below depicts correlations where the larger the coloured sphere is, the stronger the relationship. Blue relationships are positive (more of the one goes together with more of the other) and red relationships are negative (more of the one goes together with less of the other). The colour palette shows that there are all kinds of connections between the linguistic characteristics of texts that we analyzed here.

The figure is colourful and complicated and probably sends as a main message to the reader that there is 'a lot' (if not that it is incomprehensible). This is partly because direct correlations are shown with characteristics of the formal responses to reports (all values that start with 're'). We will not take this into consideration here for the time being. But even if we ignore that, there is still a picture of all kinds of positive and negative correlations between textual attributes. Does an overall picture emerge from this? We discuss some notable relationships.

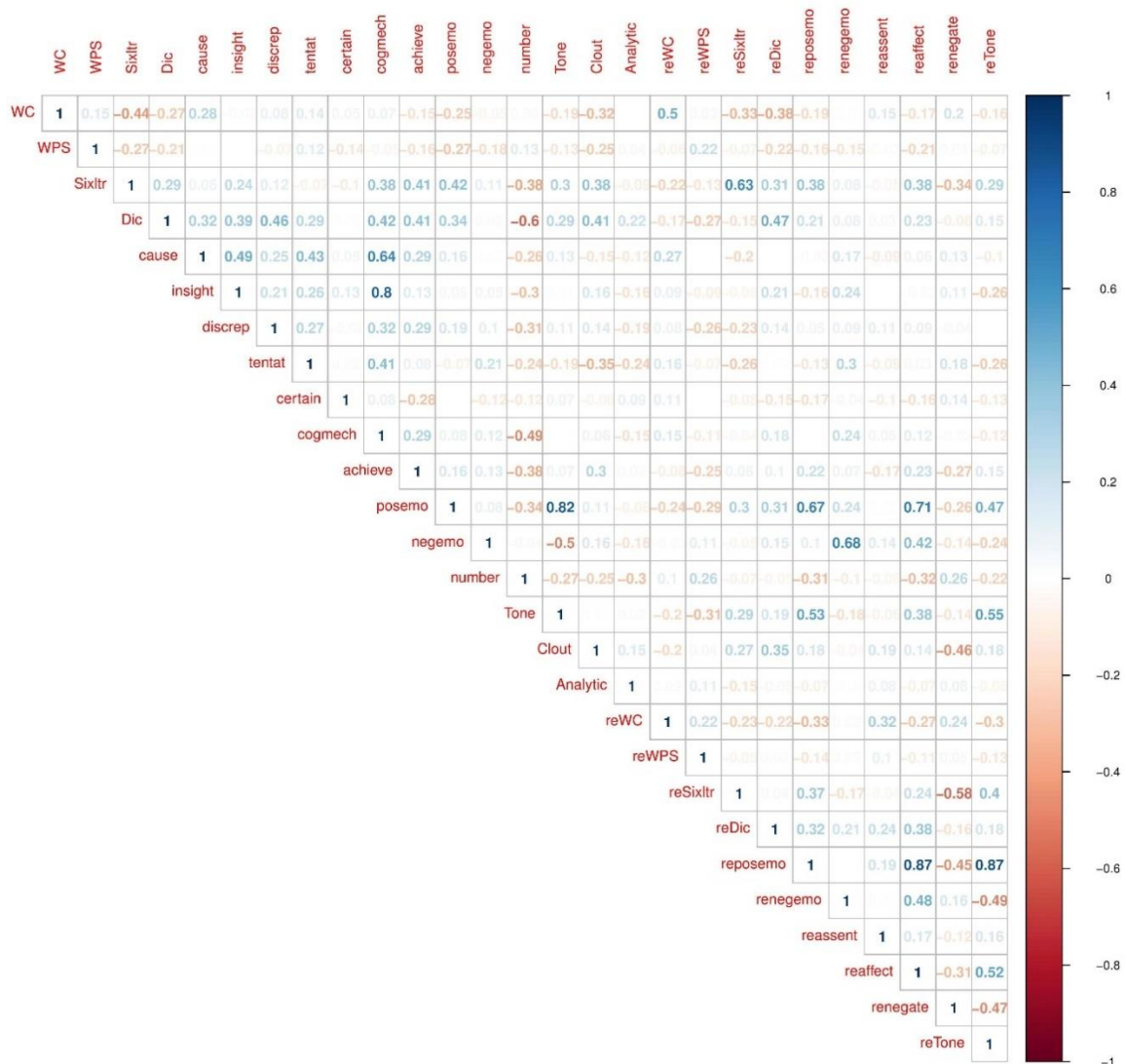
With regard to the **form characteristics**, it is first of all striking to see a cross-link between long words and long sentences; it seems to be one or the other. Longer sentences somewhat surprisingly go together with shorter words. Also, longer sentences go together with fewer positive emotions, more cognitive mechanisms, and more distinguishing of discrepancies. Less positive sentences and more complex messages apparently need more words and apparently also more simple words.

Secondly, it is striking that the **tone** of reports is strongly self-contained and not explicitly linked to other language characteristics.

Thirdly, with regard to **cognitive mechanisms**, it is particularly striking that the use of more cognitive mechanisms in reasoning is negatively correlated with the use of figures. This suggests that numbers are presented more as facts and are less part of more specific analytical reasoning.

Finally, it is striking that the two characteristics of a **dominant position**, namely 'clout' and 'achieve', are strongly interconnected. It goes together with more use of numbers, more long words, and, on the contrary, shorter sentences. This also seems to be an understandable connection.

Figure 1: correlations characteristics special reports ECA (incl. responses)



All in all, there are patterns in the reports that suggest that there are different types of reports. A cluster analysis will soon clarify that two archetypal reports can indeed be distinguished: analytical and directive. Before we elaborate on this, we first consider the features of the formal responses to the reports.

Patterns in formal responses

For many advisory, audit, evaluation, or accountability reports, a mandatory, formal response from the addressee is often required. To a certain extent, such a response is a formality and the recipient is often careful to draft a respectful response which endorses the analysis as such yet does not suggest the organization will change its

policies in responses. Nevertheless, the response is not without relevance. A more positively phrased response normally suggests a more serious engagement with the report and probably signals an intention to use the report in some way. And, conversely, a more negatively phrased response normally suggests a more reserved attitude towards the analysis which may signal limited and minimal use of the contents.

Below we first discuss striking features of responses to the special reports, also in comparison to the results for the four organizations.

Table 2: means responses to special reports ECA compared to means full sample of reports

	Form characteristics		
reWC	nr of words	5,689.241	2,067.214
reWPS	Long sentences	26.584	21.981
reSixltr	Long words	31.561	31.325
reDic	Causal reasoning	69.113	63.094
	Tone		
reposemo	Positive emotions	2.398	1.741
renegemo	Negative emotions	0.657	0.450
reTone	Tone: net positive – negative emotions	1.741	1.291
	Deny or confirm		
reassent	Expressions of assent	0.013	0.120
reaffect	Expressions of affect	3.055	2.199
renegate	Expressions of negation	0.728	0.581

With regard to the **formal characteristics**, the difference in length is particularly noticeable. For some reason, the responses to the special reports by the European Court of Auditors are more than twice as long as responses to comparable reports in the Netherlands. Furthermore, it is amusing to see that the use of longer words is almost identical for all formal responses.

With regard to the **tone** of responses, it is striking that there are clear differences between the four organizations investigated. Responses to reports from the European Court of Auditors are most pronounced both and especially in more positive terms but also in more negative terms. This suggests also a process of mirroring: the more ‘emotional’ special reports are also met with more ‘emotional’ formal responses.

Finally, the responses were analyzed as to whether they give indications of **denial** or **confirmation**. Analysis shows that there is hardly any explicit confirmation in the responses, while explicit denial is more often the case. This happens slightly more often in response to ECA-reports than to others, although the differences are small. Another indication of how reports are received is the ‘affect’ value, which brings together different expressions of feelings. What is visible here is that the responses to the special reports of the European Court of Auditors express the strongest ‘affect’.

Figure 1 already showed that various of these characteristics of responses, in particular the tone and also affect, are related to the characteristics of the special reports. This suggests that some reports are received more positively and in more positive terms than others. This does not mean that these reports are better in any meaningful way, yet receiving a more favourable response is likely to be related to having a stronger impact. This begs the question: do we see any meaningful patterns? Do reports that are received more positively in the formal response exhibit specific features that make them stand out from others? When do reports then evoke more positive responses? We can explore this by first classifying the various reports more precisely.

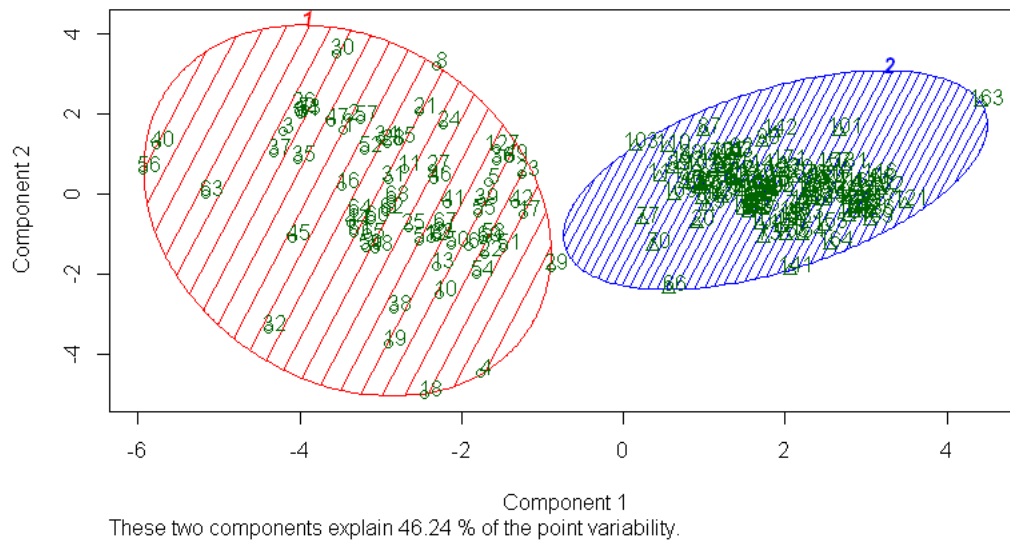
Two archetypal reports

(note: here the comparison between the special reports and the Dutch reports did no longer work properly, due to language issues. The next sections **are not** based on the special reports, but display patterns we found in other reports which may be applicable to the special reports as well, as we indicate).

Until now, the analysis has focused on specific characteristics within reports, but more important is perhaps the question of whether there are meaningful differences *between* types of reports and then on a range of characteristics. Because if that is the case, it is also possible to speak of linguistically different reports. To determine whether there are any substantive linguistic differences between the reports, we have performed an analysis of clusters of all factors at the report level for the Dutch reports. We have used the explorative method K-means clustering, with which coherence between clusters of variables can be made visible. Simply put, it is a method that minimizes the differences between clusters of observations (by reducing the Euclidean distance between observations) and maximizes the variation between clusters. Put even simpler: it is a method that helps to discover coherent patterns within the data.

The analysis then revealed that there are two very clear clusters of reports in the Dutch sample that together comprise almost half of the reports (46.24%). Figure 2 below visualizes the two clusters within which reports from the four organizations can be found.

Figure 2: two clusters of report



The reports in cluster 1 have the following characteristics. They are more complex in terms of reasoning by applying more 'cognitive mechanisms'. They are also more explicitly negative in tone and use longer sentences. The power position in these reports is much lower than in the other reports: they are low on 'clout' and also low in terms of performance focus. Cluster 1 could also be called a cluster of *analytical reports*, in which the analysis of more or less complex issues stands at the core.

The second cluster of reports contrasts on almost all points with the first cluster of analytical reports. The reports in the second cluster are longer, but with more simple sentences and significantly more positive words. Much less use is made of cognitive mechanisms. At the same time, the performance focus in the reports and the social positioning in terms of 'clout' are clearly higher. Cluster 2 could also be called a cluster of *directive reports*, which have a simpler tone and structure but are also more hierarchical.

With the exception of 'clout', on average the ECA reports display most characteristics of what we called *directive reports*.

In order to analyse whether the two types of reporting lead to different responses and at the same time to check for the differences between the organizations, both have been included in a multiple regression analysis. The results show that the formal responses to the two report types differ significantly from each other. The directive reports (cluster 2) are related to a more positive tone in responses than the analytical reports.

Conclusions and reflections

This report aimed to explore whether there are relations at a deeper linguistic level between performance accountability reports on the one hand and the formal responses to those reports on the other. In simple terms: are some types of reports received more favourably than others? At the end of the analysis, the answer is affirmative. The linguistic analysis showed that there are broadly two types of reports – analytical and directive reports – that differ considerably from each other on a fairly large number of indicators. The analysis also shows that they indeed lead to different types of reactions. This provides insight into the simple but crucial question on how the words matter for formal responses.

All in all, there seems to be a relationship between the language of a report and the response it provokes. This suggests a simple solution for all report writers: writing differently leads to a different reaction. That may sound attractive, but is unfortunately it too simplistic. To start with, our report is 'contaminated' by the variance between reports of organizations with are sometimes carried out according to a fixed methodology that leads to a certain standardisation, also in our analyses. In addition, there are several much simpler explanations for more positive responses. Perhaps the actual situation that is reported is simply much more positive, which may explain the positive tone of the performance auditor and the response to the audit? Or perhaps there is a mirror effect: the assessed body mirrors the tone of the auditor, as is common in 'normal' conversations. And it may of course also be that the connection between the words of the report and the response is indeed there, but that this is not due to the wording as such but is caused by underlying factors, which are more important but do not come into the picture in our analysis. In short, one could think of many alternative explanations for the link between the linguistic properties of a report and its formal reception.

At the same time ... at the same time it *may* just be that there are indeed relations between the wording in performance accountability reports and the effects on the recipient. In human communication this is a robust and easily understandable connection. If you express the same thing differently, you also have a different effect on the person you are addressing. If this is indeed also visible at an institutional level, then this opens up interesting strategic perspectives and also practical possibilities for public sector auditors with which they can try to make their assessments effective by articulating them in the 'correct' manner. This report provides insight into a number of factors that can play a role in this. It is much too early to translate this directly into practice, but it is certainly worthwhile to reflect, together with the auditors themselves, on the value of our findings so far.

References

- Bruijn, H. de, T. Schillemans en M. van der Steen. (2012). Framing framing. Betekenisgeving en besluitvorming over beleid. *Bestuurskunde*, 21(4):2-8.
- Cohn, M.A., Mehl, M.R., Pennebaker, J.W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15:687-693
- Conway, L. G., Conway, K. R., Gornick, L. J., & Houck, S. C. (2014). Automated integrative complexity. *Political Psychology*, 35(5):603-624.
- Esch, F. van (2014). *A matter of personality? Stability and change in EU leaders' beliefs during the Euro crisis*. In: D. Alexander & J. Lewis (eds.), *Making public policy decisions: Expertise, skills, and experience*, Londen: Routledge, pp. 53-76.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
<https://doi.org/10.1177/0261927X13502654>
- Ministerie van BZK. (2016) *Cijfers over Wonen en Bouwen 2016*. Den Haag: BZK.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5), 665-675.
- Parker, I. (2014). *Discourse Dynamics (Psychology Revivals): Critical Analysis for Social and Individual Psychology*. London: Routledge.
- Schillemans, T. (2009). *Visitaties als vorm van verantwoording*. In: Bovens, M. en T. Schillemans, 2009, *Handboek publieke verantwoording*. Den Haag: Lemma, p. 161 – 180.
- Suedfeld, P., & Tetlock, P. E. (2014). Integrative complexity at forty: Steps toward resolving the scoring dilemma. *Political Psychology*, 35(5):597-601.
- SVWN (2014), *Methodiek maatschappelijke visitatie woningcorporaties 5.0*. Zeist: SVWN.
- Tetlock, P. E., Metz, S. E., Scott, S. E., & Suedfeld, P. (2014). Integrative complexity coding raises integratively complex issues. *Political Psychology*, 35(5), 625-634.
- Twist, M. van (2018), *Woorden wisselen. Werken aan een hertaling van besturen, organiseren en adviseren*. Deventer: Vakmedianet.
- Zijlstra, H., Meerveld van, T., Middendorp van, H., Pennebaker, J.W., Geenen, R. (2004). De Nederlandse versie van de 'Linguistic Inquiry and Word Count' (LIWC). *Gedrag & Gezondheid*, 32(4):271-281.
- Zijlstra, H., Middendorp van, H., Meerveld van, T., Geenen, R. (2005). Validiteit van de Nederlandse versie van de Linguistic Inquiry and Word Count (LIWC). *Nederlands Tijdschrift voor de Psychologie*, 60:55-63.

Colophon

Prof. dr. Thomas Schillemans is a Professor of public governance at the Utrecht University School of Governance, with a focus on accountability and behaviour in institutions.

Marija Aleksosvka, Msc, is a doctoral candidate at the Utrecht University School of Governance, focusing on the effects of accountability mechanisms on the work of public sector organizations, by using experimental methods.

This research was conducted in consultation with four public organizations. We would like to thank the organizations involved for help with the selection and collection of reports. Thanks also to Maj Grove Jeppesen, Rixt Riemersma and Kees van der Wel for research assistance.

The research is part of the *Calibrating Public Accountability* research program that can be carried out thanks to an NWO Vidi grant in 2015. The research program focuses on the question of how forms of public accountability in the public sector can be improved in order to really help public organizations to perform their duties properly.

See for the project: <https://accountablegovernance.sites.uu.nl>