

Text Mining, Travel Writing, and the Semantics of the Global

An AntConc Analysis of Alexander von Humboldt's *Reise in die Aequinoktial-Gegenden des Neuen Kontinents*

Christine Peters

Literary scholars have long been arguing that an integral part of travel writing is its continuous engagement with the world. In her introduction to travel writing, Anne Fuchs even identifies *Welthaltigkeit* – a containment of the world and an engagement with the world in the text – as one of the two key characteristics of travel writing, the other being its engagement with alterity.¹ While the interest in literary figurations of the global is not exclusive to research on travel writing,² the genre appears to be an especially productive object of investigation. Studies on Alexander von Humboldt, who can be considered a central figure of 19th century travel writing, frequently address the different ways in which he inscribes a global perspective into his writing, aiming to describe the world as a coherent, wholesome entity.³ Even

1 Fuchs, Anne, Reiseliteratur, in: Dieter Lamping (ed.), Handbuch der literarischen Gattungen, Stuttgart: Alfred Kröner, 2009, 593–600.

2 Moser, Christian/Simonis, Linda, Einleitung: Das globale Imaginäre, in: Sebastian Moser/ Linda Simonis (eds.), Figuren des Globalen: Weltbezug und Welterzeugung, Göttingen: V&R unipress, 2014, 11–22.

3 Daum, Andreas, Alexander von Humboldt, die Natur als 'Kosmos' und die Suche nach Einheit: Zur Geschichte von Wissen und seiner Wirkung als Raumgeschichte, in: Berichte zur Wissenschaftsgeschichte 22 (2000), 246–250; Böhme, Hartmut, Ästhetische Wissenschaft: Aporien der Forschung im Werk Alexander von Humboldts, in: Ottmar Ette (ed.), Alexander von Humboldt: Aufbruch in die Moderne, Berlin: Akademie-Verlag, 2001, 17–32; Heyl, Bettina, Das Ganze der Natur und die Differenzierung des Wissens: Alexander von Humboldt als Schriftsteller, Berlin/Boston: De Gruyter, 2007, 179 ff.; Ette, Ottmar, Humboldt und die Globalisierung: Das Mobile des Wissens, Frankfurt a. M./Leipzig: Insel,

though these studies focus on different aspects of Humboldt's engagement with the world, methodically, they take the same traditional 'close-reading' approach, focusing on so called 'symptomatic readings' of the material. The following paper is conceived as a computationally-assisted contribution to thinking about the specific ways in which Humboldt imagines the world and globality in a broader sense. It focuses on one specific piece of travel writing, namely Herman Hauff's influential German translation of the *Relation Historique* (1814 ff.), the *Reise in die Äquinoktial-Gegenden des neuen Kontinents* (1859).⁴ I proceed from the assumption that we can indeed gain some further insight into how Humboldt's travel writing engages with the world by identifying a set of relevant lexical features throughout a relatively small-sized corpus. Using *AntConc*⁵ as a concordancing and text analysis toolkit, I argue for the interpretative productivity of combining 'distant' reading methods with a more traditional 'close' reading approach.

Firstly, the following paper employs computational methods to test a particular hypothesis derived from my earlier, non-digital research on world knowledge in Humboldt's travel writing. I previously adopted the traditional close reading approach, focusing mainly on specific text strategies that engage with the world in the text, such as practices of comparing and narrating.⁶ These studies show that in his travel writing, Humboldt primarily

2009, 17ff.; Knobloch, Eberhard, Alexander von Humboldts Weltbild, in: HiN: Internationale Zeitschrift für Humboldt-Studien X (2009): 36f.; Görbert, Johannes, Die Vertextung der Welt: Forschungsreisen als Literatur bei Georg Forster, Alexander von Humboldt und Adelbert von Chamisso, Berlin: De Gruyter, 2014, 157; Erhart, Walter, Chamissos Weltreise und Humboldts Schatten, in: Julian Drews et al., Forster – Humboldt – Chamisso: Weltreisende im Spannungsfeld der Kulturen, Göttingen: V&R unipress, 2017, 13–34.

4 This study is embedded in a dissertation project on German travel writing, including German translations of originally French, English and Russian travelogues that were available to German speaking communities in the nineteenth century. The following paper can be considered as a trial run of a computational analysis of my material, focusing on a small exemplary corpus, which could easily be repeated and adjusted for the original travelogue or even a larger corpus including a variety of different travelogues or their translations. For an exposé of the dissertation project, see: <http://www.uni-bielefeld.de/sfb1288/personen/person.html?persId=80276541> [accessed: 31.08.2019].

5 Laurence Anthony (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software> [accessed: 31.08.2019].

6 Peters, Christine, Reisen und Vergleichen: Praktiken des Vergleichens in Alexander von Humboldt's *Reise in die Äquinoktial-Gegenden des Neuen Kontinents* und Adam Johann von

imagines the world as a relational concept by means of comparison: The text usually engages with the world by describing perceived geographical entities that ‘make up’ the world and by comparatively identifying or establishing relationships between these entities. A sense of the global thereby seems to be achieved through implicit means, through a comparative text structure rather than through addressing questions of the global on an explicit, lexical level. The following paper aims to test this particular thesis by employing computational methods that allow a ‘distant’ re-reading of the material.

Furthermore, this paper uses computational methods to generate new ways of thinking about how Humboldt’s travel writing engages with questions of globality. I aim to show that concordancing toolkits such as AntConc can be used to bring attention to the different levels on which a text can deal with the world as a concept.

Corpus, approaches, preliminary decisions

Considering AntConc’s capacity to perform fast and accurate searches in small and mid-sized corpora, I chose a relatively small corpus of approximately 1250 pages, consisting of the four volumes of Hermann Hauff’s German translation of the *Relation historique*. I aim to show that with a mostly supervised analytical approach⁷ you can identify a set of significant lexical

Krusensterns *Reise um die Welt*, in: IASL 42 (2017), 441–455; Peters, Christine, Historical Narrative versus Comparative Description? Genre and Knowledge in Alexander von Humboldt’s *Personal Narrative*, in: Martin Carrier/Carsten Reinhardt/Veronika Hofer (eds.), *Narratives and Comparisons: Adversaries or Supporters in Understanding Science?*, (manuscript in preparation). While the first paper focuses on practices of comparing in the German translation of the *Relation historique* as well as in the *Ansichten der Natur*, the second paper analyzes practices of comparing and narrating in the *Personal Narrative*, the English translation of the *Relation historique*.

- 7 In general, we can distinguish between two types of approaches in the field of text mining and corpus linguistics: On the one hand, we can take an ‘unsupervised’ or ‘corpus-driven’ approach that starts the analysis with direct observations of the corpus without bringing pre-existent hypotheses to the corpus. On the other hand, we can take a more ‘targeted’ or ‘corpus-based’ approach that analyzes the corpus in order to test pre-existing hypotheses (and maybe develop some new insights based on the results). Since this study tests very specific hypotheses about globality in Humboldt’s travel writing, most analytical steps belong to the latter category, taking a more ‘targeted’ or ‘supervised’ ap-

features even in a small corpus such as this. The analysis approached the corpus on four different levels:

Firstly, as a preliminary step, I performed a short unsupervised inquiry⁸ using the word list and the n-grams tools to map out frequent lexical reoccurrences that might be relevant to the general semantic structure of the travelogue and to also gain a foundation for some first guesses on how ideas of the global might factor into the predominant lexical layout of the text. Secondly, in a significantly more extensive step, I performed a supervised corpus analysis, using a short preselected list of search words with a global resonance to map out the most frequent words and word clusters with the clusters tool. An additional goal was to identify the hermeneutic limits of this approach and the measures that need to be taken to further verify or specify the preliminary results. In a third step, I analyzed the 'distant' reading results in their specific contexts, using AntConc's file view tool for a 'close' re-reading of the results. As the following analysis shows, this combination of computational methods and a 'close' reading of the material was a crucial step in mapping out the different ways in which the text engages with the world. As a last step, I tested the particular hypotheses based on that process by comparing my corpus to a reference corpus, consisting of the five volumes of Alexander von Humboldt's *Kosmos*. The comparison proved that texts can engage with the world on different levels and to different degrees.⁹

proach. For the differentiation between a 'corpus-driven' and a 'corpus-based' approach see Anthony, Laurence, A critical look at software tools in corpus linguistics, in: Linguistic Research 30 (2013), 142. For an example of a paper that differentiates between a 'more targeted' and an 'unsupervised' approach see Erlin, Matt, Topic Modeling, Epistemology, and the English and German Novel, in: Journal of Cultural Analytics (2017). Other than Lawrence, Erlin does not reflect on how these terms are usually used in the field of corpus linguistics but rather just uses the terms 'unsupervised' and 'targeted' to describe what Lawrence calls a 'corpus-driven' and a 'corpus-based' approach.

8 A word list and n-grams analysis can be considered an unsupervised or corpus-driven analytical step insofar as neither tool requires a search word list or other devices that depend on pre-existing research. However, because my *interpretation* of the results relied heavily on my previous research, this analytical step appeared to be somewhat supervised after all, even if to a lesser extent than the following analytical steps.

9 Many of the text mining techniques used in this paper, such as frequency analysis, collocations, or n-gram clustering, have already been evaluated as valuable tools for the analysis and interpretation of data. For example, Dawn Archer demonstrates how such quantitative methods can enable the comparison of genres, the analysis of the author's ideological stance, or the attribution of authorship. See Archer, Dawn, Data Mining and

As mentioned above, I used a preselected list of search words with a global resonance for the supervised parts of the analysis. This list was built on the work of specialists in the field as well as on my own research on comparison and world knowledge in Humboldt's travel writing. The first search words I added to the list were "Neue Welt" ["New World"] and "Alte Welt" ["Old World"] as well as "neuer Kontinent" ["new continent"] and "alter Kontinent" ["old continent"] because my previous close reading analysis of Humboldt's travel writing had provided evidence that all of these terms are included in comparative practices that produce a sense of the global in the text.¹⁰ For the present study, the aim was to use computational methods to go beyond symptomatic readings of the text and instead trace these terms and their lexical variations throughout the whole length of the travelogue and potentially generate some new insights into how these terms conceptualize the world on an explicit, lexical level.

To broaden the analytical spectrum, I added the search terms "Welt" ["world"] and "Erde" ["earth"] to the initial list, not only because they seemed like obvious choices but also because experts in the field had previously relied on them to explain how Humboldt conceptualizes the world in his writing.¹¹ Concerning the term "Welt", the goal was to analyze both the occurrence of the term in its singular form as well as the wider variety of compound nouns, such as "Weltbewusstsein" ["global consciousness", "world consciousness"], "Welthandel" ["world trade"], "Weltkulturen" ["world cultures"], "Weltbegriffe" ["world concepts"], "Weltwissenschaft" ["world science"], or "Weltethos" ["global ethics"], which, as Ottmar Ette shows, play a central role in Alexander von Humboldt's different world concepts.¹² Adding the term

Word Frequency Analysis, in: Gabriele Griffin/Matt Hayler (eds.), *Research Methods for Reading Digital Data in the Digital Humanities*, Edinburgh: Edinburgh University Press, 2016, 72–92.

10 C. Peters, *Reisen und Vergleichen*, 450–455.

11 See footnote 3.

12 O. Ette, *Humboldt und die Globalisierung*; Ette, Ottmar, *Weltbewusstsein*: Alexander von Humboldt und das unvollendete Projekt einer anderen Moderne, Weilerswist: Velbrück, 2002. None of these compound nouns are easily translated since they carry very specific philosophical and epistemological connotations. For this paper, I used Ette's translations of the terms. For the translations of "Weltbewusstsein" ["global consciousness", "world consciousness"] see Ette, Ottmar, *Unterwegs zum Weltbewusstsein*: Alexander von Humboldts Wissenschaftsverständnis und die Entstehung einer ethisch fundierten Weltanschauung, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* 1 (2000);

“Erde” to the list, including compound nouns in the same manner as with the term “Welt”, proved essential to the study since it allowed a focus on two alternative figurations of the global.¹³ In both cases I decided not to add the specific compound nouns to the list of search words but rather their lexical stems “erd*” and “*welt*”¹⁴ to enable search results that include compound nouns that I had not anticipated. The final list¹⁵ included all the search terms in form of their lexical stems:

Ette, Ottmar, Languages about Languages: Two Brothers and one Humboldtian Science, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* XIX (2018), 47–61. For the translations of “Welthandel” [“world trade”] and “Weltkulturen” [“world cultures”] see *Ette, Ottmar*, The Scientist as Weltbürger: Alexander von Humboldt and the Beginning of Cosmopolitics, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* II (2001). For the translation of “Weltbegriffe” [“world concepts”] see *Ette, Ottmar*, Unterwegs zu einer Weltwissenschaft? Alexander von Humboldts Weltbegriffe und die transarealen Studien, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* VII (2006), 34–54. For the translation of “Weltwissenschaft” [“world science”] see *O. Ette*, Languages about Languages. For the translation of “Weltethos” [“global ethics”] see *O. Ette*, Unterwegs zum Weltbewußtsein.

- 13 *Stockhammer, Robert*, Welt oder Erde? Zwei Figuren des Globalen, in: Christian Moser/Linda Simonis (eds.), *Figuren des Globalen. Weltbezug und Welterzeugung in Literatur, Kunst und Medien*, Göttingen: V&R unipress, 2014, 47–72.
- 14 As most search engines, AntConc allows the truncation of search words, meaning the use of “wildcards”, such as *, #, ?, or +. The asterisk (which finds zero or more characters) helps finding different variations of the same word, such as nouns in their singular and plural forms (e.g., “Welt” and “Welten” for “welt*”) or different compound nouns containing the same words (e.g., “Erdball” and “Erdstrich” for “erd*”). However, it is important to note that even though it is possible to search for different variations of the same word, the list of results will not display these variations as different instances of the same word. In addition, the list will most likely include other random words that contain the same combination of characters but do not belong to the same lexical group. In any case, it is necessary to sort through the list of results manually to identify the relevant findings.
- 15 I excluded the search term “*erd*” after my first attempts because it rendered too many results, namely 2000 occurrences in the corpus, which in turn included too many results not connected to the term “Erde”, such as “werden” or “allerdings”. I also excluded the search term “*asi*” for the same reason: Even though it only occurs 226 times in the corpus, too many instances were not related to the term “Asien”, such as the term “Brasilien” which occurred twice as often as the term “Asien”. Also, since the German differentiation between the upper and lower case was of no significance to my analysis, in the following, I quote all my search words and results in the lower case form, the only exception being quotes including whole text passages or longer phrases.

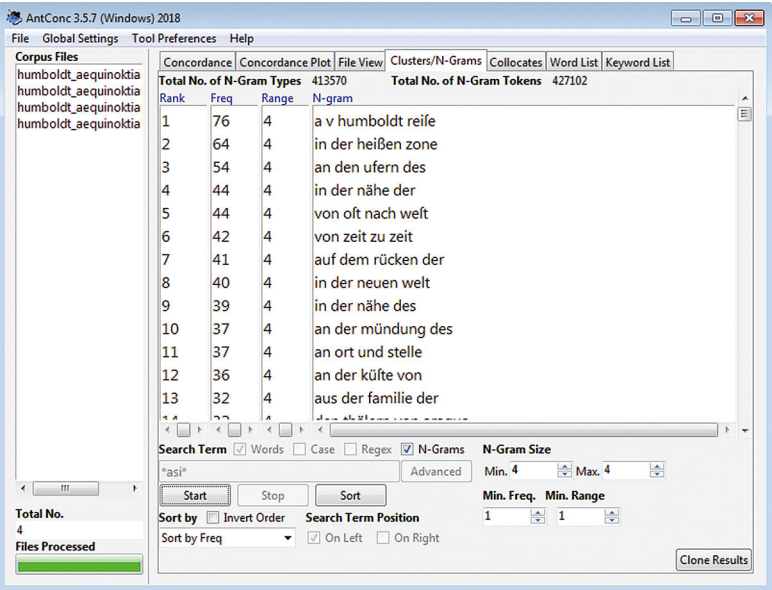
Fig. 1: Search word list for the distant reading analysis, lexical stems

welt	[world]
erde*, erd*	[earth]
neue* welt*, alte* welt*	[new world, old world]
neue* kontinent*, alte* kontinent*	[new continent, old continent]

Distant reading I: unsupervised corpus analysis

Before approaching the corpus with the preselected search word list, I conducted an unsupervised corpus analysis. I used the n-grams tool which allows the user to find common expressions in a corpus by searching the entire corpus for “N” length clusters (e. g., one word, two words, etc.). For this corpus, the analysis rendered the most significant results at an n-gram size of four words:

Table 1: Results of the n-gram analysis, n-gram size of 4 words



Apart from the highest ranking n-gram “a v humboldt reise”,¹⁶ which refers to the title of the travelogue, and the temporal n-gram “von zeit zu zeit” [from time to time], all the n-grams ranking in the top ten carry spatial meaning, either referring to a specific region or place, or describing a spatial movement in a certain direction. The only n-gram with a global ring to it is the 8th ranking n-gram, “in der neuen welt” [in the new world], which, as mentioned before, evokes a sense of the global but does in fact refer to a specific region of the world and more specifically to the region that is the object of this travelogue.

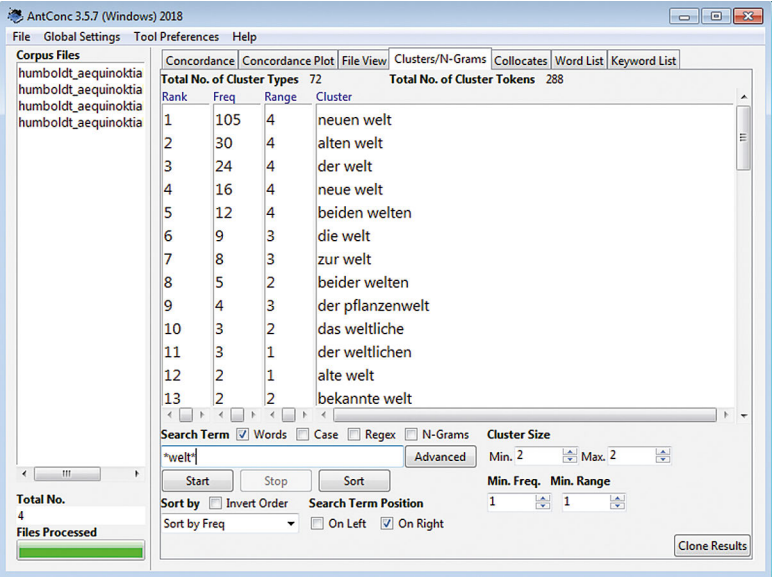
Since continuously narrating movement through space is an integral part of travel writing, it is almost to be expected that the lexical features reflect this spatial focus of the text. However, for my analysis of the imagination of the global, it is noteworthy that none of the high ranking n-grams actually feature a global perspective on a lexical level. A general wordlist points to a similar focus on local or regional space. The wordlist tool counts all the words in a corpus and then displays them in an ordered list which gives the user easy access to the most frequent words in a corpus. With 1107 instances, the highest ranking word in the travelogue is “orinoko” [Orinoco], the river that poses a major stop on Humboldt’s travels and a central object of description in the travelogue. In comparison, the term “welt” only ranks 76th with 219 occurrences. While 219 occurrences can still be considered significant in a 1250 page corpus, the difference in frequency is still remarkable. It points to the idea that on the lexical surface of the text, globality does not actually play as big a role as it does when it comes to the overall narrative and comparative structure of the text. One way to test this tentative notion is to apply a supervised approach and analyze the terms “welt” and “erde” in the concordance, clusters, and collocates tools.

16 Literally this n-gram translates as “a[lexander] v[on] humboldt voyage”, but with regard to the title of the English translation, a more appropriate translation would be “personal narrative.” See *Humboldt, Alexander von, Personal Narrative of Travels to the Equinoctial Regions of the New Continent, during the Years 1799–1804*, 7 vols., translated by Helen Maria Williams, New York: Cambridge University Press, 2011, 1814 ff.

Distant reading II: supervised corpus analysis

The supervised corpus analysis, at first, seemed to contradict the notion that the lexical features of the text do not point to the world in its entirety. After all, a search for “*welt*” in the corpus¹⁷ shows 288 instances of the word.¹⁸ However, a cluster analysis shows that the term carries mostly regional connotations. The clusters tool displays clusters containing a chosen search word, based on specific search conditions, such as the size of the cluster or the position of the search word. A cluster analysis for the search term “*welt*” with the search term positioned on the right and a cluster size of two words rendered the following results:

Table 2: Results of the cluster analysis for “*welt*”, cluster size of two words, search word on the right



17 To determine how often a specific search word occurs in the corpus, I used the concordance tool which displays the search words in their contexts but also provides the user with the total number of occurrences.

18 The earlier mentioned 219 occurrences as displayed in the wordlist refer to the actual word form “welt”, while the above mentioned 288 occurrences refer to all lexical variations of the term as displayed in the concordance tool when searching for the term “*welt*”.

The list shows that different variations of the cluster “neue welt” [new world] as well as the cluster “alten welt” [old world] indeed figure prominently in the top 5 of the 72 existing clusters. According to a search in the concordance tool, the most frequent cluster, “neue* welt*”, occurs 121 times, while the second leading cluster, “alte* welt*”, occurs 32 times. This shows that on a lexical level “welt” is rarely used as a global but mostly as a regional and relational term that divides the planet into two corresponding worlds. The third ranking cluster, “beide* welten” [both worlds], which occurs 18 times in the corpus, only reinforces this hemispheric notion of “welt”. With only 9 and 24 occurrences, the clusters “die welt” [the world] and “der welt” [declined form of “die welt”], which actually refer to the world in a generic and singular sense, occur significantly less frequent than the hemispheric clusters.

Changing the cluster size to one word shows that there are a number of compound nouns that refer to the world as a whole but that occur far less frequently than the above mentioned hemispheric clusters. A search for these compound nouns in the concordance tool, using search words such as “weltmeer*” [ocean, world sea], “weltbeschreibung*” [history or description of the world], “welthandel*” [world trade], or “weltraum*” [space, cosmos], shows that the majority of them occur less than five times across the entire length of the corpus. This supports the observation that the term “welt” appears to be mostly used as a hemispheric, regional concept rather than a global one. A comparison with another cluster analysis further corroborates this thesis. The term “kontinent*”, an explicitly regional term that appears 91 times in the corpus, mostly carries the same attributes as the term “welt”: The most frequent clusters are “neue* kontinent” [new continent], occurring 44 times, and “beide* kontinente” [both continents], occurring 17 times.

Concerning the term “erde”, the supervised analysis pointed into a slightly different direction. While the above mentioned cluster analysis for “welt” mostly cycled back to terms I had already preselected for my search term list, the cluster analysis for “erd*” supplied me with a number of terms that I had not anticipated:

Table 3: Results of the cluster analysis for “erd*”, cluster size of one word, search word on the right

AntConc 3.5.7 (Windows) 2018

FileGlobal SettingsTool PreferencesHelp

ConcordanceConcordance PlotFile ViewClusters/N-GramsCollocatesWord ListKeyword List

Corpus Files

humboldt_aequinoktia
humboldt_aequinoktia
humboldt_aequinoktia
humboldt_aequinoktia

Total No. of Cluster Types89

Total No. of Cluster Tokens575

Rank	Freq	Range	Cluster
1	125	4	erde
2	107	4	erdbeben
3	37	4	erdstrich
4	33	4	erd
5	30	4	erdoberfläche
6	24	2	erdstoß
7	15	3	erdballes
8	12	4	erdfriche
9	11	4	erden
10	11	4	erdreich
11	10	2	erdstoß
12	9	3	erdbebens
13	9	4	erdfrichen
14	8	4	erdball
15	7	4	erdigen
16	7	3	erdfriches
17	6	2	erderfchütterungen
18	5	3	erdballs
19	4	3	erdboden
20	4	2	erdfläche

Search Term

☒ Words☐ Case☐ Regex☐ N-Grams

Cluster Size

erd*

Advanced

Min. 1Max. 1

Min. Freq. 1Min. Range 1

StartStopSort

Sort by ☐ Invert Order

Search Term Position

Sort by Freq

☐ On Left☒ On Right

Clone Results

Total No.

4

Files Processed

This list of the 20 highest ranking one-word clusters contains a number of reoccurring terms in their different grammatical variations, most prominently “erde” [earth, soil, ground], “erdbeben” [earthquake/s], “erdstrich” [zone, part of the earth], “erdball” [globe], “erdoberfläche” [earth’s surface], and “erdstoß” [earth tremor], whose exact frequency can be determined by means of the concordance tool: “erde” (125), “erdbeben*” (116), “erdstrich*” (69), “erdball*” (30), “erdoberfläche*” (30), “erdst*ß*”¹⁹ (39). I used this list as a starting point for my analysis which provided me with some unpredicted

19 Because the vowel in the word “erdstoß” changes into an umlaut in the plural (“erdstöße”) an asterisk had to be used in the middle of this search word (“erdst*ß*”). This way, both the singular and the plural appear in the results.

and yet very productive analytical trajectories but also pointed to some of the hermeneutic limits of a computational approach.

At first glance, there appears to be a higher number of terms that might explicitly refer to the planet as a whole, such as “erde” [earth, soil, ground], which occurs 125 times in the corpus, or “erdball*” [globe], which occurs 30 times. A second cluster analysis focusing on these two terms corroborated this observation since they mostly appear in a generic form that is quite different from the dualistic clusters “neue welt” [new world] and “alte welt” [old world]. The two highest ranking two-word clusters with the search word “erde” positioned on the right are “die erde” [the earth] with 37 hits and “der erde” [declined form of “die erde”] with 31 hits. In a similar manner, all 30 two-word clusters with the search term “erdball*” [globe] on the right refer to the term in a generic and singular way with a frequency ranging between 14 and one occurrences, such as the differently declined forms of “der erdball” [the globe] (“des erdballes”, “dem erdball”, “den erdball”, “des erdballs”, and “dem erdballe”) or the differently declined forms of “unser erdball” [our globe] (“unseres erdballs” and “unseres erdballes”). However, since the term “erde” carries two meanings, referring to the planet as a whole as well as to the ground or a specific type of soil, some further testing was necessary. The cluster and collocates analyses did not help to determine how many of the 125 hits actually refer to the planet and how many to the ground. As the next chapter shows, in this case only an arduous use of the file view provided specific numbers, which ultimately contradicted the notion that the term “erde” refers to the whole planet more often than the term “welt”.

The initial list of clusters displayed in table 3 also contained the high ranking compound noun “erdstrich” [zone, part of the earth] which in all its grammatical variations occurs 69 times in the corpus. Semantically it refers to the earth in its totality but also points to a specific region inside this totality. A further cluster analysis showed that the term mainly refers to specific regions with regard to their climatic conditions, the most frequent two-word clusters being “heiße* erdstrich*” [torrid zone] with 31 occurrences and “gemäßigte* erdstrich*” [temperate zone] with 11 occurrences. The compound noun “erdstrich” does not only stress the regional particularity of the planet but more specifically its climatic diversity.²⁰

20 In fact, the English translation frequently refers to what is called “erdstrich” in German both as “zone” and “climate.”

Two other words displayed in the initial list of clusters, “*erdbeben*” [earthquake/s] and “*erdstoß*”/“*erdstöße*” [earth tremor/s], refer to a geological discourse on volcanic activity. With 116 and 39 occurrences they feature prominently in the corpus, but whether they refer to volcanic activity on a global scale or to singular, regional earthquakes cannot be determined by means of distant reading. Research on the matter shows that Humboldt’s geological analysis of volcanic activity has a strong global focal point,²¹ but without switching to close reading in the file view there is no way to decide whether this global focal point manifests on a lexical level. Whether these terms are used in a local or global way can only be determined by analyzing them in their specific contexts. The only tentative notion that is to be drawn from the cluster analysis is that on a lexical level the term “*erde*” has a strong geological connotation that the term “*welt*” lacks.

In sum, the supervised cluster analysis suggested that the term “*erde*” in the form of its different compound nouns points to the planet both in its totality and its particularity, unlike the term “*welt*” which mostly points to local particularity. The results also indicated specific geological and climatic connotations of the term “*erde*” which the term “*welt*”, at least on an explicit lexical level, does not carry.

In addition to the cluster analysis, I performed a co-occurrence analysis using the collocates tool, which allows the user to investigate non-sequential patterns in a corpus. The tool displays the collocations of a search word, meaning words that occur in the context of the given search word. The aim was to identify frequent co-occurring words and to see whether they confirmed the mostly regional connotations of the search terms or rather contradicted them. However, even when applying the maximum search window size of 40 words, the analysis rendered no viable results. None of the high ranking collocations allowed valid guesses on how the specific contexts interacted with the regional focus of the search words. Again, using the file view for a close reading analysis of the results appeared to be the necessary next step.

Taking previous research on Humboldt’s engagement with the world into account, the results of the supervised corpus analysis seem unremarkable

21 Kraft, Tobias, *Erdwissen im Angesicht der Berge: Die Vulkanlandschaften der Jorullo-Ebene als Heuristik der Geologie*, in: Ottmar Ette/Julian Drews (eds.), *Horizonte der Humboldt-Forschung: Natur, Kultur, Schreiben*, Hildesheim: Georg Olms, 2016, 97–124.

and remarkable at the same time. As Ottmar Ette shows, Humboldt's world concepts are based on a long tradition of perceiving the world as a divided, hemispheric space that actually consists of *two* worlds, the "new" and the "old" one.²² The fact that the lexical level of the travelogue reflects this tradition is not surprising. However, as mentioned in the introduction, scholars have also frequently pointed towards the global perspective of Humboldt's research and writing, towards his aim to describe the world as a coherent, wholesome entity. What is remarkable about the results presented above is that the travelogue's lexical level does not display such a strong focus on the world in its totality, but rather leans strongly towards a territorial usage of the terms "welt" and "erde", drawing attention to specific regions of the planet rather than to the planet as a whole.

A close re-reading of the results: approaching the global on a different level

In reaction to the distant reading results, the first purpose of the close reading approach was to further investigate the actual connotations of the search terms in their specific contexts. A second purpose was to look for patterns in these contexts, for potential text structures that go beyond the lexical layout of the text, and to see how the search terms were embedded in these text structures. The initial idea was to first approach such contextual matters by using the collocates tool. However, the tool's capacity appeared too limited for such an approach in two ways. First, the maximum search window size of 40 words proved to be too small, considering that an argument or a narrative sequence often develops over a whole paragraph or even over several pages. Second, a collocation analysis only takes re-occurrences of identical words into account, thereby only focusing on *lexical* context. Therefore, a collocation analysis does not necessarily detect recurring context in the form of *text strategies* that often vary when it comes to the specific words which are used. A close reading of the material therefore proved the necessary next step.

22 O. Ette, Humboldt und die Globalisierung, 36 ff.

As a first step, I preselected a new search word list based on the distant reading results, containing the most frequent words and word combinations from the cluster analysis:²³

Fig. 2: Search word list for the close reading analysis, lexical stems

neue* welt	[new world]
beide* welten	[both worlds]
alte* welt	[old world]
erde	[earth]
erdball*	[globe]
erdstrich*	[zone, part of the earth]

I searched the corpus for these words using the file view tool which displays the search results in the individual text files and thereby allows the user to analyze them in their specific contexts. To keep my results comparable and also to keep the analysis practicable I decided that for each search hit I would consider the whole surrounding paragraph as context. Methodologically, it should also be noted that the patterns I identified relied heavily not only on this pragmatic decision but also on the specific questions I brought to the material. As mentioned in the introduction, my previous research had indicated that comparative text practices are especially productive when it comes to dealing with the world in its entirety.²⁴ One of my explicit goals was to test this thesis and see whether I could trace those comparative practices by analyzing the search terms in their contexts. With those questions in mind, I paid special attention to the spatial structure of the paragraphs

23 I did not include the search term “erd*” in the list because it occurs 575 times in the corpus, making a thorough and continuous close reading analysis quite difficult. Instead, I decided to use frequent words and word clusters indicated in table 3 (“erde”, “erdball*”, and “erdstrich*”) as a starting point for my close reading analysis. That way the list of results was separated into significantly smaller chunks, which made a “manual” analysis much more effective. Furthermore, I did not add the geological words referring to earthquakes and seismic activity to the list, but rather focused on words that carried spatial connotations (again: “erde”, “erdball*”, and “erdstrich*”).

24 C. Peters, *Reisen und Vergleichen*; C. Peters, *Historical Narrative versus Comparative Description*.

and to whether – and if so how – they establish relationships between different geographical entities. Also, the numbers I provide relied on analytical decisions. They were not supplied by the tool itself but rather by my ‘manual’ interpretation of the results. In short, the findings of my close reading approach, even if carried out in the AntConc toolkit, are not the direct result of a computational analysis but rather of the hermeneutic attempt to organize the material in accordance to the above mentioned analytical interests.

Tracing the high-ranking cluster “neue welt” [new world] in the file view showed that at least 69 of the 121 occurrences are set in a global context. The biggest portion of these 69 occurrences, at least 50 according to my analysis, explicitly or implicitly shows a comparative context, as can be seen in the following exemplary excerpts:²⁵

“Matanza bedeutet Schlachtbank, Blutbad, und schon das Wort deutet an, um welchen Preis der Sieg erkaufte worden. In der **Neuen Welt** weist er gewöhnlich auf eine Niederlage der Eingeborenen hin; auf **Tenerifa** bezeichnet das Wort Matanza den Ort, wo die Spanier von denselben Guanchen geschlagen wurden, die man bald darauf auf den spanischen Märkten als Sklaven verkaufte.”

[“Matanza signifies butchery, or carnage; and the word alone recalls the price, at which victory has been purchased. In the **New World**, it generally indicates the defeat of the natives; at **Teneriffe**, the village of Matanza was built in a place where the Spaniards were conquered by those same Guanches, who soon after were sold as slaves in the markets of Europe.”]²⁶

“Ich bin weit entfernt, die Sprachen der **Neuen Welt** den schönsten Sprachen **Asiens** und **Europas** gleichstellen zu wollen; aber keine von diesen hat ein klareres, regelmäßigeres und einfacheres Zahlssystem als das Qquichua und

25 Typographically, the original text uses “f” for “s” in certain cases. Since typographical issues are of no significance to the present study, I adjusted the quotes, always using “s” even though AntConc displays the original characters. Furthermore, I highlighted the names of places and regions that are being compared or related to each other in order to enable the reader to follow the analysis more easily.

26 *Humboldt, Alexander von*, *Reise in die Aequinoktial-Gegenden des neuen Kontinents*, vol. 1., translated by Hermann Hauff, Stuttgart: Cotta, 1859, 69. [*A. v. Humboldt*, *Personal Narrative*, vol. 1, 134.]

das Aztekische, die in den großen Reichen Cuzco und Anahuac gesprochen wurden.”

[“I am far from placing the languages of the **New World** in the same rank with the finest languages of **Asia** and **Europe**; but no one of them has a neater, more regular, and simpler system of numeration, than the Qquichua and the Azteck, which were spoken in the great empires of Couzco and Anahuac.”]²⁷

“Nach allem, was wir vom Gleichgewicht der Meere wissen, kann ich nicht glauben, daß die **Neue Welt** später als die **Alte** dem Schoße des Wassers entstiegen, daß das organische Leben in ihr jünger, frischer sein sollte; wenn man aber auch keine Gegensätze zwischen den **zwei Halbkugeln** desselben Planeten gelten läßt, so begreift sich doch, daß auf derjenigen, welche die größte Wasserfülle hat, die verschiedenen Flußsysteme längere Zeit gebraucht haben, sich voneinander zu scheiden, sich gegenseitig völlig unabhängig zu machen.”

[“From what we know of the equilibrium of the seas, I cannot think, that the **New World** issued from the waters later than the **Old**; and that organic life is there younger, or more recent: but, without admitting oppositions between the **two hemispheres** of the same planet, we may conceive, that in the hemisphere most abundant in waters the different systems of rivers required more time, to separate themselves from one another, and establish their complete independence.”]²⁸

“In der **Neuen Welt** gingen ähnliche Wanderungen in der Richtung von Nord nach Süd. In **beiden Halbkugeln** richtete sich die Bewegung der Völker nach dem Zug der Gebirge; aber im **heißen Erdstrich** wurden die gemäßigten Hochebenen der Kordilleren von bedeutenderem Einflusse auf die Geschehnisse des Menschengeschlechtes als die Gebirge in **Centralasien** und **Europa**.”

[“In the **New World** similar migrations flowed from north to south. Among the nations that inhabited the **two hemispheres**, the direction of this movement followed that of the mountains; but, in the **torrid zone**, the temperate

27 A. v. Humboldt, *Reise*, vol. 2, 22. [A. v. Humboldt, *Personal Narrative*, vol. 3, 242.]

28 A. v. Humboldt, *Reise*, vol. 3, 256. [A. v. Humboldt, *Personal Narrative*, vol. 5, 315.]

table-lands of the Cordilleras exerted a greater influence on the destiny of mankind, than the mountains of **Asia** and **central Europe**.”²⁹

Throughout the corpus, text passages such as these compare the “neue welt” [new world] to other regions or places of the world, such as the “Alte Welt” [“Old World”], “Asien” [“Asia”], or “Europa” [“Europe”]. They do so either by describing them in a similar manner, thereby urging the reader to perform the actual comparison themselves, or by explicitly arranging them in a comparative manner, using adjectives in their comparative or superlative form, and naming specific similarities and differences.³⁰

The rest of the 69 occurrences in a global context also relate the term “neue welt” [new world] to other regions of the world. As the following list of excerpts shows, they do not necessarily compare these regions to each other but rather focus on different types of relations between them:

“Erzeugnisse der **Neuen Welt** können in die **Alte Welt** nur in hohen Breiten und in der Richtung des **Stromes von Florida** gelangen.”

[“The productions of the **new world** cannot reach the **old**, but by the very high latitudes, and in following the direction of the **current of Florida**.”]³¹

“Wie das Zuckerrohr zuerst von den **Kanarien** in die **Neue Welt** kam, so stehen noch jetzt meist Kanarier oder Isleños den großen Pflanzungen vor und geben beim Anbau und beim Raffinieren die Anleitung.”

[“If the first canes arrived in the **New World** from the **Canary islands**, it is also in general Canarians, or Islenegos, who are now placed at the head of

29 A. v. Humboldt, *Reise*, vol. 4, 234. [A. v. Humboldt, *Personal Narrative*, vol. 6, 15.]

30 Some of the examples that are not displayed here are much shorter, others too long to quote them in a paper as the one at hand. For example, Humboldt's extensive comparison of different plains across the globe stretches over a number of pages, in one paragraph alone referring to “Europa” [Europe], “Asien” [Asia], “Afrika” [Africa], “Amerika” [America], “Arabien” [Arabia], “Venezuela” [Venezuela], “Peru” [Peru], “Neuen Welt” [New World], and the specific plains and deserts “Dsungarei” [Dzungaria], “Pußten” [Puszta], “Gobi” [Gobi], “Sahara” [Sahara], and “Llanos” [Llanos]. See A. v. Humboldt, *Reise*, vol. 2, 269. [A. v. Humboldt, *Personal Narrative*, vol. 4, 295 f.]

31 A. v. Humboldt, *Reise*, vol. 1, 29. [A. v. Humboldt, *Personal Narrative*, vol. 1, 58.]

the great plantations, and who superintend the labours of cultivation and refining.”]³²

“... große Naturforscher (Cuvier) scheinen anzunehmen, daß alle Pythone der **Alten**, alle Boa der **Neuen Welt** angehören.”

[“... great naturalists † appear to admit, that all the pythons belong to the **ancient**, and all the boas to the **New World**.”]³³

“Gegenwärtig teilen sich, kann man wohl sagen, drei **Völker europäischer Abkunft** in das Festland der **Neuen Welt**: das eine, das mächtigste, ist germanischen Stammes, die beiden anderen gehören nach Sprache, Litteratur und Sitten dem lateinischen Europa an.”

[“The continental part of the **New World** is at present in some sort divided between three **nations of European origin**; one, the most powerful, is of Germanic race; the two others belong by their language, their literature, and their manners, to Latin Europe.”]³⁴

As the list shows, a sense of globality is achieved in various ways, for example by describing global trade routes, by tracing the migration of agricultural crops, by debating to which region of the world certain animal species belong, or by discussing the European colonization of the Americas. These results indicate that the corpus actually shows a very strong focus on the world as a whole, but that this focus does not manifest on the level of lexical features. As the cluster analysis has shown, the corpus does not contain many lexical entities that refer to the world in its totality, whereas in the file view we can trace a number of comparative and otherwise relational text strategies that add a global perspective to the text. A similar close reading analysis of the much less frequent clusters “alte welt” [old world] (32 occurrences) and “beide welten” [both worlds] (18 occurrences) further corroborated this thesis. Both of them appear exclusively in comparative contexts of global proportions.

32 A. v. Humboldt, *Reise*, vol. 2, 225. [A. v. Humboldt, *Personal Narrative*, vol. 4, 182.]

33 A. v. Humboldt, *Reise*, vol. 3, 172. [A. v. Humboldt, *Personal Narrative*, vol. 5, 141.]

34 A. v. Humboldt, *Reise*, vol. 4, 286. [A. v. Humboldt, *Personal Narrative*, vol. 6, 112.]

As a next step, I traced the word “erde” throughout the corpus which showed that it is actually used mostly as a non-spatial term. 86 of the 121 instances refer to the soil or ground of a specific place, for example in phrases such as “Scheiben aus gebrannter Erde” [“discs of baked earth”],³⁵ “schnee-weiße Erde” [“clay ... of a snowy whiteness”],³⁶ “eine Art Fort aus Erde und Holz” [“a kind of little fort, constructed of earth and timber”],³⁷ or “feuchte Erde” [“damp ground”],³⁸ rather than to the planet – meaning that in approximately 70 percent of the cases the term carried no spatial and especially no global meaning. That left only 35 instances of the term carrying a global connotation, for example in phrases such as “Umdrehung der Erde” [“rotation”, “the earth’s rotation”, “rotatory motion of the globe”, “rotation of the Earth”, “rotation of the globe”]³⁹ or “Krümmung der Erde” [“curve of the globe”, “curvature of the earth”, “rotundity of the Earth”].⁴⁰

A search for “erdball*” [globe] painted a slightly more complicated picture. The term does not occur very frequently either, it only appears 30 times in the corpus, but it consistently refers to the planet as a whole, addressing such things as “Geschichte des Erdballes” [“history of the globe”],⁴¹ “Umwälzungen des Erdballes” [“revolutions of the Globe”, “revolutions of the globe”],⁴² or “Verteilung der Arten auf dem Erdballe” [“distribution of various species on the globe”].⁴³ However, it is often set in a generic comparative context that stresses the regional diversity of the planet, as is best demonstrated in the following excerpt:

“Gewinnt man einen Ueberblick über die Geschichte unseres Geschlechtes, so sieht man diese Mittelpunkte antiker Kultur da und dort gleich Lichtpunkten

35 A. v. Humboldt, *Reise*, vol. 1, 120. [A. v. Humboldt, *Personal Narrative*, vol. 1, 279.]

36 A. v. Humboldt, *Reise*, vol. 2, 130. [A. v. Humboldt, *Personal Narrative*, vol. 3, 488.]

37 A. v. Humboldt, *Reise*, vol. 3, 202. [A. v. Humboldt, *Personal Narrative*, vol. 5, 206.]

38 A. v. Humboldt, *Reise*, vol. 4, 111. [A. v. Humboldt, *Personal Narrative*, vol. 5, 618.]

39 A. v. Humboldt, *Reise*, vol. 1, 23, 29, 32, and 33; vol. 3, 184. [A. v. Humboldt, *Personal Narrative*, vol. 1, 46, 58, 64, 65; vol. 5., 171.]

40 A. v. Humboldt, *Reise*, vol.1, 53; vol. 2, 38; vol. 4, 141. [A. v. Humboldt, *Personal Narrative*, vol. 1, 105; vol.3, 290; vol.5, 676.]

41 A. v. Humboldt, *Reise*, vol. 1, 68. [A. v. Humboldt, *Personal Narrative*, vol. 1, 133.]

42 A. v. Humboldt, *Reise*, vol. 1, 197; vol. 2, 57. [A. v. Humboldt, *Personal Narrative*, vol. 2, 273; vol. 3, 342.]

43 A. v. Humboldt, *Reise*, vol. 3, 189. [A. v. Humboldt, *Personal Narrative*, vol. 5, 181.]

über den Erdball verstreut, und gewahrt mit Ueberraschung, wie ungleich die Gesittung unter den Völkern ist, die fast unter demselben Himmelsstriche wohnen und über deren Wohnsitze scheinbar die Natur dieselben Segnungen verbreitet hat.”

[“In studying the history of our species, we see, at certain distances, these foci of ancient civilization dispersed over the Globe like luminous points; and we are struck by the inequality of improvement in nations inhabiting analogous climates, and whose native soil appears equally favoured by the most precious gifts of nature.”]⁴⁴

On the one hand, the generic cultural comparison adopts a global perspective by generically referring to the “Erdball” [“Globe”] and to the “Geschichte unseres Geschlechtes” [“history of our species”]. On the other hand, it stresses the cultural differences between different regions of the world without naming these regions in particular. One of the few explicitly global terms of the corpus, so it appears, is often accompanied by a reference to the regional diversity of the planet.

The last term on my search word list further proved that comparative text strategies produce a global perspective even when the lexical features of the text carry strictly regional connotations. The term “erdstrich*” [zone, part of the earth], which occurs 69 times in the corpus and lexically refers to specific regions of the planet, is seldom used in an exclusively regional context. In fact, according to my analysis the term is only used as a strictly regional term seven times. In all other cases, the contexts show a surprisingly consistent structure, as demonstrated in the following excerpts:

“Es ist bekannt, wie häufig die Leuchtwürmer in **Italien** und im ganzen **mit-täglichen Europa** sind; aber ihr malerischer Eindruck ist gar nicht zu vergleichen mit den zahllosen zerstreuten, sich hin und her bewegenden Lichtpunkten, welche im **heißen Erdstrich** der Schmuck der Nächte sind, wo einem ist, als ob das Schauspiel, welches das Himmelsgewölbe bietet, sich auf der Erde, auf der ungeheuren Ebene der Grasfluren wiederholte.”

44 A. v. Humboldt, *Reise*, vol. 4, 52. [A. v. Humboldt, *Personal Narrative*, vol. 5, 500.]

[“We know how common the glow-worm is in **Italy**, and in all the **south of Europe**; but the picturesque effect it produces cannot be compared to those innumerable, scattered, and moving lights, that embellish the nights of the **torrid zone**, and seem to repeat on the earth, along the vast extent of the savannahs, the spectacle of the starry vault of the sky.”]⁴⁵

“Es ist auffallend, wie in den **heißesten** und in den **kältesten Erdstrichen** der gemeine Mann gleich sehr die Wärme liebt.”

[“It seems remarkable, that in the **hottest** as well as the **coldest climates**, people display the same predilection for heat.”]⁴⁶

“Hier im **tropischen Erdstrich** wachen sie auf, wenn es wieder feuchter wird; dagegen in **Georgien** und in **Florida**, im **gemäßigten Erdstrich**, reißt die wieder zunehmende Wärme die Tiere aus der Erstarrung oder dem Zustande von Nerven- und Muskelschwäche, in dem der Atmungsprozeß unterbrochen oder doch sehr stark beschränkt wird.”

[“Here, in the **equinoctial zone**, it is the increase of humidity that recalls them to life; while in **Georgia** and **Florida**, in the **temperate zone**, it is the augmentation of heat, that rouses these animals from a state of nervous and muscular debility, during which the powers of respiration are suspended, or singularly diminished.”]⁴⁷

“Die **nordischen Heiden**, die Steppen an **Wolga** und **Don** sind kaum ärmer an Pflanzen und Tierarten als unter dem herrlichsten Himmel der Welt, im **Erdstrich der Bananen und des Brotfruchtbaums**, 567000 qkm Savannen, die im Halbkreise von Nordost nach Südwest, von den Mündungen des Orinoko bis zum Caqueta und Putumayo sich fortziehen.”

[“The **heaths of the north**, the steppes of the **Wolga** and the **Don**, are scarcely poorer in species of plants and animals, than are twenty-eight thousand square leagues of savannahs, that extend in a semicircle from north-east to

45 A. v. Humboldt, *Reise*, vol. 1, 189. [A. v. Humboldt, *Personal Narrative*, vol. 2, 249.]

46 A. v. Humboldt, *Reise*, vol. 2, 233. [A. v. Humboldt, *Personal Narrative*, vol. 4, 196.]

47 A. v. Humboldt, *Reise*, vol. 3, 61. [A. v. Humboldt, *Personal Narrative*, vol. 4, 501.]

south-west, from the mouths of the Oroonoko to the banks of the Caqueta and the Putumayo, beneath the finest sky of the globe, and in the **climate of plantains and breadfruit trees.**”]⁴⁸

The term “erdstrich” [zone, part of the earth] is usually used in a context of global comparison, again, adding a global perspective to the text. Methodologically, reading through these comparisons provided me with a number of new search words for a potential follow-up study, such as: zone* [zone], breite* [latitude], länge* [longitude], *west* [west], *ost* [east], *süd* [south], *nord* [north], heiße* [hot, torrid], kalte* [cold], or gemäßigte* [temperate]. The examples also pointed to the possibility of searching the corpus for specific names of countries or continents although such an analysis would require a lengthy process of preselecting possibly significant names – a process which in turn heavily depends on extensive knowledge of the text acquired through close reading methods. However, it is important to note that following the term “erdstrich” through the corpus revealed a rich practice of global comparing in the corpus which would not have been indicated by its lexical features.

In sum, the close reading analysis proved to be a crucial step to evaluate and complement the results of the distant reading process. First of all, it helped to determine the actual connotation of search terms as used in the corpus, for example when differentiating between the different meanings of the term “erde” [earth, soil, ground]. Beyond that it was necessary to trace comparative text strategies throughout the entire length of the corpus – text strategies that produce a sense of the global beyond the regional focus that the text displays on a lexical level.

Using a reference corpus: degrees of globality

The observation that the corpus shows very different ways of engaging with the world on different levels raises some hermeneutical questions that call for further investigation. Are the lexical features of the corpus actually that significant when most engagement with the world happens on the level of text strategies? Can we come to a viable first interpretation of the corpus by

48 A. v. Humboldt, *Reise*, vol. 4, 264. [A. v. Humboldt, *Personal Narrative*, vol. 6, 70 f.]

artificially isolating these lexical features from their specific contexts as I did in the first step of this analysis? Comparing the results to a reference corpus confirmed the validity of such an approach, proving that even on a lexical level a corpus can indeed show varying degrees of engagement with the world.

I chose Humboldt's *Kosmos*⁴⁹ as a reference corpus, since the text explicitly focuses on a description of the world, on a "physische Weltbeschreibung," as indicated by the subtitle. That way I hoped to compare the travelogue which rather focuses on the voyage and specific places to a text that primarily focuses on the world in its entirety. The aim was to see whether the two corpora engaged with the world to a different extent according to their different areas of focus. As a first step, I used the keyword list to identify characteristic words in the *Reise*-corpus. Whereas the word list tool only counts the words in a corpus, the keyword list tool shows which words are unusually frequent or infrequent in comparison with the words in a reference corpus (in this case the *Kosmos*-corpus). In other words, the tool compares two corpora to each other, identifying keywords with regard to the frequency of the words in the two corpora.⁵⁰

Using this tool for an analysis of the *Reise*-corpus showed that many of the terms that ranked highly in the initial word list reappeared in the keyword list, as can be seen in the following tables:

49 Humboldt, Alexander von, *Kosmos: Entwurf einer physischen Weltbeschreibung*, 5 vols., Stuttgart/Augsburg: Cotta, 1845–1862. The text is loosely based on a number of lectures held in Berlin in 1827 and 1828. It addresses a wide variety of topics, such as cosmic nebulas, stars, volcanos, plant and animal life, or human history. Overall, it aims to give a survey of all the physical phenomena found in the world and in the universe.

50 For a more distinguished introduction to and reflection on key word analysis in historical corpora see Baron, Alistair/Rayson, Paul/Archer, Dawn, Word frequency and key word statistics in historical corpus linguistics, in: *International Journal of English Studies* 20 (2009), 41–67.

Table 4: Word list for the Reise-corpus & Table 5: Keyword list for the Reise-corpus

Word List Results 1				Keyword List Results 1				
Word Types: 35313		Word Tokens: 218084	Search Hits: 0	Keyword Types: 1679	Keyword Tokens: 94987		Search Hits: 0	
Rank	Freq	Word	Lemmas Word Form(s)	Rank	Freq	Keyness (LL4)	Effect (DICE)	Keyword
1	1107	orinoko		1	1107	+ 2065.38	0.0101	orinoko
2	950	rio		2	644	+ 1200.71	0.0059	waller
3	746	m		3	950	+ 1175.29	0.0087	rio
4	661	großen		4	574	+ 939.08	0.0052	indianer
5	655	mehr		5	420	+ 782.81	0.0038	falt
6	644	waller		6	390	+ 726.86	0.0036	km
7	574	indianer		7	350	+ 652.27	0.0032	küste
8	541	ganz		8	346	+ 644.81	0.0032	insel
9	524	luft		9	412	+ 610.9	0.0038	caracas
10	513	zeit		10	746	+ 590.28	0.0068	m
11	492	san		11	310	+ 577.69	0.0028	felbt
12	478	weit		12	337	+ 547.39	0.0031	negro
13	463	jahre		13	283	+ 527.35	0.0026	millionen
14	447	boden		14	376	+ 508.43	0.0034	cumana
15	436	zwei		15	260	+ 484.48	0.0024	teil
16	420	falt		16	256	+ 477.02	0.0023	fieht
17	412	caracas		17	492	+ 466.81	0.0045	san
18	403	kleinen		18	254	+ 461.22	0.0023	gibt
19	390	km		19	245	+ 456.52	0.0022	reife

The initial word list already included many terms referring to spatiality, such as the names of specific rivers or places (“orinoko” and “caracas”), particles used in the names of rivers and places (“rio” and “san”), or metric markers (“m” and “km”). The keyword list confirmed that such a focus on local or regional space is specific to the travelogue, pointing to a number of terms that do not only occur frequently but can be considered as keywords in comparison to another corpus, in this case Humboldt’s *Kosmos*. High-ranking keywords included the aforementioned words but also others, such as “küste” [coast], “insel” [island], “cumana” [Cumaná], and “reise” [voyage], pointing to the travelogue’s specific focus on movement through regional space.

The second step was to swap the corpora, analyzing the *Kosmos*-corpus as the main corpus with the travelogue as the reference corpus. Both the word list and the keyword list rendered significant results:

Table 6: Word list for the Kosmos-corpus & Table 7: Keyword list for the Kosmos-corpus

Word List Results 2					Keyword List Results 2				
Word Types: 51423		Word Tokens: 335387	Search Hits: 0		Keyword Types: 1432		Keyword Tokens: 113493	Search Hits: 0	
Rank	Freq	Word	Lemmas Word Form(s)		Rank	Freq	Keyness (LL4)	Effect (DICE)	Keyword
1	1020	schon			1	1020	+ 776.01	0.0061	schon
2	938	fuß			2	938	+ 629.81	0.0056	fuß
3	874	großen			3	621	+ 622.6	0.0037	fast
4	766	worden			4	628	+ 616.6	0.0037	kosmos
5	732	erde			5	544	+ 545.35	0.0032	vergl
6	692	a			6	606	+ 457.8	0.0036	planeten
7	635	mehr			7	449	+ 450.07	0.0027	erscheinungen
8	628	kosmos			8	443	+ 444.05	0.0026	theil
9	621	fast			9	377	+ 377.86	0.0022	ersten
10	606	planeten			10	370	+ 370.84	0.0022	erst
11	575	beobachtungen			11	366	+ 366.83	0.0022	theile
12	572	zeit			12	363	+ 363.83	0.0022	cometen
13	558	höhe			13	362	+ 362.82	0.0022	insel
14	557	ganz			14	345	+ 345.78	0.0021	zuerst
15	554	große			15	334	+ 334.75	0.002	herschel
16	544	vergl			16	376	+ 332.9	0.0022	cap
17	536	sterne			17	318	+ 318.71	0.0019	wahrscheinlich
18	535	jahre			18	329	+ 318	0.002	anm
19	530	sonne			19	310	+ 310.69	0.0018	sagt

In comparison to the travelogue and with regard to my interest in spatiality and globality, some of the high-ranking words can be considered especially significant. Both lists include a number of words referring to the universe and celestial bodies in particular. The word list refers to frequently occurring words, such as “kosmos” [cosmos, universe], “planeten” [planets], “sterne” [stars], “sonne” [sun], and “erde” [earth, soil, ground],⁵¹ whereas the keyword list points towards similar terms that occur especially frequently in comparison to the travelogue, again including the words “kosmos” [cosmos, universe] and “planeten” [planets] but also adding “cometen” [comets] to the list. This lexical focus on cosmic space and celestial bodies seems almost complementary to the regional focus of the travelogue.

To test if this difference in focus affected the lexical use of the word “welt” [“world”] in the *Kosmos*-corpus, I searched for “welt*” in the concordance and clusters tools. Again, the analysis revealed significant differences between the corpora. While the term in all its variations occurs 1306 times in the *Kosmos*, it mostly occurs in the form of those compound nouns that hardly ever occur in the travelogue:

51 Again, only a close reading analysis of the results could determine whether the term is actually used with reference to the planet or the soil. Nonetheless, even if the term had to be excluded from the results, the list would still be considerable.

Table 8: Results of the cluster analysis for “welt*”, cluster size of one word (Kosmos-corpus) & Table 9: Results of the cluster analysis for “welt*”, cluster size of two words (Kosmos-corpus)

Clusters Results 1					Clusters Results 2				
Total No. of Cluster Types 176				Total No. of Cluster Tokens 1306	Total No. of Cluster Types 623				Total No. of Cluster Tokens 1306
Rank	Freq	Range	Cluster		Rank	Freq	Range	Cluster	
1	134	5	welt		1	37	5	physischen weltbeschreibung	
2	133	5	weltkörper		2	32	5	der weltanschauung	
3	79	5	weltbeschreibung		3	28	5	der weltkörper	
4	67	5	weltanschauung		4	27	5	die welt	
5	55	4	weltraum		5	26	3	den weltraum	
6	37	4	weltraume		6	25	4	des weltraums	
7	35	4	weltkörpern		7	23	4	im weltraume	
8	27	5	weltall		8	22	3	der welt	
9	27	4	weltraums		9	19	3	der vorwelt	
10	26	2	weltherrschaft		10	19	5	im weltall	
11	24	3	außenwelt		11	19	3	physische weltbeschreibung	
12	21	3	weltansicht		12	18	2	physischen weltanschauung	
13	20	4	tropenwelt		13	16	3	der außenwelt	
14	19	3	vorwelt		14	14	3	der tropenwelt	
15	17	3	inselwelt		15	14	2	des weltbaues	
16	17	4	weltumsegung		16	13	3	neuen welt	
17	16	3	weltganzen		17	12	4	des weltalls	
18	16	4	weltordnung		18	10	2	dem weltraume	
19	15	2	weltbaues		19	10	5	der gedankenwelt	

A number of the high-ranking words refer to the cosmic or planetary dimension of the term “welt”, such as “weltkörper*” [heavenly body], “weltraum*” [space, outer space], or “weltall*” [space, universe], while others refer to the term in a rather philosophical or political way, as in the cases of “weltanschauung*” [world view], “weltherrschaft*” [world domination], “weltansicht*” [word view], or “weltordnung*” [world order].⁵² It is important to note that the terms which refer to the world in its totality also appear as most frequent ones when changing the cluster size to two words. The regional term “neuen welt” [new world], which is the highest ranking two word cluster in the travelogue, only ranks 16th with 13 occurrences in this corpus, seem-

52 The listed terms imply Ottmar Ette’s observation that Humboldt’s world concepts include both philosophical and planetary connotations. See O. Ette, *Humboldt und die Globalisierung*, 366 ff.; O. Ette, *Weltbewußtsein*, 90 ff. As mentioned above, compound nouns as these cannot be translated easily because they tend to carry very specific philosophical and epistemological connotations. The translations provided above can merely be considered suggestions that help the reader navigate through the analysis. To provide historically accurate translations, I would have had to track those terms through the different English translations of the time, which, with regard to the number of hits, would have exceeded the limits of this case study.

ing almost insignificant in comparison to the high-ranking compound noun clusters. These results show that the word “welt” is strongly associated with the universal and cosmic focus of the text, in contrast to the regional use of the term in the travelogue.

For a more comprehensive comparison of the two corpora, some further distant and close reading analysis of the above mentioned planetary terms, including the term “erde” [earth], would surely be necessary. However, even these preliminary results point to the hermeneutic productivity of comparing the distant reading results of two corpora. While previous research has strongly associated Humboldt’s claim to describe the world in its totality with specific ways of writing, such as descriptive, narrative, or comparative writing,⁵³ this analysis shows that even on a lexical level texts can engage with the world to different degrees. While the travelogue only displays such an engagement with the world on the level of comparative and relational text strategies, the lexical level displaying a focus on regional space, the *Kosmos*’s engagement with the world (and the universe) starts at a lexical level. It could be argued that such a corpus shows a higher degree of *Welthaltigkeit* since it approaches world concepts more explicitly. A genre such as travel writing appears to take a more implicit approach, achieving a global perspective through comparing regional entities to each other and thereby promoting a rather relational concept of the world.

Whether or not specific degrees of globality are characteristic for certain genres – travel writing is only one possible object of investigation – can only be decided by text mining much larger corpora than the ones in this case study. Nonetheless, in this case combining distant and close reading as well as comparing distant reading results concerning different corpora, has proven a productive way to ask about the different ways in which literary texts can engage with the world or refrain from doing so.

53 C. Peters, *Reisen und Vergleichen*; C. Peters, *Historical Narrative versus Comparative Description*; O. Ette, *Weltbewußtsein*, 158 ff.; Lubrich, Oliver, *Das Schwinden der Differenz: Postkoloniale Lektüren, Alexander von Humboldt – Bram Stoker – Ernst Jünger – Jean Genet*, Bielefeld: Aisthesis, 2004, 87 ff.; Kraft, Tobias, *Figuren des Wissens bei Alexander von Humboldt: Essai, Tableau und Atlas im amerikanischen Reisewerk*, Berlin/Boston: De Gruyter, 2014, 15 ff.

Bibliography

- Anthony, Laurence*, A critical look at software tools in corpus linguistics, in: *Linguistic Research* 30 (2013), 141–161.
- Archer, Dawn*, Data Mining and Word Frequency Analysis, in: Gabriele Grif-fin/Matt Hayler (eds.), *Research Methods for Reading Digital Data in the Digital Humanities*, Edinburgh: Edinburgh University Press, 2016, 72–92.
- Baron, Alistair/Rayson, Paul/Archer, Dawn*, Word frequency and key word statistics in historical corpus linguistics, in: *International Journal of English Studies* 20 (2009), 41–67.
- Böhme, Hartmut*, Ästhetische Wissenschaft: Aporien der Forschung im Werk Alexander von Humboldts, in: Ottmar Ette (ed.), *Alexander von Humboldt: Aufbruch in die Moderne*, Berlin: Akademie-Verlag, 2001, 17–32.
- Daum, Andreas*, Alexander von Humboldt, die Natur als 'Kosmos' und die Suche nach Einheit: Zur Geschichte von Wissen und seiner Wirkung als Raumgeschichte, in: *Berichte zur Wissenschaftsgeschichte* 22 (2000), 246–250.
- Erhart, Walter*, Chamisso's Weltreise und Humboldts Schatten, in: Julian Drews et al., *Forster – Humboldt – Chamisso: Weltreisende im Spannungsfeld der Kulturen*, Göttingen: V&R unipress, 2017, 13–34.
- Erlin, Matt*, Topic Modeling, Epistemology, and the English and German Novel, in: *Journal of Cultural Analytics* (2017).
- Ette, Ottmar*, Languages about Languages: Two Brothers and one Humboldtian Science, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* XIX (2018), 47–61.
- Ette, Ottmar*, Humboldt und die Globalisierung: Das Mobile des Wissens, Frankfurt a. M./Leipzig: Insel, 2009.
- Ette, Ottmar*, Unterwegs zu einer Weltwissenschaft? Alexander von Humboldts Weltbegriffe und die transarealen Studien, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* VII (2006), 34–54.
- Ette, Ottmar*, Weltbewußtsein: Alexander von Humboldt und das unvollendete Projekt einer anderen Moderne, Weilerswist: Velbrück, 2002.
- Ette, Ottmar*, The Scientist as Weltbürger: Alexander von Humboldt and the Beginning of Cosmopolitics, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* II (2001).

- Ette, Ottmar*, Unterwegs zum Weltbewußtsein: Alexander von Humboldts Wissenschaftsverständnis und die Entstehung einer ethisch fundierten Weltanschauung, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* I (2000).
- Fuchs, Anne*, Reiseliteratur, in: Dieter Lamping (ed.), *Handbuch der literarischen Gattungen*, Stuttgart: Alfred Kröner, 2009, 593–600.
- Görbert, Johannes*, Die Vertextung der Welt: Forschungsreisen als Literatur bei Georg Forster, Alexander von Humboldt und Adelbert von Chamisso, Berlin: De Gruyter, 2014.
- Heyl, Bettina*, Das Ganze der Natur und die Differenzierung des Wissens: Alexander von Humboldt als Schriftsteller, Berlin/Boston: De Gruyter, 2007.
- Humboldt, Alexander von*, *Personal Narrative of Travels to the Equinoctial Regions of the New Continent, during the Years 1799–1804*, 7 vols., translated by Helen Maria Williams, New York: Cambridge University Press, 2011.
- Humboldt, Alexander von*, *Reise in die Aequinoktial-Gegenden des neuen Kontinents*, vol. 1., translated by Hermann Hauff, Stuttgart: Cotta, 1859.
- Humboldt, Alexander von*, *Kosmos: Entwurf einer physischen Weltbeschreibung*, 5 vols., Stuttgart/Augsburg: Cotta, 1845–1862.
- Knobloch, Eberhard*, Alexander von Humboldts Weltbild, in: *HiN: Internationale Zeitschrift für Humboldt-Studien* X (2009), 34–46.
- Kraft, Tobias*, Erdwissen im Angesicht der Berge: Die Vulkanlandschaften der Jorullo-Ebene als Heuristik der Geologie, in: Ottmar Ette/Julian Drews (eds.), *Horizonte der Humboldt-Forschung: Natur, Kultur, Schreiben*, Hildesheim: Georg Olms, 2016, 97–124.
- Kraft, Tobias*, *Figures des Wissens bei Alexander von Humboldt: Essai, Tableau und Atlas im amerikanischen Reisewerk*, Berlin/Boston: De Gruyter, 2014.
- Lubrich, Oliver*, Das Schwinden der Differenz: Postkoloniale Lektüren, Alexander von Humboldt – Bram Stoker – Ernst Jünger – Jean Genet, Bielefeld: Aisthesis, 2004.
- Moser, Christian/Simonis, Linda*, Einleitung: Das globale Imaginäre, in: Sebastian Moser/Linda Simonis (eds.), *Figures des Globalen: Weltbezug und Welterzeugung*, Göttingen: V&R unipress, 2014, 11–22.
- Peters, Christine*, Historical Narrative versus Comparative Description? Genre and Knowledge in Alexander von Humboldt's *Personal Narrative*,

in: Martin Carrier/Carsten Reinhardt/Veronika Hofer (eds.), *Narratives and Comparisons: Adversaries or Supporters in Understanding Science?*, (manuscript in preparation).

Peters, Christine, Reisen und Vergleichen: Praktiken des Vergleichens in Alexander von Humboldt's *Reise in die Äquinoktial-Gegenden des Neuen Kontinents* und Adam Johann von Krusensterns *Reise um die Welt*, in: *IASL* 42 (2017), 441–455.

Stockhammer, Robert, Welt oder Erde? Zwei Figuren des Globalen, in: Christian Moser/Linda Simonis (eds.), *Figuren des Globalen. Weltbezug und Welterzeugung in Literatur, Kunst und Medien*, Göttingen: V&R unipress, 2014, 47–72.

