

Tracing Changes in Thematic Structure of Holiday Picture Postcards from 1950s to 2010s

Kyoko Sugisaki, Nicolas Wiedmer, Marcel Naef, Heiko Hausendorf

German department, University of Zurich

Schönberggasse 9, 8001 Zurich

Switzerland

{sugisaki,nicolas.wiedmer,marcel.naef,heiko.hausendorf}@ds.uzh.ch

Abstract

In this paper, we present our study of the changes of thematic structures in holiday picture postcards from 1950s to 2010s. We use over 1,000 cards that we annotated manually with thematic information and apply a clustering method (principal component analysis, PCA) to analyse the thematic structure. The primary objective of our study is to group cards with similar thematic structure and to analyse changes of themes over the decades. Our PCA analyses indicate that holiday postcards have been changed in terms of (1) thematic structure, (2) function of text, and (3) language patterns of the speech act ‘greeting’.

1 Introduction

In this paper, we introduce a novel approach to the frame-semantic annotation of thematic structures from the point of view of text linguistics and provide a data-driven analysis on the development of holiday picture postcards over decades.

So far, the annotation of theme in corpora has been carried out mainly based on information structure, such as the Prague Treebank (Hajič, 1998) and the Potsdam Commentary Corpus (Stede and Mamprin, 2016). In our work, the term *theme* is distinguished from the notion of *topic* in information structure, that is, ‘aboutness’ and ‘old/given entity’. In information structure, the topic is determined mainly by its syntactic position in a sentence and by the salience of its discourse in relation to the entities mentioned in the previous sentence(s). In contrast, in our work *theme* is rather a semantic frame that constitutes the thematic coherence of a certain genre of text (holiday picture postcards). We use the term *semantic frame* in the sense of Busse (2012, p. 563)

who defines the frame as a structure of knowledge, in which the core of a frame (theme) is connected to the constituents of knowledge. Depending on the context of a concrete situation, possible constituents vary. These constituents define the conditions of the realisation of textual phrases. In the case of holiday postcards, the theme of the frame is *to be on holiday* and the constituents of knowledge (i.e. slots) are possible ways to be filled with actual text (i.e. fillers) according to the concrete situation of writing a holiday postcard. In this work, we define a set of slots for postcards that report vacation experiences (cf. Section 3.1)

We will first describe the corpus of postcards (Section 2) and then characterise the texts with regard to thematic structure before presenting our annotation schema and discussing the annotation process (Section 3). Finally, we use a clustering method to analyse our annotated postcards and present our results (Section 4).

2 Data source: Postcard corpus

The holiday picture postcard corpus ANKO (*An-sichtskartenkorpus* ‘picture postcard corpus’) consists of 12,337 cards written in Standard German (95%, 11,760 cards, 582,675 tokens) and in Swiss German (5%, 577 cards) from 1898 to the present day. They were collected in Zurich, Switzerland from 2009 to 2017. The postcards included in our corpus are only cards that were sent from vacation. The postcards were sent from private individuals by post to Switzerland mainly from Switzerland, Italy, Germany and other European countries. In the corpus, paragraphs, sentences and tokens are segmented in a XML representation (cf. Sugisaki et. al (2018)).

| # | Question | Class | Example of a possible answer in text |
|---|--|----------------------------|--|
| A | Is anything except holiday thematised? | Extra-diegetic | Thank you for your card. |
| B | How did I travel to the holiday location? How do I travel home? | Outward and return journey | The flight to Frankfurt was terrible. |
| C | How is the weather? | Weather | The weather is fabulous! |
| D | How/where do I stay? | Accommodation | We stay in a camping place near Amsterdam. |
| E | What/where do I eat or drink? | Eating and drinking | Martin eats pizza every day! |
| F | Who did I meet on holiday? | Meeting new people | We met some Italian guys in Rome and we hang around a lot. |
| G | What do I do? | Activity | Yesterday, we visited a lot of churches in Florence. |
| H | Did something unexpected happen? | Happenings | Unfortunately, we had a car accident in Spain, so that we gave up our road trip. |
| I | Where am I on my holiday? | Location | We are now in Ibiza with the kids. |
| J | What do I know about the holiday place? | Knowledge | The romans built this city about 2000 years ago. |
| K | What kind of holiday do I take? | Type | Greetings from our hiking-trip! |
| L | What do I want to achieve on my holiday? | Reason | I always wanted to learn Italian, so now I'm taking a course in Rome. |
| M | How do I feel? | Feeling | We really enjoy our vacation in Italy :) |
| N | What can I see/hear in my holidays? | General | There are so many lavender fields here. |

Table 1: Annotation scheme of thematic structures in postcards

3 Theme annotation

We manually annotated the core thematic structures in the postcards. The text of the postcards was generally structured as follows: 1) a preface that contains the date, sometimes the location (e.g. *Laax, 20/12/1977*); 2) a salutation (e.g. *Dear Mr. & Mrs. Smith*); 3) the message; 4) greetings (e.g. *greetings from Paris*); 5) the signature of sender(s). The thematic annotations concerned only 1), 3) and 4). No thematic information was found in the salutations 2) or the signatures 5).

In the following section, we describe in detail the subcategorization and annotation process.

3.1 Developing the annotation scheme

Our primary goal of the annotation presented in this paper is to find the core thematic structures of the postcards and their development over time. In reporting holiday experiences, postcards exhibit a handful of thematic patterns. These thematic patterns have been formed and remained over time because the postcards fulfilled the main purpose of this type of text, which is the function of contact (in the term of Hausendorf and Kesselheim (2008, p. 154ff) or Brinker et. al (2014, p. 118ff)). Specifically, this function is to maintain personal contact during holidays. In other words, the thematic structures of postcards were conventionalised and standardised over time by fulfilling the communicative needs of holidayers. Of course, some variations were caused by the social changes and the use of postcards as a mean of communication. Therefore, we consider that the categories of the thematic structures that we annotated in this study could be super themes that might remain consistent over time (cf. Hausendorf and Kesselheim (2008, p. 103), Hausendorf (2008, p. 333),

Hausendorf (2009, p. 13)).

To develop an annotation schema of the thematic structures in postcards, we first determined a set of main thematic categories based on the observation of hundreds of postcards. We then tested this initial schema with 14 test participants in order to refine and extend it and then to produce the final annotation schema. In the following sections, these two steps are described in detail.

3.1.1 Defining core thematic frames

To identify the categories of semantic frames in postcards, we first defined a set of questions that can be answered in text (Ziem, 2008, p. 94f). In other words, we assume that every sentence of a postcard can be read as an answer to at least one question shown in Table 1.

We divided the frame categories into two classes: 1) about the holiday (B-N in Table 1) 2) *not* about the holiday (A in Table 1). The first category is subcategorised into semantics-oriented themes. The thematic categories do not refer to individual topic entities (e.g. *snow*, *rain* and *wind*) but to the super categories of such entities (e.g. *weather*). The super category *weather*, for example, was frequently thematised in postcards. Therefore, we concluded that the weather is an important element in the frame of being on holidays. Similarly, *eating and drinking*, *meeting new people* and *accommodation* belong to this category of relevance. Furthermore, we observed that the postcards reported what the writer would do, was doing or did on the holiday (the category of *activity*). While this category includes comments on events carried out intentionally by holidayers (e.g. hiking, skiing and dancing), the category of *happenings* refers to unexpected and unintended events (e.g. car accidents, illness and

lost baggage). In addition, the postcards often began with descriptions of where the writers were (the category of *location*) with or without explanatory comments on holiday places (the category of *knowledge*), and why they were in that holiday location (the categories of *type* and *reason*). For example, the *type* of holiday could be a school trip, a shopping trip or a ski vacation, all of which are holiday prototypes. In contrast, the category of *reason* concerns what the writer wants to achieve on holiday. Treatment in a sanatorium (body fitness as scope), and language holidays abroad (language learning as scope) are prototypical in this category. Moreover, the postcard writers described their holiday with an emphasis on their emotional state (the category of *feeling*) or without any reference to emotion, they focussed on what they saw and heard on their holiday (the category of *general*). Finally, we created the extra category of *outward and return journey*, which refers to the journey to and from the holiday location. This category includes events that were not directly related to the holiday location but were part of the holiday experience.

In our annotation scheme, the thematic unit is a sentence. Compared to words, phrases and paragraphs, sentences are ideal units for thematic analysis because each question to be answered in the text contains a proposition. However, a sentence can contain more than one proposition because of coordination (cf. sentence 1) and the inherent semantic property of categories (cf. sentence 2). Therefore, each sentence is annotated as belonging to one or more thematic categories.

- (1) Frame category of *activity* and *eating and drinking*:

Jetzt gehen wir Ski fahren und nachher Appenzeller Fondue essen.

‘Now we are going to ski and then we will eat cheese fondue á la Appenzell’.

- (2) Frame category of *accommodation* and *location*:

Unser Hotel ist in der Nähe vom Genfersee.

‘Our hotel is near Lake Geneva’.

3.1.2 Testing the core thematic categories

In order to test the robustness and the comprehensibility of our thematic annotation scheme, we conducted a study at the University of Zurich in

| # | Class | Pre | Rec | MCC |
|---|---------------------|-------|--------|-------|
| A | Extra-diegetic | 88.61 | 93.23 | 89.59 |
| B | Journey | 92.86 | 100.00 | 96.30 |
| C | Weather | 97.76 | 90.34 | 93.40 |
| D | Accommodation | 88.89 | 84.21 | 86.19 |
| E | Eating and drinking | 88.89 | 94.12 | 91.17 |
| F | Meeting new people | 84.62 | 82.50 | 83.13 |
| G | Activity | 96.92 | 92.99 | 93.91 |
| H | Happenings | - | - | - |
| I | Location | 96.24 | 86.72 | 88.79 |
| J | Knowledge | 73.21 | 98.80 | 84.13 |
| K | Type | 38.89 | 50.00 | 43.53 |
| L | Reason | 73.33 | 78.57 | 75.68 |
| M | Feeling | 81.03 | 97.92 | 87.90 |
| N | General | 75.29 | 81.01 | 76.90 |
| O | X | 92.93 | 98.90 | 93.60 |

Table 2: Precision, recall and MCC (in percent) for the preliminary study with students

which 14 linguistics students individually annotated 12 cards. After a 45-minute briefing session about the annotation scheme, they were provided with a MS Excel sheet in which each sentence was displayed in a cell. The students then assigned the categories shown in Table 1 to the sentences. The category of *happenings* (H) was not part of the tag set at that moment because it is the result of this study. Furthermore, the extra category of ‘X’ was provided in the case that the students did not find any of the categories suitable for a unit. We then compared the thematic categories assigned by the students to a gold standard that was created by our four internal annotators.

The students’ overall annotation precision ranged from 83.89% to 98.58% (average: 93.30%) with recall between 85.93% and 97.20% (average: 92.61%). The students’ overall scores were satisfying considering the short instruction time. However, there were remarkable differences with regard to precision and recall in some categories. We summarised the results as shown in Table 2. As the balanced score, we used the Matthews correlation coefficient (MCC) instead of the kappa coefficient to account for the differences in frequency between the categories (cf. Powers (2012)). Table 2 shows the problematic thematic categories of *accommodation*, *location*, *knowledge*, *type*, *reason*, *feeling* and *general*. We discussed the results with the students and came to the conclusion that the relatively low recall of the categories *accommodation* and *location* could be explained by misunderstandings in the briefing session. For example, the students often did not

assign the category *location* if the location was not mentioned in the message but in the preface (e.g. ‘Paris, 07.08.1966’). Based on the discussion, we created an annotation guideline with definitions of the categories and examples of contentious cases.

3.2 Manual annotation

Based on the study described in the previous section, we carried out a sentence-based multiple-class thematic annotation with 14 categories (cf. Table 1) and 1,120 postcards. The cards were selected from our Standard German postcard corpus using a random sampling strategy with a fixed number (160) of cards for each of the 7 decades between the 1950s and 2010s. This sampling method ensured that the range of cards would include those from less frequent decades in the corpus.

The manual annotation was carried out in three steps. First, two linguistics students (annotator A and B) were asked to assign one or more predefined thematic category to each sentence displayed on a Microsoft Excel sheet similar to the preliminary study described in the previous section. In addition to the questions shown in Table 1, the annotation guideline summarised in section 3.1.1 was handed to them. Each postcard was annotated by one of the two annotators, who, if they were not sure, noted a comment. After this first round of the annotation, two additional annotators (annotator C and D) discussed problematic sentences in the annotation process, and they jointly decided which thematic categories were to be chosen.

However, these two steps were not sufficient to obtain a highly consistent annotation. The problem was that some categories were not clearly distinguishable from others, which led to the result that the first two annotators (annotator A and B) often did not agree on the categories of *general*, *feeling* and *knowledge*. Our approach was to assign the thematic categories that best answered the questions (cf. Table 1), which, however, allowed room for interpretations of the annotators. For example, the sentence, ‘The beach is really wonderful.’ was seen as an answer to the question, ‘How did they feel on their holiday?’ (the category of *feeling*) by annotator A, and as an answer to the question ‘What was the holiday place like?’ (the category of *general*) by annotator B. For this reason, we defined a set of lexical items for the cat-

egories of *feeling* and *knowledge*. For example, *geniessen* (‘enjoy’), *gut* (‘good’), *schlecht* (‘bad’), *wunderbar* (‘wonderful’) are lexical items that express *feeling*. They express the opinion, evaluation, and emotional state of the writer. In contrast, *man*, *alle*, *hier* ‘one, all, here’ are lexical items for the category of *knowledge*. They demonstrate the general knowledge, including stereotypical prejudices, of the writer. However, the occurrence of a lexical item is not a definite criterion for placing a sentence in a certain category. Thus, these two categories were revised by examining the lexical items and their adequacy in the context of every single sentence. For the category of *general*, we did not define a set of lexical items. Instead, it was chosen whenever the writer gave a clear description of something he or she could see or hear (e.g. ‘The beach is really dirty.’). After having defined lexical items for the categories of *feeling* and *knowledge* and having determined the new criteria for the category of *general*, annotator A and C examined all the instances and revised the annotation jointly in the third step using these new annotation criteria.

4 Data Analysis

Based on the manual annotation of over 1,000 cards (49,261 tokens and 6,713 sentences), we analyse our annotated texts with principal component analysis (PCA). PCA is a dimension reduction method that locates underlying latent dimensions of a collection of text by ‘eliminating the covariance while preserving most of the variance in data’ (Moisl, 2015). In linguistics, PCA has been used in corpus linguistics as explanatory method, in particular, for authorship attribution or stylistics (e.g., Baayen et. al (1996)), and factor analysis for register analysis (e.g., Biber (1995)).

For the PCA analysis, we aggregated our thematic classes *reason* and *purpose* to a new class *why* and *meeting-new-people* and *eating and drinking* to *activity* to get a better result. We counted the frequency of each 10 semantic frames (without class *x*) in a text, normalised the count for 1000 words and log-transformed it.

PCA identified four principal components that account for 62.67% of the variance of 10 variables.¹ The loading is shown in Table 3. Figure

¹Kaiser-Meyer-Olkin (KMO) factor adequacy was .6, which indicates that the sampling adequacy was acceptable. The KMO values of the items range from .49 to .67, which

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------------|--------|--------|--------|--------|
| acc(ommodation) | -0.140 | | | |
| act(ivity) | -0.595 | -0.173 | 0.307 | 0.482 |
| ext(ra-diegetic) | -0.123 | 0.962 | | |
| fee(lings) | -0.574 | | | 0.136 |
| gen(eral) | -0.126 | | | |
| hap(pennings) | | | | |
| kno(wledge) | | | 0.130 | |
| loc(ation) | -0.167 | | 0.719 | -0.644 |
| out(journey) | | | | |
| wea(ther) | -0.465 | -0.136 | -0.587 | -0.523 |
| why | | | 0.116 | 0.221 |

Table 3: PCA loading - component (Comp) 1, 2, 3, and 4

1 (a) and (b) illustrate all the cards in our data set and the directions of the variables. In the following paragraphs, we go through each of these four components in detail.

Component 1 In the first component (20.78% of variance), PCA indicates that *activity*, *feelings* and *weather* are highly correlated. The card with the highest score was (A), while that with the lowest score was (B). In the card (A), only the purpose of the holiday is mentioned, whereas the major semantic frames (*weather*, *activity*, *feeling*, *general*, *knowledge*, *accommodation*, *location*) are mentioned in the card (B).

(A) The highest score:

Ausflug 19.6.93.

‘excursion 19.6.93.’

(B) The lowest score:

Zinal, 21.7.64. \\ Unsere Lieben, nach den paar strengen Stunden der Expo - es war vom Sonntag mörderisch heiss es glich einem langsamen Selbstmord - sind wir nun doch noch für ein paar Tage hier im wirklich malerischen & sehr ruhigen Bergdörfchen Zinal (Walliser Hochthal) gelandet. Im Grande Hotel des Diablons (ganz alt, aber [unclear] mit [unclear] franz. Küche) sind wir in allen Teilen gut aufgehoben. Wir ruhen uns Beide gut aus, mein [unclear] Mann hat dies nach seiner Bruchoperation nötig und ich laufe auch schon lange zum [unclear] Jetzt bricht eben ein Sturmwind

were by and large above the acceptance value (.5). Bartlett’s test of sphericity was significant ($X^2(55)=383.42, p < .001$). All four principal components had eigenvalue above 5.69, which was well above the acceptance value (1).

mit Gewitter los, hoffentl. kommt kein Dauerregen. - Frau [NN] hat sich dann wieder fest gemeldet \\ Liebe Grüsse, Ihre H. [unclear] [NN]

‘Zinal, 21.7.64. \\ Our beloved ones, after a few hours of rigor at the Expo - it was murderously hot from Sunday it was like a slow suicide - we eventually ended up here for a few days in the really picturesque and very quiet mountain village Zinal (Valais High Valley). At the Grande Hotel des Diablons (very old, but [unclear] with [unclear] French cuisine) we are in good hands in every way. We both rest well, my [unclear] husband needs this after his fracture surgery, and I have been walking for a long time to [unclear] Now a storm wind is breaking loose with thunderstorms, hopefully it will not rain constantly. - Incidentally, Mrs [NN] has again made a firm commitment \\ Best regards, yours H. [unclear] [NN]’²

We interpret the component as the standardisation of prototype themes in postcards. In Figure 2(a), we observe that postcards have gradually evolved from rather scattered and sparse semantic-thematic contents towards prototypical ones. Furthermore, the analysis indicates that holidayers more often wrote about why they are on holiday before the emergence of mass tourism.

Component 2 In the second component (15.61% of variance), our cards are clearly grouped into two clusters, mainly depending on the occurrence of the class *extra-diegetic*. Activity and weather are negatively correlated to *extra-diegetic*. The card with the highest score (A below) was about *weather*, *location*, and *activity*, while that with the lowest score (B below) was all about the addressee.

(A) The highest score:

Aus unseren bisher sehr sonnigen und warmen Wanderferien im Berner Oberland senden wir Euch herzliche Grüsse \\ Gret + Ralph [NN] Sandra + Katja

‘From our so far very sunny and warm hiking holidays in the Bernese Oberland we send you affectionate regards \\ Gret + Ralph [NN] Sandra + Katja’

²[NN] stands for family name, and [unclear] for unreadable passages.

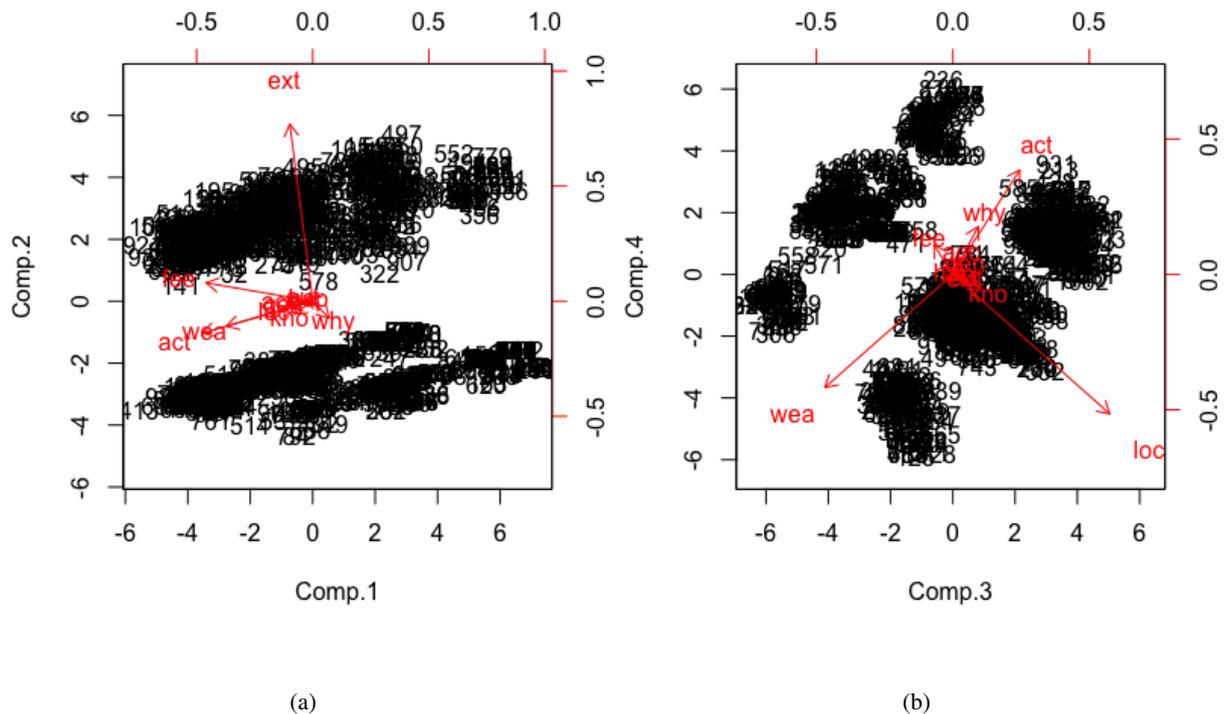


Figure 1: (a) Component 1 and 2 (b) Component 3 and 4

(B) The lowest score:

Meine liebe große Dame! Besten Dank für Ihre beiden Karten & guten Wünsche zum Geburtstag, worüber ich mich sehr freue. Sind Sie immer noch in Kreuzingen? Ich glaubte Sie wären ab 4.II. wieder im Clubhaus. - Wie geht es Ihnen - hatten Sie mit der Kur Erfolg? Ich komme Ende dieser oder Anfang nächsten Monats wieder zurück & hoffe dann auf ein Wiedersehen. Bis dahin noch weiter hin recht gute Wünsche & Liebe Grüsse auch an Herrn Sohn von Ihrer kleinen Dame.

‘My dear Grand Lady! Many thanks for your two cards & the good wishes for my birthday, that makes me very happy. Are you still in Kreuzingen? I thought you were back in the clubhouse from 4.II. - How are you - did you have success with the cure? I’ll be back at the end of this or early next month and hope to see you then. In the meantime, I send you very good wishes and best regards, also to your Mr son from your Little Lady.’

We interpret the component as changes of the

main text function of postcards. In Figure 2(a), we observe a decrease of recipient-orientation over decades. It seems that the main function of postcards has gradually shifted from a correspondence ‘how are you? I am thinking of you during my holidays’ (text function of contact) to a holiday report (‘how do I spend my holidays?’), whose text function is description.

Component 3 In the third component (14.32% of variance), PCA indicates that *location* and *activity* (slightly *knowledge* and *why*) are correlated. They are negatively correlated with *weather*. The card with the highest score was mainly about *locations* (indicating where they are in holidays) and *activities* (what they did there), while that with the lowest score was about *weather*.

(A) The highest score:

17.05.07 \\ Lieber Coni \\ Bernhard + ich befinden uns auf einer Reise von Silvanien durch Nord- griechenland, Mazedonien, Bulgarien. Wir haben die Vikos-schlucht durchwandert und am Ochridsee die Eichenwälder an den Hängen. Gestern waren wir auf dem Ohrid-See, ein Relikt aus der

Eiszeit, an der mazedonisch-albanischen Grenze.- Ganz herzliche Dank für den feinen Alpkäse u. die Güegi. Beides war über die Oster- tage bei dem vielen Besuch hoch willkommen. Mit Gruss \\ Bernhard \\ Barbara

‘17.05.07 \\ Dear Coni \\ Bernhard + I are on a journey from Sylvania through Northern Greece, Macedonia, Bulgaria. We have hiked through the Vikos Gorge and at the Lake Ochrid through the oak forests on the slopes. Yesterday we were on Lake Ohrid, a relic from the Ice Age, on the Macedonian-Albanian border. Thank you very much for the delicate Alpine cheese and the Güegi. Both were very welcome over the Easter days due to the many visitors. With regards \\ Bernhard \\ Barbara’

(B) The lowest score:

Von den viel zu kurzen, aber zu [unclear] Ferien-Tage ganz herzliche Grüße. Leider spielt das Wetter nicht mit. Sehr kalt und Regen. \\ A. [NN]

‘From the far too short, but too [unclear] holidays very affectionate regards. Unfortunately the weather is not on our side. Very cold and rain. \\ A. [NN]’

We interpret the component as changes of social- and cultural aspects in postcards. In Figure 2(b), we observe a decrease in the 1970s, the upper-bottom of the 1980s and a slight increase in the 1990s. In that period, holidayers extensively reported on weather. In 2000s and 2010s, holidayers reported with an emphasis on ‘Where are you? How many places do you visit? Why are you there (what is special about it)? What do you do there?’. We interpret that holidayers wrote about the most general topic weather in their holidays from the 1970s to 1990s. Holidayers expect to have a holiday weather in their vacation. A massive tourism might lead to the feeling of ‘one of many’ who are at the mercy of weather in holidays. Since 2000, holidayers tend to report on activity- and knowledge-oriented vacation and a round trip. Individuality and originality of a trip and travel experiences might become important for the identity of holidayers in the performance/achievement-oriented society.

Component 4 In the fourth component (11.95% of variance), PCA indicates that *location* and *weather* are highly correlated. They are negatively correlated with *activity* (slightly also with *why* and *feelings*). The card with the highest score was greeting combined with *activity*, while that with the lowest score consists of a sentence of greeting combined with *weather* and *location*. We observed this patterns in the top 10 cards in this component.

(A) The highest score:

Lieber Pius, \\ von den schönen, aber sehr sportlichen Skiferien die herzlichsten Purzelbaumgrüsse \\ [unclear] \\ Chlaus \\ Berthe [NN] Dieter

‘Dear Pius, \\ from the beautiful, but very sporty ski holidays the most affectionate somersault regards \\ [unclear] \\ Chlaus \\ Berthe [NN] Dieter’

(B) The lowest score:

Lieben Dank für Fredis Karte und herzlichen Grüße aus dem sonnigen Spanien, wo ich zwei Ferienwochen verbringe, \\ Eure Hanni

‘Thank you very much for Fredi’s card and affectionate regards from sunny Spain, where I spend two weeks of vacation, \\ Yours Hanni’

We interpret the component as a change of language patterns in the speech act ‘greeting’. In 1970s and 1980s, the greeting form combined with *weather* and *location* was more common than in other decades. Component 4 shows that *weather* and *activity* are both prototypical for greetings, but competitive semantic frames have been recurrently evoked in greetings. Again, their mention depends on time, the society and the culture.

5 Concluding remarks

In our study we showed that PCA is a suitable way to model thematic changes in holiday picture postcards over time, but also, that our annotation scheme provides an adequate basis for data driven analysis. The PCA indicated four aspects of change in postcards. Firstly, postcards have been gradually standardised with regard to themes, and evolved towards prototypical themes such as

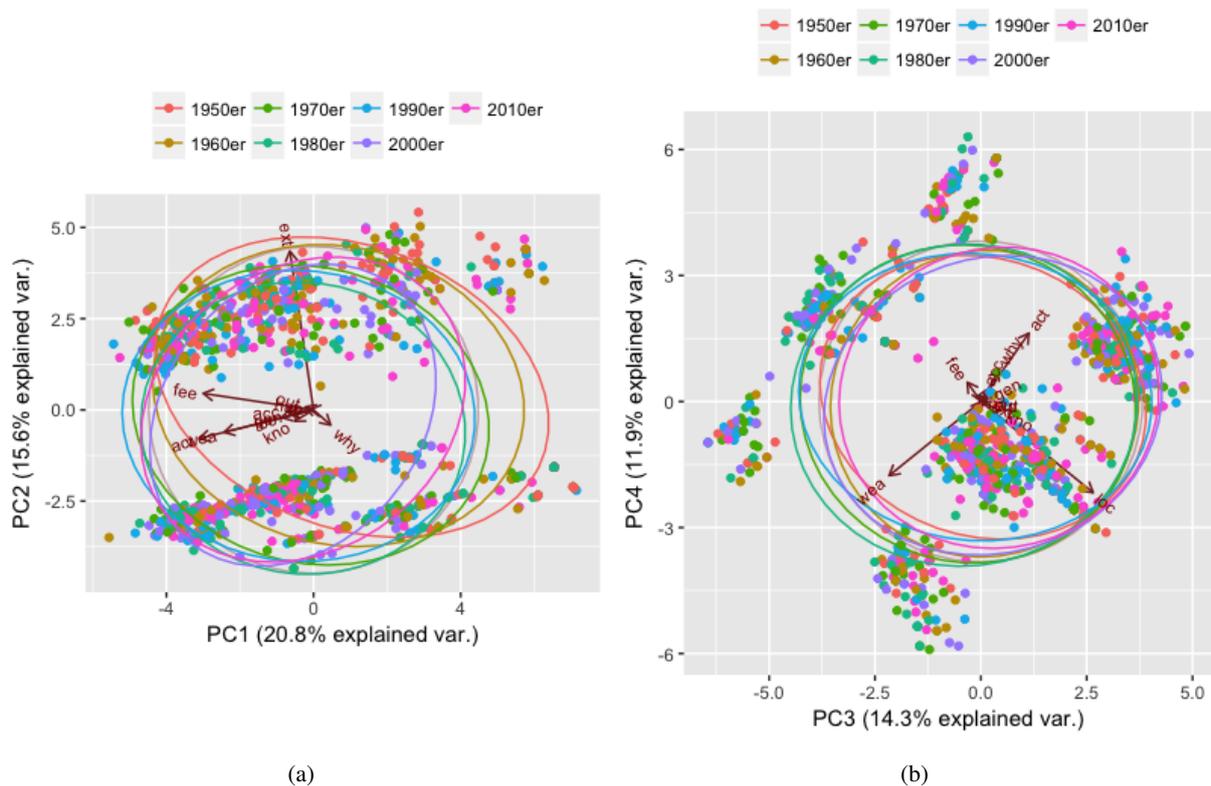


Figure 2: Trajectory over time: (a) Component 1 and 2 (b) Component 3 and 4

activity, weather and feelings (evaluation). Secondly, the PCA demonstrated that the main function of the text type ‘holiday picture postcards’ progressively shifted from the obvious function of contact to more that of a description of holidays. Thirdly, we showed that postcards have evolved in ways that can only be interpreted by further investigations into the social- and cultural backgrounds at that period in time. Lastly, PCA also identified language patterns of a prototypical speech act ‘greeting’ in postcards. We observed two patterns of greeting: weather and location, and activity are evoked prototypically in greeting. In particular, greeting with the mention of weather and location was common in the 1970s and 1980s. In future, we plan to investigate the changes of narrative structures in holiday picture postcards.

Acknowledgments

This work has been funded under SNSF grant 160238. We thank Maaïke Kellenberger and David Koch for the annotation, and Josephine Obert and Jan Langenhorst for the assistance during the annotation process, and Joachim Scharloth and Noah Bubenhofer for valuable discussions.

References

- Herald Baayen, Hans van Halteren, and Fiona Tweedie. 1996. Outside the cave of shadows: using syntactic annotation to enhance authorship attribution. *Literary and Linguistic Computing* 11(3):121–132.
- Douglas Biber. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge University Press, Cambridge.
- Klaus Brinker, Hermann Cölfen, and Steffen Pappert. 2014. *Linguistische Textanalyse: eine Einführung in Grundbegriffe und Methoden*. Erich Schmidt Verlag, Berlin, 8., neu bearb. und erw. aufl. edition.
- Dietrich Busse. 2012. *Frame-Semantik. Ein Kompendium*. De Gruyter, Berlin, Boston.
- Jan Hajič. 1998. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In E. Hajičová, editor, *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová*, Karolinum, Charles University Press, Prague, Czech Republic, pages 106–132.
- Heiko Hausendorf. 2008. Zwischen Linguistik und Literaturwissenschaft: Textualität revisited. Mit Illustrationen aus der Welt der Urlaubsansichtskarte. *Zeitschrift für germanistische Linguistik (ZGL)* 36(3):319–342.

- Heiko Hausendorf. 2009. Kleine Texte. Über Randerscheinungen von Textualität. *Germanistik in der Schweiz. Online-Zeitschrift der Schweizer Akademischen Gesellschaft für Germanistik* 6.
- Heiko Hausendorf and Wolfgang Kesselheim. 2008. *Textlinguistik fürs Examen*. Vandenhoeck & Ruprecht, Göttingen, Germany.
- Hermann Moisl. 2015. *Cluster analysis for corpus linguistics*. Quantitative linguistics. de Gruyter Mouton, Berlin.
- David M. W. Powers. 2012. The problem with kappa. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 345–355.
- Manfred Stede and Sara Mamprin. 2016. Information structure in the Potsdam Commentary Corpus: Topics. In *Proceeding of the 9th International Conference on Language Resources and Evaluation (LREC 2016)*.
- Kyoko Sugisaki, Nicolas Wiedmer, and Heiko Hausendorf. 2018. Building a corpus from handwritten picture postcards: Transcription, annotation and part-of-speech tagging. In *Proceeding of the 11th International Conference on Language Resources and Evaluation (LREC'18)*, pages 255–259.
- Alexander Ziem. 2008. Frame-Semantik und Diskursanalyse – Skizze einer kognitionswissenschaftlich inspirierten Methode zur Analyse gesellschaftlichen Wissens. In Ingo Warnke and Jürgen Spitzmüller, editors, *Diskurslinguistik nach Foucault. Methoden*, De Gruyter, Berlin, New York, pages 89–117.