

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Exploring the linguistic dynamics of contextual meaning making in microposts: a corpus linguistic study on the Memorial Day floods communication in 2015“

verfasst von / submitted by

Ivelina Georgieva, Bakk.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 899

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Angewandte Linguistik

Betreut von / Supervisor:

Univ.-Prof. Dr. Jürgen Spitzmüller

Content

Abstracts.....	i
English	i
German.....	ii
Format conventions	iii
Figures and tables	iii
1. Introduction	1
1.1 Twitter across the social media landscape	1
1.2 Weighing up the pros and cons of analysing Twitter language	4
1.3 Significance of Twitter interactions in crisis situations.....	7
1.4 State of the art from a linguistic perspective	8
1.5 Research questions	10
1.6 General theoretical framework	11
1.7 Organisation of the thesis.....	11
2. Theoretical foundations and methodology.....	12
2.1 Introduction to a social semiotic perspective within Discourse Analysis (DA)	12
2.2 Halliday's introduction to Systemic Functional Linguistics (SFL)	13
2.2.1 Ideational metafunction	21
2.2.1.1 Material processes.....	23
2.2.1.2 Mental processes.....	26
2.2.1.3 Behavioural processes	27
2.2.1.4 Verbal processes	28
2.2.1.5 Existential processes	29
2.2.1.6 Relational processes	30
2.2.1.7 The circumstantial system	32
2.2.2 Interpersonal metafunction	33
2.2.3 Textual metafunction	34
2.3 Corpus Linguistics (CL) and web-based corpora	34
2.4 Corpus linguistic tools and techniques	37
2.5 Contextualization of Twitter microblogged communication as a corpus.....	40
3. Data collection and theoretical argumentation	42
3.1 Twitter data collection.....	42
3.2 Texas and Oklahoma flooding 2015.....	44
3.3 Building the corpus	45
3.4 Theoretical argumentation and practical approach	49

4.	Analysis.....	53
4.1	Keyness and verb frequencies	53
4.2	Temporal aspects of experiential processes.....	58
4.2.1	May 22	59
4.2.2	May 23	59
4.2.3	May 24.....	62
4.2.4	May 25	64
4.2.5	May 26.....	65
4.2.6	May 27	67
4.2.7	May 28.....	68
4.2.8	Notes on the practical approach	69
4.2.9	Day-by-day analysis and summary of experiential processes.....	70
4.3	Information and message types across the corpus	76
4.4	Process variation across units and message types	83
5.	Results	87
6.	Discussion and conclusions	89
	References.....	92

Abstracts

English

This study provides a two-part linguistic analysis of microblogged communication from the Memorial Day floods in Texas and Oklahoma in 2015, generated in near real-time. Authors of microposts share semantic information which is closely related and describes the context of the situation in which they currently are, and their physical, ecological and social environment. In order to examine this information, this thesis uses a combination of corpus linguistic approaches, Halliday's conceptualisations of Systemic Functional Linguistics and experiential meanings conveyed by verbal groups as key functional resources.

The corpus linguistic analysis aims to depict the differences between the Memorial Day floods and the general discourse, both represented in the microblogging service Twitter. This examination is followed by narrowed down linguistic analysis, adapting the experiential process system. The main focus of this part is to trace how actors construe meanings in formal terms and to follow the dynamics of encoded experiential processes throughout a selected time period in which the flooding occurred. The close investigation of experiential processes will provide the foundations and facilitate the simultaneous discourse analysis of microposts in a semiotic mode, which comprises the second part of the analysis. The main focus here is laid on explicating what kinds of contextual meaning is shared during the observed timeframe and how the discourse of the flooding develops over time. The entire set of analyses will be guided by a temporal day-by-day perspective.

German

Die Masterarbeit umfasst eine zweiteilige linguistische Analyse der Mikroblogging-Kommunikation, die während der Memorial Day-Überflutungen in Texas und Oklahoma im Jahr 2015 in Echtzeit entstanden ist. Dabei übermittelten Akteur:innen semantische Informationen, welche den gegenwärtigen Kontext, in dem sie sich befanden, sowie ihre physische, ökologische und soziale Umgebung beschreiben. Methodologisch basiert diese Untersuchung auf einer Kombination von korpuslinguistischen Ansätzen sowie auf Hallidays Konzeptualisierungen der systemisch-funktionalen Linguistik und ihrer experientiellen Bedeutungen. Verben und Verbgruppen werden als Kernelemente der experientiellen Bedeutungen verstanden und stehen hierbei im Mittelpunkt.

Die korpuslinguistische Analyse zielt darauf ab, die Unterschiede zwischen den Memorial Day-Überflutungen und dem allgemeinen Diskurs, in der Mikroblogging-Kommunikation von Twitter, zu schildern. Nach dieser Auswertung erfolgt die Darlegung der linguistischen Analyse der experientiellen Prozesse. Dieser Abschnitt beschäftigt sich vornehmlich mit der Frage, wie Akteur:innen Bedeutungen in formalem Sinn von Prozessen konstruieren, und verfolgt deren Dynamik über einen bestimmten Zeitraum. Die umfassende Untersuchung dieser Aspekte stellt die Grundlage für eine parallele Diskursanalyse der Kurznachrichten aus semiotischer Sicht dar, welche den zweiten Teil der Analyse bildet. Der Schwerpunkt liegt im Zuge dessen darauf zu erläutern, welche Arten von kontextbezogener Bedeutung während des festgelegten Zeitraums realisiert werden und wie sich der Diskurs um die Überflutungen im Laufe der Zeit entwickelt. Die gesamte Auswertung erfolgt im Zusammenhang mit einer zeitlichen Einordnung.

Format conventions

To improve the readability of the theoretical part and the analysis, the following format conventions have been applied: Bold is used to introduce new topics within the chapters and for names of categories in the analysis chapters. Italics are used for quotations of the data in the text, for introducing terms in the theoretical part and if necessary, when mentioning them. In quotations, the original spelling and punctuation used by the authors of the material quoted have been kept. Mentions of personal accounts have been removed.

Figures and tables

Figure 1: Stratification (Source: Halliday & Matthiessen, 2014: 26)	15
Figure 2: Rank scale, based on Butt et al. (2000:29)	18
Figure 3: Grammar, semantics and context (adapted from Kim, 2007)	20
Figure 4: a): The configuration of situation, b) the grammar of experience: types of processes. Adapted from Halliday & Matthiessen (2014:176) and Butt et al. (2003:46)	22
Figure 5: Processes and main roles in the clause within the transitivity system network. Based on Butt et al. (2003:50-65)	23
Figure 6: Total rainfall across the Southern Plains in May 2015. NOAA Climate.gov map, based on preliminary data from the PRISM climate group at Oregon State University (NOAA, 2015).	44
Figure 7: River levels on the Buffalo Bayou (left) in Houston and the Blanco River (right) in Wimberley, Texas, in late May 2015. The Blanco registered 40.21 feet—7 feet higher than the previous record—before the gauge went offline (NOAA, 2015).	45
Figure 8: Message count Memorial Day floods, May 22 – May 28 2015	53
Figure 9: Visual representation (Wmatrix-style) of lemma-based keywords in the Memorial Day floods corpus	54
Figure 10: Normalized distribution of process types across the different days	72
Figure 11: Information and message types	80
Figure 12: Overview of the absolute distribution of process types across units	84
Table 1: The principal categories of relational clause. Source: Halliday & Matthiessen (2014:265)	31
Table 2: Overview of circumstantial elements. Source: Butt et al. (2003:65)	33
Table 3: Summary of datasets and relevant microposts	47
Table 4: Top 20 keywords ranked by log-likelihood value	55
Table 5: Absolute frequencies of verb lemmata in the corpus	56
Table 6: Frequency breakdown of verb forms in the corpus	58
Table 7: Absolute and relative distribution of process types across the different days. The percentage have been rounded to whole numbers.	71
Table 8: Relative and absolute values of information and message types across the days in the corpus. The percentage have been rounded to whole numbers	82
Table 9: Absolute and normalised distribution of process types and categories in units. The percentage have been rounded to whole numbers.	85

1. Introduction

1.1 Twitter across the social media landscape

Over the past decade, the microblogging service Twitter, along with several other microblogging services has rapidly grown in popularity and has become an influential channel, used for various communication purposes across social discourses. This has facilitated a great number of fast-paced and rapidly evolving interactions to start taking place within different contexts and domains in Twitter.

The micro-blogging platform began as an idea of an SMS mobile phone-based communication platform in 2006. Mobile carriers imposed a standard of 140 characters for short messages and so Twitter, initially being designed as an SMS based communication platform, also brought those restrictions for messages among its users. As it started gaining popularity, the limit of 140 characters remained and as a result, in order to handle that challenge, Twitter users started creating new abbreviated and jargon forms, built out of necessity. Presently, this is one of the many interesting phenomena which calls the attention of researchers in the field of linguistics. According to recent statistics, the online platform has currently 397 million users globally and 259 million daily active users, who send at least 500 million messages per day (Shepherd 2023)

This study has been conducted in the time of rebranding the Twitter platform to X (X.com). However, the data used for the purpose of the research has been collected in 2015 when the microblogging service was still named Twitter. Thus, all references here are using the terms Twitter and tweets, corresponding currently to X and posts.

In recent years, researchers have increasingly recognized the potential of Twitter as a valuable resource, containing a vast amount of data suitable for research across different domains and topics. An interesting example of this is a machine learning project initiated by Microsoft. Their objective was to learn more about the topic of 'conversational understanding', and with this aim developed a chatbot named Tay, specifically designed for engaging with humans. Tay was introduced as an experiment on the Twitter platform, engaging in automated conversations with other users and learning from their input language, essentially by mimicking it. As Tay engaged in more conversations, it gradually became more proficient. However, in less than 24 hours, the chatbot's behaviour took a turn, becoming impolite, offensive, and exhibiting racist and sexist attitudes, particularly towards women and minorities. Consequently, Microsoft had to terminate the project (Victor 2016). Although this experiment might not be very recent, it remains a valuable example showing the significance of online microblogging services. In a historical

context, it highlights how various linguistic phenomena, concerning particular issues of our society, were being reflected in Twitter conversations still in 2016.

Across the social media landscape, Twitter has evolved into a well-established online messaging platform and a mediated way to share different kinds of information. Nowadays it receives significant media attention, as it has become a hub for discussions on a wide range of pressing matters, with politics and health issues being just two examples. In addition to the just mentioned topics, Zappavigna (2012) stated that many different kinds of events, like natural disasters, weather, but also other types of events, call forth an overall communal response in Twitter.

In the past decades, traditionally the dissemination of information related to emergency situations has been taking place particularly in established media sources such as newspapers, radio, and television as stated by Martínez-Rojas et al. (2018). This form of communication is primarily based on the principle of one-way communication, where the content is produced and delivered by governmental organizations, authorities or similar sources to individuals.

Whereas present day social media communication introduces a different communicative paradigm, it allows each individual to contribute content, and thus content is delivered by a network of participants, creating a collective flow of information. In comparison with traditional sources of information sharing, the nature of interactions on social media is versatile and can be multidirectional depending upon the number of participants (Martínez-Rojas et al. 2018). In her classification of computer-mediated discourse, among other components, Herring (2007) differentiated three variables of participant structure in social media communication: one-to-one, one-to-many and many-to-many. When applied to Twitter interactions, many-to-many participation structures like discussion threads under microposts for instance, stand out as collective forms of communication. Such a dynamic format enables multiple users to engage with and contribute to discussions, while it simultaneously motivates a conversational space of different social groups that is shared and inclusive. In contrast with these dynamics, a one-to-one communication pattern occurs through direct messages for example, and unlike the communal nature of thread discussions, it facilitates private conversations between two individuals. On the other side, one-to-many communication unfolds through broadcasting updates on a user's timeline, referred to as Twitter updates. Through this mode, users can widely disseminate information to their follower networks.

Nowadays, the dissemination of information or news about emergency situations predominantly occurs on social media channels first rather than on traditional news sources as reported by Martínez-Rojas et al. (2018). This shift emphasizes the growing role of social media as a primary

source for the rapid dissemination of information during emergency events or other unforeseen situations.

Twitter is a rich source of valuable naturally occurring data; its public Firehose API generates a great number of messages per minute. During large-scale events with broad public interest, micropost activity rises even higher. Research has also shown that Twitter has proven to be a valuable source of information and a disseminating service during various types of emergencies (see sections 1.3, 1.4). Below the key concepts and main features of Twitter are outlined:

- Tweet messages or status updates. Users currently deliver their messages up to the limit of 280 characters. Since its inception, Twitter has been supporting a 140-character limit, which was doubled to 280 characters later in 2017.
- Hashtag symbols, retweets and mentions. Over time Twitter users introduced several communication patterns, which have been adopted by the platform later and now serve a central role in Twitter conversations (X Developer Platform 2023). The hashtag symbol is for instance presently used as universal tool across different microblogging social networks. It typically has the function of a conversational marker, indicating the main topic of a micropost, and is employed by users to convey meanings across different discourses in Twitter. Structurally, hashtag labels can have different lexical compositions, comprised of single or multiple lexical items. Another user-driven practice adopted by Twitter is a retweet. Initially, users applied the practice of forwarding content by manually copying and pasting microposts, adding a "RT @" pattern in the beginning. The process of manual forwarding or resharing information was officially recognised and automated by Twitter, when in 2009 they introduced a 'retweet' button on their interface (X Developer Platform 2023). Furthermore, a mention denoted with the symbol '@' defines another user or name of an account as a recipient of the message.
- Some of the native properties of Twitter are its openness for public consumption, accessible content and the possibility to use the service free of charge at present. Users participate in social interactions creating a wide scope of networking practices and building microblogging communities. Twitter provides the possibility to access this information.
- Tweets are bundled with metadata like a timestamp and geo-location.
- Twitter data are ephemeral – they can be changed or removed at any time by its authors or automatically set to disappear.

1.2 Weighing up the pros and cons of analysing Twitter language

There is a growing body of conducted research studying how language is employed on social media across academic disciplines such as education, psychology, anthropology, sociology, crisis management studies, computer science, media studies, journalism, literature, language and applied linguistics and more, summarised by Page et al. (2022:47). In order to shape fully fitting frameworks and approaches for examining language in social media domains, linguistic research can be facilitated by methodologies adapted from across these disciplines. However, navigating such a trans-disciplinary approach of investigating language within social media contexts may also become challenging. Differences and boundaries across the various disciplines sometimes might turn to barriers, making the research process difficult, for instance discrepancies such as mismatch of methods used for collecting data, or differences in research design and more, as discussed by Page et al. (2022:48).

The asymmetric nature of Twitter discourses and rapid interactions and the non-standard vocabulary and symbols employed by users, might cause difficulties in the analysis of streaming data. Correspondingly, these inherent characteristic features of Twitter constitute its distinctive communicative norms. Page et al. (2022) stated that there is a constant change within interactions and communication in social media, and that this change sets a dynamic rather than stable communicative setting. In this sense, as mentioned, the texts produced are not static entities with well-defined boundaries as understood within formal terms. Instead, they go through constant changes.

At the other end of the scale, the openness of the platform makes it easier to find and follow conversations and topics, through its searchability across Twitter discourses and through its public accessibility – microposts appearing in Google search results for instance. Adding many new perspectives in the field of linguistics, the immense volume of available data gives possibilities for quantitative and qualitative research on either micro level language phenomena, or on any language practices more broadly. In addition, hashtag conversational norms aid the processes of retrieving, gathering and sorting linguistic data, since new topics and events tend to revolve around hashtags (cf. Zappavigna 2012). Admittedly, two major advantages of conducting linguistic research on Twitter are the extensive amount of naturally occurring authentic data and its near real-time nature. Twitter's question 'what's happening', encourages users to share information which reflects the current events and happenings from their outer or inner world, and creates an event-centred environment. Furthermore, the available limited number of characters makes the linguistic practices used in Twitter an interesting phenomenon to explore. It can provide a

different linguistic perspective on how language is employed by users and how meanings are conveyed in such constrained environments of 280 characters (140 before the switch).

Another characteristic of social media interactions in general is that they typically incorporate images, videos and audio alongside text, providing plenty of material for analysis. Handling all these manifold aspects requires carefully considered multimodal methodology, tools and software packages to organise and analyse the corpus (Page et al. 2022:95).

Moreover, working with public data collected from Twitter can present some further challenges. Along with all its benefits, one should consider most importantly the credibility of the shared data. Referring to cases when studying Twitter use during hazards or natural disasters, it should be considered that not all the shared information is reliable or describes users' behaviour precisely. A study conducted in 2017 from the University of Southern California and the Indiana University (Varol et al. 2017)¹ outlined that between 9% and 15% of active Twitter users are not humans but social bots. In their study, Varol et al. (2017) defined social bots as accounts operated by software, algorithmically generating content and initiating interactions. Depending upon the research question and purpose of the study, spam messages and microposts generated by non-human services can be identified with the aim to be kept or removed from the data.

There are two possibilities for working with social media data. One is to make use of already available corpora and the other is to collect it individually. According to the research question and depending on the volume of data foreseen for the research, often analysing social media microposts requires working with tools that partially automate preceding-the-analysis processes like text annotation. However, going through all steps of cleaning and preparing the data can be rather complicated and time-consuming. For instance, several main textual characteristics of social media communication which can pose obstacles when trying to organise the data digitally in order to automate analysis via tools and software were summarised by Zappavigna (2012:19), as follows:

- 'Omitted characters and white space, ignored casing rules, and nonstandard abbreviation due to the highly condensed [spelling] required for text economy.
- Frequent oral markers and non-standard or omitted punctuation due to the conversational nature of the posts.

¹ The collected corpus for the analysis in this thesis originates from 2015, as explained in chapter 3. With this in mind, a relevant study closer to this year is provided.

- Spelling errors since posts are generated using mobile devices on-the-fly.' (Zappavigna 2012:19)

While analysing Twitter discourses appears to be very worthwhile and presents many advantages, one should consider several factors on the nature of the Twitter language before starting with the research. In their interactions, users often employ non-standard spelling of different types, deviating from the English conventional norms of e.g., capitalisation and punctuation (cf. Page et al. 2022). This process of lexical disconnecting from formal linguistic conventions is a distinctive feature of communication in Twitter. Typical examples are prolonged 'caps lock' patterns in which the entire message or a specific part of it is capitalised, such as 'I PASSED MY EXAM', 'I PASSED my exam', 'I PaSsEd My ExAm' and further variations, discussed in detail by Heath (2018).

Alongside this, common characteristic features are further spelling variations like *ur*, *dont*, instead of *your*, *don't* etc., or other linguistic deviations like compressing two or more lexical items into one unit without spacing. The brevity imposed by character limitations and the informal nature of microposts contribute to these deviations, allowing for a unique linguistic environment that distinguishes Twitter discourse from other more formal communication services. In particular, character limitations significantly influence users' linguistic choices. All outlined dynamic features which make Twitter conversations distinctive bring out further challenges for the technical software tools used for annotation and corpus querying, as these have usually been developed for traditional text genres of conventional orthographic character. As a result, very often the data requires additional preliminary procedures, which usually involve manual work for preparation.

Page et al. (2022) outlined, that in particular the presence of non-verbal material like emoticons, emojis and hashtag symbols adds complexity to pre-processing the data. How effective the concordancing software works may be influenced by its current settings, in particular how to treat characters employed in emoticons. In some cases, emoticon characters may not be recognized as authentic letters by the system, and thus be excluded from the concordance. Besides this, the software might also interpret these characters as indicating word breaks or attribute other special meanings to them. Furthermore, although emojis are standardised in Unicode, their processing may fail in the different systems depending on the current Unicode version. It is crucial to preserve accurate information about the specific operating system, platform, or device used to generate the text to be able to keep the exact rendering of emojis. This is because the representation of emojis can differ across various media (Page et al. 2022:180).

1.3 Significance of Twitter interactions in crisis situations

Compared to other social platforms, Twitter stands out specifically as an event-centred service, it facilitates real-time conversations and allows for high connectedness among its users, as highlighted above. These are some of the reasons why it is widely used as a communication platform in which actors actively share different kinds of information about different ongoing events, matters or problems in real-time. In addition, this provides users with immediate access to what is being said and to the latest reported updates within their networks. These properties prove particularly valuable during emergency or unforeseen situations, in which the rapid dissemination of information may be critical.

As outlined by Martínez-Rojas et al. (2018), emergency situations arise when the typical dynamics of economic, cultural, social, or political environments in a location are disrupted. In response to such disruptions, there is an increase in communications in the various communication and information services. Effective communication becomes crucial during such events, serving the purpose of conveying related information and also of reporting on complex scenarios which might require immediate response (Martínez-Rojas et al. 2018). The use of social microblogging services like Twitter during crisis events by private individuals, organisations or other actors can serve multiple communicative purposes, alongside dissemination of on-topic information. Zade et al. (2018:195:2) noted that affected individuals, emergency and humanitarian responders, as well as remote audiences use Twitter to exchange information and offer informational or emotional support during disaster events. Prior and recent studies confirm that users actively tend to post messages in times of mass emergencies, no matter if they are man-made or natural disasters like wildfires, floods, tornadoes, terrorist attacks and more (Vieweg et al. 2010, 2012; Zade et al. 2018)².

Twitter's event-centric nature and real-time capabilities make it a valuable tool for individuals to share and receive timely updates. Mihunov et al. (2020) pointed out the widespread practice of individuals turning to social media channels to call for rescue or communicate information when formal channels are overloaded. In that sense Twitter communication contributes to the information environment from which affected communities or authorized organizations might be able to benefit by collecting additional on-topic information.

² Further research on the topic is listed in section 1.4.

However, while such sharing practices generate big volumes of data, Mihunov et al. (2020) and Zade et al. (2018) have discussed that they also raise many questions in terms of what information is shared, how to make sense of unstructured information, how effective social media functions in such cases, how its information mechanisms can be improved, and so forth.

1.4 State of the art from a linguistic perspective

In the past decade, research on the language used in Twitter communication has gone through a significant surge, mostly because the platform has gained in popularity as a medium for everyday interaction. Subsequently, it currently provides plenty of real-time data, which presents a big potential to researchers for analysis. Karami et al. (2020) for instance, conducted an extensive literature review of Twitter-based research to identify the dominant research topics and observe the trends and evolution in the past ten years.

Multiple studies delve into various micro or macro linguistic phenomena in Twitter and aim to explore and analyse its specific discourse by adopting cross or interdisciplinary methodological practises. In the field of linguistics, a substantial volume of analysis concerning mass emergencies stems from the areas of computational linguistics and natural language processing. Within the domain of crisis, there is an extensive work relying on computational methods for classifying thematic content, for improving part-of-speech (POS) tagging techniques to obtain the most relevant data, for developing automated classifiers, for performing sentiment analysis and more (Hughes & Palen 2009; Cresci et al. 2015; Panem et al. 2014; Purohit et al. 2013; Khattri et al. 2015). For instance, Vo & Collier (2013) focused on tracking crowd emotions during earthquakes. They have proposed classification methods to identify relevant tweets, employing sentiment analysis as a tool to reveal and categorize emotional content within the Twitter discourse during such critical events.

The primary focus of research in the domain of emergency situations extends beyond the scope of linguistics, with the majority of the research originating from disciplines such as crisis management, crisis informatics, and crisis communication. Their overarching objective is to identify, automate and classify timely and accurate information shared before, during or after the crisis. This research will not be discussed here.

Vieweg et al. (2010) and Vieweg (2012) have carried out simultaneous analysis of Twitter communication on three emergency situations and have aimed to find mechanisms for identifying relevant information that contributes to situational awareness in times of crisis, drawing on a number of fields within linguistics, sociology, crisis informatics and more. In her work, Vieweg (2012) performed linguistic analysis on different levels to reveal relevant communicated meanings

drawing upon a methodological combination of coding processes, pragmatic aspects, discourse analysis and further approaches. Her efforts were aimed to contributing to the development of natural language processing techniques that aid in the automatic extraction of messages containing information related to situational awareness.

In recent years there has been an increasing number of studies concentrating on hashtag labels. La Rocca & Boccia Artieri (2023) for instance have studied the nature and characteristics of hashtags including the changeability of their meanings. The authors proposed a model for analysing hashtags' changes over time and across contexts, underpinned by the sociological concept of cultural morphogenesis through the dimensions of time and agency. Their work identified the different types of hashtags and the different types of information that they can convey (such as topics, events, opinions or emotions). In addition, their work provided an extensive review of past and current research in the field of hashtags and distinguished some common themes and challenges.

Zappavigna (2015, 2015a, 2018) has made major contributions in the field of hashtag research from a linguistic perspective. She categorised hashtags as linguistic devices enabling online communities to build relationships through social tagging and as instruments to enact a range of linguistic functions (e.g., realising different meanings in different contexts). Adopting a social semiotic view on language, Zappavigna integrated and applied semiotic practical approaches in her interdisciplinary research. In her early work on the discourse of Twitter and social media (Zappavigna 2011, 2012), she adopted the notion of linguistic metafunctions from Systemic Functional Linguistics and presented Twitter communication as a communicative space where alignments among online participants and social bonding through shared values were created. Using a large corpus of over 100 million micromessages and various case studies she investigated how individuals employ language to create affiliations and express opinions on the web. She has defined online social bonding as 'ambient affiliation' that Twitter users employ when broadcasting information.

Furthermore, she also analysed the linguistic features and patterns of online discourse, including evaluative language, slang language used on the Internet, memes, and punctuation. Her approach combined components from social semiotics, systemic functional linguistics, and corpus linguistics, providing a theoretical framework and methodological tools for understanding and interpreting the discourse of Twitter and social media. Based on her definition of 'ambient affiliation' in a recent study, she has focused on identifying and categorizing temporal meanings,

and how individuals use meanings to create social bonds and to express their values on Twitter during the COVID-19 pandemic (Zappavigna 2022).

The above-described work aimed to emphasize the multifaceted nature of linguistic research on Twitter, and how various methodologies are applied to analyse and understand the nuances of microblogging communication.

1.5 Research questions

In this study, the linguistic perspectives of Twitter data shared during the floods in Texas and Oklahoma in 2015 are explored. It aims to deliver an in-depth description and analysis from the Memorial Day floods in 2015, making statements about what kind of contextual information has been shared during the course of the floods and how contextual meanings have been realised by participants. Pragmatically, the overarching purpose of this thesis is to investigate the relationship between language and its functions in the social context of the flooding. To achieve this, on the one hand a temporal day-by-day perspective will be employed to guide the analysis, through which the experiential meanings of Systemic Functional Linguistics construed by actors in the days between May 22 2015 and May 28 2015 will be examined.

On the other hand, a combination of a qualitative and quantitative linguistic approach will underpin the analysis to enable following the dynamics of how the discourse of the flooding unfolds over time and how actors express meanings from a semiotic perspective. Furthermore, the objectives of the study are to describe how this communication is linguistically constructed and what linguistic devices actors employ in order to convey their meanings. The main research questions addressed here are:

- How do lexical choices and frequency distributions in the Memorial Day floods discourse differ from those in the general discourse on Twitter? How do verb frequencies vary across the corpus?
- Through which verb types, as categorized by Systemic Functional Linguistics do actors lexically realise events, states and happenings during the Memorial Day floods and what are their dynamics from a temporal perspective?
- What contextual meanings, in terms of message and information types are exchanged and how does the discourse of the flooding develop over time?
- How is this microblogged communication linguistically constructed?

1.6 General theoretical framework

To address the above-mentioned research questions, this study employs a combination of qualitative and quantitative methods to create good grounds for the analysis. It relies strongly on corpus linguistic approaches, mainly to manage the data, build the corpus and to analyse the frequency distributions of keywords and verbs. A combined corpus-driven and corpus-based methodology will serve as the basis for the empirical part.

Furthermore, as theoretical fundamental for the analytical part, discourse analysis informed by a socio-semiotic approach to language through a temporal perspective is used. The analyses will be extended to experiential and contextual meanings in a socio-semiotic sense, their realisations and their shifts across the given timeframe of the collected corpus.

Such approach will facilitate describing the discourse approaches and linguistic devices employed by the authors of the microposts and it will help following the dynamics of the Memorial Day floods discourse, as represented in the microblogging platform. It will also favour the discussion about social language practices occurring in Twitter during the Memorial Day floods.

1.7 Organisation of the thesis

This master's thesis further consists of two main parts. The following two chapters discuss the theoretical principles and methodologies and build the methodological framework which underpins the analyses. Here included are also the applied methods and techniques for data management in order to build the on-topic corpus of microposts. Discussed are general data collection possibilities, the topic of the collected data and the approach of building the corpus. The building of the corpus has been facilitated and realised with the technical support of Christoph Perger from Spatial Focus, a software development company based in Vienna.

The second part lays out the different types of linguistic analysis carried out and the inferred classifications. It also includes the results from the analyses, a discussion, and a set of conclusions.

2. Theoretical foundations and methodology

The aim of this section is to provide a theoretical overview and to show the synergies and linkages between the combined methods in order to answer the research questions laid out above. In different contexts, speakers construe multiple kinds of meanings through their linguistic utterances, which in turn, enact different kinds of contextual information. In order to find out how and what meanings and experiences individuals construe in the microblogged communication of Twitter during the floods, underpinning the analysis is a socio-semiotic approach within the architecture of Systemic Functional Linguistics theory (SFL) in combination with Corpus Linguistics (CL) analysis.

As a starting point, the following sections provide an introduction to the approach of Systemic Function Linguistics (SFL), a concise overview of the relevant concepts within the theory and then narrowing down to the concept of the language metafunctions and experiential meanings.

These are followed by an overview of Corpus Linguistics (CL), explications about Twitter data as a corpus and finally theoretical argumentation is put forward regarding the framework and the employed pragmatical approach to proceed with the analysis.

2.1 Introduction to a social semiotic perspective within Discourse Analysis (DA)

The purpose of this chapter is to briefly outline the notion of a social semiotic perspective of Discourse Analysis (DA) as an introduction to the following theoretical part of SFL. The overarching concept of DA is vast and complex and will not be discussed here in detail, as the scope of this thesis is too narrow.

Discourse Analysis (DA) delves into natural language use of any kind to examine social interactions and relations by identifying linguistic themes and patterns in data to learn more about discourse. In its nature, DA applies qualitative and interpretative techniques to uncover details about social phenomena like what participants do and how they do it, and thus relies strongly on naturalistic data (Schleppegrell 2012). Approached from a semiotic perspective, DA involves the study of how meaning is constructed and communicated through various semiotic resources within discourses (cf. Kress & van Leeuwen 2021). According to Spitzmüller and Warnke (2011:104), Kress and van Leeuwen perceived language as a social practice, in which meaning is constituted through processes of negotiation, and as a resource of *meaning potential* drawing upon Halliday's conceptualisations.

According to Kress and van Leeuwen (2001:4), discourses are defined as ‘socially constructed knowledges of (some aspect of) reality’, shaped in specific social contexts. They have many modes of appearance, and all available semiotic modes within a specific culture are utilized for constructing and shaping discourses in that culture (Kress & van Leeuwen 2001:24). Both authors expanded their exploration of discourse to encompass these various modes of communication beyond just language and argued that communication is inherently multimodal, involving the integration of verbal, visual, gestural, and spatial elements. Thus, meaning is co-constructed by the multiple modes and the different semiotic resources align together to produce specific meanings and ideologies (Kress & van Leeuwen 2001). Each mode of the complex semiotic landscape offers additional affordances for communication and contributes to the overall meaning making process.

Systemic Functional Linguistics (SFL) conceptualizes language as a semiotic entity, aiming to explore the social contexts, interactions, and the meanings they convey, and it also provides the means to examine these. The semiotic and social traditions are deeply rooted in SFL, and there is a reciprocal relationship between the linguistic process of meaning making and the context in which language is used in the sense that through language, meaning making shapes contexts and is similarly shaped by the contexts in which it is used (Halliday 2004, Halliday & Matthiessen 2014). As discussed by T. Bloor & M. Bloor (2004), within SFL Halliday developed several functional concepts and mechanisms that serve as a foundation for conducting meaning-based discourse analysis. The following theoretical part will concentrate entirely on concepts from Systemic Functional Linguistics, which are employed afterwards for the analytical part of this study.

2.2 Halliday’s introduction to Systemic Functional Linguistics (SFL)

Systemic Functional Linguistics (SFL) as modelled by Halliday (1994, 2004; Halliday & Matthiessen 2014) is the fundamental theory which informs this study. SFL, also known as Systemic Functional Grammar, has a holistic nature and in its core is looking at language through a socio-semiotic perspective. It is focused on the meanings of language in use within social contexts and aims, above all, at providing applicable methods for text analysis across the different linguistic realms of language as a semiotic system (Halliday 2004). In broad terms, SFL is functional because it looks at the functions of language and what language does, rather than what language is. The concept of systemic linguistics as ‘functional’ is fairly complex since functionality extends over multiple linguistic units and layers in language, so it will be referred to at different stages in the sections below.

In a historical context, the origin of SFL can be traced back to the work of Saussure, aligned with the linguistic traditions of the Prague School and French functionalism, bringing out the functional and semantic side of language in the foreground, rather than the formal and syntactic (Halliday 1994a). The foundations of the theory, which have been laid down by the English linguist and Halliday's teacher J. K. Firth, were further developed and cultivated mainly in Halliday's work; however, more linguists, including scholars from different areas, have contributed to it over time. SFL builds upon a number of previously known ideas and concepts. Some of these ideas stem for instance from the anthropologist Malinowski (e.g., social functioning of language is reflected in its linguistic system) and are later affiliated with and elaborated by Firth and further developed by Halliday. SFL has been constantly developing and evolving since its creation. Matthiessen (2007:505-561) provided a concise overview of the evolution of fundamental concepts across the different domains and aspects of the theory since the 1970s.

Halliday examines language through different perspectives, but the focus in his approach lies on the conception of language as a semiotic entity, through which human's culture and society are framed. He sees language as a semiotic system, out of which meanings arise and in which they are exchanged. In order to be communicated, these meanings are incorporated or encoded in the lexicogrammatical structures of a language. In that sense, Halliday does not see language in a sheer function as of means of communication, but as a *meaning potential* or an ample set of resources for meaning making (Halliday 1975). The different options available in the meaning potential of an utterance represent the concept of *linguistic choice* – speakers select one option from a set of alternatives in the language system. More specifically, meaning making is modelled by systemic patterns of linguistic choices on different levels (semantic to phonetic), which humans make when they use language. Linguistic choices represent the speaker's intentions, they are not necessarily conscious and some of them are adequate only for one or another specific language environment.

Functional linguistics is systemic as a network of coupled options, in which speakers choose either one or another, and these choices serve as a starting point for the next set of choices in a next linguistic system (Halliday 1994). These choices are made across all language levels (Figure 1) and gradually add to the meaning.

Stratification system

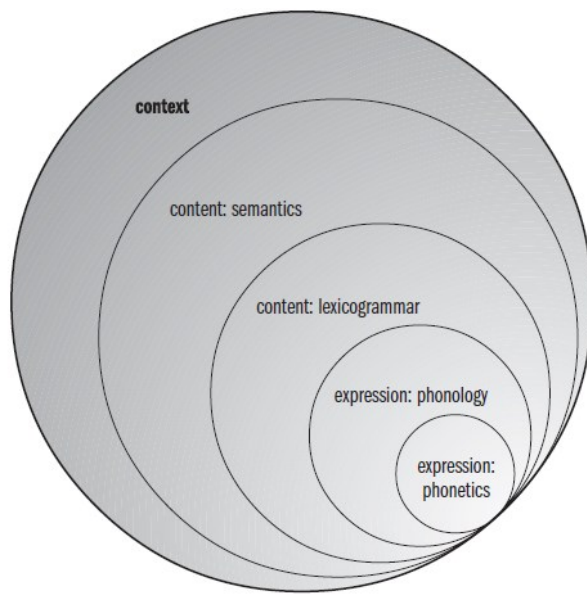


Figure 1: Stratification (Source: Halliday & Matthiessen, 2014: 26)

As a complex semiotic system, language has a wide scope and multiple levels of operability called *strata*, set out in Figure 1. The notion of stratum is one of the most distinctive features of SFL as a framework, and it serves to enable us to depict and analyse the relationships within the different levels of language - linguistic form, utterance meaning and social interaction (Bartlett & O'Grady 2017). To explain these interrelations, SFL constitutes four strata stretching from the abstract to the concrete level: semantics, lexicogrammar, phonology and phonetics, and it positions context on the extralinguistic level. As a component of the semiotic system, context relates language to its situational use and stratified language is embedded in the context.

The relationships between the different layers of language are expressed through *realisations*, which function in both directions from bottom-up and top-down across the hierarchy of language strata (Halliday 2004, Halliday & Matthiessen 2014). This construes a system of mutual relations in which the lower stratum encodes the higher, and the higher stratum activates the lower, i.e., the choice of semantic meaning reflects the choice of lexicogrammatical wording (language form), and the choice of lexicogrammatical wording reflects the choice of phonological speech. These processes can also be followed bottom-up the hierarchy of strata, e.g., the clustered lexicogrammatical elements as a whole unit encode the semantics of the text. Meaning is dynamic and according to Halliday (2004:530) it develops gradually while a text unfolds in the discourse. The concept of meaning creation as texts unfold is theorized as *logogenesis*. It is an active process and applies to all language strata and across all language metafunctions. The logogenetic patterns created at the level of lexicogrammar are at the same time semantically significant. Language

stratification, in that sense, represents the functionality of the grammar in SFL, giving priority to grammar units as a resource for meaning making.

In the hierarchy of strata (Figure 1), the context is positioned on the top of the architecture, on the extralinguistic level. The central principle of SLF is understanding language as a system of choices; thus, in line with the theory, linguistic choices cannot be justified without respect to the context and vice versa, a context cannot be explained without taking into account linguistic choices (cf. Halliday 1994). *Contexts of situation* act upon the circumstantial linguistic choices; however, in certain situations linguistic choices in reverse can trigger changes in contexts, for instance the choice of using formal or informal language forms will inevitably affect the context. This presupposes the idea, that the way in which the whole language system is organized is related to the way in which the contexts of situations are organized (Bartlett & O'Grady 2017).

The contexts of situation convey all the happenings of the world outside the text³, which give substance to the words and grammatical patterns that speakers or writers choose consciously or unconsciously to build different texts. A text is what it is, because of the context of situation in which it has emerged (Butt et al. 2000). The situational differences between texts can be described through a conceptual framework for contexts of situation by means of three parameters, referred to as *Field*, *Tenor*, and *Mode* of discourse. These parameters show how context and language are interrelated.

The 'discourse semantics' or the semantic stratum is the level on which we explore 'how grammatical units are constructed into discourse' (Halliday 1994:15). It is not only concerned with the meanings of words, but also with the entire system of meanings (meaning potential), expressed by means of grammar and vocabulary. The meaning of language is organised within three basic components, called *metafunctions*. The *ideational metafunction* encapsulates experiences, the *interpersonal metafunction* enacts relationships and the *textual metafunction* provides the structure of the text. The metafunctions are uniform to all situations of language use and encompass all the relevant meaning potential. They serve to organise meaning in language and operate across all strata. The concept of metafunctions is essential for the analysis of this study and will be discussed in detail at a later stage in this chapter.

³ Resting upon Halliday's conceptualisations, Eggins (2004:23-24) has defined text as any authentic form of a social interaction (spoken or written), of any length, which constitutes a unified whole. A text has texture, it is coherent (it has a relationship with its social and cultural context) and it is cohesive (its elements link together to create unity). Both written and spoken interactions are perceived as a text.

The lexicogrammatical stratum includes the structures of grammar (syntax and morphology) and lexis (vocabulary) and positions them on the same plane. The lexicogrammar is a central processing unit, and by its means, speakers use and organise words to realise their semantic choices. Sounds are combined into words, which in turn come together in different grammatical structures and create meanings. The lexicogrammar has a natural, nonarbitrary relationship to meanings. In a functional sense, lexicogrammar is the study of wording from a standpoint of what meanings words convey. On the other hand, the lexicogrammatical wording is realised by sound and writing, moving down the stratification model (Halliday 1994, 2004). As one way of understanding functionality, T. Bloor & M. Bloor (2004:10) described the change in the communicative functions of words and utterances depending on the context in which they are used, i.e., the same words can serve different communicative functions in different contexts. Furthermore, different utterances can serve the same communicative function, for instance using direct or indirect linguistic ways of communicating the same meaning. SFL puts a focus on revealing and describing the ways in which types of semantic meaning and linguistic form of utterances in language are interrelated, and this relationship is established on the fact that speakers adjust what they say to the context in which they are saying it (Bloor & Bloor 2004:10).

The stratum of phonology encompasses the sounds, their organisation in patterns and systems and their relation to meaning. In phonology we identify several key units as follows: tone group (melodic line), foot (rhythm group), syllable and phoneme (Halliday & Matthiessen 2014:11). Certain grammatical systems can be realised through prosodic means, such as rhythm and rhyme of speech, stress patterns and so forth. The available choices at the phonological level include for instance choices among intonational patterns which can be associated with the expressed meanings. Thus, phonology is regarded as a subsystem within the language's meaning-making system, responsible for encoding and manifesting lexicogrammatical meanings through choices in specific sound production.

The concept of phonetics is understood in a rather traditional way, enabling the physical production and perception of speech sounds. The stratum of phonetic is the lowest level in the model of stratification; thus, each distinctive form within the multiple levels is composed of phonetic material. From a bottom-up perspective, phonetics realises phonology, phonology – lexicogrammar, lexicogrammar – semantics, and semantics realises context, with some exceptions

to this general rule⁴ (Bartlett & O'Grady 2017). Furthermore, Halliday (2004, 2014) introduced the notion of graphology to SFL.

Graphology relates to the graphical resources in language like complexities of spelling and variations of punctuation (Halliday 2004:7). As discussed by Kress & van Leeuwen (2021), graphology refers to the study of visual elements and their compositional arrangement within texts, particularly in relation to their communicative and semiotic functions. Graphology focuses on analysing the graphic features of visual images or texts, including aspects such as typography, layout, colour, texture, and their spatial organizations. As semiotic resources, these visual elements carry meaning and are considered as part of the meaning making process since they interact with language and contribute to the overall conveyed meaning in a message. Graphology primarily contributes to the *textual metafunction*, which is concerned with how meaning is organized and structured in discourse.

The notion of clause

To explain how the system of human language functions, Halliday introduced the concept of a *rank scale* within the functional grammar. Each unit has a rank, and each higher rank is composed of one or multiple units from the lower level, illustrated in Figure 2. The highest rank in the scale is the *clause complex*, comprised of one or more clauses. Butt et al. (2000:29) further pointed out the fact that in a similar way clause complexes combine to build paragraphs, which accordingly combine to build texts; however, these linguistic units refer to the semantical and rhetorical levels of language, rather than to the grammatical discussed here.

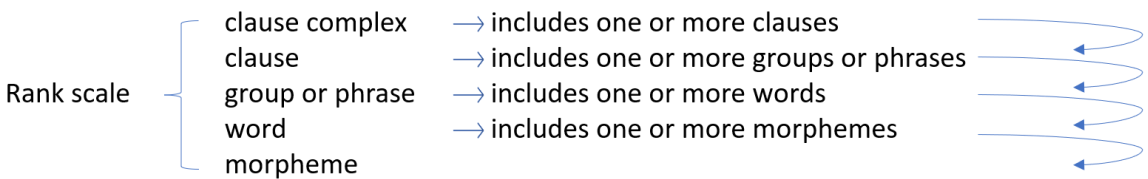


Figure 2: Rank scale, based on Butt et al. (2000:29)

Within SFL, the clause is discussed as a major unit of grammatical analysis, revealing how the ideational, interpersonal, and textual metafunctions simultaneously incorporate meanings expressed in clause structures. A clause is a basic unit of grammar, which is usually composed of a subject and a verb phrase and is thus the most integral and fundamental message structure in a language, which enables us to syntactically combine words into meaningful message structures in a complete way. It is at the rank of the clause where we create and convey meanings about how

⁴ Phonetics and phonology units are not discussed in detail here.

things exist, how things happen and what humans' attitudes are. In these lines, through grammatical and syntactical means English clauses enact the human experience of the world with account for things, events and various circumstances that surround and influence those events (Butt et al. 2003).

Field, Tenor, Mode

Field, *tenor*, and *mode* illustrated below in Figure 3 are abstract categories, which enable the representation of the contextual configuration of any context of situation. They serve to depict the relationship between language and context and to cluster together the relevant features of language in a certain context of situation. The features of these contextual parameters have been described by Bartlett & O'Grady (2017:394) as follows:

- Field – what is happening; the nature of the social action in which language is an essential component.
- Tenor – who is taking part; the nature of the participants, their social statuses and roles vis-à-vis one another and the types of speech role that they are taking on in the dialogue.
- Mode – what part the language is playing; the symbolic organisation of the text, and its status and function in the situation, including the channel (for example spoken or written mode) and the rhetorical mode.

Social contexts are a configuration of these three components, their characteristics are relevant features for the context of situation, and they relate to the three metafunctions of language. In addition, particular aspects of a context are realised through respective grammatical aspects in the system of language. Following Bartlett & O'Grady (2017:393), ideational meanings overlay the category of *field*, both reflecting the topic or the content and grammatically realised through the system of transitivity; the interpersonal meaning echoes the category of *tenor*, both enacting relationships and grammatically realised through choices in the systems of mood and modality; and both the textual meaning and *mode* reflect the role language plays in a context and are expressed through choices in cohesion. These interrelations are illustrated in Figure 3.

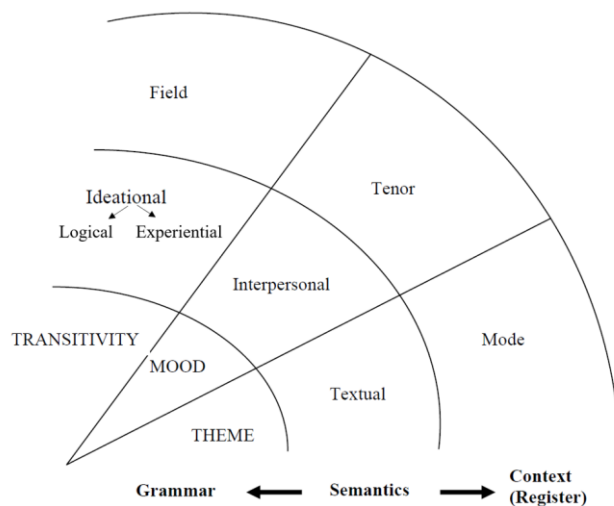


Figure 3: Grammar, semantics and context (adapted from Kim, 2007)

Metafunctions

Relying on previous theories, which he further developed, the fundamental concept of Halliday's theory revolves around the notion of language functions and their relation to the language itself in a semiotic sense. To accommodate human linguistic needs, according to Halliday (1973:331) 'language has evolved in the service of social functions', and he calls these functions metafunctions. Language has many uses, may serve different purposes under different circumstances and is thus deposited with an indefinite number of meanings. The concept of metafunctions represents a finite set of functional components of the semantic system of a language, which are general to all language uses and therefore encompass all the meaning potential related to them.

Within the human social and ecological environment, the indefinite language uses are expressed entirely and simultaneously through ideational, interpersonal and textual metafunctions (Figure 3), which in line with the theory play a pivotal role for the interpretation of language as a system (Halliday 1973). A meaning potential is the language, envisioned as a source of options and alternatives of meanings that extend to all facets of a language, and are available to a speaker and a hearer. The three metafunctions of language are overarching, have the meaning potential incorporated into them and thus serve to enact and organise meanings in language. When we say, 'meaning potential', we think of the available semantic options, and of what a speaker 'can mean' (Halliday 1973:301).

The three metafunctions are described by Halliday (1994:36) as follows:

- ‘the ideational metafunction - serves to construct experience and is the function of meaning,
- the interpersonal metafunction - serves to enact social relationships,
- the textual metafunction – serves to create relevance to the context’ (Halliday 1994:36).

Language encodes all three aspects of meaning simultaneously. Butt et al. (2003:39) gave the following example: ‘The high school students put on a noisy protest’ - describing what is happening, interacting with somebody to whom you are speaking and at the same time organising the structure of the message. All aspects of this message are result of the choices that a speaker makes.

2.2.1 Ideational metafunction

The ideational metafunction conveys information about what the content of language is about and is underlying all language uses as noted above. A language construes the human experience, which is then transformed into meanings according to Halliday (2004). The ideational meaning conveys information about representing these human experiences and is conceived as the major element signifying meaning in the linguistic system. It entails the greatest extent of the meaning potential. Two main components can be distinguished within the ideational metafunction – the *experiential* and the *logical* as shown in Figure 3. The experiential subfunction largely encodes experiences, contents, or ideas and the logical subfunction mostly carries the semantic load and manifests semantic relationships. In the case at hand, the experiential meaning will be described in detail.

As a carrying unit of experiential meanings, the clause conveys a group of human experiences comprised of occurrences describing a change or no change in our external or internal environment. The experiential meaning is made up of three kinds of functional elements: (i) a *process* - event, (ii) *participants* - individuals or things involved in the process; and (iii) *circumstances* (Halliday 2004:79), seen in Figure 4a below. Language represents the landscape of human experiences in terms of things and chains of events, which flow and thus change. While a *process* unfolds over time, the *participants* are directly involved in this process, whereas associated *circumstances* of time, space, cause, manner etc. might only be attendant on it. These three major semantic categories of human experience occur together, building the configuration of a situation. In line with Butt et al. (2003), processes (the expression of event) are core elements to construe meanings and are positioned in the centre of the system as in Figure 4a. The latter are

accompanied by the different participants that function within these processes. In practice, participants realise processes and are affected by them, and as Halliday (2004:176) outlined ‘the nature of participants will thus vary according to the type of process’.

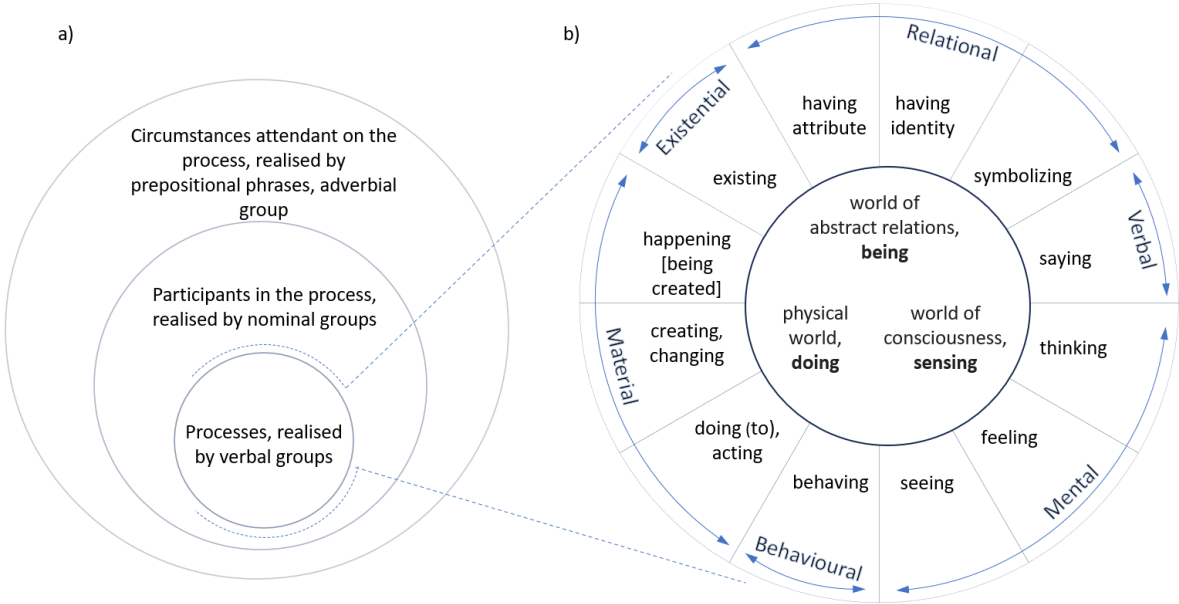


Figure 4: a): The configuration of situation, b) the grammar of experience: types of processes. Adapted from Halliday & Matthiessen (2014:176) and Butt et al. (2003:46)

In functional terms, a practical way of explaining how experiential meaning is represented is through asking the question ‘Who does what to whom under which circumstances?’ (Butt et al. 2003:46). Experiential meanings are reflected through the grammatical system of *transitivity*, which organises them in a set of *process types*, outlined in Figure 4b. The different process types are represented in a circular way constructing the model of experience as a continuum in the semiotic space, where processes fuse into one another. The processes (e.g., actions, events, states) are realised through verbal groups, which may consist of one word, or several words and they account for the described events or ‘goings on’ in the clause. The participants in these processes are realised through nominal groups and circumstances, often through an adverbial group or a prepositional phrase (Butt at al. 2003, Halliday 2004). An example of how experiential meaning is encoded (Butt et al. 2003:48) is shown below:

1	The shop	closes	at six
	Participant	Process	Circumstance

2	The shop’s closing time	is	six
	Participant	Process	Participant

As illustrated in the centre of Figure 4b, all verbs conveying processes fall under three main categories: verbs in the sense of doing words, which describe actions and happenings in the material world; verbs in the sense of expressing feelings, thinking, and saying words, which essentially refer to cognition, perception and emotions; and verbs as being or having words. Following Halliday (2004) and Butt et al. (2003) these main types of processes can be subdivided into the following subtypes: *material*, *behavioural*, *mental*, *verbal*, *existential* and *relational*. Relational processes, in turn, subdivide into *relational attributive* and *relational identifying* (see Figure 5). With respect to expressing experiences, processes are associated with particular kinds of roles and configurations and involve at least one participant who can be a human, a non-human or a non-animate entity. All process types follow practical rules and mandate a particular set of participant roles. Once the type of the process has been identified, the different roles arrange accordingly. For instance: when the process is material, its doer is *actor*, when the process is mental, its doer is *sensor*, when the process is verbal, its doer is *sayer*, and when the process is behavioural, its doer is *behave* (Butt et al. 2003). In the following sections below, each of these types of processes and their functional labels (roles) will be clarified with examples. All the following examples of processes 1-46 are taken from Butt et al. (2003:48-63) if not indicated otherwise.

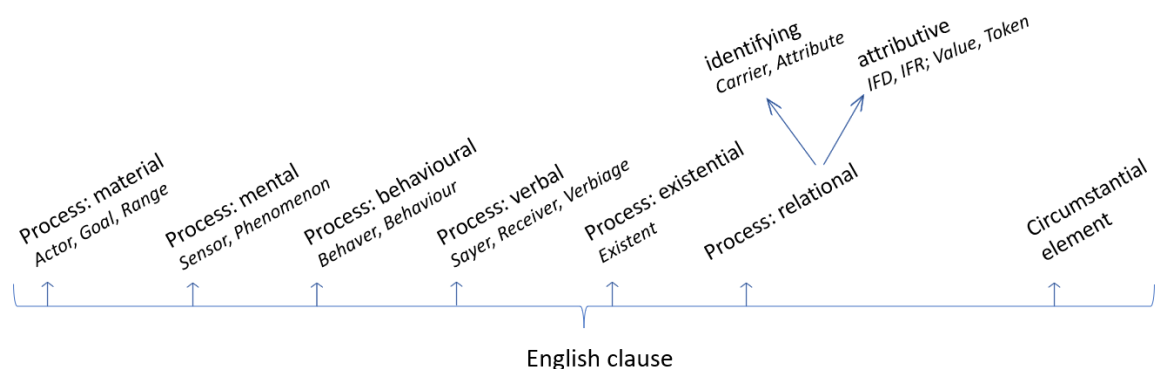


Figure 5: Processes and main roles in the clause within the transitivity system network. Based on Butt et al. (2003:50-65)

2.2.1.1 Material processes

Material processes involve experiences related to concrete doings or happenings from the external material world. They thus include verb forms like: ‘jump’, ‘arrived’, ‘collapsed’, ‘works’ etc. and are most of all very common in action-oriented interactions or text narratives. Verbs or verbal groups conveying material processes should answer questions such as: ‘What did X do?’ or ‘What happened?’ (cf. Butt et al. 2003:52), as shown in examples 3 and 4 below: what did John do, or what did they do?:

3	John	arrived
	Actor	Material process

4	They	ate	at noon
	Actor	Material process	Circumstance

Material processes are about doing an action and involve an *actor* (doer of the deed/performer of the action), a *goal* (the entity affected by the process), a *range* (the entity not affected by the process), and a *beneficiary* of the process. The above examples involve a single actor as the participant of the process; examples of a material process involving an *actor* and a *goal* are listed below.

5	The dog	was chasing	the cat
6	Mother	made	a fruit cake
	Actor	Material process	Goal

Material processes with a *goal* can be active, as in the example above, or passive as in the example below. The *goal*, which is a direct object in terms of traditional grammar, is the entity affected by the process and to which the process is directed. The *goal* traditionally answers the question ‘What did X do to Y?’. In a passive construction the *goal* usually changes to the subject of the clause, and the *actor* of the process can be left out, as in the examples 7 and 8 below.

7	The cat	was being chased	(by the dog)
	Goal	Material process	Actor

8	The fruit cake	was made	(by mother)
	Goal	Material process	Actor

In passive constructions, the *actor* of the passive clause is called an *agent* instead, (however both terms are used interchangeably), and as indicated, passive constructions enable speakers to omit mentioning by whom a certain action was done; see the examples above where the *actor* is placed in brackets. Such passive constructions are called ‘agentless passive’ (Butt et al. 2003:53) and considering the nature of SFL as a system of choices, it is up to the speakers whether they will aim to neglect such additional information depending on how and what meanings they intend to ascribe through linguistic structures. Within material processes, a *beneficiary* role represents a

participant that is benefiting from the performance of the process and might be subdivided into *recipient* and *client* (Butt et al. 2003, Halliday 2004). The *recipient* is in the role of receiving something from someone and the *client* is in the role of getting something done by someone⁵. Both might occur with or without a preposition in the text (respectively *recipient*: to; *client*: for), depending on their position in the clause as shown in examples 9-11.

9	I	posted	a letter	to a friend
	Actor	Material process	Goal	Beneficiary: recipient

10	I	posted	my friend	a letter
	Actor	Material process	Beneficiary: recipient	Goal

11	The architect	built	a house	for his mother
	Actor	Material process	Goal	Beneficiary: client

The *Range* as participant can be a single structure, a form or an entity which is not affected by the process. It can represent an independent entity from the process, only signifying the essence of the process itself or showing its domain. In example 13 below, the verb ‘did’ does not necessarily have a lexical meaning, it can be thought as of semantically weak or empty in this context, and it takes its substance from the *range* (‘some research’), which serves as an extension of the process.

12	Jack	is climbing	the fence
13	She	did	some research
	Actor	Material process	Range

It is essential to point out, however, that the above outlined process types and their attached participant roles are discussed in a rather fuzzy way, where sometimes no obvious, clear-cut distinction among the different processes or other categories is possible in practice. Such an example is the very close similarity between a *range* and a *goal* (Butt et al. 2003). In fact, Halliday & Matthiessen (2014) discussed the functional elements that configure transitivity categories as inherently fuzzy to a certain extent and noted that the different process types (cf. Figure 4b) naturally have blurred borders, i.e., instead of being precisely delineated entities, they should be

⁵ T. Bloor & M. Bloor (2004:113) noted that *beneficiary* should not be understood literally in a sense of the positive connotation of the word ‘beneficial’, as here it is not necessarily positively connotated.

seen as continuous ones, one merging into the other. Systemic terms should in principle be perceived as ‘fuzzy sets rather than ‘crisp’ ones’ (Halliday & Matthiessen 2014:218).

2.2.1.2 Mental processes

In contrast with material processes, mental processes do not describe physical actions, but abstract doings and happenings as part of humans’ cognition, perception, inclination, or liking/disliking (also called affect), which are grammatically treated as actions. These types of sensing constitute the system of sensing in which they shade into one another. Accordingly, we can probe mental processes by asking ‘what do you think/know/feel about x?’ (Eggins 2004:225), which shows that here doings and happenings are envisaged and presented as mental operations and perceptions rather than as physical entities. The potential participants of mental processes are a *sensor* (doer of the process), fulfilled through a human or a conscious participant and a *phenomenon*, fulfilled by a nominal group or embedded clause which implicates what is experienced, thought, wanted, perceived and so forth (Butt et al. 2003). In cases when mental processes project a separate clause, the mental process is reflected in the projecting clause (e.g., I know, I see etc.). The projected clause, in turn, can reflect any kind of verb process, as shown in examples 20-23 below (Butt et al. 2003:56). Mental processes are typically expressed through verbs such as: see, know, want, like, please, think, feel, smell, hear, hate, admire, enjoy, fear, frighten and so forth, outlined below in examples 14-23.

14	Annie	knows	the answer
	Sensor	Process mental: cognition	Phenomenon

15	Jessica	likes	ice-cream
	Sensor	Process mental: affect	Phenomenon

16	Austin	hears	the ice-cream truck
	Sensor	Process mental: perception	Phenomenon

17	Henry	wants	an ice-cream
	Sensor	Process mental: inclination	Phenomenon

Listed below are some examples of mental process with embedded clause. Conventionally, double brackets symbol marks the embedded clause.

18	Annie	knows	[[what she wants]]
	Sensor	Process: mental: cognition	Phenomenon

19	Austin	hears	[[the ice cream truck coming]]
	Sensor	Process: mental: perception	Phenomenon

Examples of mental process projecting direct thought as a separate clause:

20	'Why am I tired?'	she	wondered
	Projected clause	Sensor	Process: mental

21	I	thought	'he looks ridiculous'
	Sensor	Process: mental	Projected clause

22	She	wondered	'Why am I tired?'
	Sensor	Process: mental	Projected clause

An example of a mental process projecting an indirect thought as a separate clause is shown below:

23	She	wondered	why she was tired
	Sensor	Process: mental	Projected clause

2.2.1.3 Behavioural processes

Behavioural processes are usually processes of physiological and psychological behaviour and unlike the other process types, have no clearly identified characteristics on their own that make them distinct, as noted by Halliday (2004:250). Thus, they realise meanings in the grey area between material and mental processes. Typical verbs which might express behavioural processes are for example: laugh, frown, breathe, dream, look over, worry, argue, smile, sniff, cry, stare, taste, watch etc. However, Eggins (2004:233) noted that although there are obvious similarities and a close relationship between mental and behavioural processes, sometimes lexical items of behavioural processes may contradict synonyms for mental processes, like the verbs: *look at* – behavioural, *see* - mental; *listen to* – behavioural, *hear*- mental. They typically have one obligatory participant - a *behave*r and express a form of doing which does not intervene with another

participants, for instance: ‘Don’t breathe!’, ‘No one’s listening’, ‘He’s always grumbling.’ (Halliday 2004:251). As a rule, the *behave*r is a human being (if not the clause is seen as a personification) and takes a similar role here like a *sensor* takes in a mental process clause (Egins 2004:233). Rarely however, a second participant called *behaviour* is possible, fulfilling a *range*-like function, to restate or extend the process. Examples of behavioural processes and their different roles follow bellow in examples 24-27.

24	The woman	laughed	
	Behaver	Process: behavioural	

25	The cat	sleeps	on the back veranda
	Behaver	Process: behavioural	Circumstance

26	Betty	cried	bitter tears
	Behaver	Process: behavioural	Behaviour

27	The volcano	slept	
	Behaver	Process: behavioural personification	

2.2.1.4 Verbal processes

Verbal processes are processes of verbal actions, fulfilled respectively through verbs such as saying and telling and their synonyms (call, compliment, tell, ask etc.). The person who produces the utterance is in the role of the *sayer* - the doer of the action and the one responsible for the verbal process. The process is directed to the *receiver* as addressee of the speech, a *target* is the object of the talk or act, and the *verbiage* is a noun expressing the type of verbal behaviour, such as a statement, questions, an answer etc. (Egins 2004:235). According to Butt et al. (2003:57), a verbal process usually reflects what is said in a separate clause. Examples of verbal processes and their roles follow below:

Example of verbal process with nominal group:

28	She	said	her piece
	Sayer	Process: verbal	Verbiage

Example of verbal process with embedded clause:

29	Chris	said	[[what he had to say]]
	Sayer	Process: verbal	Verbiage

Below is an example of a verbal process projecting direct (30) and indirect (31) speech as a separate clause. The clause reflecting the verbal process is the projecting clause, and in the projected clause, there might occur any other type of verb process.

30	They	asked	'Where are you going?'
31	She	said	that she was tired
	Sayer	Process: verbal	Projected clause

Example of verbal process with *receiver*:

32	Isabella	told	the secret	to her best friend
	Sayer	Process: verbal	Verbiage	Receiver

Example of verbal process with *target*:

33	Marcus Antonius	praised	Julius Caesar
	Sayer	Process: verbal	Target

To the group of verbal processes belong verbs like suggest, claim, report, manifest, document, request etc. In addition, 'saying' should be understood in a broad way, and it can include verbs addressing any kind of symbolic exchange of meaning or metaphoric expressions, like 'the notice tells you to keep quiet', 'my watch says it's half past ten' (Halliday 2004:253).

2.2.1.5 Existential processes

All the processes described above until now reflect actions, happenings or events of some kind. Existential and relational processes encompass the rest of the verbs in English, which encode meanings about states of being. The existential processes describe things that are simply stated to exist, and there is only one known participant, who is the *existent*. As shown in examples 34-35, this is achieved through introducing the structure 'there is/was something', which occur at the beginning of the text or in other parts, which mark new phase of the text. Within the system of transitivity 'There' does not have a representational function and is thus left empty.

34	There	's	a strange smell
35	There	are	several difficulties
		Process: existential	Existent

2.2.1.6 Relational processes

As explained previously, experience might be construed through material clauses, concerned with experience from the physical world, through mental clauses, concerned with experience of our consciousness and through relational clauses, representing experience through the world of abstract relations (Figure 4b, inner circle). In line with the theory, any experience can be represented with either of the three main process types, depending on the choices made by the speaker. There are two main modes of relational clauses: one that relates a participant to its general features or description, called *relational attributive*, and the other, which relates a participant to its identity, role or meaning, called *relational identifying*.

Following Butt et al. (2003:58), in relational attributive mode, characteristics and qualities are attributed to an entity. These characteristics and qualities are called *attributes* and are ascribed to a *carrier* (the person or entity which carries them). The carrier is typically realised by a noun or a nominal group. For instance, illustrated in example 36, there is the *attribute* (very heavy) and the *carrier* (the bookcase) and the process, construed through the linking verb 'looks'. Attributive processes are primarily realised by any form of the verb 'to be' or by other copulative verbs, linking the subject of a clause with the subject complement. Such verbs are for instance: seem, appear, look, become, taste, sound, smell, feel etc., but sometimes as well by other verbs like have, possess, etc.

36	The bookcase	looks	very heavy
37	The bookcase	seems	a fine piece of furniture
38	The office	is	sumptuous
	Carrier	Process: relational	Attribute

The other mode of relational clauses is *relational identifying*, which establishes an identity, role or meaning of participants. The grammatical and semantic difference between the two modes is that a *relational identifying* clause defines entities whereas *relational attributive* ascribes classes.

Halliday & Matthiessen (2014:265) provided examples of how both modes differentiate, in Table 1. These examples show one of the most distinct features of identifying clauses, namely that they are reversible in contrary to attributive clauses.

Relational Attributive 'a is an attribute of x'	Relational Identifying 'a is the identity of x'
Sarah is wise	Sarah is the leader - the leader is Sarah
Peter has a piano	the piano is Peter's - Peter's is the piano
the fair is on a Tuesday	tomorrow is the 10 th - the 10 th is tomorrow

Table 1: The principal categories of relational clause. Source: Halliday & Matthiessen (2014:265)

Relational identifying processes have two sets of functional labels which are not exclusive but interrelated: *identifier* and *identified* and the other set: *token* and *value*. Token and value can both be used to identify the other, as shown in examples 39-40 adapted from Halliday & Matthiessen (2014:279). In the first example, Tom is identified by assigning him to a value, and in the second example, Tom is identified by assigning a value to him. The more generalised entity is always the value, whereas the more specific one is the token.

39	Tom	is	the treasurer
	Token	Process: relational identifying	Value

40	Tom	is	the tall one
	Value	Process: relational identifying	Token

Some verbs that can relate a *token* and its *value* are for example: be, mean, spell, express, play, act, show, represent and so forth. When one of these verbs is in the active voice, the clause structure should be defined as: *token, process, value*. When the verb is in the passive voice, the clause structure should be defined as: *value, process, token*. Unlike relational attributive processes, the participants in identifying processes are usually definite or proper nouns (Butt et al. 2003:59).

The other set of labels assigns a new identity. One entity is used to identify another entity, namely 'x is identified by a', or 'a serves to define the identity of x', where x is the *identifier* and a is the *identified* entity. In the case of 41, 'the room on the left' identifies 'your office' and in example 42, the structure is reversed. Regardless of its position in the clause, the nominal group about which the question is asked is the *identified* and the new identity is the *identifier* (Butt et al. 2003:59). Both sets of labels can be applied together in an analysis.

41	Your office	is	the room on the left.
	Identified	Process: relational identifying	Identifier

42	The room on the right	is	your office
	Identifier	Process: relational identifying	Identified

2.2.1.7 The circumstantial system

Along with process types and participants, circumstantial components concerned with the setting, or the temporal and physical manner in which a process is implemented, are realised in clauses. As illustrated on Figure 4a, circumstances in which meanings are construed are positioned in the outer part of the experiential system. As shown in examples 43-46, circumstantial elements provide additional information about the given context of a situation and answer questions like where, when, how, why, with whom or as what process occurred. Speakers might choose to present circumstantial information in a separate clause; however, it is typically built in the same clause through a prepositional phrase, an adverbial group or sometimes a nominal group (Butt et al. 2003:64). Any additional information provided in the clause can be considered as a *circumstance*.

43	The premier	arrived	at the end of the meeting.
44	Harry	was walking	along the road.
45	My mother-in-law	lives	in Perth.
	Participant	Process	Circumstance

In contrast to other elements of the experiential system, *circumstantial elements* occur freely in any types of processes and can be rather flexible. There might be more than one structure of circumstantial elements in an utterance, illustrated in example 46. Circumstantial elements can shift in a different order without causing change in the semantic load of the utterance; the example below, for instance, can be reversed to: Every week Charles sings in a choral group; or Charles sings in a choral group every week.

46	Charles	sings	every week	in a choral group
	Participant	Process	Circumstance	Circumstance

Circumstances serve to enhance the process, provide additional details about it and tie it with the situation in which the experiential meanings are created. They facilitate us to better understand the essence and the purpose of the message being conveyed. Through circumstances, a process can be defined in terms of space, time, manner, degree, and frequency. Butt et al. (2003:65) summarised the types of circumstantial elements:

Type of circumstance	Answer the question	Examples
Extent	How long? How far? How many times?	For two hours, for two miles, five times a week
Location	Where? When?	In the yard; after dinner
Contingency	In what?	In case of rain; in spite of rain; in the absence of the fine weather
Cause	Why? What for?	Because of the rain; for a rest
Accompaniment	With whom? And who else? But not who?	with a friend; as well as Henry; instead of Michael
Matter	What about?	about suffering
Role	What as?	as a clown
Manner, Means, Quality, Comparison	How? What with? How? What like?	by car; with a stick; quietly; like a trooper
Angle	According to whom?	to Mary; according to Luke

Table 2: Overview of circumstantial elements. Source: Butt et al. (2003:65)

2.2.2 Interpersonal metafunction

Except for enacting their experiences through language, humans also engage in communicative acts with each other, take on different roles, and convey and experience feelings, attitudes, and judgements to express interpersonal meanings. The interpersonal metafunction is concerned with any form of social and personal relationships among humans. According to Butt et al. (2003:86), there are two main domains of the interpersonal meanings, depending on whether language is currently used to exchange information, or depending on whether it is used to exchange goods and services. In a similar way, there can be two different types of interactions within these domains: when speakers demand services or information and when speakers give services or information. In the linguistic environment of a clause, the interpersonal meanings find expression through the grammatical systems of mood and modality. The two main grammatical entities in interpersonal structures are the subject and the finite, and they combine to make the mood of a clause, as explained by Butt et al. (2003:88).

As discussed above, in terms of experiential structure of the clause, nominal groups serve as participants and verbal groups serve as processes. In an experiential sense, verbal groups within declarative clauses function as finite followed by optional auxiliaries followed by an event verb,

where the event verb is the main item, and in an interpersonal sense, verbal groups function as finite followed by a predicator (includes auxiliaries) (Butt et al. 2003:89).

Unlike the experiential system (i.e., expression of what happened) where the event verb is the most important part of the verbal group, in the interpersonal system (expression of how the speaker relates to what happened), the finite element is the most important part. It performs two main interpersonal functions – shows time in relation to the speaker (conveying information about tense), or it can show a speaker's opinion through its modal function. In terms of mood choices, there are three basic interpersonal structures for clauses: interrogative, declarative and imperative.

2.2.3 Textual metafunction

The third, textual function serves to relate between what has been said or written in a text and the context. It shows the structure of the communicated information and reveals the relationship among discourse, the whole and the setting, i.e., it accounts for the cohesiveness and coherence of the text. It is intrinsic to language and a prerequisite to allow for language to fulfil the other two metafunctions efficiently. *Theme* and *rheme* are the main two concepts introduced by Halliday, which define the thematic structure in text organization (cf. Butt et al. 2003:134-150).

Methodologically, this study concentrates primarily on experiential processes realised by verbs and verbal groups in terms of the transitivity system. Since the interpersonal and textual metafunctions are not centralised, theoretical foundations for these are not exhaustive.

2.3 Corpus Linguistics (CL) and web-based corpora

Corpus linguistics is a powerful methodology of using computational capabilities to analyse large samples of language from different varieties and genres. It introduces a wide scope of applications across different areas and theories of linguistics and is not restricted to a particular domain of language such as syntax or semantics etc. but can be applied to various kinds of studies. The term 'corpus' literally translates to body (cf. Latin), so corpus linguistics is the study of a body of language, which can be comprised of a large number of written or electronic texts representing a specific topic or a subject. Nowadays the developments in technology and the prevalence of internet forums, web applications and social media account for the abundant growing body of textual and visual information available to researchers to study different language and social phenomena using corpora from the web. In terms of volume and in terms of means for analysis, corpus linguistic tools, above all, make the task of using such corpora feasible. Zappavigna (2012) envisaged the World Wide Web as a system of interconnected hypertext

documents, which uses various types of links to establish semantic and pragmatic relations, and thus can be seen as a vast multimodal set of texts. By its nature however, raw and not manipulated, the web is not a corpus. A corpus is usually carefully designed and collected for a specific purpose and aims to represent an infinite set of language, based on a certain typology. McEnery et al. (2006:5) defined a corpus as:

‘a collection of (1) *machine-readable* (2) *authentic* texts (including transcripts of spoken data) which is (3) *sampled* to be (4) *representative* of a particular language or language variety’

As with any other corpora, web corpora have to conform to fundamental criteria of the corpus linguistic methodology, such as representativeness and balance, authenticity of the data and the sampling design used, which are closely intertwined with each other. However, using web-based data as a corpus carries advantages and disadvantages. Building such a corpus is not a trivial procedure and requires careful considerations of the different stages of data collection and processing. Moreover, depending on the purpose of the study, this task can become rather cumbersome, as firstly, a complete or sufficient set of web data for research purposes may not always be easily accessible, and secondly, web data, in particular data from social media, can be quite unsystematic, requiring additional efforts in comparison with conventional data sources and formats. Some of these challenges have been discussed in section 1.2.

Returning to the definition above, the function of the corpus and the purpose that it will serve define the frame for its sampling design according to McEnery et al. (2006:A7.1). The concept of data authenticity is in principle rooted in the traditions of corpus linguistics, exploiting genuine communication which occurs in a natural setting, i.e., language in use. Representativeness is an essential principle of a corpus. In that sense, a sample is considered representative if the findings from this sample are also valid for the whole population; or results can be extrapolated from the corpus to the whole entity (language users, variety etc.). Typically, a corpus is a sample of a considerably greater population. In the perspective of McEnery et al. (2006:19), population can be explained as ‘language production, language reception or language as a product.’ Leech (2007:135) defined a corpus as representative when ‘the study of a corpus (or combination of corpora) can stand proxy for the study of some entire language or variety of a language.’ In practice however, Zappavigna (2012) pointed out that language is a complex phenomenon, and on many occasions its scope cannot be unequivocally defined. Similarly, Paquot & Gries (2021:4) suggested that a compiled corpus should strive to be representative at a maximum possible level, which they define to account for ‘acceptably representative’ in practical terms.

Representativeness is a cardinal characteristic that differentiates a corpus from any random collection of texts, like text archives for instance. General corpora are such that reflect the overall diversity of a given language by being balanced in terms of the scope of language categories that they represent. The notion of balance is relative and depends on the research question, and provided that a corpus is general in nature, it should accordingly include a range of genres and domains and hence be balanced, whereas a specialized corpus on the other hand, is 'specialized relative to a general corpus' McEnery et al. (2006:60). The criteria used to collect a specialized corpus determine whether this corpus is domain or genre specific. A domain specific corpus focuses on specific content types or subject fields of the texts, e.g., academic disciplines (biology, history), professional fields (law, business), or like in the case of this study, the topic of crisis situations and specifically the flooding event in Texas and Oklahoma in 2015. A genre specific corpus stands for the genre of language data in the corpus, for instance newspapers, poetry, social media data and so forth. In these lines, specialised corpora are such that they use narrowing sampling parameters and are carefully composed collections of data, representing a sublanguage or a form of subtype (McEnery et al. 2006). Due to the fact that language varies across its domains and genres, specialised corpora are very effective instruments which contribute to a deeper understanding of the domains and genres which they represent. Ensuring representativeness of a corpus is a question of taking into account the two parameters of reaching a decent level of balance in the corpus and the choice of relevant sampling techniques. To recapitulate: in most cases natural language and its uses cannot be described entirely and exhaustively; hence, to be able to make statements and answer research questions about language, we need to sample it according to the relevant research question, and in that way achieve reasonable levels of balance and representativeness (McEnery et al. 2006:2:2.5).

Tognini-Bonelli (2001) laid out a two-lined pathway within corpus linguistics: corpus-based and a corpus-driven approach. From a corpus-driven standpoint to corpus linguistics, an analysis of the corpus and the emerging descriptions and categories aim to be entirely comprehensive and analytical with respect to the data. Such an approach presupposes that the analysis is driven by corpus procedures and evidence, all theories and arguments stem entirely from the data, and are based on the data alone. In accordance with Tognini-Bonelli (2001:87), a corpus-driven approach 'aims to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context'. Furthermore, McEnery et al. (2006) pointed out in this regard, that this approach has been claimed to be in line with the corpus and strictly committed to it; however, on the other hand it might sometimes be perceived as idealistic and fairly extreme in making general statements, which are valid for a whole language. Corpus-based

studies, in contrast, use corpora to improve, test and refine pre-existing theories, descriptions or hypothesis and imply carrying out targeted analysis on pre-selected attributes in the corpus. Within a corpus-based analysis, the researcher's commitment to the data is not entirely strict or systematic and corpora often serve to 'validate certain categories or to supplement the theory with a probabilistic dimension' as described by Tognini-Bonelli (2001:81). Apart from being criticised as sometimes disregarding unfitting the researcher's pre-existing theories textual evidence from the corpus, a corpus-based methodology has also been seen as occasionally favouring certain textual proof over other.

However, McEnery et al. (2006) argued that the differences between both approaches are in fact less obvious and fuzzy (see McEnery et al. 2006:A 1.7 for further discussion). This view is also adapted and supported by Baker & Egbert (2016:11), who agreed 'that it is very difficult to carry out a truly 'naïve' corpus-driven analysis and that most forms of analysis appear somewhere on a cline between the two extremes.'

In the case at hand, the nature of the study and the multilayered analysis presuppose a combination of a corpus-based and corpus-driven approach. On the one hand it is corpus-based due to the nature of the corpus and the steps taken to collect and annotate the data, but also due to the functional qualitative analysis in the frame of SFL which will be carried out. While certain levels of the analyses which will be performed consider the entirety of the corpus, other levels take advantage of researching smaller sets of language features and linguistic categories. On the other hand, a purely inductive, corpus-driven approach is used to create semantically fitting categories of information inferred from the corpus. Studies incorporating a combination of both approaches have been outlined by Baker & Egbert (2016:11-12).

2.4 Corpus linguistic tools and techniques

By means of various corpus techniques, corpus linguistics can provide empirical evidence in a quantitative way about language within a collected corpus. In this section some of the most widely used techniques which are applied in the frame of this master's thesis are described.

In order to extract linguistic information, to facilitate complex analysis and to examine different language phenomena in an electronic corpus, additional interpretative linguistic information can be encoded in a corpus. The process of adding such information is typically achieved through automatic or manual tagging of the corpus and is known as linguistic annotation. Usually, corpora appear to be too large to achieve linguistic annotation manually, so automatic annotators are often used. Depending on the purpose of the research, tags can be attributed to multiple elements from the corpus, e.g., words, phrases, paragraphs etc. through other linguistic units.

Tags are usually assigned through codes and can account for grammatical, semantic, or textual categories (Paquot & Gries 2021). Linguistic coding or linguistic annotation is a practical step in carrying out linguistic analysis within large electronic corpora. There are several main types of linguistic annotation: part of speech tagging (POS), lemmatisation, syntactic parsing, and semantic annotation (Paquot & Gries 2021:26). The choice of annotation type introduced to a corpus is associated with the specific research question for the study. In the case at hand, the first two and most common types of linguistic annotation will be applied to the collected corpus.

Part of speech (POS) tagging refers to categorizing words in the corpus through assigning grammatical categories (e.g., articles, nouns, verbs, adjectives etc.) to them. Lemma annotation or lemmatisation is used to reduce inflated lexical items, word compositions and other morphological forms to their basic lexical forms: lexemes or lemmata. In a lemmatised corpus, one can make search requests for a base form of a lexical item and obtain all derivative forms of this item, e.g., go, went, going, etc as results. Lemmatisation allows for locating the searched information more easily, especially in corpora where words may appear in multiple forms (Paquot & Gries 2021:30).

Admittedly, one of the benefits of exploiting electronic corpora is that using the appropriate corpus linguistic tools will facilitate the analysis exposing a quantitative aspect of the data. McEnery & Hardie (2012:49) pointed out that the majority of corpus linguistic studies usually generate basic descriptive statistics to facilitate corpus analysis, even if the latter are rudimentary and serve to describe the data in a simple way. Such quantitative analysis are typically followed by qualitative analysis and functional interpretations of the counts, in order to explore the importance of these findings in a corpus.

Underpinning the analysis of corpora, frequency of occurrence is one of the fundamental statistics in corpus linguistics. Frequency lists exist for various language structural units and classes of a language and are applied to address a wide variety of research questions in a number of linguistic domains. The latter can shed light on multiple interesting phenomena and are an effective starting point when investigating any corpus (Baker 2006). McEnery et al. (2006:52) explained frequencies as simply the 'arithmetic count of the number of linguistic elements (i.e., tokens) within a corpus that belong to each classification (i.e., type) within a particular classification scheme'. Frequency lists of words usually consist of listings of the words in a corpus, their frequencies, their percentages as units of the total corpus and serve to reveal the most common or uncommon lexical items in a corpus. The majority of the concordancing tools can generate frequency lists which can be alphabetically sorted (employed for instance to quickly locate

particular items, or to check for spelling variations) and ascending or descending frequency sorted (shows the general features of a corpus which can be primarily selected for further analysis or highlights unusual phenomena in the corpus). Frequencies are relative (McEnery & Hardie 2012:48). Considering a word list, for instance, the relative position of a word on the list indicates its frequency in a given corpus. Notably, frequencies will vary across different corpora depending on various factors (such as genre, register, topic, or time period). Therefore, it may be beneficial to compare the frequency of a specific word in this corpus with its frequency in another corpus that has a different linguistic profile.

Although frequencies have proved to be a very valuable measure, they need to be analysed sensitively as comparing frequency lists (their profiles) from different corpora based on raw numbers alone can be misleading. A very important factor is the size of the corpora that are compared; if one corpus is larger than the other, the frequencies need to be normalised to make a meaningful comparison. In such cases it is more sensible to calculate raw frequencies as a percentage of the occurrences per number of words (McEnery & Hardie 2012:50).

The principle of comparison of frequency lists from different corpora is referred to as keyword or keyness analysis. As explained by Baker (2006:125), this procedure involves comparing frequency values from wordlists to identify which words or lexical items occur statistically more often in a certain wordlist, compared to another one. Afterwards, the lexical items which occur more frequently than expected in one corpus in comparison to another, are compiled into another list, which is called a keyword list.

Thus, a keyword or a key item is defined as a lexical item which occurs more often in a given *target corpus* when compared to a *reference corpus* of some kind (cf. McEnery 2016). *Keyness* shows the quality of words or phrases to serve as 'key' in the specific context in which they are employed and thus provide a linguistic representation of socially important and significant conceptualisations within this given context (cf. McEnery 2016, Baker 2006, Gabrielatos 2018). Furthermore, the terms *keyness* and *aboutness* are closely related in a corpus linguistic sense and they both refer to characteristic concepts, topics or attitudes discussed in a given corpus (Baker 2006, Gabrielatos 2018). In the sense that any lexical item in a corpus can be considered potentially of interest, McEnery (2016) stated that the principle of keyness analysis can also be purely corpus driven, depending on the analytical approach.

The keyword procedure usually uses statistical metrics to determine if a lexical item is significantly more frequent or not. One of the several statistic metrics which can be used to measure frequency differences is the log-likelihood significance test (Baker 2006:125, Scott 1999), which

assigns each lexical item a probability value, called *p value*. The *p value* indicates the confidence that a word is a key word on a random basis. Thus, the smaller the *p value* is set, the higher the likelihood that the word's strong presence in any of the compared corpora isn't due to chance, but a result of the author's choice to use that word repeatedly (Baker 2006:125).

According to Scott (1999:72), among other tests, the log-likelihood test provides better estimations of keyness when two corpora of different sizes are contrasted, which is the approach in this study. This will be further discussed in the analysis section 4.1.

To further examine the significant words or phrases in a corpus qualitatively, one might look into the semantic or lexical contexts in which these words have been employed. Concordance analysis is another technique for exploring occurrences and lexical behaviour of different word forms in their lexical contexts, here context relates to the linguistical environment of words or co-text as referred to by Weisser (2015). As output, concordance lines generate listings of individual or a bundle of lexical items, in fact serving as search queries, and their co-text. A concordancing tool typically displays the results in a single line format. The co-text of a lexical item can fluctuate and might be adjusted to fit the desired extent correspondingly to whole sentences or a paragraph or it might be reduced to a fixed number of characters to the left and the right of the searched word form(s). The latter approach is one of the most widely-used methods and is known as a keyword-in-context (KWIC) structure. The search term might be a single lexical item or a series of such items in a row (McEnery & Hardie 2012).

2.5 Contextualization of Twitter microblogged communication as a corpus

If multiple language units above and beyond a single utterance are put together, and if these reveal the connection of a specific context of social interaction, then we talk about discourse according to Martin & Ringham (2006), aligning with Halliday's views. Combined units constitute texts and interactions, which in turn shape discourse and reflect social contexts and their social practices. As noted, in a semiotic sense, a text (written or spoken) is defined broadly as a semantic unit, and a harmonised collection of thematically binding meanings, which reflect and are reflected by the context of a certain situation. From a social semiotic perspective Halliday and Hasan (1985) defined text as any instance of naturally occurring language, which is functional, i.e., it plays a role in a specific context of a situation in which it is serving to convey meanings.

In a similar way, the compiled microblogged data for this study are thematically and temporally bound to the particular context of the Memorial Day floods and represent participants' information exchange, attitudes, opinions and judgements in the specific context of the flooding. As set forth in section 3.3, the relevance of the data in the compiled corpus has been further

improved by means of hashtags, which, above all, serve as topic-markers (Zappavigna 2015a) in different discourses and contextual situations in Twitter. The data illustrates the gist of the crisis and the urgency of the situation occurring at that time, and through its peculiar writing (e.g., #, @) conveys different meanings as the discourse develops over time.

McEnery & Hardie (2012:3) pointed out that 'Corpora may encode language produced in any mode'. Bernardini et al. (2006) proposed four ways towards understanding and utilizing the concept of 'web as a corpus', also adopted by Zappavigna (2012:17). The latter are summarised below as follows:

1. Web as Corpus Surrogate: the web is used as a substitute to traditional linguistic corpora due to practical constraints, e.g., when no access to suitable existing corpora is possible or when available corpora are insufficient for the purposes. For instance, researchers in the field of translation might make use of web corpora as a reference tool in translation tasks.
2. Web as a Corpus Shop: the web is treated as a source of creating and assembling specialized corpora. Researchers query search engines to extract combinations of search terms, which are afterwards assembled into specialised corpora, e.g., in the fields of translation, terminology or linguistic analysis. This process can be automated at different levels.
3. Web as Corpus Proper: an approach, utilized as a means to study the nature of language on the web itself, by creating a sample which represents a set of the infinite language. This research paradigm contributes to the understanding of web-based language usage and provides insights into it.
4. Mega-Corpus/Mini-Web: this approach involves the creation of an extensive corpus containing a vast volume of web data, combining both a 'mini-Web' and a 'mega-corpus' perspective. It is characterized by its up-to-date content, basic annotation categories (such as POS and lemma tagging), advanced query capabilities, and relative stability. This approach serves researchers seeking to investigate language through web-based resources and those interested in exploring aspects of the web through language analysis.

As a source of written communication, microblogged communication introduces the opportunity to investigate authentic social data and to explore different social and environmental phenomena shaping and shaped by human's everyday lives. So far, multiple linguistic or cross disciplinary studies have employed Twitter as a corpus for studying various linguistic phenomena, as pointed out in section 1.4. The most relevant are listed here as follows: Zappavigna (2011, 2012, 2015, 2018, 2022), and Page (2012, 2014).

3. Data collection and theoretical argumentation

3.1 Twitter data collection

Various methods exist for acquiring Twitter data, and several are discussed here to provide an overview of the available options. A commonly employed approach is collecting data through the Twitter Streaming Application Programming Interface (API), which allows for the retrieval of near real-time short messages. The Streaming API provides a randomized sample of data from the broader live stream of Twitter data (Littman 2017). To control the relevance of the output information, filtering techniques can be applied. For example, if the goal is to collect geotagged data, the search can be narrowed down based on specified latitude and longitude coordinates. Alternatively, when focusing on subject-based data collection, the search can be refined through relevant keywords, such as hashtags. This approach of filtering ensures that the acquired data aligns with specific predefined criteria.

Using search terms to navigate the collection of relevant data is typically a result of a preliminary qualitative analysis conducted on the Twitter stream. This may involve for instance identifying frequently used words or hashtags within the domain of interest to guide the input of relevant information. Additionally, it should be mentioned that Twitter's API access is subject to certain limitations and constraints, which include for example restrictions on the volume or frequency of data requests.

Besides the Streaming API, Twitter offers the Search API as another way to retrieve data. While the Streaming API provides a continuous flow of live stream data, the Search API is restricted to providing a sample of historical data, generated within the past seven days before the current date. In this sense the search API prioritises more the relevance of the data than its completeness, so certain messages or users may be missing from the search results. In addition, more recently Twitter has started charging for using its data APIs and provides less amount of data to users and researchers (Jingnan 2023).

In addition, an alternative method for acquiring information involves purchasing historical tweets directly from Twitter. Very often this is filtered data - subject based, geotagged or aligned with any other data queries specified by the user. However, for academic research purposes, this option can be relatively expensive.

The microposts selected for the purpose of this analysis originate from an existing archived data collection of Twitter data available on the internet. This collection comprises unfiltered data encoded in JSON format, extracted from the general Twitter stream for research, historical or

testing purposes. It encompasses random data obtained from the Streaming API, known as the 'Spritzer' version from the entire stream, representing approximately a 1% sample of all near real-time microposts. It is a random sample, collected without specific filters like geotags, hashtags or keywords. In this context no claims for completeness of the data are made for this research, which is also not the objective. The data used are published and freely available in the 'Internet Archive'⁶. The aim of this archive is to build a digital library of rapidly disappearing internet sites and various cultural artifacts in digital form and to provide researchers, historians, and scholars with free access to all materials. The curated archive comprises a diverse range of materials, including books and texts, web pages, as well as audio and video recordings, among other digital artifacts. This comprehensive collection serves as a valuable resource for those interested in studying and exploring digital content across various domains.

In order to proceed with the analysis, when data from Twitter have been collected, they need to be cleaned and organised, as they come fully bundled with metadata. In addition, a number of decisions depending upon the research question of interest, types of analysis and expectations need to be taken at this stage. With regards to crisis situations, collecting a new geographically bound and subject-based corpus through the Streaming API might be quite a time-consuming process and is currently not free of charge. Furthermore, as a result of a predefined search, the immense volume of data that the Streaming API will return will exceed the scope and analytical focus of this master's thesis.

Depending upon how data for the research are collected and upon the aim of the research, they can take different narrative forms and can be analysed in several ways. Two possibilities are either to aim at collecting threaded conversations, or at any randomized volume of data obtained from the Streaming API. Both types of data can be worthwhile in different ways, however, they can also have disadvantages, depending upon the research question. In the case at hand the obtained data from the archive come from the Streaming API, and its sampling algorithm generated a random selection of microposts as an output. Due to this randomization, most of the sequential nature of the streaming data is lost, as reported by Zappavigna (2012:23).

In this case it is not possible to trace threaded conversations. Nevertheless, the collected data are deemed suitable for this thesis's objectives and the scope of the analyses undertaken. The approach of using standalone snippets of data was also adapted from Vieweg (2012) and Zappavigna (2012, 2022). Single microposts of streaming data produced in a sequence can provide

⁶ Freely available and accessible from: <https://archive.org/details/twitterstream>

a wide range of information and message types related to the Memorial Day flooding discourse and it is assumed here that this approach will not limit the overall thematic content of the corpus significantly. The data are considered as sufficient for the exemplary purpose of this master's thesis.

It should be noted however, that according to the research question, analysing threaded conversations can provide a deeper and systematic semantic understanding of conversations and thus facilitate focused analysis to gain insights on these conversations and their participants.

3.2 Texas and Oklahoma flooding 2015

During the time from May to June 2015, record rainfalls and strong wind outbreaks caused flooding and chaos in Texas and Oklahoma, USA. Other states like Arkansas, Colorado, Ohio, Louisiana, Georgia and South Carolina reported heavy rains and severe weather storms as well (NOAA 2023). For many cities this was reported as one of the wettest months of May on record.

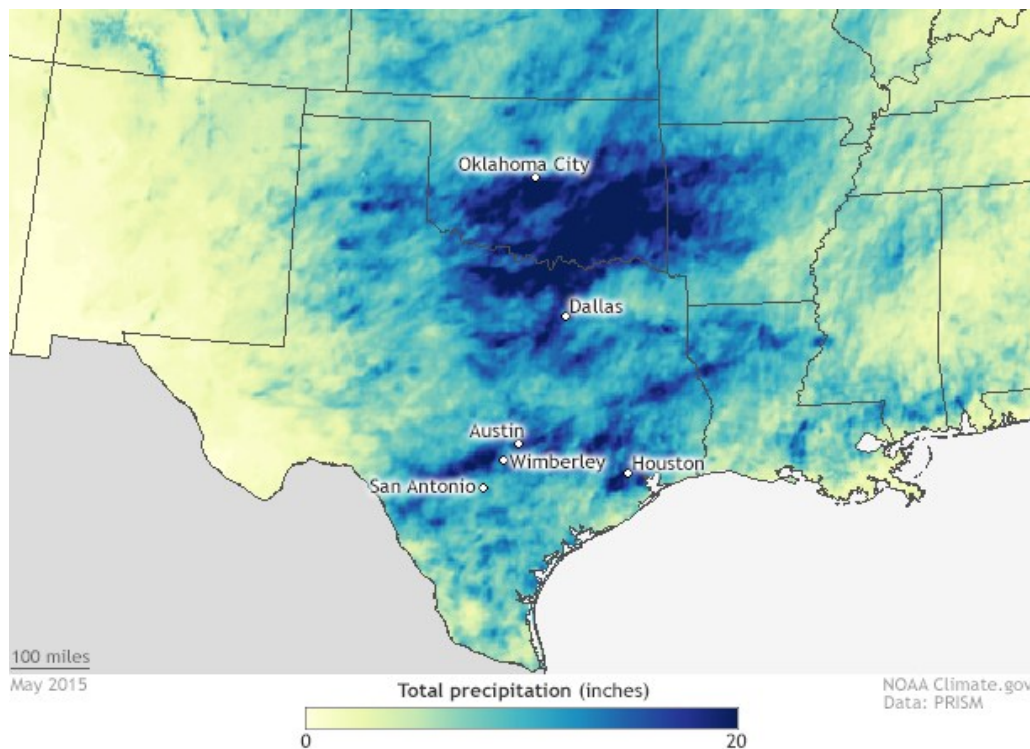


Figure 6: Total rainfall across the Southern Plains in May 2015. NOAA Climate.gov map, based on preliminary data from the PRISM climate group at Oregon State University (NOAA, 2015).

As illustrated on Figure 6, the entire states of Oklahoma and Texas received a high amount of precipitation in May 2015. Badly affected have been, in particular, major cities like Houston, Austin and Dallas with values of 4-12 inches of solid rainfall, received in a period of only one to two days in late May. Areas around Oklahoma city received 19.48 inches of rain during the month of May, making it the wettest month on record (the previous record was 14.52 inches of rain in

2013). Overall, in the month of May, there was approximately 200% to 400% more precipitation than normal for Texas and Oklahoma (NOAA 2015). The rain caused rivers and streams to swell and subsequently widespread flooding rapidly begun to unfold.

The critical period of rainfall was reported to be in late May, during May 23 and May 26 2015 (NOAA 2023), referred to as the Memorial Day floods. Initially badly affected were areas in eastern Texas along the Blanco River, San Marcos River and Buffalo Bayou (see Figure 7 for river levels) and eastern Oklahoma, where due to high river levels, water started flooding the streets, causing great damage and chaos among residents. Consequently, high amounts of water spread to the central parts of the two states and a flash flood emergency was issued.

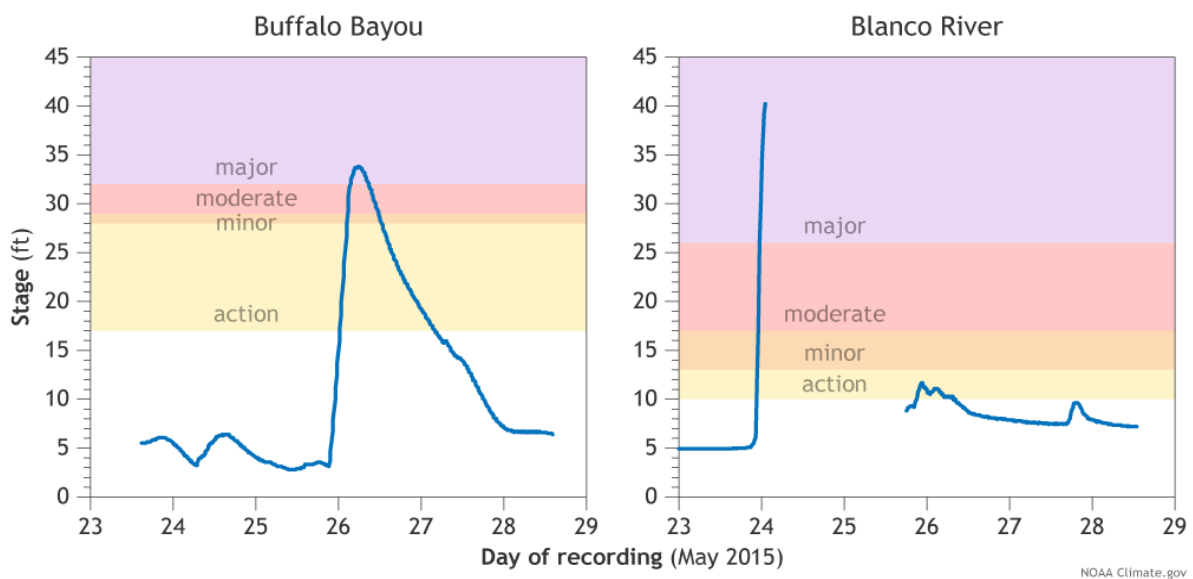


Figure 7: River levels on the Buffalo Bayou (left) in Houston and the Blanco River (right) in Wimberley, Texas, in late May 2015. The Blanco registered 40.21 feet—7 feet higher than the previous record—before the gauge went offline (NOAA, 2015).

3.3 Building the corpus

According to the former report of NOAA (2023), the most intensive period of rainfall and flooding in Texas and Oklahoma was estimated to be from May 23 until May 26 2015. Considering this timeframe, the relevant timespan of the data for the research selected from the Twitter sample discussed above, is limited to one day before and two days after the reported intensive period, from May 22 until May 28 2015. Since the data are collected on a random principle and are not filtered to fit this study, several stages of preparation had to be carried out in order to build the on-topic corpus.

The initial phase involved the implementation of various data management techniques, realised with the technical support of a colleague from Spatial Focus. Initially a tweet Extractor and Analyzer toolbox was developed in order to clean the aggregated data and to identify the relevant

tweets from the available dataset. Each raw tweet was bundled together with a number of attributes, from which the following appear as the most significant when collecting the data of interest: message id, datetime (created at), text, author screen name, GPS coordinates, location, reposted status and the reposted text. All these attributes are very important for assembling the relevant corpus. Beyond the content of microposts, metadata offers simple contextual variables which provide complementary information and can be used to either aid or make the analysis possible depending upon the research questions (cf. Page et al. 2022). Very important for the analysis of this corpus is, e.g., alongside the text and reposted text attributes, the datetime attribute, which points to the exact date and time when a micropost was created. This is because the temporal perspective plays a central role for addressing the research questions in this master's thesis.

The raw dataset contained more than 26.5 million tweets, produced in the days between May 22 until May 28 2015. Since Twitter is a global public network, users can post micromessages in a range of different languages. The first step was to filter out all messages in other languages except English, resulting in approximately 8.5 million microposts, out of which around 1.7 million contained hashtags. In the data, more than 350 thousand unique hashtags were found to occur. As follows, hashtag labels appearing more than 18 times in the corpus have been reviewed and several spreadsheets of microposts containing these hashtags were produced.

Originally, five distinct groups of hashtags were created based on their semantic or syntactic differences. This pragmatic approach was adopted to facilitate access, further processing, and a description of the data. In a subsequent phase of the corpus development, these five datasets were coupled together:

1. The first set contained 9 311 microposts, which included rather abstract hashtags in the sense of not obviously being related to the event of flooding, such as for instance: *#climatechange*, *#faith*, *#missing*, *#prayer*, *#running*, *#fear*, *#escape*, *#storms*, and multiple kinds of derivatives and lexical units combined into longer constructions without spacing.
2. The second set comprised 2 713 microposts, and includes hashtag labels such as *#houstonflood*, *#sanmarcosflood*, *#atxflood*, *#txflood*, and other composita, containing information about physical locations. This dataset stands out as the richest in terms of content-related information in the context of this research.

3. The third set features 224 microposts with hashtag symbols, including entities such as *#warning*, *#tornadowarnig*, *#severethunderstormwarning*, *#emergency*, *#rain*, *#blancoriver*, and further lexical items combined into single units.
4. Furthermore, a fourth set of 21 651 microposts containing hashtags with place names like *#hounews*, *#okwx*, *#texas* and more originated from the collected data, although it didn't appear closely related to the context of floodings eventually.
5. The final fifth set accounted for 15 728 microposts, and encompassed all posts generated between May 22 and May 28 2015, with respect to their location (areas in Texas and Oklahoma). Unlike the four other datasets, the primary criteria for data collection in this set were not hashtag labels but GPS coordinates and location. Using location and GPS coordinates is a frequently applied technique to gather on-topic data from Twitter and to investigate geographical relevance. However, as anticipated in this case, this subset appeared to be the least related to the floodings compared to the other four, for reasons explained above, but primarily because the data were not filtered to fit the current contextual criteria when they were obtained from Twitter.

Following this, all five sets were inspected for duplicate tweets, ensuring that a tweet was not selected more than once based on various attributes or hashtags. Subsequently, manual reviewing of the tweets was carried out to determine the relevant ones, as outlined below in Table 3.

<i>Dataset</i>	<i>Raw data</i>	<i>Relevant messages</i>
1	9 311	53
2	2 713	2 524
3	224	5
4	21 651	489
Geo tagged	15 728	35

Table 3: Summary of datasets and relevant microposts.

Afterwards all tweets have been merged into a unified dataset and a further iteration of removing duplicates and spam posts was carried out. To refine the textual content of the corpus, additional measures were taken. This included the systematic extraction of hyperlinks from the text of the tweets. The hyperlinks were subsequently uploaded as metadata in the CQPWeb software for further reference and consulting during the analyses if necessary. Additionally, the corpus contained a limited number of emojis, approximately 35 of the microposts in total included emojis. Upon manual review, micropost with emojis were investigated. Although that graphical and visual resources are considered to contribute to the meaning making of a message or a text in

a semiotic sense, in this case they did not prove to influence the communicated meaning of these microposts significantly but were rather employed to put visual emphasis on what was already communicated by language. Analysing emoticons, emojis or any other graphical resources requires very inclusive methodological framework, which is out of the scope of this master's thesis. Because of these reasons emojis had to be excluded from the analysis here.

When messages were obtained from the Twitter API, they came with encoded or escaped special characters (cf. Zappavigna 2012:20). These had to be decoded to their initial form before uploading them to the concordance software.

Furthermore, retweeting in the context of this research is considered as an important device for sharing contextual information. Not surprisingly, retweeted or reposted messages dominate the data. From a textual perspective, microposts in the corpus were typically constructed of the symbol '@', followed by a username in the beginning, which denoted a reference to this person or account. Next, the main text followed, often concluding with hyperlinks at the end of the message.

Depending on the research questions, the types of analysis and the nature of the data, decisions on how to organise, sort and prepare the data are typically made on a case-by-case basis. Such decisions have a significant impact on the size of the corpus and on any subsequently carried out quantitative analysis (Page et al. 2022). Below the main steps underpinning the corpus development of the flooding events are summarised; however, some of these steps should be rendered as iterative processes:

- temporal filtering: from May 22 until May 28 2015,
- language filtering: English messages,
- hashtag identification: sorted out related hashtag labels and relevant messages and compiled them,
- data cleaning: multiple stages of cleaning up the data and removing duplicates or messages that seemed to be spam.

The final composed corpus consisted of 1692 carefully selected on-topic messages. Following Bubenhofer's overview (2006-2023) and CQPWeb System Administrator's Manual⁷, the corpus has been uploaded, set up and linguistically annotated in the web-based corpus analysis system CQPWeb. Several relevant attributes of the data described above, most importantly message ID

⁷ CQPweb System Administrator's Manual available at:
<https://cwb.sourceforge.io/files/CQPwebAdminManual.pdf>

and timestamp, were included as metadata. Additional layers of basic linguistic annotation with lemma and part-of-speech (POS) information were added, using the Tree Tagger (Schmidt 1994) tool⁸. As Page et al. (2022) outlined, many POS taggers do not work very well when applied to social media data and cannot manage the properties of non-standard spelling (like symbols #, @, punctuation) very well. Thus, the related POS categories for this study were additionally reviewed.

Furthermore, in order to identify which are the keywords in the collected corpus, and to provide reliable cues for its *aboutness*, a reference general corpus is needed. Scott (1999:70) suggested that keywords generally offer a reasonable indication of the content of the corpus. For this purpose, a freely available corpus of Twitter data collected by Hans Christensen⁹ is used. It contains approximately 2.3 million tweets, mainly collected between 2011 and 2012. The data were gathered by a web crawler¹⁰ and each micropost was tagged with the date and the time in which it was generated. Once the raw corpus has been created, it was further sorted and cleaned from duplicates, split into separate lines and made available for downloading. No processes of data preparation or further data iterations have been done here. In order to upload the reference corpus in CQPWeb, the same procedure as previously of lemma and part-of-speech (POS) tagging was done, using the Tree Tagger (Schmidt 1994) tool. In CQPWeb automatic frequency lists have been created for both corpora. The general reference corpus will serve to identify any lexical items which are more frequent and characteristic for the specific corpus of the Memorial Day floods in 2015.

3.4 Theoretical argumentation and practical approach

McEnery et al. (2012:167) reported on an increasing tendency within corpus linguistics, namely the coupling of corpus linguistic methods, theories and results along with a functional approach to language. Functional and cognitive linguistics have been used by corpus linguists as a framework from which to explicate the findings of a corpus analysis and to further inspire novel approaches of exploring corpus data. McEnery & Hardie (2012:167) pointed out that there is an obvious relation and 'rapprochement' between corpus linguistics and functional approaches in linguistic research. Their relationship is reciprocal in a way: not only have functionalist studies employed corpus techniques, but corpus studies have also adopted functionalist frameworks (e.g., McKoon & Macfarland (2000), McEnery et al. (2003), Xiao & McEnery (2004), McEnery & Xiao (2010),

⁸ Available at: <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

⁹ Available at: <http://corpora.epizy.com/index.html>

¹⁰ A search engine bot which systematically browses the World Wide Web for the purpose to index or collect content.

Stefanowitsch & Gries (2003, 2008)¹¹). Based on the previously laid-out theory, SFL is concerned with the functions of language used to communicate meanings and it poses questions such as: Who are the participants, what are the processes, what attitudes and orientations are expressed by speakers. Since their primary motive is to explore language in use, in that sense CL and SFL are rather compatible and that is why they work in synergy so well.

In a similar tenet, this study makes use of a functional perspective of language, which will be applied to make observations about the data collected, informed by fundamental principles of corpus linguistics. Corpus linguistic techniques are primarily applied for developing the corpus and for conducting keyword analysis, which then enables a quantitative orientation and an overview of respective aspects of the corpus. A framework combining these two approaches and employed to analyse Twitter messages can be essentially associated with that used by Zappavigna (2012, 2015, 2022). To restate briefly, as Butt (2003:46) concluded 'Language has an experiential function, so it has experiential meanings. We use the experiential function to encode our experience of the world; that is, we use it to convey a picture of reality.' According to Halliday (2004:176) 'processes are the most central element of the structure of situation.' (cf. Figure 4a). An approach drawing upon the concept of experiential meanings, more specifically exploring patterns of process types (cf. Figure 4b), allows for investigating how and what experiences are encoded in language through processes, grammatically realised by verbs or verbal groups.

Given the discussion mentioned in section 2.2 from Bloor & Bloor (2004:10) that 'What speakers are saying is closely related to the context in which they are saying it' and that meanings are not rigid but continuously adjusted to the context, here the assumption is adopted that in the current flood situation, the participants share such types of semantic information, which is closely related and describes the context of the situation in which they currently are, and their physical, ecological and social environment. In his Introduction to Functional Grammar, Halliday stated that one of the basic functions of language is to make sense of human experience and that while 'language construes human experiences' and names things it has the function to 'construct things into categories' (Halliday 2004:29). Turning back to the notion of processes, according to Halliday & Matthiessen (2014:213), each 'process type constitutes a distinct model or schema for construing a particular domain of experience as a figure of a particular kind'. In line with this reasoning, the aim of this study is to explicate, through linguistic discourse analysis and categorisations, constructed by means of processes experiential meanings, and the different

¹¹ Listed not exhaustively, for further discussion and references see McEnery (2012:171-189)

message types encoded by authors in their utterances. In practical terms, the analysis will be developed in three stages as described below.

Initially, corpus linguistic techniques will be employed to conduct keyword analysis of the corpus and to provide an overview of the keywords in the context of the Memorial Day floods, as represented in the online messaging service. In addition, emphasis will be put on the distribution of verbs and verb forms in the corpus. These will serve to establish first impressions of the verbs contained in the corpus.

Secondly, the structured analysis of experiential processes will facilitate the development of discourse analysis from a semiotic stand point. Thus, this analysis is composed of two parallel layers, which complement each other. While investigating aspects of experiential meanings through encoded processes in a formal sense (in terms of events, states and happenings), it will be also paid attention how this discourse is linguistically constructed, i.e., what types of language devices actors used in their communication to convey meanings during the flooding in Texas and Oklahoma, 2015. As a result, this chapter will start establishing different message types inferred from the corpus and will provide an overview and comparison of experiential processes during the selected timeframe.

Finally, a discourse analysis aims to aid the grouping together of the different kinds of information and message types, which will help to trace the dynamics of the discourse and how it unfolds. According to Zappavigna (2012: 39), the variable of time plays an important role in microblogging communication. Although focusing on hashtags, Bruns (2013) similarly has suggested several types of metrics, including temporal metrics with the purpose of a more systematic analysis of microblogged messages in Twitter. The approach of this study is strongly based on a temporal perspective. Employing a day-by-day frame in the analysis will aid the development of categories of types of information inferred from the corpus, which will be further analysed. Categories are discussed from a semiotic view, as patterns of ideational meanings directly derived from the data throughout the analysis. The combination of SFL and CL will allow for grouping together semantically fitting coherent instances of meaning making. Furthermore, information units will be established to better be able to trace and discuss experiential processes. As a set of methods, this configuration will assist in illuminating construed meanings.

For the purpose of the analysis, any single short post will be treated as a separate textual unit, and a clause will be discussed as any piece of language, which is centred around a verbal group, following Thompson (2014). For example, the following post is organised within three clauses: *'Sending lots of love to my friends in #houstonflood. Remember all too well being caught in 2001*

flood over US 59 just near downtown.' Below are outlined several further practical explanations with respect to the applied approach of analysing the corpus.

Verbs as part of long hashtag compositions, as well as hashtags which are constituted of full clauses and have unconventional spelling like *#StaySafeTexas*, *#StaySafe*, *#Wegotflooded*, *#IAmSurroundedByIdiots*, *#TurnAroundDontDrown*, *#RainRainGoAway*, etc. have not been considered in the analysis. However, single verbs employed as hashtags such as in '*#Pray for us in #Texas*' have been included. This is because long hashtags are typically employed as standalone and limit or lack the context, which will ignore the developed methodological framework of SFL, in comparison to single words employed as hashtags.

Due to the nature of the data, sometimes there are verbal groups occurring in incomplete utterances or such that are considered as not cohesive or semantically not coherent. These verbal groups were not taken into consideration in the analysis of verbs, since they are considered as ambiguous for one of above-listed reasons. One such dubious expression in the data is: *Loving all things Oklahoma*. After reviewing examples in which it was contained, it did not become evident in which cases it was employed by authors, or what meaning it should convey and was thus left out. Further practical explanations with examples of linguistic phenomena are discussed within the analysis and in the notes section afterwards.

Within the analysis of conveyed experiential meanings, leading elements are experiential process types. Because language is a complex social phenomenon, the system of verbal groups usually derives from the different metafunctions. In order to better understand and explicate construed meanings and to be able to provide argumentation for process classification and rendering, occasionally participants in the processes and other linguistic phenomena have been discussed.

4. Analysis

The analyses will start off with an overview of the corpus, employing the described above corpus linguistic tools and techniques in order to explore the corpus. Secondly, with a more specific focus on verbs, analysis of experiential verb processes in terms of SFL is carried out and finally these are followed by building the inferred from the corpus categories of information and message types.

4.1 Keyness and verb frequencies

The whole corpus consists of 1692 short posts, which contain information considered as relevant for the Memorial Day flooding. In the corpus, a significant number of posts are bundled with hyperlinks and 113 of these provide relevant information for the current situation of flooding only through their embedded links. For this reason, these posts have been left out from the analysis below, as they are out of the methodological scope employed here. They are, however, registered as containing related information in the corpus.

The working corpus consists of 1579 microblogged messages, stemming from the days between May 22 and May 28 2015. Figure 8 provides a breakdown of the posts in a day-by-day timeframe, which will henceforth become the fundamental and leading principle for the further analyses of experiential process types and information types in the data.

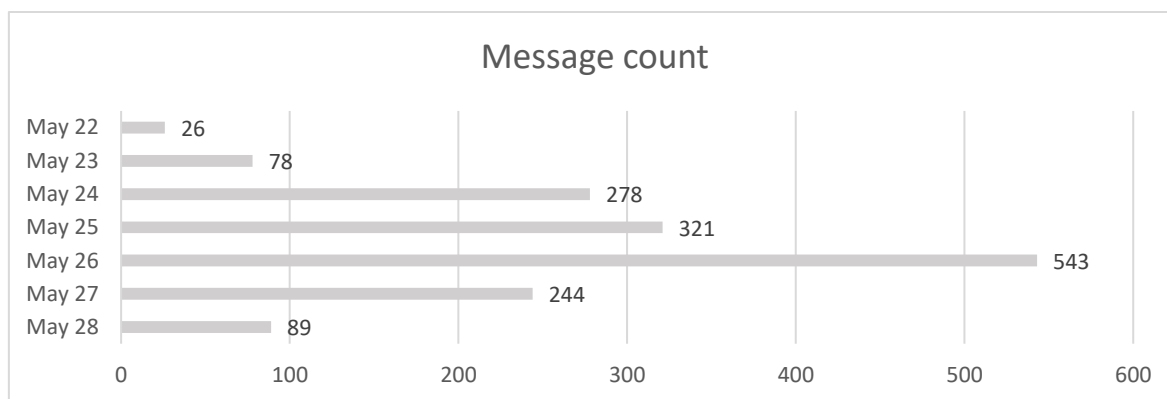


Figure 8: Message count Memorial Day floods, May 22 – May 28 2015

In order to highlight the characteristics and the *aboutness* of the collected corpus, as a starting point of the analyses a keyness comparison with a general reference corpus will be performed. For this purpose, both corpora - the investigated and the reference corpus - have been uploaded to CQPWeb (see chapter 3.3). The comparative keyword analysis aims to reveal any distinctive lexical items, through which the context of the flooding in Texas and Oklahoma is represented.

One of the statistical metrics that CQPWeb uses to measure keyness is the log-likelihood test for significance. Following Scott (1999:96), to calculate the keyword list, the minimum frequency for

the reference corpus was set to 10 and to 3 (by default) for the investigated corpus. This decision was motivated by the fact that first the reference corpus is very large in comparison with the investigated one, and second, it was more sensible to restrict the number of keywords that will be calculated to a manageable number. The implications of this decision are that a comparison between the two corpora will ignore lexical items which do not appear at least 10 times in the reference corpus and respectively 3 times in the collected corpus.

Furthermore, the confidence interval width (*p value*) was set to 0.00001, which is the lowest supported by the software. As Scott (1999) explained, when the value threshold is lower, the comparison is more selective and the keyness confidence is higher. With $p < 0.00001$, CQPWeb calculated the adjusted log-likelihood threshold to 44.7.

#ATX #Austin #breaking **#Houston** #Oklahoma **#Texas** #tornado **#txst** #WEATHER @card@ Advisory affect
Alert am area at Austin Bastrop bayou bend blanco by CDT central Co Co. continue county creek damage dead disaster Drown due
East effect Ellis emergency evacuation flash Flood flooding further Gust Hail Harris Hays heavy Houston
hwy I-35 in issue lake Lamar Marcos may missing move mph ne Near NWS ok Oklahoma PM prayer radar
Rain rainfall Relief report rescue resident river road safe San Severe shelter stay Stephens Storm sw Texas
thunderstorm tornado Travis TX until update victim Warning Water weather west Williamson
Wimberley wind WX

Figure 9: Visual representation (*Wmatrix*-style) of lemma-based keywords in the Memorial Day floods corpus

Figure 9 provides an illustration and first impressions of the most significant keywords from the corpus, based solely on positive keywords and representing the items with the highest significance values. The entry @card@ is a collective tag of the software, under which cardinal numbers with high keyness scores from the corpus are organised.

What makes an impression on the first sight is that many of the illustrated significant words are proper nouns like *Austin*, *Houston*, *Texas*, *Blanco (river)*, *Lamar*, *Oklahoma*, *TX (which equals to Texas)*, *Bastrop*, *Wimberley*, *NWS (National Weather Service)* and so forth. The reason why so many proper nouns are keywords here is that many of the microposts in the corpus include names of specific places, institutions and in some cases of people, mostly political figures. In particular, these location names are significant for the corpus because they provide precise geographic reference to the areas which are very likely to be affected in some way by the flooding events. Furthermore, they add important spatial details and information around the happening events.

Apart from proper nouns, in the illustration as other keywords occur multiple lexical items such as nouns, verbs, adjectives, and adverbs, e.g., *flood*, *flooding*, *severe*, *storm*, *warning*, *thunderstorm*,

tornado, emergency, safe and more. These lexical items show the *aboutness* of the corpus (Scott 1999:71), provide a good indication of what the corpus is about and depending upon the context are principally the most interesting for analysis.

Typically, in a keyword list, some high-frequency grammatical words will occur as keywords (Scott 1999:71). In the illustration above, no grammatical words occurred which might be a result of the nature of the microposts, e.g., the imposed character limitations, as discussed in the introductory chapters, but also due to the size of the sample shown. Grammatical words, however, are more reflective of the style of a text rather than the content (Scott 1999:71).

As a result of the corpus settings, a keyword list of 132 entries at the lemma level was produced. The majority of these words are *aboutness* words. For the purpose of this analysis here proper nouns discussed previously and few grammatical items were left out and Table 4 shows the top 20 entries with the highest keyness score.

No.	Lemma	Memorial Day floods corpus		Reference corpus		Log-likelihood
		Frequency (absolute)	Frequency (per mil)	Frequency (absolute)	Frequency (per mil)	
1	flood	365	13,182.13	473	12.90	4106
2	warning	334	12,062.55	821	22.39	3418.48
3	severe	160	5,778.47	330	9.00	1682.95
4	until	324	11,701.40	12,253	334.17	1670.25
5	county	213	7,692.59	2,124	57.93	1641.1
6	tornado	171	6,175.74	827	22.55	1546.87
7	thunderstorm	128	4,622.77	443	12.08	1234.03
8	flooding	98	3,539.31	85	2.32	1156.82
9	issue	122	4,406.08	5,886	160.53	571.25
10	water	118	4,261.62	5,394	147.11	564.66
11	storm	91	3,286.50	2,131	58.12	552.2
12	effect	66	2,383.62	1,623	44.26	394.23
13	emergency	47	1,697.42	659	17.97	331.38
14	rain	90	3,250.39	7,375	201.14	331.29
15	safe	64	2,311.39	4,899	133.61	243.68
16	rainfall	21	758.42	28	0.76	235.09
17	shelter	32	1,155.69	437	11.92	227.22
18	rescue	30	1,083.46	555	15.14	195.57
19	affect	34	1,227.92	1,002	27.33	191.22
20	report	48	1,733.54	4,134	112.75	172.18

Table 4: Top 20 keywords ranked by log-likelihood value

Here the keywords are ordered by the strength of their keyness score, i.e. the log-likelihood value. The third and the fourth columns present the absolute and normalized (per million) frequencies in the investigated corpus and columns five and six list these values for the reference corpus.

The most frequent lexical item in the corpus is *flood*, followed by *warning* and the adjective *severe*. In order to be able to explain the significant results, keywords can be examined in their contexts. When carrying out concordance analysis, it can be observed that the lemma *flood* is used in the context of flood advisories, warnings, flash flooding, flood victims, and flood forecasts. It is primarily associated with weather alerts and the potential dangers and consequences of the flooding. As follows, further consulting the concordance lines of *warning* show that in the corpus, it is mainly associated with weather alerts, particularly flood warnings, flash flood warnings, and tornado warnings, classifying them as *severe*. In the corpus, *until*, which functions as a preposition, is employed by authors to indicate the duration of these warnings and advisories.

From the table it can be observed that the keywords above reveal valuable information about the thematic content of the corpus and manifest its overarching central topics of the flooding and the severe weather events. This is evident from the high frequencies of keywords such as *flood*, *warning*, *severe*, *tornado*, *thunderstorm*, *water*, *storm*, *emergency*, *rain*, *rescue*, and so forth. Alongside these, other outstanding lexical items here such as *safe*, *shelter*, *rescue* and more highlight other aspects and implications from the flooding within the corpus.

Furthermore, narrowing down the analysis to verb forms, the results obtained from a corpus query produced a total of 2771 verbs and 703 unique verb types occurring in all microposts of the Memorial Day floods corpus.

<i>Lemma</i>	<i>Frequency</i>	<i>Lemma</i>	<i>Frequency</i>	<i>Lemma</i>	<i>Frequency</i>
be	555	help	38	find	25
have	148	see	37	say	24
issue	122	continue	36	look	22
stay	63	close	35	pray	21
move	62	affect	34	make	21
flood	61	come	32	turn	20
do	60	take	28	need	19
get	50	can	27	call	19
will	46	miss	26	keep	17
go	40	drive	25	confirm	16

Table 5: Absolute frequencies of verb lemmata in the corpus

Beyond that, a lemma-based frequency list summarizes the highest-ranking thirty lemmata verb forms in the data which have absolute frequency values between 555 and 16, set out in Table 5. The initial top two lemma positions of *be* with 555 occurrences followed by *have* with 148 occurrences are not surprising, because these verbs along with *do* typically perform multiple linguistic functions, the most common of which operate as auxiliary verbs. Interestingly, the verb *issue* and its forms rank third among all verbs, underscoring its significance in the dataset. A concordance lines analysis shows that in the corpus, *issue* is employed to express the idea of distributing, formally sending out, or making known different weather announcements like warnings, advisories, statements etc.

Another interesting example appearing among the highest-ranking with 21 occurrences in the corpus is the verb *pray* and its associated forms. Concordance lines show that it is often employed by authors in two ways: either directed towards the audience asking them for prayers or to express authors' personal thoughts and prayers in relation to a specified group of individuals or locations. Furthermore, the remaining verbs listed above also provide valuable information about what verbs authors employed to express meanings in the corpus. Broadly, these lemma forms convey states of being, doing or sensing. For instance, the presence of verb forms like *go*, *see*, *close*, *stay*, *continue*, *move*, *help*, *affect*, *call* and *come* reflects a wide range of actions and happenings which authors communicated in their microposts.

A POS query of verb forms carried out on the corpus returned 2771 results in 1190 different microposts. A closer look into verb frequencies (see Table 6) points out that overall, seven of the positions of this ranking are filled by multiple conjugated and contracted forms of the verb *be* (is, are, be, been, was, the contracted 're and partially 's). Two of these are the most frequent verb forms in the corpus, namely the conjugated *is* and *are* with respectively 199 hits, making up 7.18% of all verbs and *are* with 114 hits, making up 4.11% of all verbs in the corpus. Full verbs with high frequencies like *issued*, *stay*, *moving*, *flooded*, *help*, *continue* etc. employed by authors in the context of discussing or describing the flood events show aspects of the most common lexical choices of individuals for how they convey information about their own or other participants' activities.

<i>Verb form</i>	<i>No. of occurrences</i>	<i>%</i>	<i>Verb form</i>	<i>No. of occurrences</i>	<i>%</i>	<i>Verb form</i>	<i>No. of occurrences</i>	<i>%</i>
is	199	7.18	's	46	1.66	missing	24	0.87
are	114	4.11	do	44	1.59	're	23	0.83
issued	114	4.11	was	39	1.41	go	21	0.86
has	76	2.47	affected	32	1.15	being	17	0.61
be	65	2.35	help	29	1.05	coming	17	0.61
stay	54	1.95	closed	28	1.01	continues	17	0.61
moving	51	1.84	get	27	0.97	drive	17	0.61
flooded	49	1.77	take	26	0.94	continue	16	0.58
have	47	1.7	been	24	0.87	see	16	0.58

Table 6: Frequency breakdown of verb forms in the corpus

Two dimensions of time can be considered here: while the corpus is limited to a certain timeframe in the past within which the messages have been broadcast, verbs here also show tense variation (e.g., past: flooded, affected; present: is, go; continuous: moving, coming). This indicates that authors communicate information about events, actions and happenings at various points in time during the flooding.

Another interesting but expected aspect of the corpus which the frequency breakdown uncovered is the high proportion of low-frequency verbs, most of them *hapax legomena* (appearing once in the corpus). Among these, there are many lexical verbs which play an important role revealing the authors' lexical choices and encoded through experiential processes meanings (for example, forms of: *approach, arrive, abandon, assist, climb, deny, escape, mitigate, prepare, wait* etc.). The following analysis will concentrate on verbs and most of these meaning-bearing verbs will be discussed with the purpose of explicating experiential processes and construed types of meanings in the corpus.

4.2 Temporal aspects of experiential processes

To structure the following experiential analysis of process types, below is extracted a sample of two to five microposts from each day of the observed time period. This number of microposts has been identified for discussion in order to show how authors communicated meanings and what discourse devices they employed to realise these meanings. The messages also provide a starting point to the reader to understand what actors talked about in the collected corpus.

The main criterium to select the posts was because they best illustrate the presence of single or multiple experiential process types and demonstrate the above listed objectives.

This experiential analysis has been extended to all microposts comprising the corpus and in chapter 4.2.9 temporal summary and analysis of the processes follow.

4.2.1 May 22

601810385532899000 – ‘Latest Oklahoma lake levels as of 11am today. SW OK lakes keep rising<Material>. Hugo is<Relational> 30 FEET above normal. #okwx @NEWS9’

In this micropost information about rising water levels is communicated, encoded by two experiential processes. The verbal group ‘keep rising’ represents a material clause, describing the ongoing process of the lakes in southwestern Oklahoma, continuing to increase their water level. Semantically, the finite verb ‘keep’ attaches an additional layer to the meaning of the event of increasing and serves the function to denote its continuation. Furthermore, in the case at hand the copular verb ‘is’ signifies a relational attributive process, indicating the state of Hugo Lake, being 30 feet above its normal water level.

601840571926315000 – ‘#SalvationArmy disaster services team is<Relational> on standby, prepared to respond<Verbal> to emergency needs in Texas #Austin #storms’

The purpose of this message is to inform the audience about the readiness of the Salvation Army disaster services team to assist with emergency needs in Texas, specifically in Austin, during any emergency situations.

The verb ‘is’ materialises the relation between both entities in this clause, attributing the quality of being ‘on standby’ to the ‘carrier’ in the clause - Salvation Army disaster services team. Thus, the verb here is in relational attributive mode. Furthermore, within the verbal complex ‘prepared to respond’, the main event verb ‘respond’ reflects a verbal process, which is modified by ‘prepared’, emphasizing the readiness, the ability to respond on time.

4.2.2 May 23

602200833296466000 – Flooding at Foss Lake campgrounds from @NEWS9 viewer Jennifer Staggs. PLEASE be<Relational> safe! #okwx @tornadopayne

In this micropost, the author has communicated important information about a specific location, namely, the Foss Lake campgrounds, which are currently being affected by the flooding. Such information can be crucial for those who are in or near the area and need to take precautions or stay informed about how the circumstances develop.

From a functional perspective, the copular verb 'be' represents a relational attributive process. In terms of transitivity, verbal groups such as 'be safe!' may have an exclamative, but also an imperative status and can be classified between major and minor clauses. They are common for the corpus and include expressions like alarms or warnings, e.g., 'Keep off! or Stay off!' but also appeals, such as 'Help!'. In this case through the alert 'please be safe!', the author communicates an advice or a recommendation to the readers to stay safe and is implicitly urging them to take the necessary precautions, or to avoid unnecessary risks. This expression shows empathy and is a common way to convey a message of caution or care, especially in situations where there are potential dangers or risks, as indicated in the beginning of the message.

602244336588206000 – 'If you know<Mental> anyone whose not paying<Material> attention to the weather tell<Verbal> them to check<Material> it asap!! #okwx'

In this message, the author communicates advice or a recommendation to the readers, to stay informed and safe.

From a functional perspective, in this example verbs realise four experiential processes. The verb 'know' encodes a mental process of cognition, by means of which the writer is trying to raise awareness, asking to inform individuals who are not aware of the upcoming weather situation to inform themselves. Both verbal groups 'paying' and 'to check' realise material processes in terms of transitivity analysis, but there is one significant difference. In its literal sense, the verb "paying" fulfils a material process, and would be used in a rather different context to denote the transfer of money or value in exchange for goods or services. The expression 'paying attention' is a common functional formulaic language of a metaphorical nature and in its congruent meaning¹², it signifies a cognitive or a mental action describing the act of being focused and receptive to information or stimuli. Metaphors are some of multiple language facets, functioning as well as a resource for expanding the meaning potential of language. The last verb here 'tell' is an instance of a verbal process.

602147636917834000 – '(TX) Flood Warning issued<Relational> May 23 at 10:55AM CDT until May 26 at 1:00PM CDT by NWS #txwx'

In its essence, the message has an informative function, providing important temporal details regarding when a flood warning was issued and how long it will remain in effect. In the corpus,

¹² In that case, this is an incongruent expression which stands as the metaphorical variation of the clause. In SFL Halliday & Matthiessen (2014) discussed congruent realisations as 'unpacking' metaphorical language to construct congruent wording.

posts conveying information about issuing different kinds of advisories, warnings or statements, and containing additionally location and temporal descriptions are common. This is also evident from the observed high frequencies of the verb 'issue' in the corpus as shown in the keyword analysis above. In this particular utterance, it serves to convey the action taken by a relevant authority, the National Weather Service (NWS), to formally send out a flood warning.

From a functional perspective, the verb form 'issued' realises a relational attributive process. Gerund forms or verbal nouns like 'warning' are not analysed as belonging to the verb class here. Although they are derived from verbs typically, they have all of the properties of nouns and can function as subjects, objects, or adjectives (Thompson 2014).

602259259946996000 – 'Watching<Behavioural> these jagaloons on the news driving<Material> in water... I can't #IAmSurroundedByIdiots #TurnAroundDontDrown #okwx'

The overarching function of this post is to communicate the emotional response of the author to the current flooding events. The message exemplifies the author's pejorative attitude to individuals who engage in the act of driving in water during the flooding, describing them as 'jagaloons'.

This micropost contains three verbs, among which two are analysed as event verbs in experiential sense. The modal verb 'can't,' functions as a finite verbal operator, which, by itself, does not convey any experiential meaning in terms of transitivity. The verb 'watching' encodes a behavioural process, while 'driving' denotes a physical action or a material process.

602257808717779000 – 'Sirens are going off<Material> in Moore. #okwx'

In this post the author warns individuals of potential emergencies or severe weather events in the area of Moore. The message conveys information about emergency warning sirens, which are currently being activated and sounding.

Within the clause 'Sirens are going off in Moore', the verbal group here realises a material process. Like in the case at hand, the process sometimes might encompass an additional constituent beyond the verbal group. This occasion becomes particularly evident while examining prepositional, phrasal-prepositional and phrasal verbs like here 'going off', where the particle is analysed as filling the predicator, i.e., as expressing part of the meaning of the process. The reason for this is that within multi-word verbs, the verb and the particle build a semantic unit, which often serve an idiomatic purpose and thus have an arbitrary figurative or a metaphorical meaning.

4.2.3 May 24

602285449164362000 – ‘805pm - new tornado warning for SE Cleveland County. Does NOT include<Relational> OKC, Moore or Norman. #okwx’

Within this utterance, there is one relational attributive process, realised by the verbal group ‘does not include’. Notably, in the beginning of the clause ‘[It] Does not include’ the subject ‘it’ - taking up the role of a carrier - is ellipsed. Outstanding for social media domains is their informal register, in which a linguistic phenomenon such as elliptical constructions is a common characteristic feature.

In the context of this message, the author provides an explicit clarification, indicating that the areas Oklahoma City, Moore, and Norman are exempted from the recently issued tornado warning. As a whole, the post contains important information about the newly issued tornado warning and its geographic scope.

602295532275540000 – ‘if you're<Relational> in norman, tell<Verbal> me how the roads are<Relational> near you. i'll add<Relational> it to a live map of the area on @oudaily! #okwx’

The message realises two different functions: simultaneously requesting and providing information regarding the road conditions in Norman, Oklahoma. Here the author employs a cooperative approach to sharing information about the current local road conditions.

In the first sentence, there is a combination of interrelated relational and verbal processes. The clause ‘if you're in norman’ encodes a relational attributive process and establishes a conditional relationship between the reader and their physical location. Subsequently, through the verbal clause ‘tell me’, a request for information is addressed to the reader, asking them to report the condition of the roads nearby their location. The verbal group ‘I’ll add’ realises a relational attributive process as well and reveals the author's intention to include the information received on a live map.

602550604662067000 – ‘NWS Weather Radio service knocked out<Material> over most of Oklahoma. People will not hear<Mental> warnings until it's<Relational> fixed! Stay< Relational > alert! #OKWX’

The author of this micropost communicates important information by warning readers not to rely on this specific radio service for further information at that moment. In terms of experiential meanings, four processes are realised through verbal groups.

The phrasal verb ‘knocked out’ encodes a material process, describing that the National Weather Service Radio service has been disrupted in most parts of Oklahoma. As next, a process of sensing, which will be experienced by the sensor ‘people’ in this case, is construed by the verbal group ‘will not hear’. The verb ‘hear’ is a perceptive verb i.e., it realises an experience through perception. Both verbal groups in the following two small clauses encode relational attributive processes.

602341707351396000 – ‘North #Austin getting<Material> substantial damage. NWS survey team will have to determine<Mental> if it was<Relational> a tornado. #txwx’

The verbal group ‘getting’ realises a material clause, with an actor of the process – ‘north Austin’ and the goal ‘substantial damage’. Furthermore, within the verbal complex ‘will have to determine’, the lexical verb ‘determine’ realises a mental clause of inclination¹³ and the copular form ‘was’ encodes a relational process in an attributive mode, questioning whether the nature of the event which caused the damage was a tornado.

As observed, processes that encode human experience can be represented through single verbs or verbal groups. In the corpus, both cases are common, including complex structures of verbal groups, as in the case at hand ‘will have to determine’. In verbal groups treated as complexes combinations of multiple verbal groups are present, typically in the order of a finite group followed by a predicator. However, there is just one process encoded per verbal group, which is realised through the last (main lexical) verb in the clause, like in this case, where the mental process is realised by the lexical verb ‘determine’.

The primary function of this message is to convey information about significant damage in North Austin, and that the National Weather Service survey team will need to assess whether this damage resulted from a tornado caused this damage.

602311747454665000 – ‘Remember<Mental> when the lakes, rivers, and ponds were<Relational> all separated in Oklahoma?! Yeah, me neither. #OklahomaFlood2015’

The verb ‘remember’ represents a mental cognitive process, asking readers to recall past experiences. In mental clauses, ‘remember’ is typically used to report ideas or introduce an anecdote. As follows, the copular verb ‘were’ realises a relational attributive process. The underlying nature of this post can be rendered as sarcastic, the author is referring the current situation in a sarcastic way, highlighting the significant impact the flooding is having, causing

¹³ Halliday & Matthiessen (2014) call ‘inclination’ type of sensing verbs within mental clauses ‘desiderative’.

water bodies to merge. From a rhetorical stance, sarcasm is a linguistic device, typically used to convey criticism or a commentary in an indirect way.

4.2.4 May 25

602660713543377000 – *‘Curfew for San Marcos and Wimberley is<Relational> now in effect. Head<Material> home and stay<Relational> dry! #TXST #SMTX #SMTXFlood’*

The message informs the audience that a curfew has been in place in San Marcos and Wimberley, providing essential information about a change in the local regulations. It offers practical instructions and an advice to readers to head home and stay dry.

As described, relational identifying clauses have the quality of reversibility. A probe of the possibility to reverse the first clause to ‘In effect is curfew for San Marcos and Wimberley’ shows that the clause is in relational identifying mode, realised by the copular verb ‘is’. Within the imperative structure ‘head home’ the lexical verb ‘head’ implies a material process, instructing individuals to leave their current location or any potentially unsafe areas and seek safety in their homes. The following imperative structure ‘Stay dry’ encodes a relational attributive process.

602831354594722000 – *‘Oh Mr. Golden Sun, please come back!<Material> We love<Mental> you too! #Oklahoma #RainRainGoAway Loving all things Oklahoma! #ionok #oklahoma...’*

The message contains two verbal groups which encode different types of processes and experiential meanings. The first clause is a material clause, and the event is encoded through the phrasal verb ‘come back’, indicating some process of doing. It expresses the meaning of returning in this case, referring to the personified ‘Mr. Golden Sun’, treated as the actor of the process in the clause. A mental process of affect is realised by the lexical verb ‘love’, representing the author’s affection or admiration towards the sun. Through this post the author conveys emotion.

602908202640871000 – *‘The water was<Relational> so intense, these foot wide concrete pillars snapped<Material> like twigs. #WimberleyFlood’*

The message contains two verbs that encode different types of experiential processes. The copular verb ‘was’ depicts a relation of attribution – the two participants in the clause the ‘water’, treated as the carrier and its attribute ‘intense’ do not have a reversing potential. The lexical verb ‘snapped’ encodes a material process, representing the action of the concrete pillars breaking or fracturing. Furthermore, the comparative figurative language employed by the author in ‘concrete pillars snapped like twigs’ is a simile, used to emphasize how destructive the force of the coming

water is. The main function of the text is to convey information about potentially dangerous situations caused by the flash flood that affected Wimberley.

602917472052720000 – ‘Large damaging #tornado is<Relational> still in progress with this storm SW of Gatesville moving<Material> NE at 30, take<Material> cover! #txwx’

In this micropost, the first clause is rendered as a relational identifying clause, in which the process is realised by the copular verb ‘is’. Furthermore, the verb ‘moving’ encodes a material process, indicating that the storm is advancing from its current location – southwest of Gatesville in a direction towards northeast.

As a component of the metaphoric expression ‘take cover’, the lexical verb "take" represents a material process likewise. Grammatically this construction is built in an imperative mood and is common in the corpus, particularly employed in messages which serve as warnings for severe storms, tornadoes, or floods. Thus, this message communicates recommendations or general advisory to the audience.

602931585898263000 – ‘Patrol officers reporting<Verbal> traffic signal issues and very heavy rain making<Material> driving conditions in Waco very hazardous.#txwx’

In the short text above two processes are encoded by the verbal groups. Firstly, ‘reporting’ signifies a verbal clause, communicating information about patrol officers notifying on traffic signal issues and tough weather conditions in Waco. Secondly, within the clause ‘making driving conditions in Waco very hazardous’, the lexical verb ‘making’ encodes a material process and indicates a causative relationship between the two participants: heavy rain and hazardous driving conditions.

Structurally, this relationship is expressed in the two distinct clauses, the causing and the caused action are manifested separately. Causative constructions are commonly associated with the verb 'make,' which is frequently employed for that purpose, likely due to its inherent syntactic adaptability and flexibility. The main message here conveys information about patrol officers reporting the potentially dangerous road conditions to raise awareness and potentially advise caution or action to those in the area.

4.2.5 May 26

602990327104831000 – ‘Build<Material> a student living complex on the river, they said<Verbal>. It'll never flood<Material>, they said<Verbal>. #TXST #SMTXFlood’

Each of the sentences above is comprised of two separate clauses through projection. In both small clauses ‘they said’, the verb ‘said’ encodes a verbal process and one of the main characteristic features of verbal processes i.e., reporting verbs is that they can project. Like in the case above, where the two lexical verbs ‘said’ are employed in a projecting mode by the author.

Within the transitivity system, transitivity units in the projected clause are analysed separately from these constituting the projecting clause. The relationship between the projected and the projecting clause is that of two separate entities, both constituents from a more complex unit. Returning to the message above, in the first sentence the verb 'Build' realises a material process. Within the second sentence, the verbal complex ‘‘Il...flood’ is interrupted by other lexical entities, but the event verb is ‘flood’ which realises a material clause.

Through employing sarcastic figurative language in their commentary, the author of the message conveys the idea, that the claim that building a student living home on the river will be safe, is apparently not true.

*602997788792730000 - Looks<Relational> like most of Lamar now re-open<Material> to traffic.
#atxfloods @statesman*

This message conveys general information and updates about the road or transportation infrastructure in Lamar, which after being closed has reopened and is already accessible to vehicles. In the utterance above ‘Looks’ is a copular verb, rendered as a relational attributive process. Furthermore, the lexical verb ‘reopen’ encodes a material process.

*603008169707635000 – ‘Bastrop blocks<Material> roads, closes<Material> offices,
cancels<Material> school activities #atxfloods’*

In this utterance the author employs a figure of personification. While the city of Bastrop itself doesn't possess human-like qualities, the employed by the author verbs ‘blocks’, ‘closes’ and ‘cancels’ imply that the city is taking these deliberate actions, as if it were a human being. The purpose of the message is to convey information about the various measures taken in Bastrop in response to the flooding.

Within the first clause the finite verb ‘blocks’ realises a material process, outlining that measures to restrict access to certain roads within the city of Bastrop have been taken. The finite verb in the followed clause ‘closes offices’ encodes a material process as well, reporting that administrative or government offices in the city are temporarily shutting down. The third clause is also a material process clause, realised by the finite verb ‘cancels’, referring to schools suspending planned activities.

603269873288937000 – ‘Comfort<Mental> those who mourn<Behavioural>, strengthen<Material> those who are<Relational> weary, encourage<Verbal> those in despair, and lead<Material> us all to fullness of life. #TexasFlood’

The current post is an instance of six experiential processes represented by verbal groups, as follows: The verb ‘comfort’ construes a mental process of perception and in this case signifies the action of providing emotional support to those who are mourning. The lexical verb ‘mourn’ is rendered as a behavioural verb of grieving or expressing sorrow. As follows, ‘strengthen’ represents the material process of making someone physically or emotionally stronger and the copular verb ‘are’ establishes attributive relationship between entities. The lexical verb ‘encourage’ represents a verbal process, and the lexical verb ‘lead’ also represents a material process. It implies the action of guiding or showing the way ‘to fullness of life’. This message is highly concerned with the author’s emotions and conveys empathy and compassion, expressing their wish for the well-being of those affected individuals by the floods and going through difficult times.

603095612532240000 – ‘Please call<Verbal> 713-207-2222 to report<Verbal> downed power lines and avoid<Material> water where a power line might be<Relational> down. #houwx #hounews #besafe’

The main function of this message is to provide instructions and safety information to the audience and simultaneously urge them to take action by calling to report downed power lines.

This meaning is mainly realised by the three experiential processes, two of which encoded by the lexical verbs ‘call’ and ‘report’, both realising verbal processes. In the following context, both verbs serve the purpose of instructing individuals to contact a specifically indicated telephone number in order to report downed power lines. The following verb ‘avoid’ is analysed as encoding a material process.

4.2.6 May 27

603371656455717000 - 806pm: FLASH FLOOD WARNING until 11pm for Eastland/N. Erath counties. Never drive<Material> through flooded roads. Turn<Material> Around Don’t Drown<Material>! #txwx

Three experiential meanings are realised in the domain of material processes here. The first material clause is encoded by the lexical verb ‘drive’, employed by the author as warning not to drive through floodwaters. The verbal group ‘turn around don’t drown’ features two lexical verbs

and an auxiliary verb which does not realise any experiential meaning. Both main lexical verbs 'turn' and 'drown' relate material processes.

This expression is very common for the corpus and has initially originated as a campaign, initiated in 2003 by the National Weather Service with the purpose to warn individuals of the hazards when crossing or driving a vehicle through flooded areas. Drivers are urged to refrain from passing through flooded roadways and instead, to turn around their vehicle (NWS 2023). Overall, this message fulfils an informative function, providing information about a flood warning in effect for certain locations and advises individuals not to drive through floodwaters.

603406695700799000 – 'Floods in TX cause animal shelters to evacuate<Material>; supplies destroyed. Please help<Material>! #TexasFlood'

In this short post the verbal complex 'cause...to evacuate' is interrupted by other entities, however a material process is fulfilled by the lexical verb 'evacuate'. As part of the verbal complex, the verb 'cause', although not the main event verb, plays a crucial role in illustrating the causative relationship between the actor who has influence over the unfolding process and the process itself. Additionally, the verb 'help' realises another material process, signifying the action of asking others for assistance and help. Together, the linguistic choices of the author convey a message highlighting the urgent need for support to facilitate the evacuation of animals from shelters.

603536022861844000 – 'The #houstonflood may be<Relational> over but the rescues continue<Material>. This car crashed<Material> here this AM. Driver plucked<Material> to safety.'

Within the verbal complex 'may be', the copular verb 'be' realises a relational attributing process, indicating the possible end of the flooding. On the other hand, the event verb "continue" construes a material process, describing the ongoing action of rescues. Furthermore, the lexical verbs "crashed" and "plucked" both realise material process clauses. The message provides an update on a potentially dangerous situation, emphasizing that despite the assumption that the Houston flood is at its end, rescue operations are still running.

4.2.7 May 28

603720744179114000 – 'The floods covered all when it was<Relational> dark Sun AM and haven't seen<Mental> the damage till today. #Speechless #WimberleyFlood'

Two verb processes are encoded in this message. The copular verb 'was' is a verb of relational attributive process, while the verbal complex 'haven't seen' realises a mental cognitive process.

The main purpose of this message is to express the speaker's emotional state, their attitude and response towards the damage caused by the flooding.

603735663314342000 - There is<Existential> a monster storm with a confirmed tornado headed<Material> towards Belding, #TXWX. Large hail coming<Material> for Fort Stockton

In the construction 'there is' the verb form 'is' introduces an existential process. Furthermore, the lexical verb 'headed' encodes a material process, describing the course of the storm moving in the direction of the city of Belding. In the last clause, the third, material process is realised by the verbal group 'coming', indicating that a specific location – the city of Fort Stockton, will be likely affected by a large hail. Overall, the message conveys information about a severe storm with a confirmed tornado and the threat of a large hail, threatening the mentioned locations.

The analyses on the selected sample microposts showed how authors realised the different process types of experiential meanings in their communication. All process types and their distributions across the different days in the corpus are discussed in detail in chapter 4.2.9.

This analysis also shed light on the various discourse devices employed by authors. These include imperative structures, metaphorical language, similes, personifications, sarcasm and other figurative expressions. Authors employed these figures across a broad spectrum of topics within the corpus, encompassing support and safety advisories as well as infrastructure impacts, community resilience and more. These topics will be further elaborated and discussed in chapter 4.3.

4.2.8 Notes on the practical approach

There are messages in the data which reveal variations of relevant and important information for the current state of the flooding, warnings, advice etc. although only through an attached video or image. Such messages include, for example: *'1125pm - Every green polygon is a flash flood warning. #okwx #texomawx'* containing a link with a map, where one could see all regions where a flash flood warning is issued, or messages like: *'In case you need a reminder of why you should keep your kids out of the flood waters. #HoustonFlood'* containing an attached image showing snakes in the floodwater. Such messages are left out from this analysis, as they require a multimodal methodological approach of analysis, which is beyond of the scope of this study.

A further note regarding verb forms in terms of process types needs to be added here. In traditional terms, participial forms ending in -ed or -ing, may in certain cases have functional status of adjectives or verbs depending on the context. Following Thompson (2014), the

distinction between the two cases is not obvious and it is not always clear where one can draw the line between the two syntactic functions.

On these grounds, in the analysis here only verb forms that are responsive to supporting standard tests are considered, and such that do not cause syntactic or semantic ambiguity. Such standard tests are for example: i) probing their acceptability of the modifiers 'very/rather' i.e., if it makes sense in the given context, the lexical items are considered as adjectives, ii) probing their acceptability of affixation 'un-' for building adjectives (Borer 1990). Besides this, participle adjectives usually address a quality or attribute of the noun they modify and tend to indicate psychological states.

Although the above-mentioned resources have proven useful, it should be noted that standard probes do not always work faultlessly in different lexical environments (Borer 1990). Thus, in cases of ambiguity, participial forms have been excluded from this analysis (like in several cases the lexical items 'affected' and 'closed'). In the following instances, the participial forms are considered to function as verbs: 'Lots of low water crossings closed in Williamson County', '4+ people trapped on roof and in home at 3008 S. Old Stagecoach', 'Residents trapped on roofs in Wimberley & near Kyle', 'Evacuations ordered in San Marcos;'.

4.2.9 Day-by-day analysis and summary of experiential processes

In total, the entire corpus is comprised of 1579 short posts, within which a total of 1656 processes have been identified. All six experiential process types are present in the data. A breakdown of the distribution of process types shows 747 instances of material, 166 instances of mental, 136 instances of verbal, 36 instances of behavioural, 22 instances of existential, and 549 instances of relational processes.

Table 7 provides the summary and comparison statistics for the total number processes and process types per day as absolute and relative values. In an early stage of building the corpus, the raw data for May 26 appeared to be the richest in terms of relevance and hence in volume, compared to the rest of the days. Therefore, it is not surprising that the highest number of processes and process types as shown in Table 7 were identified on May 26. The time span between May 23 and May 26 is reported as the critical period of heavy rainfall (NOAA 2023), which can also be observed from the trend in the increasing number of relevant messages in the corpus.

	May 22	May 23	May 24	May 25	May 26	27 May	28 May	Total
Material	9 (36%)	28 (43%)	131 (48%)	133 (42%)	280 (46%)	119 (45%)	47 (46%)	747 (45%)
Mental	2 (8%)	3 (5%)	20 (7%)	27 (9%)	72 (12%)	28 (11%)	14 (14%)	166 (10%)
Behavioural	-	4 (6%)	5 (2%)	7 (2%)	15 (2%)	5 (2%)	-	36 (2%)
Verbal	1 (4%)	6 (9%)	19 (7%)	29 (9%)	46 (8%)	27 (10%)	8 (8%)	136 (8%)
Existential	1 (4%)	1 (2%)	2 (1%)	5 (2%)	7 (1%)	4 (2%)	2 (2%)	22 (1%)
Relational	12 (48%)	23 (35%)	96 (35%)	115 (36%)	190 (31%)	81 (31%)	32 (31%)	549 (33%)
Total	25	65	273	316	610	264	103	1656

Table 7: Absolute and relative distribution of process types across the different days. The percentage have been rounded to whole numbers.

From the results in the table above and the representation in Figure 10, the difference among the relative distribution of the six process types across different days can be observed. May 22 exhibits the lowest number of identified processes, a total of 25. Among the five process types identified on the day, relational processes dominate with 48%, followed by material processes, which make up 36% of all processes. Furthermore, in contrast to May 22, on May 26 the highest number of experiential processes across all days have been identified, accounting for a total of 610. Among them, material processes appear as the most significant, making up 46% from all process types, followed by relational (31%), mental (12%) and verbal (8%) processes. Across all days, the largest proportion of material processes in the entire corpus (48%) has been identified on May 24 and the largest proportion of mental processes (14%) on May 28. All discussed process types demonstrate the thematic variation in a formal sense (of what process types represent) over the observed period of time in the entire corpus.

The statistics in Table 7 show that the majority of verbal groups employed by authors construe material processes, making up nearly half (45%) of all processes, followed by relational processes, making up approximately one third (33%) of all processes. Relational processes are followed by mental processes (10%), which indicates that the three principal process types within systemic functional linguistics (material, relational, mental) and its transitivity system, occurring most often across different discourses are also the most frequently occurring process types in the discourse of flooding in Twitter microblogged communication. As shown, they describe three very different kinds of structural configurations in the semiotic system of experiences.

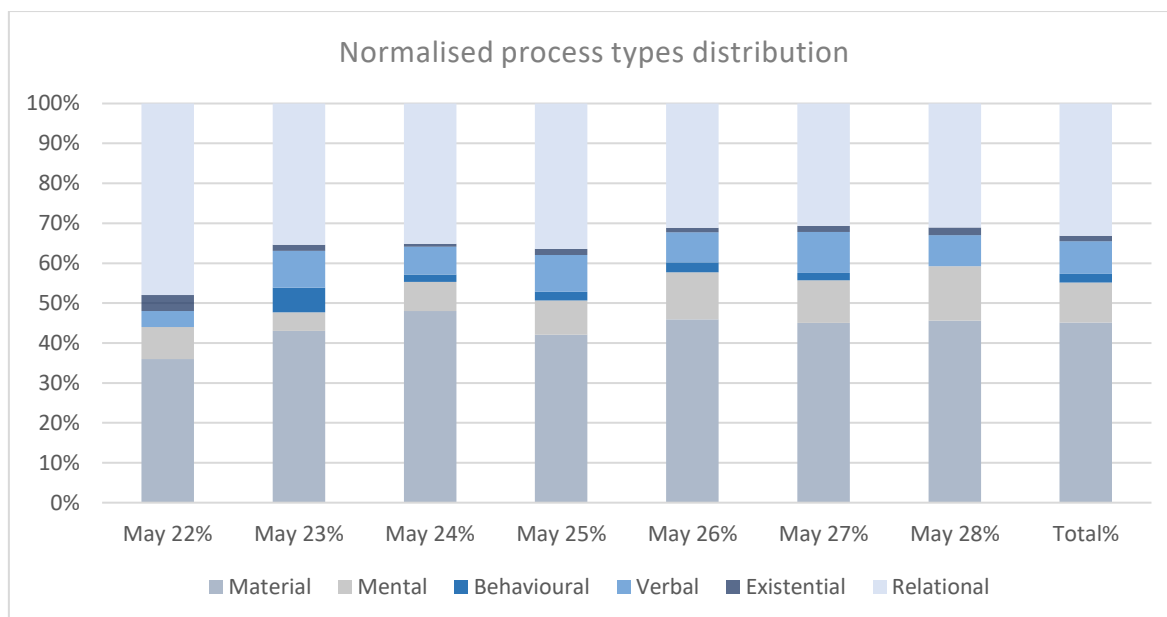


Figure 10: Normalized distribution of process types across the different days

Considering the prevalence of material and relational clauses in the corpus, it can be inferred that the underlying focus on experiences is related to concrete doings or happenings from the outer, material world, as well on identifying, describing and classifying entities, as it also became evident from the analysis carried out above. Mental clauses represent authors' choices of how they construe meanings about the flooding through their cognition, emotion, or attitudes towards humans or entities.

The results also illustrate that verbal processes, referred to as verbs of reporting, have consistent values across the days between May 23 and May 28. The various process types are consistently disseminated across all days, showing a relatively stable pattern. Within the sampled data, none of the processes constitute more than half of all process types per day. Behavioural processes, which are intermediate on the border between material and mental processes and are not the most frequent types of processes in the transitivity system, are nevertheless present in the most analysed days, except May 22 and May 28 2015.

Material clauses show some fluctuation from 36% to 48% across the different days, and as discussed above, they constitute a little less than half of all processes in the corpus. In addition, material processes rank first place on each of the days, except for May 22. This implies that authors overall put focus on talking about actions and happenings from the outer, material world, i.e., the flooding events and their possible effects in the physical environment. In verbal groups or constructed verb complexes, verbs such as make, take, get and have occur very frequently in different morphological forms, and because of their ability to adapt within different linguistic

structures, they construe various meanings in the corpus. Furthermore, within material clauses, verbs like move, help, recover, drive, rise, flood, turn, destroy, send (sending prayers, help), deny, escape, suspend, deal, look for and so forth are also very common and can be observed in different verb forms. A few examples of clauses with verbs conveying processes as physical actions or doings have been selected and are listed below. Verbs conveying material processes are underlined.

thunderstorms approach; turn around don't drown; residents prepare; storm is moving; take cover; tornado is heading; driving in water; will set a new record; seek shelter; consider leaving your home; illegal to remove barricades; flooding forces evacuations; homes were flooded by; troops helping; have decided to evacuate; to survey storm damage; tornadoes touched ground; Red Cross is looking for volunteers to help flood victims; flash floods destroy hundreds of homes; you can donate or sign up to volunteer; people trapped on their roofs;

Furthermore, in the corpus, one in ten clauses has been identified as a **mental clause**. Such clauses are construed from verbs of sensing which show states of mind, for example, in the case at hand of verbs such as know, think, see, etc. and from verbs expressing psychological events or emotions, like observe, expect, notice, brace, pray. Examples of verbs construing mental clauses are underlined below.

thinking of folks in Oklahoma; starting to think that Oklahoma; #Pray for us in #Texas; If you see high water; survey team will have to determine; Some Ppl Think That; I bet Parking Services; People will not hear warnings; 10 things you need to know; let's all hope; a tornado was spotted; Are you freaking kidding me; Hard to believe; I hope; Notice clouds are building; Praying for the city of Houston; Feeling horrible; Hard to imagine; we are saddened by the heartbreak; Stop hating; Brace yourself, folks wake up and realize the crazy night we just went through, God bless all those stranded in this #houstonflood, @AnniseParker urges people not to go sightseeing during #HoustonFlooding and so forth.

Behavioural clauses make up only 2% of the messages in the corpus. Identified verbs that realise behavioural processes are verbs like watch, see, worry, awaiting, experiencing, look after, look around, and so forth, which semantically signal both material and mental activities. The verb 'experience', for instance, is analysed within the domain of behavioural processes, due to the fact that it can simultaneously refer to both dimensions of feeling, or of being affected by something physically and emotionally. Examples of clauses containing behavioural processes are listed below:

Watch for local #flooding; Worried About Weekend Flooding; on their roofs awaiting rescue; watch massive flood waters in downtown; We are experiencing flooding; I go see Mad Max; Texas is experiencing horrible extreme weather; #poor who already struggle in "good weather."; "We never lose hope"; they'd rather watch entire country drown;

Verbal clauses constitute a little less than one tenth (8%) of all analysed clauses in the corpus. Within these, the verb ‘issue’ is very commonly employed in the context of reporting that a type of warning or statement has been issued. Verbal groups or group complexes containing ‘issue’ have been occasionally treated as verbal or relational processes depending upon their textual environment. In cases like ‘NWS has issued a Flood Warning for Dallas County’ or ‘FWD issues Severe Thunderstorm Warning’ it is analysed as construing verbal process, since the doer of the action i.e., the sayer in this case is explicitly mentioned (although it is not a human sayer). While in cases like ‘(OK) Severe Weather Statement issued May 25 at 4:39PM CDT’ and ‘Flood Warning issued for Madison and Walker Counties’ it has been rendered as relational process. Other verbs considered within this category are verbs like report, warn, tell, say, declare, ask, advise, document, and so forth, including the verb ‘tweet’, which is a coinage meaning ‘send a message via the Twitter platform’. In the corpus the verb ‘declare’ has been rendered as reflecting processes of both verbal and relational domains, depending on its meaning in the context and the structure of the clause. Instances for verbal clauses from the corpus are listed below:

OK where rescues from vehicles and water in homes have been reported; A 55mph wind gust was reported at KRBD/Dallas; forecast for Wichita Falls TX warns; #ATXFloods reporting 90 road/crossing closures; @NWStulsa issuing their first-ever Flash Flood Emergency; NWS Norman has issued 70 flash flood warnings; Commish advised residents along Blanco River; as we speak due to high water; Evacuations ordered in San Marcos; we are now told I-35 N.bound is open; are asked to report damage to 211; Please call 211 to report; National Guard called in to airlift 13 trapped; Abbott said; will tweet video later; WacoPolice just posted on FB; For those who want to contact officials; @CityofSanMarcos is advising everyone to; #flood Governor says deadly flooding; PA Announcer telling fans at #ToyotaCenter to; a rescue boat responding to motorist trapped; The news even inquired; TranStar map denote only some of the areas with high water; @erincargile can you help spread the word?; President Obama has declared Texas a disaster area; We documented a large tornado;

Relational clauses are just over one third (33%) within the whole corpus. Very common verbs in the relational domain of meanings are any forms of the verbs ‘be’, ‘stay’, ‘continue’, ‘issue’, ‘keep’, ‘add’, ‘equate’ and so forth. In the corpus, the verb ‘stay’ has been mostly employed in

various types of imperative structures, most frequently in constructions such as ‘stay safe’, ‘stay weather aware’, ‘stay clear’, ‘stay dry’ etc., and has been analysed as verb encoding relational processes, as its meaning equates a state of being careful or being safe, etc. Furthermore, similar to the verb ‘issue’, the verb ‘continue’ has been analysed occasionally as realising relational or material processes according to its meaning and textual environment in the current context. Typically, when these verbs are rendered to perform the function of relating abstract nominal participants, they have been identified to belong to the group of relational verbs. Because relational processes typically exhibit relations of either characterising or identifying entities, they also appear to be very revealing the context and connected with the genre of the specific discourse of flooding. Examples of relational clauses are selected below:

Ample moisture remains in place and will result in heavy rainfall; (TX) Flood Advisory issued May 22 at 8:23AM CDT; Oklahoma City is now just 0.11 inch; Tulsa radar shows the rain has moved out; Be informed!, Stay safe if Memorial Day plans include camping or activities on lakes; Flash #flood warnings continue; Tornado is west of Pocasset; Stay safe Oklahomies!; Flooding is now the main though MAJOR concern; Tornado warning includes the city of Austin; Please view to remain up-to-date!; Tornado Watch for areas NW of DFW metro allowed to expire; Rescue boats now being requested; all the telephone lines are down; Blanco River reached 40ft, 7ft over 1929 record!; If you have a truck; Stay weather aware!; They're in need of dry towel donations and animal fosters ASAP; Stay away from downed power lines and flooded areas; Make sure you keep your cars in the highest place possible; @TexansCheer are you all safe?; Be careful, y'all; That used to be grassy trails in the back!; Harris County added to list of counties under disaster declaration; rainfall across Harris County equates to 162 billion gallons of water;

Existential clauses are the least common within all process types. They make up an insignificant amount of 1% of all analysed clauses in the corpus. Essentially, all constructions with ‘there is/are’ have been treated as existential processes. Examples from the corpus are listed below.

There are reports of some cars trapped; There is definitely a tornado coming; There was a tornado spotted; There are currently 342 low water crossings closed; There were about 220 this afternoon before the storms rolled in; There's no point in even trying; there are 69K+ outages in Houston; There was some intense lightning around #SMTX last night!; There's no sunshine now but svr storms are possible again;

4.3 Information and message types across the corpus

In terms of what actors talk about and how they convey meanings, it can be observed from the corpus that microposts reveal different semantic content types and language patterns. Based on these observations, twenty-seven distinctive types of messages stand out.

In category one to four belong messages dealing with different types of warnings, as follows:

flood warnings (165 messages), **thunderstorm warnings** (138 messages), **tornado warnings** (120 messages), **general warnings** (47 messages). The first three groups share a common characteristic: they contain posts providing a spatial and temporal information about the different kinds of warnings. These posts typically include details about the location for which the warning has been issued, the type of warning, and its duration. Many of the messages in the aforementioned groups are sent out by institutions and a large majority of the posts in the corpus, which are syntactically organised without verbal groups belong to these categories. Structurally, these messages exhibit similarities, with the verb 'issue' being particularly prevalent. For example: *'(OK) Flood Warning issued May 24 at 3:26AM CDT until May 25 at 11:20AM CDT by NWS'*, *'Severe Thunderstorm Warning for Palo Pinto and Stephens County until 7:00pm.'*, *'Tornado Warning - Freestone; Navarro County(s), TX until 05/25/15 5:00PM (CDT)'*. Category four contains messages of general warnings and unlike the first three categories, these do not necessarily originate from official institutions. In this category there are very few verbless posts. Examples of posts belonging here are: *'Wx Alert Severe Weather Statement issued May 23 at 6:39PM CDT until May 24 at 4:00AM CDT by NWS'* and *'There are 142 low water crossings closed in Central Texas right now. Turn Around Don't Drown!'*

Several messages in which users discuss in narrow sense **preparation** measures for an upcoming dangerous events comprise the smallest in number category (5 messages) in the corpus. Such posts are for instance: *'#Texas residents prepare for a #flood'* and *'#SalvationArmy disaster services team is on standby, prepared to respond to emergency needs in Texas'* etc.

Posts providing highly generalised information compared to the rest of the corpus and reporting on the flooding, heavy rain, storms, climate change, insurance issues are coupled together in the category **reporting general matters** (213 messages). It includes posts like: *'San Antonio Texas is getting some of the heaviest of the rain & storms'*, *'#Floods turn #deadly in #TX and right now more rain and severe weather hitting'*, *'After the devastating storm. Floods like this not seen for decades in Austin'*, *'You're screwed if you dont have car insurance & tried to deal with the #houstonfloods'*, *'Climate change is messing with Texas.'* In contrast, posts providing personalised information which relates to specific individuals, their behaviour and social or physical

environment constitute the category of **reporting individual matters** (39 messages). It ties together posts sharing authors' personal experiences, including individuals reporting that they are safe, like *'Barely made it across my driveway to find water on the street too deep to drive through', 'We're safe, but closing the store. Expect to open tomorrow at 9am. Stay safe', 'Water is actually seeping into our studio. #KHOU Flooding will continue. Don't drive #Houston'.*

Furthermore, in the corpus present are messages containing information about **evacuation** activities (12 messages), such as: *'Village shopping center, US-75 and 138st in Glenpool, flooded and being evacuated...'*. Further coded categories relate to **trapped** (18 messages) individuals, like: *'Blanco River flooding by our apartment on River Rd. The irony does not escape me. For now we're trapped'* or *'#flooding These guys have been stuck on the Katy Freeway @ Washington since before midnight.'*; to **rescue** activities (14 messages) like: *'Hays County: national guard troops helping relief effort. Helicopters rescuing people stranded on their roofs right now'* or *'This man in Orange just saved three people from their car on memorial'*; to **missing** individuals (82 messages), like: *'#Firefighter #Killed as #Flooding #Wreaks #Havoc in #Texas, #Oklahoma', 'Hays officials: 30 'unaccounted for' people have been contacted - 13 still confirmed missing #centexfloods'*. To the latter group of missing individuals belong as well messages reporting about injured, victims of the floods and fatalities. While the category of rescue activities contains microposts describing actual rescue efforts, it also includes messages praising individuals or institutions involved in rescue operations. Into the aforesaid coded categories are categorised several messages syntactically organised without verbal groups.

Two additional categories can be identified in the corpus: **flood advisory** (44 messages) and **general advisory** (54 messages). The distinction between these two categories is evident, the latter includes posts of general advisory such as: *'Tornado is west of Pocasset. People in Pocasset and Minco take cover now! #okwx', 'There are reports of some cars trapped in high water in Irving on North Belt Line Road. Be extremely cautious while driving! #dfwwx #txwx', 'If you smell gas, do not turn the lights on or off or operate anything that might cause a spark. #houstonflood'.*

Whereas flood advisory primarily contains messages providing details about issued flood advisories and flood safety recommendations such as: *'Midnight: dangerous flash flooding continues from Lawton to Norman to Tulsa. Avoid travel!!'*. Both above discussed categories encompass posts composed by various organisations, authorities, and individuals. A semantic feature and the key difference between the messages coded as advisory and such coded as warnings is that the advisory texts mostly employ imperative clauses. By their nature imperative grammatical structures aim to achieve a certain outcome or response from the reader or listener. Hence, advisory messages are preliminary focused on raising awareness and urging readers to

take necessary actions immediately and explicitly, while warning messages are also focused on raising awareness however require actions only implicitly.

Prominent in the corpus are posts expressing authors' emotions through a sarcastic or non-sarcastic language towards i) the current weather conditions and their effect, ii) personal and collective experiences, such as fear, loss, damage, or discomfort caused by the floods, iii) the causes or implications of the floods including climate change, political issues, or social justice and iv) flood management. These categories of messages are labelled respectively **conveying emotional state** (156 messages), which include posts like: *'#SanMarcosFlood got trees in the road, cars underwater. This has got to be a joke'*, *'OMG the floods in #Texas and #Oklahoma so many prayers for all of those dealing with loss'*, *'Oh hey @AnniseParker We pay \$150 Million+ annually for drainage fee's in Houston. How's that working? #TexasFlood'*; and **sarcastic lens on flood management** (37 messages), which includes posts like: *'God I know you said you are going to wash away our sins ... But got damn!! For 2 months straight !?#HoustonFlood'*, *'when you think it's a river but it's actually a highway #houstonflood'*, *'In Oklahoma, you don't go to the lake. The lake comes to you.'*, *I bet Parking Services still out tryna write tickets lol #SanMarcosFlood #TXST'*.

Distinguished can be a category of posts which convey information about **weather conditions** (66 messages), describing various details such as temperature, precipitation, wind, humidity, cloud cover etc. In a similar way, coded are messages which discuss the impact of the rain on the water levels and flood risk information about river levels in the category **information river levels** (34 messages). These two categories include several posts not containing verbal groups. Besides the former two categories, two more of describing road hazards and risks, road closures and openings, flooded building structures or areas, infrastructure local impact and conditions have been coded as follows: **infrastructure of roads** (119 messages) and **infrastructure of areas** (19 messages). These posts differ from the rest in the corpus because they contain geospatial information, providing respectively name of a street or building structure or name of an area (cities, counties) and describing an event: *'Serious flooding at Red River looking north from Texas side I-35 1:46pm 5-22-2015 @tvnflood @TWCBreaking #txwx #okwx'*, *'IH-35 is closed in both directions from Aquarena Springs north to Blanco River Road. #TXST #SMTX #SMTXFlood'*, *'Water flowing over the twin bridges at Lake Thunderbird this morning #okwx'* and areas: *'Lots of low water crossings closed in Williamson County, including Harrell Pkwy in Old Settlers Park. #txwx'*.

Additional categories inferred from the corpus include **offering and requesting help** (29 messages) and **requesting information** (10 messages) during the floods, related to rescue actions, finding shelter, cleaning and salvaging individuals' belongings, or offering and requesting financial

or material support, donating goods, etc. A further defined category of **information channels** (67 messages) couples together all messages in the corpus which give helpful recommendations about various available channels and point to different resources of where latest updates can be found. Another set of coded message types are as follows: **reporting damage** (43 messages), **danger** (8 messages), **federal aid** (10 messages), **animal care** (8 messages) and **donations** (19 messages). The category of reporting damage includes microposts related to damage and various impacts of the flooding situation, ranging from individual property damage to broader community and environmental effects. Example of these are: *'City officials just raised the estimate of buildings damaged in #houstonflood to 4,000, from 1,400.'*, *'Several new trees planted after 2011 fire @ #bastropstatepark injured in #memorialdayflood'*, *'#houstonflood the entrance to my apartment home has been swallowed by White Oak Bayou'*, *'Cars becoming disabled in road flooding in Lawton, OK @ Sheridan'*, *'Wastewater overflows from 10 Austin treatment plants: #atxfloods'* and more. The category of danger in general contains microposts of hazards associated with the severe weather and flooding events, e.g. *'Oil and other chemicals leaking out of 1000 N. Lamar #atxfloods'* or *'Patrol officers reporting traffic signal issues and very heavy rain making driving conditions in Waco very hazardous.#txwx'*.

It should be noted that some of the messages can relate to more than one of the labelled categories above and a differently motivated classification will diverge from the current. Despite that other classifications may appear to be equally valid, here the approach was to take into account the primary subject of each message and use it as the fundamental criteria for the groupings. Figure 11 provides illustration of all coded message types inferred from the corpus.

Except the above listed categories, the corpus also includes microposts referring to insurance practices (2 messages), stolen belongings (1 message), lost belongings (1 message), messages describing that a certain threat is gone, like the dam is not going to break (2 messages) and such discussing save acts (4 messages). Condition for developing the categories is that each category should contain a minimum of five microposts from the collected corpus. Thus, these message types have been merged to other closely related categories.

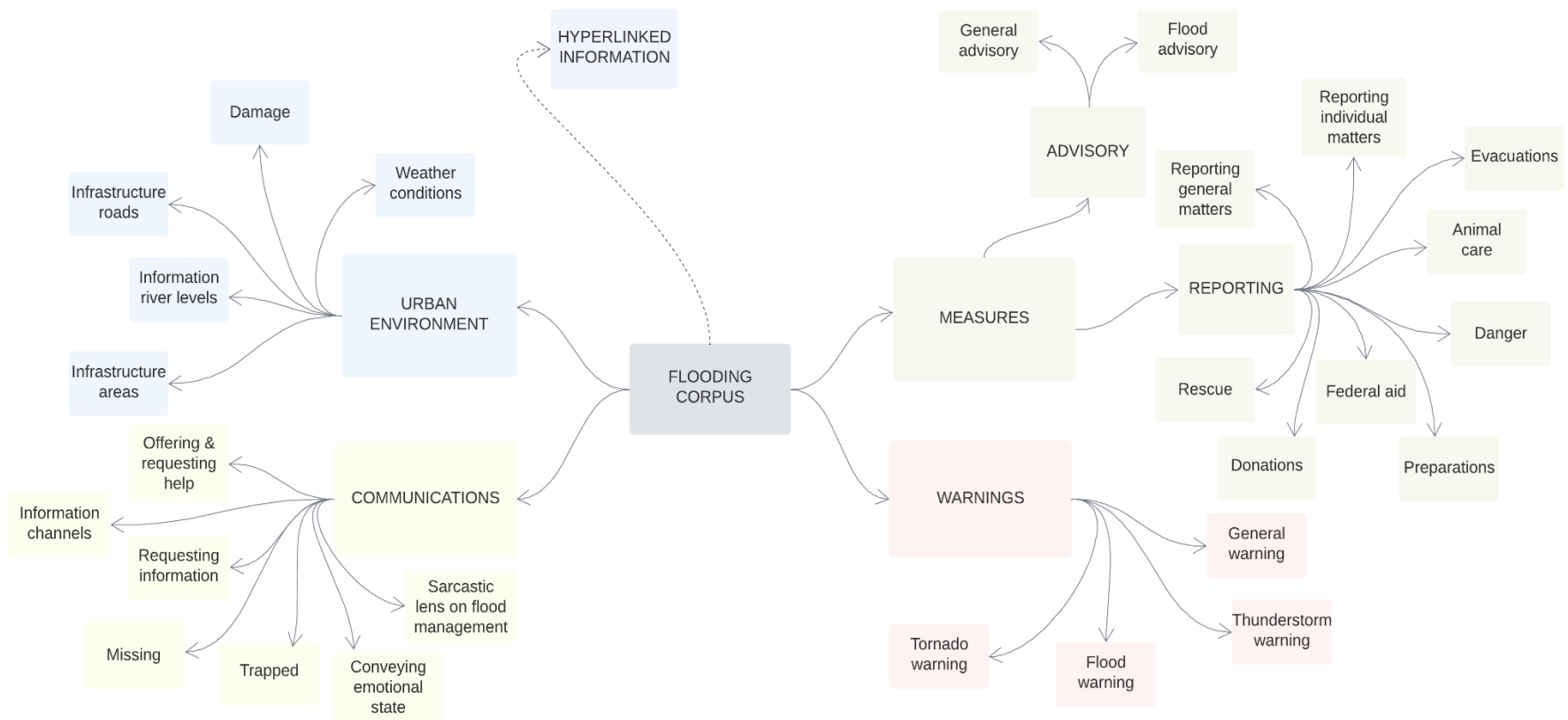


Figure 11: Information and message types

Included in the figure above is the category of **hyperlinked information**, which is not taken into consideration for further analysis.

Outlined in Table 8 below is a day-by-day (May 22 to May 28) breakdown of the entire corpus, showing the dynamics of the different categories of message types across the seven days. It provides an overview of how microblogged communication unfolds in the different days of the Memorial Day floods and shows the activity patterns of sharing information during these days.

The statistics (Table 8) reveal the shifts in emphasis of shared information throughout the observed time period. On May 22, individuals and organisations predominantly shared information about flood warnings (01) - 25%, flood advisory (10) - 14% and weather conditions (15) - 14%, raising awareness within their audiences of an upcoming potential danger or risk weather conditions that may cause harm or damage. Among other categories, the emphasis on the topic of flood advisory (10) implies that actors use the microblogging platform as means to convey important and timely informational guidance or recommendations, which require direct actions. Interesting is also that the category of preparation (05) has its highest value of 11% on May 22 in comparison with other days.

Furthermore, in addition to the category of tornado warning (03), on May 23 individuals started broadcasting information about thunderstorms (02), each of these two categories representing 15% of all message types in that day, followed by flood warning (01) and general advisory (09) respectively with 13% and 12%. While May 22 contains 12 of all 27 message types comprising the corpus, May 23 contains 15 message types. Respectively 16% and 12% of all message types make up flood warnings (01) and reporting on general matters (06) on May 24. Starting from May 24 there is a significant increase in the number of message types in the following days of approximately 24/25 types from all 27 message types, and except reporting about general matters (06), here individuals start reporting about any individual matters (07) as well.

As shown on Table 8, the main types of information represented on May 25 are reporting on general matters (06) and tornado warnings (03) each of which constitute 13% of all message types on that day, followed by the categories of thunderstorm warnings (02) with 12% and infrastructure roads (16) with 11%. In addition, users started sharing information about donations (26), animal care (25), and possible dangers (23). Followingly, the highest proportion of communicated information on May 26 is related to conveying emotional states (14), representing 16% of all message types in the corpus. It is followed by the category of representing general matters (06) with 14%, infrastructure roads (16) and missing individuals (13), each accounting for

8%. Furthermore, on May 27 there is a shift towards the categories of reporting general matters (06) - 18%, followed by thunderstorm warnings (02) - 13%.

<i>Category</i>	<i>May 22</i>	<i>May 23</i>	<i>May 24</i>	<i>May 25</i>	<i>May 26</i>	<i>May 27</i>	<i>May 28</i>
01 Flood warning	7 (25%)	10 (13%)	45 (16%)	31 (10%)	39 (7%)	24 (10%)	8 (9%)
02 Thunderstorm warning	-	12 (15%)	11 (4%)	39 (12%)	20 (4%)	32 (13%)	21 (24%)
03 Tornado warning	1 (4%)	12 (15%)	29 (10%)	44 (13%)	15 (3%)	9 (3%)	9 (10%)
04 General warning	1 (4%)	5 (6%)	8 (3%)	11 (3%)	13 (2%)	7 (3%)	2 (2%)
05 Preparation	3 (11%)	1 (1%)	1 (1%)	-	-	-	-
06 Rep. general matters	3 (11%)	5 (6%)	35 (12%)	43 (13%)	79 (14%)	44 (18%)	4 (5%)
07 Rep. individual matters	-	-	1 (1%)	7 (2%)	28 (5%)	3 (1%)	-
08 Evacuations	-	-	9 (3%)	1 (1%)	-	1 (1%)	1 (1%)
09 General advisory	1 (4%)	9 (12%)	7 (2%)	15 (5%)	18 (3%)	3 (1%)	1 (1%)
10 Flood advisory	4 (14%)	3 (4%)	9 (3%)	8 (2%)	15 (3%)	4 (2%)	1 (1%)
11 Trapped	-	-	6 (2%)	3 (1%)	9 (2%)	-	-
12 Rescue	-	-	2 (1%)	2 (1%)	6 (1%)	3 (1%)	1 (1%)
13 Missing	1 (4%)	-	11 (4%)	10 (3%)	46 (8%)	19 (8%)	5 (6%)
14 Conv. emotional state	-	4 (5%)	11 (4%)	18 (5%)	88 (16%)	21 (8%)	14 (16%)
15 Weather conditions	4 (14%)	8 (10%)	11 (4%)	11 (3%)	18 (3%)	13 (5%)	1 (1%)
16 Infrastructure roads	1 (4%)	3 (4%)	24 (8%)	35 (11%)	43 (8%)	11 (4%)	1 (1%)
17 Infrastructure areas	-	1 (1%)	6 (2%)	2 (1%)	10 (2%)	-	-
18 Information river levels	1 (4%)	1 (1%)	16 (6%)	6 (2%)	8 (2%)	2 (1%)	-
19 Offering & requesting help	-	-	2 (1%)	8 (2%)	6 (1%)	5 (2%)	7 (8%)
20 Information channels	1 (4%)	2 (3%)	13 (5%)	10 (3%)	24 (4%)	12 (5%)	5 (6%)
21 Reporting damage	-	-	12 (4%)	3 (1%)	18 (3%)	9 (4%)	1 (1%)
22 Sarcastic lens on floods	-	2 (3%)	7 (2%)	3 (1%)	22 (4%)	4 (2%)	-
23 Danger	-	-	-	3 (1%)	3 (1%)	1 (1%)	-
24 Federal aid	-	-	1 (1%)	-	3 (1%)	4 (2%)	2 (2%)
25 Animal care	-	-	-	2 (1%)	2 (1%)	3 (1%)	1 (1%)
26 Donations	-	-	-	5 (2%)	5 (1%)	7 (3%)	2 (2%)
27 Requesting information	-	-	1 (1%)	1 (1%)	5 (1%)	3 (1%)	-
<i>Total</i>	<i>28</i>	<i>78</i>	<i>278</i>	<i>321</i>	<i>543</i>	<i>244</i>	<i>87</i>

Table 8: Relative and absolute values of information and message types across the days in the corpus. The percentage have been rounded to whole numbers.

On the last day of the observed timeframe, the number of message types decreases significantly, with represented 19 message types of all 27. Microposts mostly revolved around the categories of thunderstorm warning (02) and conveying emotional states (14), representing 24% and 16% of the messages on this day. These are followed by communicated information about tornado (03) and flood warnings (01) and offering and requesting help (19), which has its highest values on that day in comparison with other days.

Overall, the analysis show that while nine of all message types: flood (01), tornado (03) and general warnings (04), reporting general matters (06), general (09) and flood advisories (10), weather conditions (15), infrastructure roads (16) and information channels (20) are fully present throughout the whole timespan, other message types are concentrated around definite periods.

Actors begin reporting regarding individual matters (07), evacuations (08), trapped people (11) and rescue activities (12) on the second day of the noted critical period (May 23 – May 26) from the Memorial Day floods - May 24. On the same day discussions about offering and requesting help (19), about caused damage (21) and about missing people or fatalities (13) are initiated. In the sampled data, messages expressing types of emotions such as conveying emotional states (14) and sarcastic lens on flood management (22) remain relatively constant throughout a longer timeframe and, in particular, the category of conveying emotional states (14) has a strong representation throughout the days of May 23 – May 28 2015.

By contrast however anticipated, posts revolving about preparation measures (05) for the floods compile the only one category which occurs from the beginning of the period on May 22 until May 24. Message types like donation activities (27), animal care (25) and danger (23) follow quite the reverse pattern and occur from May 25 until May 27 and 28 in the corpus. The occurrence of the category danger (23) as late as May 25 may seem a little counterintuitive, but this can be attributed to the fact that it is constituted of posts describing dangerous situations that emerged as a result of the flooding.

4.4 Process variation across units and message types

Based upon their meanings and what is talked about, the laid-out message types can be summarised into four overarching units, represented on Figure 11 above.

The unit of **warnings** (465 messages) includes messages which deal with any types of alerts that warn people of tornado, thunderstorms, flood or general warnings. Messages in which authors communicate matters about trapped and missing people or convey emotions including sarcastic utterances constitute the unit of **communications** (409 messages). **Urban environment** (280

messages) comprises of posts in which users describe their own or other physical environment and the unit of **measures** (425 messages) is a collection of messages in which users provide advice or report about particular matters. The latter classification has emerged through coupling together the inferred from the corpus message types and serves as a pragmatic approach for the further analysis of process types and their dynamics across the corpus. Followingly, the distribution of experiential process types across these units and categories is shown below (Figure 12 and Table 9). As illustrated in Figure 12, all process types within the experiential system are represented in each of the four units.

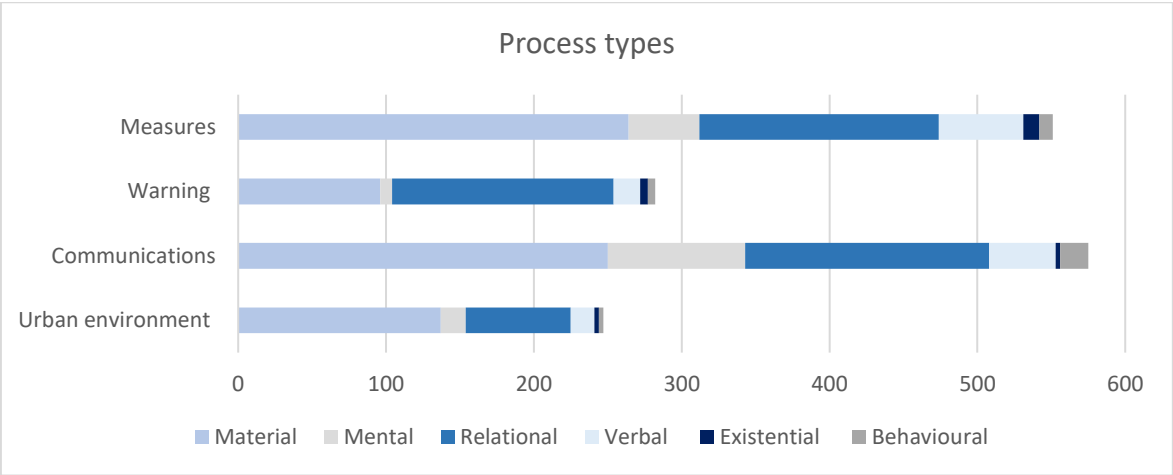


Figure 12: Overview of the absolute distribution of process types across units

Table 9 provides an overview of the variation of process types across the units and their categories. In each unit, highlighted in grey is the total number and percentage of processes for each process type. For each process type (calculated per column) are the number and percentages of processes in each category.

The unit of **urban environment** includes 247 experiential processes (see Table 9), of which a significant number are material processes (55%), followed by relational processes (29%). The highest values of material and relational process types across all categories originate from the category of infrastructure roads (16), respectively 34% and 37%. These are followed by mental and verbal processes with respectively 7% and 6%. Material and relational processes are leading the rank within all the five categories which constitute this unit (15, 16, 17, 18, 21).

	<i>Category</i>	<i>Material</i>	<i>Mental</i>	<i>Behavioral</i>	<i>Verbal</i>	<i>Existential</i>	<i>Relational</i>
WARNINGS	01 Flood warning	21 (22%)	-	-	6 (37%)	1 (20%)	64 (43%)
	02 Thunderstorm warning	20 (21%)	1 (13%)	-	2 (13%)	1 (20%)	34 (23%)
	03 Tornado warning	29 (30%)	1 (13%)	3 (60%)	4 (25%)	-	40 (27%)
	04 General warning	26 (27%)	6 (74%)	2 (40%)	4 (25%)	3 (60%)	12 (8%)
	<i>Total 280</i>	96 (34%)	8 (3%)	5 (2%)	16 (6%)	5 (2%)	150 (54%)
MEASURES	05 Preparation	3 (1%)	-	-	1 (2%)	-	2 (1%)
	06 Rep. general matters	112 (42%)	28 (58%)	4 (45%)	34 (60%)	6 (55%)	63 (39%)
	07 Rep. individual matters	28 (11%)	6 (13%)	2 (22%)	3 (5%)	2 (18%)	12 (7%)
	08 Evacuations	6 (2%)	-	-	2 (4%)	-	4 (2%)
	09 General advisory	43 (16%)	5 (10%)	-	4 (7%)	3 (27%)	36 (22%)
	10 Flood advisory	19 (7%)	2 (4%)	2 (22%)	2 (4%)	-	29 (18%)
	12 Rescue	11 (4%)	3 (6%)	-	2 (4%)	-	6 (4%)
	23 Danger	6 (2%)	-	-	1 (2%)	-	1 (1%)
	24 Federal aid	8 (3%)	2 (4%)	-	3 (5%)	-	4 (2%)
	25 Animal care	8 (3%)	-	1 (11%)	2 (4%)	-	2 (1%)
	26 Donations	20 (8%)	2 (4%)	-	3 (5%)	-	3 (2%)
	<i>Total 551</i>	264 (48%)	48 (9%)	9 (2%)	57 (10%)	11 (2%)	162 (29%)
URBAN ENVIRONMENT	15 Weather conditions	34 (25%)	4 (23%)	-	2 (12%)	2 (67%)	17 (24%)
	16 Infrastructure roads	46 (34%)	7 (41%)	2 (67%)	6 (38%)	-	26 (37%)
	17 Infrastructure areas	10 (7%)	-	-	-	-	5 (7%)
	18 Information river levels	19 (14%)	3 (18%)	-	1 (6%)	-	13 (18%)
	21 Reporting damage	28 (20%)	3 (18%)	1 (33%)	7 (44%)	1 (33%)	10 (14%)
	<i>Total 247</i>	137 (55%)	17 (7%)	3 (1%)	16 (6%)	3 (1%)	71 (29%)
COMMUNICATIONS	11 Trapped	12 (5%)	2 (2%)	2 (11%)	3 (7%)	-	4 (2%)
	13 Missing	51 (20%)	4 (4%)	1 (5%)	10 (23%)	-	18 (11%)
	14 Conv. emotional state	83 (33%)	69 (74%)	13 (68%)	9 (20%)	1 (33%)	80 (48%)
	19 Offering & requesting help	27 (11%)	1 (1%)	-	1 (2%)	1 (33%)	19 (12%)
	20 Information channels	46 (18%)	8 (9%)	-	17 (39%)	-	22 (13%)
	22 Sarcastic lens on floods	26 (10%)	7 (8%)	2 (11%)	4 (9%)	1 (33%)	14 (8%)
	27 Requesting information	5 (2%)	2 (2%)	1 (5%)	-	-	8 (5%)
	<i>Total 574</i>	250 (44%)	93 (16%)	19 (3%)	44 (8%)	3 (1%)	165 (29%)

Table 9: Absolute and normalised distribution of process types and categories in units. The percentage have been rounded to whole numbers.

The unit of **communications** entails a total of 574 processes, which makes it the unit containing the greatest number of processes compared to the rest three units. Material (44%), relational (29%) and mental (16%) processes comprise the majority of all process types. It is interesting that from all included in this unit categories (11, 13, 14, 19, 20, 22, 27), the highest value and one third of all material processes (33%) stem from the category conveying emotional state (14). Moreover, a closer look into the former category reveals that almost three quarters of all mental processes (74%) and a large proportion (68%) of all behavioural processes within this unit originate from it. Thus, this category appears to be one of the richest on processes among others, which can be attributed to the fact that it contains messages conveying authors' emotions and attitudes. In addition, the results also show that 39% of all verbal process in this unit stem from the category of information channels (20). In this category individuals provided details about where timely and accurate information regarding the flooding can be found, talked about communications, information sharing, and reporting.

The unit of **measures** comprises of 551 experiential processes, with material processes accounting for nearly half (48%) of all processes. Following material processes, relational processes make up a significant portion (29%) of all, followed by verbal processes (10%). Table 9 shows significant differences and shifts among processes across categories, namely the category consolidating posts which deal with reporting general matters (06) is the richest category on processes. It has the highest values of verbal processes making up almost two-thirds (60%) of all verbal processes, 58% mental and 42% material processes of all processes in the unit accordingly.

While in the above discussed three units process types are organised in the schema of material - relational - mental/verbal processes, what stands out in the unit of **warnings** is that this general pattern is interrupted. The process system in this unit reveals different dynamics than those observed above. On the first position rank relational processes making up just over the half of all processes (54%), followed by material (34%) and verbal (6%) processes. The majority of all relational processes in this unit come from the category labelled flood warning (01) - 43%, showing the tendency of conveying information about how authors relate and create discursive ties among various entities in their microposts. Among all four warning categories material processes are fairly evenly distributed.

Overall, the analysis point out that across the four units experiential processes show a repetitive pattern of material-relational-mental-verbal-behavioural-existential processes, with some shifts of the last three process types.

5. Results

The conducted keyword and frequency analyses aimed to investigate the relationship between the content and the lexical choices of authors, as well as to illustrate distinctive features of the Memorial Day floods corpus. These analyses revealed that certain lexical items, such as 'flood', 'warning', 'tornado', 'thunderstorm', 'water', and more, exhibited a notably high keyness score in the collected corpus compared to the general Twitter corpus.

Additionally, the keyness analyses highlighted a specific characteristic of the corpus: the prevalence of proper nouns, particularly location names like: 'Houston', 'Oklahoma', 'Austin', 'Texas', 'Lamar', 'Blanco River', 'San Marcos' etc., which dominated the corpus in contrast to the general one. Furthermore, a deeper examination of verb frequencies within the corpus identified the most commonly used verbs, providing valuable insights into the language patterns employed by authors in discussing the flooding.

The experiential analysis revealed that in 1579 analysed posts, authors realised a total number of 1656 experiential processes in their utterances, encompassing all six experiential process types in the semiotic transitivity system: material, mental, verbal, behavioural, existential, and relational. The comparison of process types across the seven days uncovered the predominant linguistic choice and construed meanings of users, through which they requested or provided specific information in their microposts about the Memorial Day floods in Texas and Oklahoma, 2015. The most frequent experiential processes throughout the observed days built the following two striking patterns of verbs in the corpus:

- i. material processes - relational processes - verbal and mental processes, where verbal and mental processes have been represented in even proportions and
- ii. relational processes - material processes - mental processes.

The findings showed that material processes, through which authors explicated concrete actions and happenings from their physical environment related to the floods constituted the majority of all processes in the corpus. Actors talked about the flow of the flooding in the world around them, about other actors who do things, or make things happen, and about doing or acting.

Furthermore, through relational processes authors conveyed meanings about relationships of either describing or identifying individuals or objects in the context of the specific discourse of flooding. On the third place ranked the last of the three principal processes, mental processes, through which authors represented their emotions, thoughts, attitudes in the context of the flooding events. In addition, as a result of the classification combining the different message and

information types from the corpus in overarching units (warnings, communications, urban environment, and measures) an overview of the distribution of process types across the different units and categories has been produced and analysed.

In order to find out what kinds of contextual meanings actors deposited in their utterances, Systemic Functional Linguistics has been applied in discourse analysis. Firstly, the focus of this part was put on aligning and classifying semantically fitting construed meanings exchanged during the Memorial Day floods into categories of information and message types. The analysis revealed twenty-seven distinct message types of communicated information. Followingly, several of these message types have been reflected in larger proportions in the corpus in comparison with others. Secondly, in order to examine the dynamics of these message types and their arrangement across the days, a comparison in a day-by-day perspective was carried out. Some observed fluctuations of categories on the different days underscored the evolving nature of microblogged communication and discursive patterns during the Memorial Day floods.

The shifting emphasis of broadcast information within the days - from different types of warnings and advisories to wider discussions in the domains of reporting any general matters (including generalised information about the weather and types of emergency alerts, community impact and support, official response) and reflecting actors' emotions provided a nuanced perspective on how actors engaged with the current flooding situation over time. Each of the inspected days exhibited distinct characteristics in terms of the focus and prevalence of its categories, providing insights into the dynamics of shared microposts during the flooding.

Overall, nine of the inferred information and message types showed a stable pattern throughout the whole timespan of seven days, while other categories appeared to be concentrated around certain periods of time, e.g. the period before the flooding and during the flooding towards its ending.

Lastly, this study aimed to reveal and discuss the linguistic devices and mechanisms which authors employed to convey contextual meanings. From a semiotic perspective, the discourse analysis provided examples in which authors employed a number of linguistic and literary devices to construe meanings, such as metaphors, personifications, similes, sarcasm, ellipses, functional and formulaic language.

6. Discussion and conclusions

Depicting language as a social semiotic system, this study has explored the language of microblogging through the episodic status updates which users broadcast in a specific timeframe during the Memorial Day floods. To address the predefined research questions, while delving into aspects of construed contextual meanings and their dynamics encoded by actors, the study also highlighted the non-standard conversational language norms which users employed to communicate meanings. Furthermore, it also showed and discussed the complex and multi-layered linguistic substance of shared textual information during the floods. The developed methodological framework served as lens to:

- i. examine the lexical choices and frequency distributions across the corpus,
- ii. study the experiential processes which authors encoded to talk about actions, happenings and states during the floods,
- iii. show what kinds of contextual meanings users deposited in their utterances and how they changed over time, and
- iv. uncover the linguistic nature of microposts.

Viewed through a semiotic prism three key interrelated aspects guided the research parallelly to address the research questions: keyness analysis, analysis of experiential processes and classification of contextual meanings through a temporal perspective.

The keyword analysis aimed to explore what distinct characteristics does language used in the discourse of the Memorial Day floods exhibit in comparison to the communication in the broader context of Twitter. By comparing frequencies and distribution of its keywords, several linguistics features unique to the flood-related discourse were identified. This analysis provided insights into how language varies across different domains and what are the specific linguistic markers of the crisis context in the Memorial Day floods.

To study the experiential processes of the transitivity system, this analysis has drawn on the fact that the linguistic choice of actors is encoded in the lexicogrammatical structures of their utterances. The different types of verbs and verbal groups through which authors realised types of experiential processes and meanings have been explicated and analysed from a semiotic stand. Practically, verbal processes of Systemic Functional Linguistics and their frequencies have shed light on what individuals talked about and what kinds of events, happenings and states have been reflected in the corpus. The results and findings indicate also, that employing the system of experiential processes of Systemic Functional Linguistics has proved to be fitting, used as lens to

aid describing the experiential meanings of packed up language in the concise linguistic environment (140 characters) of Twitter microposts. To show the fluctuation and frequencies of types of experiential processes across the units and categories in the corpus, four overarching message units have been created and primarily served the purpose of a practical approach for the analysis. In order to carefully align these message units in the domain of crisis management a more informed approach is needed, for which, the scope of this paper was too narrow.

Furthermore, the close investigation of experiential processes provided the foundations and facilitated the simultaneous discourse analysis of single utterances in a semiotic mode. Based on this analysis, a total of twenty-seven specific information and message types have been identified within the Texas and Oklahoma flooding corpus, which show what kinds of communicated contextual meanings users deposited in their utterances when they talked about the flooding. In her work, Vieweg analysed Twitter data collected during four different mass emergency events (2009 and 2010 Red River floods and the 2009 Oklahoma grass fires, 2010 Haiti earthquake) and identified thirty-two categories of information types inferred from the data (cf. Vieweg, S. E. 2012:261). In particular, she has explored how Twitter is used in mass emergency situations and established a scheme of specific information types that contribute to the concept of situational awareness during crisis situations.

Interpretative classification approaches like the one employed in this study might turn subjective, no matter how consistently a researcher tries to follow the characteristics of the data and the beforehand set-up rules and principles of the analysis. However, although Vieweg's classification differs in many ways from the one performed here, the inferred categories from the present data broadly align with those found in Vieweg's collected dataset. This also implies, that the nature of shared information in Twitter during such unforeseen events remained stable over time.

In their work Zade et al. (2018) have discussed the concept of situational awareness introduced and operationalised by Vieweg (2010, 2012) as such that shaped further research and design in the topic of identifying related information to situational awareness during emergency situations, broadcast in microblogging services like Twitter. They have also outlined a more recent paradigm shift of research from this general concept of situational awareness towards the narrower concept of 'actionability', recognizing that different responders may have different information needs in different time. The authors defined the concept of 'actionability' as any information containing a request or suggestions that a person could act upon (Zade et al. 2018:195:5). Within the grammar of experience, Halliday (2004:172) identified the three principal units which comprise experiential meanings: experiences related to and realised through doings, sensing and

being (as discussed in this thesis). In the domain of material processes, experiences can encompass and refer to events, happenings, doing and acting and in these lines, adapting an analytical framework which includes a semiotic perspective, and its functional elements may further aid research aiming towards identifying 'actionable information'.

In terms of the data used for this study, it should be highlighted that due to constraints within the scope of this thesis, comprehensive analysis of entire threaded conversations within the Memorial Day floods communication was not possible given the nature of the collected data. Threaded conversations can have the potential to offer a richer semantic understanding of the inferred message and information types, their dynamics and participants' linguistic realisations.

In this thesis, the approach of employing a temporal perspective on the analyses played a central role in better demonstrating the dynamics of the shared textual data. The integrated day-by-day frame has enabled the exploration of linguistic realisations of experiential contextual meanings on the one hand, and on the other hand the exploration of the different kinds of information, actors sought and provided over time in the context of the Memorial Day floods 2015. Overall, this comparative approach has aided in reflecting the shifts of individuals' reactions, attitudes and responses throughout the different days of the crisis.

References

- Archive Team: The Twitter Stream Grab. Available: <https://archive.org/details/twitterstream>
[accessed 2022, September 19]
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Baker, P., & Egbert, J. (2016). *Triangulating methodological approaches in corpus linguistic research*. New York; London: Routledge.
- Bartlett, T., & O'Grady, G. (2017). *The Routledge handbook of systemic functional linguistics*. New York London: Routledge, Taylor & Francis Group.
- Bernardini, S., Baroni, M. and Evert, S. (2006). A WaCky introduction. In *Wacky! Working Papers on the Web as Corpus*, ed. M. Baroni and S. Bernardini. Bologna: GEDIT
- Bloor, T., & Bloor, M. (2004). *The Functional Analysis of English. A Hallidayan Approach*. (2. ed.). Arnold. <https://doi.org/10.4324/9780203774854>
- Borer, H. (1990). V + ing: It Walks like an Adjective, It Talks like an Adjective. *Linguistic Inquiry*, 21(1), 95–103.
- Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2), 91–108. doi: 10.1080/13645579.2012.756095
- Bubenhofer, N. (2006-2023): Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge. Elektronische Ressource: <http://www.bubenhofer.com/korpuslinguistik/>.
- Butt, D., Fahey, R., Feez, S., Spinks, S., & Yallop, C. (2003). *Using Functional Grammar*. (2. ed.) National Centre for English Language Teaching and Research, Macquarie University.
- Cresci, S., Tesconi, M., Cimino, A., & Dell'Orletta, F. (2015). A Linguistically-driven Approach to Cross-Event Damage Assessment of Natural Disasters from Social Media Messages. Proceedings of the 24th International Conference on World Wide Web - WWW '15 Companion, 1195–1200. Doi: 10.1145/2740908.2741722
- Eggins, S. (2004). *An introduction to systemic functional linguistics* (2. ed.). New York: Continuum.
- Fawcett, R. P. (2000). In place of Halliday's 'verbal group'. *Word (Worcester)*, 51(2), 157–203. doi: 10.1080/00437956.2000.11432500

- Gabrielatos, C. (2018). Keyness analysis. In *Corpus Approaches To Discourse* (1. ed., pp. 225–258). Routledge. doi: 10.4324/9781315179346-11
- Halliday, M. A. K. (2003). Towards a sociological semantics (1972). In Halliday M.A.K., & Webster, J. *On language and linguistics*. Vol 3. (pp. 323-354). Continuum
- Halliday, M. A. K. (2003). The functional basis of language (1973). In Halliday M.A.K., & Webster, J. *On language and linguistics*. Vol 3. (pp. 298-322). Continuum
- Halliday, M. A. K. (1975). Language as Social Semiotic: Towards a General Sociolinguistic Theory. In Halliday, M. A. K., & Webster, J. J. (2007). *Language and Society: Vol. 10*. (pp. 169–202). Continuum
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2. ed.). London: Arnold.
- Halliday, M. A. K., Webster, J. J., & Halliday, M. A. K. (2003). Appendix: Systemic theory (1994a). In *On Language and Linguistics*. United Kingdom: Bloomsbury Publishing Plc.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3. ed., 1. publ.). London: Hodder Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4. ed.). London: Routledge.
- Hasan, R. (1985). 'The structure of a text'. In Halliday, M. A. K. & Hasan, R. (Eds.), *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. (pp. 52-69). Geelong, V.: Deakin University Press
- Heath, M. (2018). Orthography in social media: Pragmatic and prosodic interpretations of caps lock. *Proceedings of the Linguistic Society of America*, 3(1), 55. doi: 10.3765/plsa.v3i1.4350
- Herring, S. C. (2007). A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet*, 4, article 1, available <https://www.languageatinternet.org/articles/2007/761> [accessed 2023, November 15]
- Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3/4), 248. doi: 10.1504/IJEM.2009.031564
- Jingnan, H. 2023. "Twitter's new data access rules will make social media research harder.". Available: <https://www.npr.org/2023/02/09/1155543369/twitters-new-data-access-rules-will-make-social-media-research-harder> [accessed 2023, November 13]

- Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review Through Text Mining. *IEEE Access*, 8, 1–1. doi: 10.1109/ACCESS.2020.2983656
- Khattari, A., Joshi, A., Bhattacharyya, P., & Carman, M.J. (2015). Your Sentiment Precedes You: Using an author's historical tweets to predict sarcasm. *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 25–30). Lisboa, Portugal: Association of Computational Linguistics.
- Kim, Mira. 2007. *A Discourse Based Study on Theme in Korean and Textual Meaning in Translation*. Sydney, Macquarie University, dissertation.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold Hodder Education.
- Kress, G., & van Leeuwen, T. (2021). *Reading images : the grammar of visual design* (3. ed.). London: Routledge.
- La Rocca, G. & Boccia Artieri, G. (2023) Interpreting the changeable meaning of hashtags: Toward the theorization of a model. *Front. Sociol.* 7:1104686. doi: 10.3389/fsoc.2022.1104686
- Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In *Corpus Linguistics and the Web* (Vol. 59, pp. 133–149). doi: 10.1163/9789401203791_009
- Littman, J. 2017. "Where to get Twitter data for academic research.". Available: <https://github.com/sfm-ui/posts/2017-09-14-twitter-data> [accessed 2023, November 13]
- Martin, B., & Ringham, F. (2006). *Key Terms in Semiotics*. (1. ed.). London: Bloomsbury Publishing Plc.
- Martínez-Rojas, M., Pardo-Ferreira, M. del C., & Rubio-Romero, J. C. (2018). Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. *International Journal of Information Management*, 43, 196–208. doi: 10.1016/j.ijinfomgt.2018.07.008
- Matthiessen, C. M. I. M. (2007). The 'Architecture' of language according to systemic functional theory: developments since the 1970s. In R. Hasan, C. Matthiessen, & J. J. Webster (Eds.), *Continuing discourse on language: a functional perspective* (pp. 505–561). Vol. 2. Equinox Publishing.

- McEnery, A., Xiao, Z., & Mo, L. (2003). Aspect Marking in English and Chinese: Using the Lancaster Corpus of Mandarin Chinese for Contrastive Language Study. *Literary and Linguistic Computing*, 18(4), 361–378. doi: 10.1093/lc/18.4.361
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book* (1. publ.). London: Routledge.
- McEnery, T., & Xiao, R. (2010). *Corpus-Based Contrastive Studies of English and Chinese*. (1. ed.). London: Taylor & Francis Group.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge; New York: Cambridge University Press.
- McEnery, T. (2016) Keywords. In: *Triangulating methodological approaches in corpus linguistic research*. (pp.20-33) New York; London: Routledge.
- McKoon, G., & Macfarland, T. (2000). Externally and Internally Caused Change of State Verbs. *Language (Baltimore)*, 76(4), 833–858. doi: 10.2307/417201
- Mihunov, V., Lam, N. S. N., Zou, L., Wang, Z., & Wang, K. (2020). Use of Twitter in disaster rescue: lessons learned from Hurricane Harvey. *International Journal of Digital Earth*, 13(12), 1454–1466. doi: 10.1080/17538947.2020.1729879
- NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2023). Available: <https://www.ncdc.noaa.gov/billions/events/US/2015> [accessed 2023, June 24]
- NOAA *Flood disaster in Texas and Oklahoma* (June, 2015). Available: <https://www.climate.gov/news-features/event-tracker/flood-disaster-texas-and-oklahoma> [accessed 2023, June 24]
- NWS National Weather Service. National Oceanic and Atmospheric Administration. Available: https://www.weather.gov/tsa/hydro_tadd#:~:text=What%20Is%20Turn%20Around%20Do%20n,to%20water%20over%20the%20roads [accessed 2023, September 21]
- Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication*, 6(2), 181–201. doi: 10.1177/1750481312437441
- Page, R. (2014). Saying ‘sorry’: Corporate apologies posted on Twitter. *Journal of Pragmatics*, 62(Feb), 30–45. doi: 10.1016/j.pragma.2013.12.003

- Page, R., Barton, D., Lee, C., Unger, J. W., & Zappavigna, M. (2022). *Researching language and social media: a student guide* (2. ed.). Abingdon, Oxon; New York, NY: Routledge.
- Panem, S., Gupta, M., & Varma, V. (2014). Structured Information Extraction from Natural Disaster Events on Twitter. Proceedings of the 5th International Workshop on Web-Scale Knowledge Representation Retrieval & Reasoning - Web-KR '14, 1–8. doi: 10.1145/2663792.2663794
- Paquot, M., & Gries, S. T. (2021). *A Practical Handbook of Corpus Linguistics*. Cham: Springer International Publishing AG. doi: 10.1007/978-3-030-46216-1
- Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, 29(6), 2438–2447. doi: 10.1016/j.chb.2013.05.007
- Vieweg, S. E. (2012). Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications. Doctoral dissertation, University of Colorado at Boulder. 1–300
- Schleppegrell, M. J. 2012. Systemic Functional Linguistics: Exploring meaning in language. In James Paul Gee and Michael Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 21-34). New York: Routledge.
- Schmid, H. (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing, Vol. 12*.
- Scott, M. (1999), *WordSmith Tools Help Manual*. (Version 3.0.) Oxford: Oxford University Press.
- Shepherd, J. (2023). 23 Essential Twitter Statistics You Need to Know in 2023. Available: <https://thesocialshepherd.com/blog/twitter-statistics> [accessed 2023, November 6]
- Spitzmüller, J., & Warnke, I. (2011). *Diskurslinguistik: eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin Boston: De Gruyter.
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A. & Gries, S. T. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen & R. Dirven (Ed.), *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems* (pp. 129-152). Berlin, New York: De Gruyter Mouton.

- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.
- Thompson, G. (2014). *Introducing Functional Grammar*. (3. ed.). London: Routledge.
- van Leeuwen, T. (2006). *Introducing social semiotics*. London/New York: Routledge.
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. doi: 10.48550/arxiv.1703.03107
- Victor, D. 2016. The New York Times, 2016. Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk. Available: <https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html> [accessed 2023, November 6]
- Vieweg, S., Hughes, A., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events. *Conference on Human Factors in Computing Systems - Proceedings*, 2, 1079–1088. ACM. doi: 10.1145/1753326.1753486
- Vieweg, S. E. (2012). Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications. ProQuest Dissertations Publishing.
- Vo, B.-K. H., & Collier, N. (2013). Twitter Emotion Analysis in Earthquake Situations. *International Journal of Computational Linguistics and Applications*, 4(1), 159–173.
- Weisser, M. (2015). *Practical Corpus Linguistics* (1. ed.). Hoboken: Wiley. doi: 10.1002/9781119180180
- X Developer Platform. Available: <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets> [accessed 2023, November 20]
- Xiao, Z., & McEnery, A. (2004). A corpus-based two-level model of situation aspect. *Journal of Linguistics*, 40(2), 325–363. doi: 10.1017/S0022226704002543
- Zade, H., Shah, K., Rangarajan, V., Kshirsagar, P., Imran, M., & Starbird, K. (2018). From Situational Awareness to Actionability. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–18. doi: 10.1145/3274464
- Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, 13(5), 788–806. doi: 10.1177/1461444810385097

- Zappavigna, M. (2012). *Discourse of Twitter and social media* (1. ed.). London, England: Bloomsbury.
- Zappavigna, M. (2015). Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, 25(3), 274–291. doi: 10.1080/10350330.2014.996948
- Zappavigna, M. (2015a). Searchable talk: The linguistic functions of hashtags in tweets about Schapelle Corby. *Global Media Journal*. Australian Edition, (9).
- Zappavigna, M. (2018). *Searchable talk: hashtags and social media metadiscourse*. London/New York: Bloomsbury Academic, an imprint of Bloomsbury Plc.
- Zappavigna, M. & Dreyfus, S. (2022). “In these pandemic times”: The role of temporal meanings in ambient affiliation about COVID-19 on Twitter. *Discourse, Context & Media*, 47, 100595–100595. doi: 10.1016/j.dcm.2022.100595