Available online at www.sciencedirect.com

ScienceDirect

Lingua 277 (2022) 103401

Lingua

www.elsevier.com/locate/lingua

Emotion at the end of life: Semantic annotation and key domains in a pilot study audiovisual corpus

Clara Inés López-Rodríguez

Department of Translation and Interpreting, University of Granada, Buensuceso, 11, 18002 Granada, Spain

Received 5 March 2022; revised 10 June 2022; accepted in revised form 4 July 2022;

Abstract

This article focuses on emotion talk in English and the semantic annotation of emotions in a pilot study corpus about the end of life. It describes the process of compiling and annotating a corpus containing the transcript of the verbal component of audiovisual material regarding end-of-life care. The paper also aims to present a lexico-semantic analysis of emotion talk based on the combined use of two corpus processing tools: Wmatrix and Sketch Engine. The findings indicate that the limitations of semantic annotation can be overcome by concordance and collocational analysis. They also reveal that the lexis of emotion is commonly present at the end of life and show the main keywords and key concepts, the predominant semantic categories of emotion and the most frequent emotion words in the corpus. The results suggest that the most frequent emotions in the corpus are SADNESS, FEAR, LIKING, LOVE, HAPPINESS/RELIEF, WORRY, CALMNESS, ANGER, HOPE and CONFIDENCE.

Keywords: Emotions; Emotion talk; Health care communication; Corpus analysis tools; Semantic annotation

1. INTRODUCTION

Emotions and how they are expressed have been a focus of human introspection from the times of Plato and Aristotle: our survival, thoughts and cultural activity are the result of experience and emotions (Damasio, 2003, 2018). Cognition, emotion and affect are bound to our lived bodily experience and allow for decision-making, moral evaluation and the ability to understand others and ourselves (Maiese, 2011: 3). The ability to identify, label and discuss our emotions, as well as emotional regulation, that is, the efficient management of both positive and negative emotions (Fernández-Berrocal and Ramos-Díaz, 2005), are very important for our health and well-being. Emotions and emotional abilities are fundamental to keeping us alive and physically and mentally healthy, and are particularly beneficial when the end of life is approaching, since expressing emotions can provide solace in such moments. Consequently, interdisciplinary empirical research on the expression of emotions as cognitive-subjective, physiological and behavioral phenomena (Lang, 1995) is relevant to linguistics, psychology and medicine alike.

E-mail address: clarailr@ugr.es

https://doi.org/10.1016/j.lingua.2022.103401
0016-7037/

This paper aims to analyze the feelings verbalized by people and their families when life is coming to an end due to a terminal disease or old age. The objectives of this article are twofold. First, I describe methodological issues related to the automatic semantic annotation of corpora compiled from audiovisual material, as well as their lexico-semantic analysis using two corpus management tools: Wmatrix (Rayson, 2009) and Sketch Engine (Kilgarriff et al., 2014). Second, I present a semantic study of the emotions articulated by people faced with the prospect of death due to a terminal illness or old age. The use of corpus linguistics and audiovisual resources to identify the lexical expressions and non-verbal cues that are present in these poignant moments can complement emotion research in the fields of psychology, palliative medicine and gerontology. It can also provide data for both lexicographic resources and English for Specific Purposes materials aimed at health professionals and caregivers with limited proficiency in English. In short, this paper focuses on the lexical expression of emotion in audiovisual sources, explores the interface between linguistics and psychology, and provides a methodology to obtain lexical data for materials intended for psychologists, health practitioners and caregivers assisting terminal patients.

To this end, I review the literature on emotion that is pertinent to this study and analyze an English-language pilot study corpus about the end of life that was compiled from the following audiovisual sources: 4 Netflix documentaries, 11 TED talks, and the summaries and transcripts of 40 interviews with patients and their caregivers. The corpus is called *Terminal corpus* and contains 80,274 words. The videos deal with the topic of death caused by terminal illness or old age, and were selected mainly on the basis of the relevance and influence of the organizations behind them, more specifically, Netflix, TED: Ideas Worth Spreading, and Healthtalk.org (supported by the University of Oxford). The individuals appearing on the screen were real health professionals, patients, caregivers and relatives.

The objectives of this paper have been guided by the following research questions (RQs):

- RQ1. What keywords and key concepts are found in the *Terminal corpus*, a pilot study corpus about the end of life that has been semantically annotated with the UCREL Semantic Analysis System (USAS)?
- RQ2. How relevant are emotion words in relation to other semantic categories in this corpus?
- RQ3. What are the predominant general categories of emotion in the *Terminal corpus*?
- RQ4. What are the most frequent emotion words and the most common emotions in the corpus once the inaccuracies of automatic semantic annotation have been corrected?

2. LITERATURE REVIEW: EMOTIONS, MOVING OBJECTS OF STUDY IN PSYCHOLOGY, LINGUISTICS AND MEDICINE

Although there is no consensus on what an emotion is, most scientists will agree that it is a set of physiological, cognitive, subjective and motor changes derived from the conscious or unconscious assessment of a stimulus in a given context and in relation to the purposes of an individual in a particular moment of their life (Cotrufo and Ureña, 2018: 17). Emotions are often used as synonyms for feelings, moods and personality traits in spite of their differences, and all of them fall under the umbrella term of 'affect'. In this article, the word 'emotion' is used as the general term that encompasses all of them.

Emotive communication is a complex verbal, vocal and kinesic phenomenon (Bednarek, 2008: 9): emotions are expressed both verbally and non-verbally. The present study focuses on the verbal facet of emotion, mainly the lexis. However, non-verbal communication is equally important and is intertwined with the linguistic expression of emotions. There are paralinguistic aspects (such as intonation, prosody, pitch, loudness, pauses or rhythm of speech), involuntary physiological responses (pupillary dilatation, perspiration, etc.) and behavioral reactions such as smiling, laughing or embracing. Paralinguistic features are usually described in the closed captions of audiovisual material and their time codes facilitate the process of linking the lexis of emotion with the visual and auditory elements that accompany the verbal expression of emotion. Moreover, paralinguistic and visual signals are usually consistent with the message or emotions that speakers wish to communicate, but this is not always the case: silence can express an emotion as much as a word, and words can contradict non-verbal cues.

Bednarek (2008:10) distinguishes between (i) *language about emotion* or emotion talk (linguistic expressions denoting affect/emotion such as love, hate, joy, etc.) and (ii) *language as emotion* or emotional talk, which encompasses linguistic expressions as conventionalized reflexes or signals of speakers' emotions (intonation, interjections, diminutives, paralinguistic resources, etc.). These terms correspond respectively to Kövecses' (2000) distinction between *descriptive* and *expressive emotion words*.[1]

[1] However, Kövecses identifies a third category to include metaphor and metonymy and names it 'figurative expressions denoting particular aspects of emotion'.

As this article deals with emotion talk, a review of some emotion models from the fields of psychology and linguistics is needed. Bizquerra Alzina (2020: 90) provides a thorough review of these models and identifies 35 different emotions. The literature normally arranges and classifies emotions according to either 'discrete' or 'dimensional' models (Cowen and Keltner, 2017), but authors such as Scherer (2009) identify 3 models: basic theories of emotion, constructivist emotion theories and appraisal theories. The universal or culture-specific nature of emotions is also discussed in the literature.

Discrete or categorical models of emotion recognize unique and discrete categories within emotions and propose inventories of basic emotions. For example, Johnson-Laird and Oatley (1989) analyzed 5 basic emotions: HAPPINESS, SADNESS, FEAR, ANGER, and DISGUST. Ekman's proposal (1992) includes six universal basic emotions: SURPRISE, DISGUST, FEAR, JOY, SADNESS and ANGER. Frijda et al. (1995) distinguished 5 universal emotions in 11 languages: HAPPINESS, SADNESS, ANGER, FEAR, and LOVE.

As far as dimensional theories are concerned, they organize emotions around dimensions, mainly, affective valence (pleasant vs unpleasant), arousal intensity (relaxed vs excited) and dominance (in control vs dominated). Ortony et al. (1988) grouped emotions around 3 axes that reflect the reaction of people to the consequences of events (pleased-displeased), the actions of people (approving-disapproving) and the aspect of objects (liking-disliking). These theories configure the semantic space of emotions around affective dimensions and emotion families (Russell, 2003; Scherer, 2005) on the basis of experimental studies. In the Geneva Emotion Wheel (Sacharin et al., 2012), emotion families or affect categories include concepts such as ANGER, INTEREST, AMUSEMENT, PRIDE, JOY, PLEASURE, CONTENTMENT, LOVE, ADMIRATION, RELIEF, COMPASSION, SADNESS, GUILT, REGRET, SHAME, DISAPPOINTMENT, FEAR, DISGUST, CONTEMPT and HATE. Also dimensional is the approach of Bednarek (2008). Adapting Martin and White's (2005) appraisal theory and corpus linguistics methods, she proposed a fuzzy system in which emotions can be grouped under 5 affect types: un/happiness, dis/satisfaction, in/security, dis/inclination and surprise (Bednarek, 2008: 169).

Some authors have combined both categorical and dimensional perspectives (Plutchik, 2001; Cowen and Keltner, 2017), thus recognizing the dynamic nature of emotions and their fuzzy boundaries. Alba-Juez and Mackenzie (2019: 15) also consider this dichotomy in terms of categorical theories of emotion (emotions as a set of basic, universal emotions) and theories that conceive emotions as a process that changes according to the appraisals made. They define emotion as 'a (dynamical) system of language which interacts with the system of evaluation but whose main function is the expression of the speaker's feelings, moods and affective experience' (Alba-Juez and Mackenzie, 2019: 18). They also consider emotion as a multimodal discourse process with both linguistic and non-verbal manifestations.

In the field of Computational Linguistics, some of the above-mentioned categories of emotion have been applied to research on sentiment analysis and emotion mining, resulting in lexicons such as WordNetAffect, SentiWordNet and the NRC Word-Emotion Association Lexicon. WordNetAffect (Valitutti et al., 2004) is based on Wordnet and classifies affective knowledge according to 4 valence values (positive, negative, ambiguous or neutral) and 11 labels (emotion, mood, trait, cognitive state, physical state, hedonic signal, emotion-eliciting situation, emotional response, behavior, attitude, sensation). SentiWordNet was used to determine the objective or subjective nature of texts and the positive–negative polarity of opinions (Esuli and Sebastiani, 2006). Regarding the NRC Emotion Lexicon (Mohammad and Turney, 2013), it is based on manual annotations and lists English words and their associations with: (i) a negative or positive valence; and (ii) Plutchik's eight basic emotions (ANGER, FEAR, ANTICIPATION, TRUST, SURPRISE, SADNESS, JOY, and DISGUST). A revised version of this lexicon is distributed by ELRA[2] (Zad et al., 2021).

This line of research has been followed in the identification of emotions in medicine. Sokolova and Bobicev (2020) analyzed messages posted on a medical forum on infertility treatments. On the basis of the HealthAffect Lexicon (Bobicev et al., 2015), annotators manually assigned the emotions expressed in the messages to five categories: GRATITUDE, ENCOURAGEMENT, CONFUSION, FACTS and FACTS + SENTIMENTS. Moreover, emotions in the domain of cancer were the focus of CancerEmo, a dataset of 8500 sentences from an online cancer network that were manually annotated with Mechanical Turk on the basis of Plutchik's taxonomy of 8 basic emotions (Sosea and Caragea, 2020). The authors found that the most relevant emotions in the corpus were JOY, FEAR, SADNESS, TRUST, SURPRISE, DISGUST, ANGER and ANTICIPATION. Finally, the project MuSE (Multimodal Dataset of Stressed Emotion) provides an annotation model for recordings aimed at observing the interplay between the presence of stress and the expressions of affect (Jaiswal et al., 2020). It also defines relevant multimodal features for the analysis and classification of emotion: acoustic, lexical, thermal and visual features of the subjects' facial actions as shown by close-ups.

The importance of affect in medical settings has also been highlighted in research using experimental procedures and corpus-based methods to study familiarity (Alarcón Navío et al., 2016), sensitivity to medical images (Prieto-

[2] NRC Emotion Lexicon – Revised version: https://islrn.org/resources/007–544-786–822-8/.

Velasco, 2017) and the language and semantic prosody used in patients' medical forums (Láinez Ramos-Bossini and Tercedor-Sánchez, 2021; Tercedor-Sánchez and Láinez Ramos-Bossini, 2017). Tercedor-Sánchez and Láinez Ramos-Bossini (2017) found that the use of emotional language was much higher in a corpus of patients' medical forums as opposed to its use in a general language corpus.

The work of Semino et al. (2018) is also relevant to the current study of emotions at the end of life. In the framework of the Metaphor in End-of-Life Care project, they identified the metaphors used to talk about cancer and the experience of the end of life. Their findings were based on a 1.5 million-word corpus containing the views of patients, unpaid family caregivers and healthcare professionals as expressed in interviews and online forum posts or blogs. The book includes a chapter on the metaphors used to describe emotions associated with bereavement. More specifically, negative emotions related to sadness (grief, anguish), uncertainty, helplessness, physical fragmentation, confusion, disappointment, failure, releasing emotions, being unable to control negative emotions and sudden emotional changes. The authors emphasize the need to verbalize thoughts and feelings: 'the importance of emotional expression, acknowledgement of the reality of the loss and sharing of thoughts and feelings with others' (Semino et al., 2018: 213).

Furthermore, different psychology models describe the process of grief associated with impending death or the death of a beloved person (Tyrrell et al., 2021). Numerous institutions provide guidelines aimed at helping people cope when they are near the end of life by specifying the emotions that they may encounter. For example, the American Cancer Society[3] includes fear, anger, guilt and regret, grief, anxiety and depression, loneliness and seeking meaning. Most of the emotions highlighted in these clinical guidelines are lexicalized in our pilot study corpus and will be shown in the Results and discussion section.

3. RESEARCH METHODS FOR THE COMPILATION AND ANALYSIS OF A PILOT STUDY CORPUS ON THE END OF LIFE

The current research involved the compilation of a pilot study audiovisual corpus, its semantic annotation and the use of different corpus analysis tools. It required the following stages:

- Establishing criteria for corpus design.
- Building a corpus from audiovisual texts on the end of life and palliative care: (a) text selection; (b) register of text metadata; (c) extracting verbal language from audiovisual material and converting it into plain text.
- Semantic annotation of the corpus with USAS tagger and Wmatrix versions 4 and 5 (see Section 3.3).
- Uploading both the plain text corpus and the semantically annotated corpus to Sketch Engine [https://www.sketchengine.eu/].
- Corpus analysis of words and multiword units with Wmatrix and Sketch Engine (SkE) focusing on the lexis of emotions). I benefited from affordances such as the automatic identification of key semantic domains, the keyness method, frequency lists and concordances.

3.1. Corpus design criteria

Bearing in mind the focus of the current study –the verbal expression of emotions in English audiovisual material about the end of life, the following considerations were taken into account when selecting texts:

- Only material that had been publicly broadcast or published was selected in order to ensure compliance with ethical issues (see Section 3.2).
- The material had been broadcast by a quality and influential source, as evidenced by the organizations behind the videos or the number of views (more than 1 million in the case of the TED talks). The aim of this criterion was to ensure that the texts were reliable and representative.
- Audiovisual texts showed different varieties of English (US, UK, South Africa, etc.), as well as different views on the subject of death.
- The content of the texts should focus on the topic DEATH CAUSED BY TERMINAL ILLNESS/OLD AGE. The indexing terms used to search for audiovisual materials on video on-demand platforms and health portals were a combination of the expression *end of life*, and one or more of the following: *care, palliative care, terminal illness, old age, deterioration, death* or

[3] American Cancer Society. (2019). Emotions and Coping as You Near the End of Life. https://www.cancer.org/treatment/end-of-life-care/nearing-the-end-of-life/emotions.html.

bereavement and grief. Semino et al. (2018) used the following search terms to select texts in online forums: *death, dying, end of, hospice, palliative, life, support.* As will be seen in section 4.1, these words appear very frequently in the pilot study corpus.
- All the materials were visualized in order to judge whether their content was appropriate and relevant for the purpose of the study.

3.2. Compilation of the corpus: Text selection, corpus details and conversion to plain text

In recent years several corpora have been compiled from the subtitles of audiovisual material. Cases in point are the OpenSubtitles corpus (Tiedemann, 2016), and some corpora available from Mark Davies' site, in particular, the TV corpus, the Movies corpus, and more recently, the SOAP corpus of American soap operas (Davies, 2021). These are notable initiatives for lexicogrammar research, but they do not allow for semantic annotation and users cannot access the audiovisual information of the corpora. Furthermore, even when selecting medical dramas, these corpora do not focus specifically on the topic of death caused by terminal illness or old age. Therefore, they were not used for the compilation of the *Terminal corpus*.

After searching different online sources fitting the criteria of the previous section, the corpus was finally composed of 3 types of audiovisual material: (i) documentaries broadcast on the video on-demand platform Netflix, (ii) talks from the well-known platform TED: Ideas Worth Spreading and (iii) interviews of patients and their caregivers, as well as summaries of interviews published on the website Healthtalk.org. The details about each text were stored in an Excel file with data categories such as the following: language and language variety, author/director, year, country, duration, genre, number of words, keywords, etc.

The corpus is called *Terminal corpus* and is made up of a total of 55 texts in English and 80,274 words (95,105 tokens). It includes the following components:

- Closed captions of 4 documentaries broadcast on Netflix and produced in the United States (13,115 words, 134 minutes).
- Subtitles of 11 TED Talks (19,704 words, 138 minutes).
- Summaries and transcripts of interviews with caregivers and patients published on HealthTalk.org (47,555 words).

3.2.1. Closed captions of documentaries broadcast on Netflix

From the video on-demand platform Netflix, I searched the options Films > Genres > Documentaries and narrowed down the search to find those dealing with end-of-life care and terminal diseases. After watching them, I finally selected 4 documentaries produced in the United States with a total duration of 134 minutes. Their titles, directors and keywords, as provided by Netflix, were the following:

- *Cristina* (Ohayon, 2016): breast cancer, liver metastasis, battle, live the moment.
- *End Game* (Epstein and Friedman, 2018): terminally ill patients, medical practitioners.
- *Extremis* (Krauss, 2016): wrenching emotions, end-of-life decisions, doctor, patients and families.
- *Ram Dass: Going Home* (Peck, 2018): love, life and death, end of life.

Since the present study focuses on the lexis of emotion and does not explore the interplay between linguistic, auditory and visual cues in the expression of emotions, I downloaded the English closed captions. They are quite literal in relation to the audio of the video and include both the linguistic elements of the dialogues and some paralinguistic information (contained within brackets). However, I understand that the accuracy in the identification of emotions is limited to some extent without an intermodal analysis (visual-auditory-verbal), even if these three facets are not necessarily consistent in real life. For example, the words of a patient stating that she is relaxed will not necessarily indicate CALMNESS if her lips are quivering. At the same time, a doctor may talk about grief objectively and not exhibit any physical signals of SADNESS.

In order to download the subtitles, I first used Google Chrome to access Netflix, selected the video of my choice and paused it. Second, I pressed Ctrl + Shift + I to open Chrome Developer Tools and its Network tab. Then, I went back to the video display, deactivated the subtitles, refreshed the page, and activated the closed captions for English. After that, in the Network tab of Chrome Developer Tools, I searched the files that included the following characters: *?o=*.[4]

[4] This can be done by using the search bar or by clicking on the Filter icon.

Considering that the files containing the subtitles are only those beginning with*?o*=, I double-clicked on the *Name* column to sort the files alphabetically, and selected the appropriate file. Then, I right-clicked on the file, selected the option Open in new tab, and the file was automatically downloaded.

Finally, I changed the extension of that file to .xml and opened it with Subtitle Edit (Lynge Olsson, 2021), free software under the GNU Public License. With Subtitle Edit, I converted the xml file to a plain text file (.txt) and a SubRip subtitle file (.srt). The time codes in the.srt files allowed for cross-checking the coherence between linguistic, paralinguistic (articulation and voice) and visual signals (facial expression, body language, eye contact, etc.). However, the study was not meant to analyze the link between verbal and non-verbal signals of emotion.

3.2.2. Subtitles of TED talks

From the website TED: Ideas Worth Spreading, I explored their talks by topic (Discover > Topics) and selected the heading Death. Thirty-six videos were retrieved.[5] I watched them carefully to verify that their content fell under the scope of the study. Some of the talks were ruled out; for example, those dealing with ecological burial practices or feelings derived from a violent death.

I finally selected 11 talks and extracted their subtitles, which were word-for-word transcripts of the speech of the speakers, most of whom were from the USA, although some came from Indonesia, Australia and the UK, and their talks were aimed at an international audience. The persuasive and non-spontaneous nature of TED talks had an effect on the expression of emotions, which was more vivid and lexically rich. The transcripts of the selected talks were obtained from the website https://ted2srt.org.

3.2.3. Transcripts of videos and summaries of interviews published on HealthTalk.org

Texts from the website HealthTalk.org were added to the corpus in order to increase the presence of British English, as well as to make the corpus more representative of the feelings and testimonies of patients and their families. Healthtalk.org provides information about 100 health topics based on the testimonies of British people interviewed in their homes under the supervision of the charity DIPEx (Database of Individual Patient Experiences) and the Health Experiences Research Group of the University of Oxford. The interviews were carried out between May 2010 and December 2017. The texts were retrieved from the following sections of the website: *Caring for someone with a terminal illness* and *Living with dying*.

This sub-corpus (47,555 words) is made up of 2 components: (a) 30 summaries of interviews with patients and their caregivers (24,861 words); and (b) 10 transcripts of excerpts from 40 interviews conducted with 33 patients and 11 caregivers of different ethnic backgrounds, aged 32–84, and living in the United Kingdom (22,694 words). The corpus includes different varieties and registers of English, so as to ensure a broader discussion of 'emotion'. However, the size of the corpus does not allow for the comparison of sub-corpora, for example, the study of the difference between the expression of emotion in British vs American English or in spontaneous speech (interviews) as opposed to non-spontaneous speech (documentaries and TED talks).

3.3. Semantic annotation of the corpus: USAS tagger and Wmatrix

In order to recognize emotion words in the corpus I used a system for the automatic semantic annotation of texts, that is, for assigning the words of a text to different semantic fields. More specifically, in line with the methods of Semino et al. (2018), I benefited from the following tools developed by the University Centre for Computer Corpus Research on Language (UCREL) of the University of Lancaster: the UCREL Semantic Analysis System and Wmatrix.

3.3.1. UCREL Semantic Analysis System (USAS)

USAS is an online semantic tagger that assigns a semantic field label to each word or multiword unit (MWU) in a text from a set of tags [https://ucrel.lancs.ac.uk/usas]. Its latest version is based on a lexicon of 56,316 words and a template list of 18,971 multiword units, each of them with a list of potential tags. Thanks to the lexicon, USAS tags the meaning of the word or MWU in context. Tags are organized in the following top-level hierarchy of 21 semantic domains (Archer et al., 2002; 2):

[5] To access these playlists, just type one of these numbers (241, 505, 511, 526, 580, 590) after the URL https://www.ted.com/playlists/.

A GENERAL AND ABSTRACT TERMS
B THE BODY AND THE INDIVIDUAL
C ARTS AND CRAFTS EMOTION
E EMOTION
F FOOD AND FARMING
G GOVERNMENT AND PUBLIC
H ARCHITECTURE, HOUSING AND THE HOME
I MONEY AND COMMERCE IN INDUSTRY
K ENTERTAINMENT, SPORTS AND GAMES
L LIFE AND LIVING THINGS
M MOVEMENT, LOCATION, TRAVEL AND TRANSPORT
N NUMBERS AND MEASUREMENT
O SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT
P EDUCATION
Q LANGUAGE AND COMMUNICATION
S SOCIAL ACTIONS, STATES AND PROCESSES
T TIME
W WORLD AND ENVIRONMENT
X PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES
Y SCIENCE AND TECHNOLOGY
Z NAMES AND GRAMMAR

Emotions are grouped under the semantic category E *Emotional Actions, States & Processes* and subdivided into 6 subcategories: (E1) Emotional Actions, States & Processes (General); (E2) Like & Dislike; (E3) Serenity/Composure/Anger/Violence; (E4) Happiness & Sadness; (E5) Bravery & Fear; and (E6) Worry/Concern/Confidence.

Once the text is introduced into USAS, it is tagged so that each word or phrase appears with an underscore character and a tag, as shown in this excerpt from the documentary *Cristina* (Ohayon, 2016).

In_Z5 my_Z8 experience_X2.2+,_PUNC it_Z8 took_A9 + me_Z8mf several_N5 years_T1.3 before_Z5 I_Z8mf stopped_T2- comparing_A6.1 my_Z8 new_T3- body_B1 to_Z5 the_Z5 old_T3 + body_B1._PUNC.

But_Z5 when_Z5 I_Z8mf did_Z5 stop_T2- comparing*--_Z99 When_Z5 this_Z8 became_A2.1 + the_Z5 whole_N5.1 + me_Z8mf,_PUNC not_Z6 me_Z8mf missing_A3- stuff_O1,_PUNC I_Z8mf stopped_T2- suffering_E4.1-._PUNC*.

In the text above, there are Z tags indicating function words such as prepositions (Z5) or pronouns (Z8), as well as punctuation tags. Other tags refer to time (T), general & abstract terms (A), psychological actions, states & processes *(experience_X.2.2)* or emotions *(suffering_E4.1)*.

3.3.2. Wmatrix

Wmatrix [https://ucrel.lancs.ac.uk/wmatrix/] is a corpus analysis tool (Rayson, 2009) that provides a web interface to the corpus annotation tools of the University of Lancaster (USAS and CLAWS). It claims 92% accuracy for the semantic tagging of USAS and 96–97% for its part-of-speech tagger.[6] Wmatrix enables corpus linguistic methodologies such as frequency lists, concordances and keywords. It also extracts key grammatical categories and key semantic domains. To do so, the frequency list of the corpus under study is compared to reference corpora (the BNC sampler, British English 2006, American English 2006) or other corpora previously uploaded by the user. This affordance is found under *Keyness analysis > Key concepts compared to: Name of the corpus > Go,* and is based on keyness, a common method in corpus linguistics aimed at discovering what is distinctive about a text in terms of vocabulary (Scott, 1997; Rayson, 2008, Prentice et al., 2021). Keyness involves, firstly, comparing frequency lists for 2 corpora (the study corpus and the reference corpus) and, secondly, identifying lexical units which are significantly more frequent than expected in one corpus than in the reference corpus on the basis of statistical measures such as log-likelihood significance (Semino et al., 2018: 72). In the present study, I used the BNC Sampler spoken corpus (982,712 words) as a reference

[6] https://ucrel-wmatrix5.lancaster.ac.uk/cgi-bin/wmatrix5/help.pl#norm.

corpus, and the log-likelihood measure in order to find out the key semantic domains of the corpus (research question 1).

3.4. Sketch Engine (SkE) as a control tool

To complement the results of Wmatrix, I uploaded both the plain text corpus and the USAS semantically annotated corpus to Sketch Engine. SkE is a platform for corpus compilation, management and analysis (Kilgarriff et al., 2014) and provides rich lexical and combinatory information for linguists, lexicographers, terminologists, translators, and foreign language teachers and learners.[7] It has been widely used for the analysis of medical texts (Jiménez-Crespo and Tercedor-Sánchez, 2017; López-Rodríguez, 2019; López-Rodríguez and Sánchez-Cárdenas, 2021, *inter alia*).

SkE can also compare a study corpus with a wide variety of reference corpora (enTenTen, COCA, BNC, to name a few), and extract the most significant single words and multiword expressions in the corpus. The Keyword function of SkE analyzes the frequency of words and multiword units against the backdrop of one of the previously mentioned corpora and calculates a keyness score for each lemma using the 'simple maths' method (Kilgarriff, 2009). This metric is appropriate to contrast corpora of unequal sizes and is based on the normalized frequencies (per million words) in the focus and reference corpora.[8] In this study I compared the Terminal corpus with the English Web corpus 2020 (enTenTen20), which contains nearly 38 billion words collected from the Internet to represent a wide variety of domains (.com,.org,.net,.edu,.gov,.info) and varieties of English (.au,.ca,.ie,.nz,.uk,.us, *inter alia*).

3.5. Combining Wmatrix and Sketch Engine to triangulate results

By using two different tools it is possible to check the consistency and validity of the data generated and to compensate for the limitations of the tools. For instance, Wmatrix does not allow the lemmatization of word forms under their corresponding lemmas, even though it provides raw and relative frequencies of the word types belonging to the same USAS semantic field and is extremely useful in the identification of key domains. Consequently, for a quantitative lexical analysis of the most frequent emotion lemmas in the corpus, I searched the semantic categories of interest (Emotions) in the annotated corpus uploaded to SkE. For this, I used the concordance function and performed an advanced search for the character *_E*. This retrieved the expressions that had been automatically tagged as an emotion in the corpus, more precisely 1365. When the concordances were sorted by keyword, emotions under categories E1 to E6 appeared together. Fig. 1 shows some examples retrieved when searching the category E4 (Happy/Sad).

S2.2m makes _A1.1.1 plans _X7+ . And God _S9m laughs	_E4	.1+/X3.2 . Michele _Z1f . [Michele _Z1f] Tell _Q2.2 me _Z8
.1 n't _Z6 even _A13.1 ... I _Z8 just _A14 feel _X2.1 happy	_E4	.1+ , that _Z8 I _Z8 got _A9+ to experience _A2.1+/A3+ tha
_N5.1+[i68.3.2 way _N5.1+[i68.3.3 . [Bruce _Z1mf giggling	_E4	.1+] -[Bruce] _Z99 Okay _A5.1+ . -[Cristina] _Z99 Are _A3-
_Z1m[i70.2.1 Bond _Z1m[i70.2.2 ? [Bruce _Z1mf giggling	_E4	.1+] Nothing _A6.1-[i71.2.1 like _A6.1-[i71.2.2 being _A3+
/ing _B1/E4.1-] I _Z8 am _A3+ just _A14 so _A13.3 happy	_E4	.1+ . Wow _Z4 . [Michele _Z1f] So what _Z8 does that _Z8
" Cristina _Z99 ... I _Z8 am _A3+ more _N5++ than happy	_E4	.1+ to tell _Q2.2 you _Z8 my _Z8 story _Q2.1 . I _Z8 am do
4.2.1 absolutely _A13.2 fine _A5.1+/A8[i104.2.2 . [relieved	_E4	.1+ sigh _Q2.2/X3.2] [Doctor _B3/I3.2/S2mf] You _Z8 nee
S2.2m] Yes _Z4 . -[Cristina _Z99 and Bruce _Z1mf rejoice	_E4	.1+] [little _N3.2- boy _S2.2m squeals _X3.2] [Bruce _Z1

Fig. 1. Sketch Engine concordance for the semantic category *E4 Happy/Sad.*

[7] SkE is very powerful for compiling corpora, finding the most frequent words and multiword units in a corpus, identifying semantically similar terms and collocations of a word, and generating concordances to explore the linguistic context of a lexical unit. Its most original feature is that it shows a summary of a word's lexico-grammatical behavior *(word sketch)* based on its distributional behavior in the corpus and grammatical relations (subject, object, etc.) with other words.

[8] For more information about the simple maths method: https://www.sketchengine.eu/documentation/simple-maths/.

Additionally, with the Frequency button of Concordance, SkE can calculate the frequency of the lemmas to the left of the tag, and this allowed us to obtain complementary results such as the most frequent lemmas under the category *E4 Happy/Sad* (Fig. 2).

		Lemma	Frequency ↓	Relative [?]	% of conc. [?]
1	☐	suffer	29	102.86	7.65 %
2	☐	laughter	26	92.22	6.86 %
3	☐	merrily	21	74.49	5.54 %
4	☐	suffering	20	70.94	5.28 %
5	☐	happy	20	70.94	5.28 %
6	☐	laugh	14	49.66	3.69 %
7	☐	upset	13	46.11	3.43 %
8	☐	laughing	11	39.02	2.90 %
9	☐	glad	11	39.02	2.90 %
10	☐	relief	8	28.38	2.11 %
11	☐	grief	8	28.38	2.11 %
12	☐	joy	8	28.38	2.11 %

Fig. 2. Frequent lemmas under category *E4 Happy/Sad* (Sketch Engine).

In Sketch Engine, forms of the same stem that belong to a different part of speech are not grouped together. Therefore, *suffering* as a verb is included under the lemma 'suffer' (together with other verbal forms such as *suffers* or *suffered*), but the instances of *suffering* as a noun (e.g. *unnecessary suffering*) are grouped under the lemma 'suffering'. In any case, there is a Corpus Query Language option to refine the searches.

4. RESULTS AND DISCUSSION

In this section, I describe the results of combining Wmatrix and Sketch Engine to analyze the language and feelings regarding death contained in the Terminal corpus, a pilot study corpus about the end of life that was annotated with the UCREL Semantic Analysis System. I also try to answer the research questions formulated in the Introduction.

4.1. Research question 1: Keywords and key concepts in the Terminal corpus

Keyness analysis was used to find out the keywords and semantic domains in the corpus. The most relevant single words and multiword units were extracted with Sketch Engine by comparing the Terminal corpus with the English Web corpus 2020. As mentioned in Section 3.4, Sketch Engine calculates the keyness score with the 'simple maths' method (Kilgarriff, 2009). I 'fine-tuned' SkE to a value of 100 so that it focused on words that are neither rare (for example, unusual proper nouns) nor common in the reference corpus. Table 1 lists the lemmas for the most frequent single words, except for proper nouns, pronouns and words that point to the oral nature of the Terminal corpus *(yeah, okay, yes, Cristina, gonna, oh, she)*.

The identification of keywords enabled the collocational analysis of some of these frequent words in order to understand nuances of meaning in context and lexical preferences when talking about the end of life. For example, in the Terminal corpus the lemmas *die, death* and *dead* frequently co-occur with words such as *imminent, assisted, spirit, birth, prefer, actual, fear, frightened, near, without, afraid, dignity, after, natural* or *home*, as measured by their Mutual Information score. Concordances around the word forms of the verb *die*, the noun *death* and the adjective *dead* (retrieved with the simple search *die*|*death*|*dead*) show key areas of interest and priorities in those difficult moments, as shown in the following sample of sentences (1–10).

Table 1
Keywords (single words) of the Terminal corpus.

		Frequency per million		
	Lemma	Focus corpus	Reference corpus (English Web 2020)	Keyness score (simple maths method)
1	Hospice	1,661.32	4.87	16.8
2	Carer	1,282.79	6.55	13.0
3	Die	3,385.73	183.02	12.3
4	Illness	1,345.88	32.85	10.9
5	Cancer	1,955.73	102.79	10.1
6	Death	2,807.42	197.26	9.8
7	Nurse	1,114.56	43.76	8.4
8	Care	3,154.41	307.40	8.0
9	Hospital	1,345.88	120.86	6.5
10	Husband	935.81	66.53	6.2
12	Talk	2,176.54	271.63	6.1
13	Euthanasia	515.22	1.91	6.0
14	Pain	1,072.50	97.30	5.9
15	Doctor	1,093.53	104.74	5.8
16	Want	4,100.73	674.61	5.4
17	Feel	2,534.04	398.04	5.3
18	Terminal	599.34	36.37	5.1
19	Ill	525.73	23.43	5.1
20	Palliative	410.07	3.19	4.9
21	Patient	1,482.57	220.99	4.9
22	Think	4,058.67	746.74	4.9
23	Diagnosis	483.68	24.84	4.7
24	Wife	872.72	111.02	4.6
25	Life	3,228.01	624.05	4.6

(1) She found the doctor very clinical, but it helped her not to break down when they discussed her dad's **imminent** *death*.

(2) Faced with **imminent** *death*, all he wanted was forgiveness.

(3) Watching a patient suffer in the terminal stages of their disease made several carers wonder whether **assisted** *dying* would have been appropriate in their case.

(4) One Christian had reluctantly concluded that quality of life could be so poor that **assisted** *death* would be appropriate.

(5) I have been privileged, I think is the right word, although it was a bit scary at the time, each of my husbands has died in my arms and there is no doubt whatsoever in my mind that as someone *dies* that **spirit**, that essence, you physically see it go.

(6) She was convinced that when someone *dies* the **spirit** leaves the body.

(7) Imagine the book of your life, its covers, its beginning and end, and your **birth** and your *death*.

(8) In talking about being a midwife. . . . .she talked about **natural** *death* and natural **birth**, and linking those two things.

(9) 70% of people would **prefer** to *die* **at home**.

(10) A woman with breast cancer said she would **prefer** to *die* **in a hospice** because she worried about pain, didn't want to be a burden on the family, and didn't want her family to think of death when they visited the family home, a place they had all loved.

Among multiword terms, some of the most relevant keywords were *terminal illness, palliative care, breast cancer, intensive care, support group, multiple sclerosis, hospice care, nursing home, bad news, motor neurone disease* or *serious illness.*

Furthermore, a list of lexical bundles containing 3–5 words was generated with the N-gram tool of Sketch Engine in order to retrieve additional meaningful multiword expressions: *do not want to, be able to, do not know, want to be, would like to, the end of, a man with, a woman with, (at) the end of, do not think, friend or relative, die at home* or *we talk to.*

			Item	O1	%1	O2	%2		LL	LogRatio	
1	List1	Concordance	B3	1118	1.45	1024	0.10	+	3050.35	3.80	Medicines and medical treatment
2	List1	Concordance	L1-	859	1.11	501	0.05	+	2788.91	4.45	Dead
3	List1	Concordance	B2-	860	1.12	959	0.10	+	2137.02	3.52	Disease
4	List1	Concordance	L1+	343	0.45	51	0.01	+	1502.23	6.42	Alive
5	List1	Concordance	S8+	867	1.13	2020	0.21	+	1321.54	2.45	Helping
6	List1	Concordance	S2	791	1.03	2728	0.28	+	808.18	1.89	People
7	List1	Concordance	Z99	1116	1.45	5684	0.58	+	637.26	1.32	Unmatched
8	List1	Concordance	S4	775	1.01	3699	0.38	+	496.72	1.42	Kin
9	List1	Concordance	H4	322	0.42	1130	0.11	+	322.09	1.86	Residence
10	List1	Concordance	A13	45	0.06	0	0.00	+	235.91	10.16	Degree
11	List1	Concordance	S9	274	0.36	1106	0.11	+	227.87	1.66	Religion and the supernatural
12	List1	Concordance	X2.6+	131	0.17	261	0.03	+	226.68	2.68	Expected
13	List1	Concordance	E4.1-	151	0.20	400	0.04	+	204.86	2.27	Sad
14	List1	Concordance	S3.1	210	0.27	752	0.08	+	204.85	1.83	Personal relationship: General
15	List1	Concordance	A2.1+	382	0.50	2031	0.21	+	201.00	1.26	Change
16	List1	Concordance	X6+	163	0.21	511	0.05	+	185.97	2.02	Decided
17	List1	Concordance	X7+	728	0.94	5302	0.54	+	174.41	0.81	Wanted
18	List1	Concordance	X2.1	895	1.16	7031	0.72	+	164.54	0.70	Thought, belief
19	List1	Concordance	S1.1.1	185	0.24	746	0.08	+	154.08	1.66	Social Actions, States And Processes
20	List1	Concordance	E5-	126	0.16	398	0.04	+	142.56	2.01	Fear/shock
21	List1	Concordance	X9.2	37	0.05	12	0.00	+	141.23	5.30	Success and failure
22	List1	Concordance	B1	520	0.67	3703	0.38	+	133.76	0.84	Anatomy and physiology
23	List1	Concordance	A13.3	701	0.91	5457	0.56	+	133.35	0.71	Degree: Boosters
24	List1	Concordance	E3+	87	0.11	207	0.02	+	130.22	2.42	Calm
25	List1	Concordance	E6-	168	0.22	733	0.07	+	124.56	1.55	Worry

Fig. 3. Wmatrix Keyness analysis: Top 25 key concepts in the Terminal corpus.

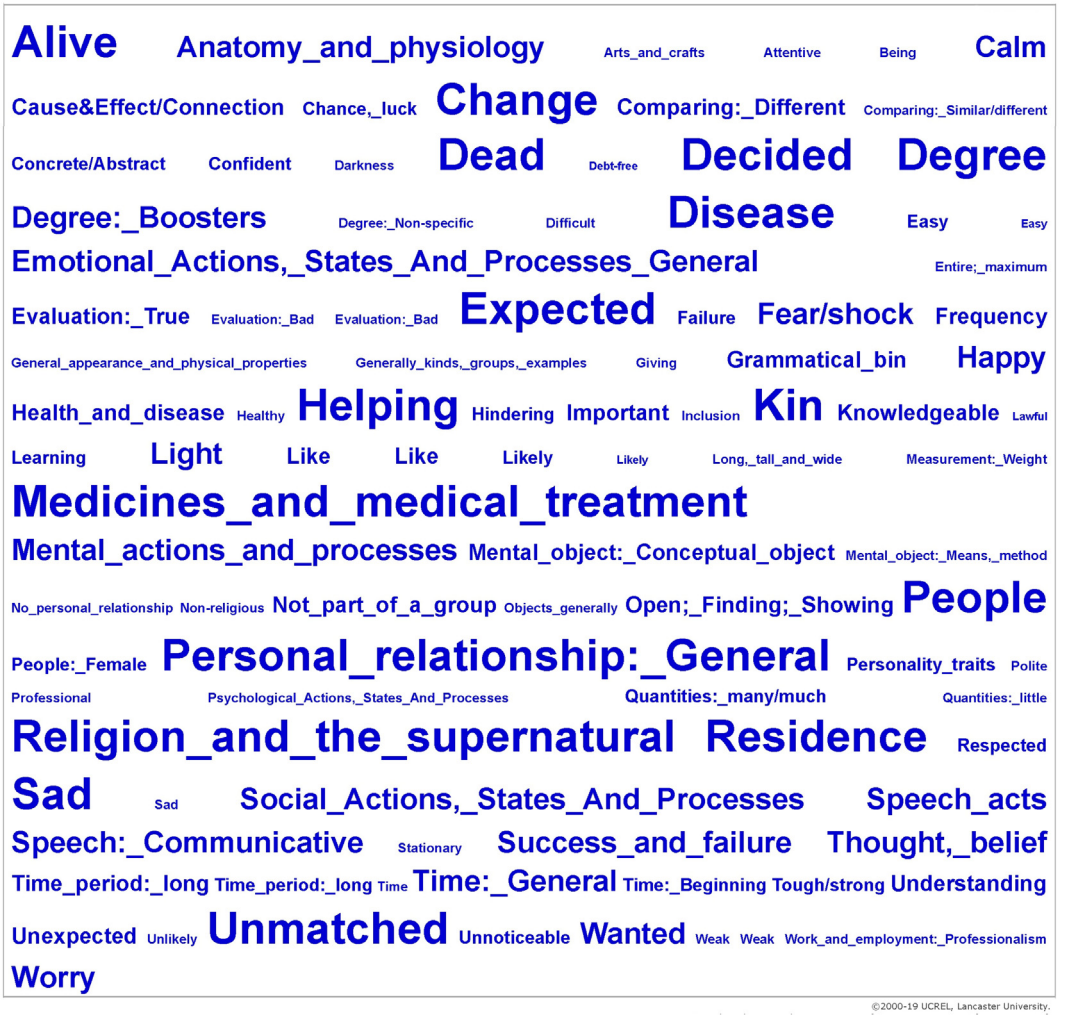

Fig. 4. Key domain cloud in the Terminal corpus as opposed to BNC Sampler Spoken (Wmatrix).

These multiword expressions point to the wishes, limitations, uncertainties, preferences and need of support and company experienced by terminal patients.

Regarding the main conceptual domains, Wmatrix was the tool used to contrast the frequency and relevance of key semantic categories in the Terminal corpus with their presence in a similar reference corpus, the BNC Sampler spoken corpus (982,712 words). Fig. 3 includes the 25 most frequent semantic categories of the Terminal corpus.

If one compares the frequencies of semantic categories in the Terminal corpus (column O1 for raw frequency; column %1 for relative or normalized frequency) to their frequencies in the BNC Sampler Spoken (O2; %2), it can be observed that some semantic categories are overused in the Terminal corpus as opposed to the reference corpus. This is indicated in the LL (log-likelihood significance) column with a +sign. Corpus adequacy to the purpose of the study (emotions at the end of life) seems to be validated by the list of the 25 most activated semantic fields, which points to some areas related to the end of life: Medicines and medical treatment (under category B The Body and the individual); Dead; Disease; Alive; Helping; People; Kin; Residence; Degree; Religion and the supernatural; Expected; Sad; Personal Relationship; Change; Decided; Wanted; Thought, Belief; Social actions, states and Processes; Fear/shock; Success and Failure; Anatomy and physiology; Degree: Boosters; Calm; and Worry. To determine the accuracy of Wmatrix when assigning the tags, I opened the frequency list for the category Medicines and medical treatment (by clicking on the link List 1). All the words included under this semantic category are relevant and consistent with the keywords obtained with Sketch Engine. Some examples with a relative frequency higher than 0.02% are *hospice, hospital, doctor, treatment, medical, diagnosis, nurse, palliative, GP, palliative, resuscitate, morphine,* and *nursing home.*

The results in Fig. 3 can be represented visually with the Key domain cloud of Wmatrix, which shows the 100 most relevant semantic domains (Fig. 4). Larger words are those with higher log-likelihood significance in the keyness analysis so that size is indicative of the significance of the concept in the corpus.[9]

4.2. Research question 2: Relevance of emotional actions, states and processes in relation to other semantic categories

The semantic tags provided by Wmatrix (Frequency list > Semantic > USAS Tags Only > Sorted by: Frequency) were very useful to measure the relevance of emotion in relation to other semantic categories. Wmatrix retrieved 384 types of semantic tags and a total of 77,060 semantically tagged words. Fig. 5 includes some of the top categories: grammatical bin (Z5), pronouns (Z8), being (A3), definite and modals (A7), negative (Z6), medicines and medical treatment (B3), unmatched (Z99), getting and giving/possession (A9), terms relating to communication (Q2.1), discourse bin—*oh, yeah, yes*, etc. (Z4), general and abstract terms (A1), thought, belief (X2.1), location and direction (M6), helping/hindering (S8), health and disease (B2), life and living things (L1), people (S2), kin (S4), wanting, planning, choosing (X7), and speech act terms (Q.2.2).

By double-clicking on the Semtag column, tags belonging to the same semantic category (for example, *E Emotion*) are grouped together. Since there was a total of 77,060 tagged words, the raw frequency of tags beginning with the same letter was sufficient to calculate the percentage for each general semantic category. Table 2 includes the 11 most important categories arranged by frequency, representing 96.26% of the words in the corpus.

Of the 21 general categories identified by the UCREL Semantic Analysis System and Wmatrix, the semantic category *Emotional Actions, States & Processes* occupies position 11, with a presence of 1.63 per cent in the corpus. As is usually the case in most corpora, the lexemes corresponding to proper nouns and grammatical words (Z) represent almost 50% of the corpus. The lemmatized list generated with SkE contains the following stems of the Z category, each of them appearing at least 500 times in the corpus: *the, and, to, I, a, of, it, that, you, not, in, we, for, with, they, she, that, her.* Both wordlists and concordance lines indicate that the pronouns *I, you, we, she* and *her* are highly significant since they indicate personal attachment and the importance of people, especially women, as care providers. *General & abstract terms* (A) represents nearly 16% with lemmas such as *be, have*, *do, would, can, make, process, busy, prepare, deal, activity*, etc.

The third most activated category is *S: Social actions, states & processes* (6.36%), in which the lemmas appearing at least 20 times denote:

- Family and personal relations: *people, family, child, woman, carer, friend, man, husband, person, mother, wife, relative, visit, daughter, love, brother, father, human, comfort, son, partner, relationship, kid, member, power, parent, personal.*

[9] By default, this functionality only computes items with an LL higher than 6.63 so that there is a 99% chance of non-randomness ($p < 0.01$).

Semtag	Frequency	Relative Frequency		
Z5	20995	27.25	Concordance	List
Z8	12137	15.75	Concordance	List
A3+	2976	3.86	Concordance	List
A7+	1381	1.79	Concordance	List
Z6	1149	1.49	Concordance	List
B3	1118	1.45	Concordance	List
Z99	1116	1.45	Concordance	List
A9+	1093	1.42	Concordance	List
Q2.1	1079	1.40	Concordance	List
Z4	1069	1.39	Concordance	List
A1.1.1	934	1.21	Concordance	List
X2.1	895	1.16	Concordance	List
M6	893	1.16	Concordance	List
S8+	867	1.13	Concordance	List
B2-	860	1.12	Concordance	List
L1-	859	1.11	Concordance	List
S2	791	1.03	Concordance	List
S4	775	1.01	Concordance	List
X7+	728	0.94	Concordance	List
Q2.2	707	0.92	Concordance	List

Fig. 5. Frequency measure of USAS semantic tags with Wmatrix.

Table 2
USAS Semantic tags and conceptual categories.

	Frequency	Relative freq.
Z: Names & Grammatical words	37,572	48.75%
A: General & Abstract Terms	12,141	15.76%
S: Social Actions, States & Processes	4902	6.36%
X: Psychological Actions, States & Processes	4140	5.37%
N: Numbers & Measurement	3240	4.20%
B: The Body & the Individual	2633	3.42%
T: Time	2479	3.22%
Q: Linguistic Actions, States & Processes. Communication	2301	2.98%
M: Movement, Location, Travel & Transport	2200	2.85%
L: Life & Living things	1326	1.72%
E: Emotional Actions, States & Processes	1258	1.63%
	74,192	**96.26%**

- Support, community and social norms: *help, care, group, support, service, social, together, meet, must, fight, manage, allow, assist, attend, share.*
- Religious concerns: *God, pray, religious.*

Under category S, lemmas referring to females are more frequent than those referring to men, and again, this confirms that women are usually more active in professional and unpaid patient care (Table 3). Moreover, many words in

Table 3
Women and men in the Terminal corpus.

Lemmas (including singular and plural forms)	Female	Male
Woman	115	
Man		106
Husband		89
Mother	67	
Wife	83	
Daughter	38	
Brother		37
Father		36
Son		27
Mum	26	
Dad		25
Mom	13	
Sister	10	
Mama	9	
Uncle		5
Grandfather		4
Granddad		3
Grandma	3	
	364	**332**

the semantic fields of spirituality and funeral rites appear in the corpus: *religion, counselor, spiritual, soul, ritual, afterlife, ceremony, prayer, church, spirit, heaven, incarnation, remembrance, omnipotent, holy, hell, agnostic, memorial, Buddhist, paradise, resurrection, ritually, eulogy, Christians, atheist, Lord.*

In fourth and fifth position, we find words indicating *Psychological Actions, States & Processes* (5.37%) such as *want, know, think, see, feel, way, able, believe, decision, information, idea, hope, try*... and the category *Numbers and Measurement* (4.2%).

Then, we find categories that are obviously important in those final moments and the following lemmas with a frequency of more than 100 per million words:

- The Body & the Individual (3.42%): *cancer, hospice, illness, patient, nurse, hospital, pain, doctor, treatment, body, medical, disease, ill, diagnosis, palliative, health, sick, breast.*
- Time (3.22%): *time, now, go, day, year, never, start, long, gone, still, moment, II, week, night, end, ever, month, stop, old, young.*
- Linguistic Actions, States & Processes. Communication (2.98%): *say, talk, tell, ask, call, mean, story, explain, describe, write, question, suggest, point, discuss.*
- Movement, Location, Travel & Transport (2.85%): *go, come, here, where, there, out, end, stay, put, place, get, leave, in, back, move on, bring, away, sit.*
- Life & Living things (1.72%): *die, death, life, terminal, euthanasia, alive, live.*
- Emotional Actions, States & Processes (1.63%): *like, love, care, fear, caring, suffer, worry, faith.* The words in this category will be analyzed in Sections 4.4 and 4.5.

4.3. Discussion of findings related to research questions 1 and 2

The findings related to research questions 1 and 2 confirm that some key domains of the corpus are related to our social, psychological and embodied nature: social processes, family and community support, together with psychological processes, numbers and measurements, time, and body and health issues. Moreover, emotions are significant in the Terminal corpus and represent 1.63 per cent of its words. When contrasting this percentage with the results of general language corpora, emotions are more relevant in our pilot study corpus (1.63 per cent) as opposed to other corpora: BNC Sampler Spoken (1.52), BNC Sampler Written (1.03), American English 2006 (1.12), British English 2006 (1.15). This outcome is consistent with the richness of emotion expressions found in a corpus of online patient forums compared to a general language corpus (Tercedor-Sánchez and Láinez Ramos-Bossini, 2017).

Moreover, the most activated semantic areas are consistent with the key domain cloud of the study corpus (see Fig. 4) in which certain semantic fields of emotions such as SAD, CALM, FEAR/SHOCK, HAPPY, LIKE and WORRY were much more significant in relation to a control corpus of general language (BNC Sampler Spoken). It might be argued that the reference corpus for the semantic analysis is a component of the British National Corpus, and the Terminal corpus includes different varieties of English. However, the Terminal corpus is mainly based on oral texts (filmed dialogue, talks, interviews), and this is consistent with the composition of the BNC Sampler Spoken.

Some key semantic areas in the Terminal corpus coincide with those identified in the Metaphor in End-of-Life Care corpus (Semino et al., 2018, chapter 4), although their corpus is larger and they compare key topics for patients as opposed to professionals and unpaid caregivers. In both corpora, personal pronouns are very frequent, as are words relating to the body, positive and negative emotions, kinship, time, medicine and treatments, death and dying, being alive, people, and religion and spirituality. Finally, the Terminal corpus data indicate an important presence of women as caregivers and supporters of terminal patients.

4.4. Research question 3: Predominant general categories of emotion resulting from automatic semantic recognition. Discussion of the findings

This section shows the importance of certain general emotions or affect types in relation to other emotions as revealed by the use of USAS and Wmatrix. The research question was formulated as follows: *What are the predominant general categories of emotion in the Terminal corpus?*

Table 4
Positive and negative evaluation of emotions in the Terminal corpus (Wmatrix).

Semantic tags and meaning	Freq	Rel. freq.
E1 EMOTIONAL ACTIONS, STATES AND PROCESSES GENERAL: GENERAL TERMS DEPICTING EMOTIONAL ACTIONS, STATES AND PROCESSES: E1+ (EMOTIONAL) AND E1− (UNEMOTIONAL)		
E1	60	0.08
E2 LIKING: TERMS DEPICTING FONDNESS/AFFECTION/PARTIALITY/ATTACHMENT, OR THE LACK OF		
E2	1	0.00
E2+ Like	**272**	**0.35**
E2++	32	0.04
E2+++	3	0.00
E2- Dislike	17	0.02
E3 CALM/VIOLENT/ANGRY: TERMS DEPICTING (LEVEL OF) SERENITY/COMPOSURE/ANGER/VIOLENCE		
E3+ Calm	**87**	**0.11**
E3++++	1	0.00
E3− Violent/Angry	78	0.10
E4 HAPPINESS AND CONTENTMENT		
E4.1+ Happy	**154**	**0.20**
E4.1++	2	0.00
E4.1− Sad	**151**	**0.20**
E4.1− −	3	0.00
E4.2+ Contentment	23	0.03
E4.2− Discontent	14	0.02
E5 BRAVERY AND FEAR: TERMS RELATING TO (LEVEL OF) TREPIDATION/COURAGE/SURPRISE, ETC.		
E5+ Bravery	9	0.01
E5+++	1	0.00
E5− Fear/shock	**126**	**0.16**
E6 WORRY/CONCERN/CONFIDENCE: TERMS RELATING TO (LEVEL OF) APPREHENSION/CONFIDENCE, ETC.		
E6+ Confident	56	0.07
E6− Worry	**168**	**0.22**

As seen in Section 3.3.1, USAS divides the semantic field *E Emotional Actions, States & Processes* into 6 subcategories. Additionally, USAS and Wmatrix[10] provide a positive or negative evaluation of words according to their meaning, in line with the affective valence towards pleasantness or unpleasantness proposed in dimensional approaches to emotion. For instance, within category E6 (Worry, concern, confidence), the system assigns a positive appraisal (+) to expressions such as *happy go lucky* or *confidently*, and a negative one (–) to *apprehensive* or *anxious*.

With Wmatrix, the nuanced data about emotions were calculated (Table 4). The categories with a relative frequency higher than 0.1 per cent are highlighted in bold.

The USAS semantic labels shown in Table 4 refer to some of the universal emotions proposed by psychologists and linguists such as HAPPINESS, SADNESS, LOVE, DISGUST, WORRY, ANGER or FEAR, as well as four of Bednarek's (2008) affect types: UN/HAPPINESS, DIS/SATISFACTION, IN/SECURITY, and DIS/INCLINATION. The most activated are UN/HAPPINESS and DIS/INCLINATION. SURPRISE does not appear because Wmatrix tags this word under category *X2.6– Mental actions & processes > Unexpected*.

From the frequencies of semantic tags, it can be concluded that the most activated categories are positive emotions (LIKING, HAPPINESS and CALMNESS/SERENITY), as well as negative emotions (FEAR/SHOCK and WORRY/APPREHENSION). Since these findings are derived from automatic semantic annotation, a deeper analysis was needed to provide more nuanced insight about emotions at the end of life.

4.5. Research question 4: Most frequent emotion words and most common emotions in the corpus. Discussion of the findings

Wmatrix's 92% accuracy in automatic semantic annotation might seem insufficient to discover the most frequent specific emotions in the corpus. Apart from assigning some semantic tags incorrectly, some words can be assigned more than one tag depending on context. Consequently, concordances and manual analysis were used to disambiguate and to correct the inaccuracies of automatic semantic tagging.

4.5.1. Correcting the inaccuracies of automatic tagging

On the basis of the list obtained with Wmatrix (Semantic frequency list > Word and USAS [sorted by: USAS Tag]), I identified the most frequent words and MWUs under each semantic category. When in doubt, concordances offered valuable contextual information for determining whether the examples had been tagged appropriately. For example, the expressions retrieved under category *E3+ Calm* generally fit well in the category *(peace/peaceful/peacefully, patient/patience, comforting, relaxation/relax/relaxed, softly, respite, calm/calmness, resting/rest, gentle/gently, calm (ed) down, serene, equanimity),* though there are some exceptions: *laid back*, which occurs once, and *patient*, which is tagged as an adjective of emotion (E3+) on 9 occasions. Concordances show that the word *patient* indicates emotion only once. In any case, the singular form *patient* occurs 63 times in the corpus, and is correctly tagged under Disease (B2–) in 85.7% of the cases.

Another example in which concordances were useful for a fine-grained analysis is the case of *E4.1+ Happy*. Some of the concordances contained the paralinguistic information coded in the closed captions: (Laughter),[11] (Everybody laughing), (chuckles), etc. Although these cases might not be categorized as a linguistic expression of emotion, they do indicate happiness in context (i.e. their co-occurrent frames and sounds), which I analyzed thanks to the time codes of the subtitles. Therefore, these instances were not discarded and I computed them under the HAPPINESS category.

By generating concordances with SkE (Concordance > Frequency of lemmas to the left of the tag _E[number]), I obtained details about the collocational context of the words categorized as emotion, disambiguated cases of polysemy, and discarded instances when necessary. For instance, the lemma *care* (noun and verb) appears 293 times, and USAS assigns it to categories S8 (Helping), E6– (Worry), B3 (Medicines & treatment), and H1 (Architecture).[12] After checking the concordances, I found that on 49 occasions, the 53 words tagged under *E6– (Worry)* were not emotion talk; their meaning was 'to look after a person and keep them in a good condition' in relation to *palliative care*. As a result, I ruled out those 49 cases and eliminated them in the computations.

Additionally, some emotions were not tagged by USAS as emotions despite being considered as such in some emotion taxonomies: HOPE, ACCEPTANCE, GRATITUDE, LONELINESS, DENIAL, SURPRISE and GUILT. Even though the small size of the cor-

[10] In Wmatrix, under the option *Semantic frequency list > USAS Tags only*, I sorted the results by *USAS tag*, and then searched the shortcut *Emotion*.

[11] In fact, all the examples of 'laughter' except for one are descriptions of paralinguistic information included in the closed captions.

[12] Wmatrix version 5 includes a Broadsweep search option that allows for a word to be searched for anywhere on the list of possible tags, and this also helped in the disambiguation process.

pus hinders any quantitative inferences, these emotions were ultimately included in the analysis on the basis of their 0.02 per cent frequency.

4.5.2. Most frequent emotion words and most common emotions in the corpus

The combination of both tools yielded a myriad of lexical expressions of emotions and quantitative data. For example, the following expressions were found under category E6– WORRY: *worry/worried/worries/worrying, stress/stressful, concern/concerned with, anxiety, distressing/distress/distressed, trouble/troubles, apprehensive, tension/tensions, bother/bothered, anguish, disturbing, fuss, insecure, nervous, caring.*

The analysis indicates that the most frequent stems under the category of emotion are the following: *Like* (110), *Love* (71), *Suffer* (49), *Fear* (46), *Worry* (37), *Faith* (31), and *Happy* (22). Moreover, after correcting some of the inaccuracies indicated above and associating USAS tags to emotion terms used in psychology, I present a visual representation of the predominant emotions in the *Terminal corpus* (Fig. 6).

The fine-grained analysis seems to indicate that the positive emotions LIKING, LOVE, HAPPINESS, RELIEF and JOY are the most frequent emotions in the texts of the Terminal corpus, amounting to 32% of all instances of emotion. These emotions represent the support structure and foundation necessary to cope with the despair brought on by disease and death, and are the most highly-valued emotions. These are followed in importance by some negative emotions that altogether have a 33% presence in the corpus: SADNESS (11%), FEAR (9%), WORRY (8%) and ANGER (5%). This negativity is counterbalanced by additional positive emotions: HOPE (5%), CONFIDENCE-TRUST (4%), ACCEPTANCE (3%), GRATITUDE (3%), CONTENTMENT (2%) and BRAVERY (1%). Finally, we find some peripheral emotions such as LONELINESS (2%), DISLIKE/HATE, DENIAL, SURPRISE, FRUSTRATION and GUILT, each representing 1% of the corpus.

Furthermore, a reflection on the experience of a terminal illness or the prospect of death led me to consider the possibility that the relevance of expressions denoting PLEASANTNESS and HAPPINESS in my corpus is a consequence of their frequencies in the English language. To verify this, I generated keyness data for emotions with Wmatrix using BNC Sampler Spoken as a reference corpus (Fig. 7).

The key concepts obtained from a keyness analysis are the following: SAD, FEAR/SHOCK, WORRY, CALM, EMOTIONAL ACTIONS, HAPPY, LIKE, CONFIDENT, DISCONTENT, CONTENT, BRAVERY, DISLIKE and VIOLENT/ANGRY.

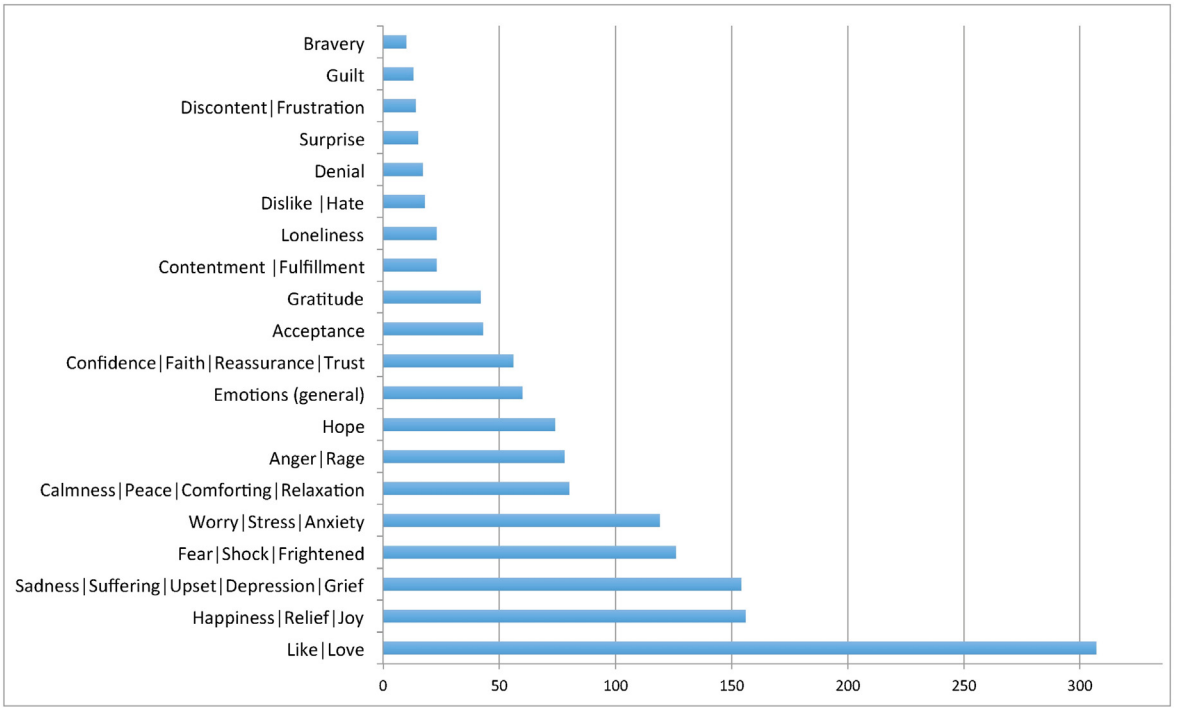

Fig. 6. Predominant emotion categories in the corpus.

Item	O1	%1	O2	%2		LL	%DIFF	
E4.1-	151	0.20	400	0.04	+	204.87	381.42	Sad
E5-	126	0.16	398	0.04	+	142.56	303.73	Fear/shock
E6-	168	0.22	733	0.07	+	124.56	192.29	Worry
E3+	84	0.11	207	0.02	+	121.87	417.50	Calm
E1	59	0.08	137	0.01	+	90.19	449.20	Emotional Actions
E4.1+	153	0.20	782	0.08	+	86.80	149.51	Happy
E2++	32	0.04	76	0.01	+	47.97	436.96	Like
E2+	271	0.35	2255	0.23	+	39.46	53.26	Like
E6+	55	0.07	292	0.03	+	29.02	140.21	Confident
E4.2-	14	0.02	88	0.01	+	5.09	102.88	Discontent
E4.2+	23	0.03	178	0.02	+	4.47	64.78	Content
E5+	9	0.01	55	0.01	+	3.51	108.68	Bravery
E2-	16	0.02	286	0.03	-	1.91	-28.66	Dislike
E3-	77	0.10	959	0.10	+	0.04	2.39	Violent/Angry

Fig. 7. Keyness data and key domain cloud for emotions (Wmatrix).

4.5.3. Discussion of findings related to research question 4

Polysemy and the limitations of automatic semantic tagging required a manual analysis of the data and the combination of two tools for corpus analysis and comparison. Wmatrix and SkE were key to discard the words that had been improperly tagged, and also to add some emotions described in the literature about the end of life, but not categorized as an emotion by Wmatrix. For example, HOPE was tagged under X2.6– (*Mental actions & processes > Unexpected)* and GRATEFULNESS under S1.2.4+ *(Social actions > Deserving).*

However, the findings arising from the manual analysis do not differ significantly from the initial data generated with the USAS semantic tagger. The analysis conducted with regard to research questions 3 and 4 shows that LIKING, HAPPINESS and SADNESS are the most important emotions. The initial relevance of *Worry (E6–)* was the result of the automatic tagging of some cases of 'care' under this category. After the corrections, the frequency of this emotion decreased.

The positive emotions identified in the Terminal corpus (Fig. 6) do not maintain their prominent position when this corpus is compared to the reference corpus (Fig. 7). SADNESS and FEAR/SHOCK outweigh HAPPINESS and LIKING. Nevertheless, the presence of positive emotions continues to offset the negative ones. Moreover, as mentioned before, the size and relevance of WORRY must be tempered due to the improper tagging of the word *care*, which is not used in the sense of an emotion in the corpus. This more realistic interpretation of the emotions found in the corpus is more in line with the results of Láinez Ramos-Bossini and Tercedor-Sánchez (2021). In a corpus of forums on mental health, these authors noticed a tendency towards negativity in both emotions and semantic prosody in relation to a control corpus of general language.

Finally, the results shown in Figs. 6 and 7 and the detailed analysis suggest that the most frequent emotions in the corpus are SADNESS, FEAR, LIKING, LOVE, HAPPINESS/RELIEF, WORRY, CALMNESS, ANGER, HOPE and CONFIDENCE. This result is consistent with Sosea and Caragea's (2020) analysis of the CancerEmo dataset, which revealed that SADNESS, JOY and FEAR accounted for more than 75 per cent of the emotions identified, and that other frequent emotions were TRUST, ANGER, and ANTICIPATION.

5. CONCLUSIONS

Corpus methods can shed light on the way we conceptualize life and the prospect of death by empirically measuring linguistic phenomena that are not easy to outline, such as the lexis of emotion. The compilation of a pilot study corpus (the Terminal corpus) and the combination of two online platforms for corpus analysis (Sketch Engine and Wmatrix) have contributed to quantitative research on the emotions felt at the end of life. This has also allowed for the triangulation of data and an exploration of the interface between psychology and linguistics. The methodology proposed for the

extraction and corpus analysis of text from audiovisual material poses opportunities for further research in cognitive linguistics, lexical semantics and grammar, based on the part-of-speech and semantic tagging affordances of corpus tools.

The automatic semantic annotation provided by Wmatrix has proven fairly accurate when assigning words and multiword expressions to semantic categories, and it is feasible to confirm the 92% accuracy claimed by Wmatrix's developer, Paul Rayson. The limitations of automatic semantic annotation have been overcome by a manual analysis. Likewise, the findings are limited by the fact that they are based on a small corpus, and this might reduce empirical validity. However, the Terminal corpus is the result of a careful selection of English-language texts from platforms with noticeable social influence: Netflix, TED: Ideas Worth Spreading and HealthTalk.org. Moreover, the keywords and key concepts generated by both Sketch Engine and Wmatrix coincide with the topic under study. They are also quite similar to those found by Semino et al. (2018) in the corpus they compiled to investigate metaphors around cancer and the end of life.

In this paper, there is no distinction between the emotions expressed by patients, family caregivers and health professionals in line with studies that examine the disparate feedback and perceptions of these different groups (Semino et al., 2018; Baker et al., 2019), and this might be an interesting thread for future research with a much larger corpus. Additional insight may also be derived from a future qualitative study of the communicative and linguistic context of corpus examples. This complementary analysis could overcome some of the limitations of the present research, in which the correspondence between words and paralinguistic and visual elements was not verified in every lexical instance. This type of multimodal analysis beyond the lexis of emotion can illustrate both the interplay between semantic, social and pragmatic variables, and between verbal, visual and auditory information.

Furthermore, although this research cannot help patients directly, it provides linguistic evidence of the emotions felt at the end of life, most of which are described in both clinical practice guidelines and resources intended for patients. The linguistic data of the Terminal corpus and the methods described in this paper can be useful for lexicographic resources and English for Specific Purposes materials aimed at health professionals and caregivers with limited proficiency in English. For instance, semantic annotation and the extraction of keywords, n-grams and concordances facilitate the identification of frequent and non-frequent expressions of emotion. In the case of HAPPINESS, corpus analysis tools retrieve not only basic vocabulary *(happy, smile, laugh, joy, pleased, glad),* but also additional expressions: adjectives *(cheerful, gleeful, thrilled, overjoyed, uplifting, happy-go-lucky),* verbs *(rejoice, giggle, rapture),* adverbs *(merrily, joyfully)* or phrases and idioms such as *to have a sense of humor, to feel uplifted, to celebrate the life you have had, music to my ears* or *happy camper.* Some of these expressions will be represented in the multimodal lexical resource being developed within the LEXEMOS project (Lexico-semantic characterization of emotions in multimodal communication): https://varimed.ugr.es/lexemos.

In this regard, this study might pave the way for the publication of linguistic resources aimed at non-native speakers of English who work with terminal patients, as well as palliative psychology guidelines to facilitate the communication of emotions and the construction of narratives in one of life's most poignant moments. Finally, worldwide, many people have experienced a COVID-related loss of a relative or friend, as well as the complex emotions this entails. The analysis of the language and feelings around death in the pre-pandemic era might help us understand and verbalize emotions in the current COVID context.

Data availability

The data that has been used is confidential.

ACKNOWLEDGEMENTS

This research was carried out within the framework of the following projects: (1) A-HUM-131-UGR18, Lexico-semantic characterization of emotions in multimodal communication (LEXEMOS), funded by FEDER Andalucía 2014-2020 and the Andalusian Regional Government (Junta de Andalucía-Consejería de Economía y Conocimiento); and (2) PID2020-118775RB-C21, Heritage for all through Translation and Simplified Language: Developing a Query and Analysis Tool (TALENTO), funded by the Spanish Ministry of Science and Innovation (MICINN). A research stay at the University of Bologna (Dipartimento di Interpretazione e Traduzione) was also important for the elaboration of this paper.

Funding for open access charge: Universidad de Granada / CBUA

DECLARATION OF COMPETING INTEREST

No conflict of interest.

ABOUT THE AUTHOR

Clara Inés López-Rodríguez is a tenured professor of the University of Granada, and teaches scientific and multimedia translation. She has participated in various research projects that focus on medical translation such as VariMed (http://varimed.ugr.es), CombiMed-Combinatory Lexis in Medicine: Cognition, Text and Context (http://combimed.ugr.es/) or Lexemos (http://varimed.ugr.es/lexemos/). She is the author of more than 70 publications on scientific, audiovisual and medical translation, corpus linguistics, terminology and knowledge representation.

References

Alarcón Navío, E., López-Rodríguez, C.I., Tercedor-Sánchez, M., 2016. Variation dénominative et familiarité en tant que source d'incertitude en traduction médicale. Meta 61 (1), 117–144. https://doi.org/10.7202/1036986ar.

Alba-Juez, L., Mackenzie, J. L., 2019. Emotion processes in discourse. In: Mackenzie, J. L., Alba-Juez. L. (Eds.), Emotion in Discourse. John Benjamins, Amsterdam, pp. 3–26. https://doi.org/10.1075/pbns.302.

Archer, D., Wilson, A., Rayson, P., 2002. Introduction to the USAS Category System. Benedict Project Report, October 2002. http://ucrel.lancs.ac.uk/usas/usas_guide.pdf.

Baker, P., Brookes, G., Evans, C., 2019. The Language of Patient Feedback: A Corpus Linguistic Study of Online Health Communication. Routledge, London.

Bednarek, M., 2008/2015. Emotion Talk Across Corpora. Palgrave Macmillan, Basingstoke.

Bizquerra Alzina, R., 2020. Psicopedagogía de las emociones [Psychopedagogy of Emotions]. Síntesis, Madrid.

Bobicev, V., Sokolova, M., Oakes, M., 2015. What goes around comes around: Learning sentiments in online medical forums. Cognitive Computation 7 (5), 609–621. https://doi.org/10.1007/s12559-015-9327-y.

Cotrufo, T., Ureña Bares, J. M., 2018. El cerebro y las emociones. Sentir, pensar, decidir. [Brain and emotions: feeling, thinking, deciding]. EMSE EDAPP, Barcelona.

Cowen, A.S., Keltner, D., 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc. Natl. Acad. Sci. 114 (38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114.

Damasio, A., 2003. Looking for Spinoza: Joy, Sorrow and the Feeling Brain. Heinemann, London.

Damasio, A., 2018. The Strange Order of Things. Life, Feeling, and the Making of Culture. Pantheon Books, New York.

Davies, M., 2021. The TV and Movies corpora: Design, construction, and use. Int. J. Corpus Linguist. 26 (1), 10–37. https://doi.org/10.1075/ijcl.00035.dav.

Ekman, P., 1992. An argument for basic emotions. Cogn. Emot. 6 (3–4), 169–200. https://doi.org/10.1080/02699939208411068.

Epstein, R., Friedman, J. (Dirs.), 2018. End Game [Documentary]. Netflix.

Esuli, A., Sebastiani, F., 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. Proceedings of the 5th Conference on Language Resources and Evaluation (Genoa, 24-26 May 2006), pp. 417–422.

Fernández-Berrocal, P., Ramos-Díaz, N., 2005. Evaluando la Inteligencia Emocional. In: Fernández-Berrocal, P., Ramos, N. (Eds.), Corazones inteligentes [Intelligent Hearts]. Kairós, Barcelona, pp. 35–38.

Frijda, N.H., Markam, S., Sato, K., Wiers, R., 1995. Emotions and emotion words. In: Russell, J.A., Fernández-Dols, J.M., Manstead, A.S.R., Wellenkamp, J.C. (Eds.), Everyday Conceptions of Emotion. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 121–143. https://doi.org/10.1007/978-94-015-8484-5_7.

Jaiswal, M., Bara, C.P., Luo, Y., Burzo, M., Mihalcea, R., Mower Provost, E., 2020. In: MuSE: a Multimodal Dataset of Stressed Emotion. European Language Resources Association, (Marseille, France), pp. 1499–1510.

Jiménez-Crespo, M.Á., Tercedor-Sánchez, M., 2017. Lexical variation, register and explicitation in medical translation: A comparable corpus study of medical terminology in US websites translated into Spanish. Translat. Interpret. Stud. 12 (3), 405–426. https://doi.org/10.1075/tis.12.3.03jim.

Johnson-Laird, P.N., Oatley, K., 1989. The language of emotion: An analysis of a semantic field. Cogn. Emot. 3, 81–123. https://doi.org/10.1080/02699938908408075.

Kilgarriff, A., 2009. Simple maths for keywords. In: Mahlberg, M., González-Díaz, V., Smith, C. (Eds.), Proceedings of Corpus Linguistics Conference CL2009 (University of Liverpool, July 2009). https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V., 2014. The Sketch Engine: ten years on. Lexicography ASIALEX 1, 7–36 http://www.sketchengine.eu.

Kövecses, Z., 2000. Metaphor and Emotion: Language, Culture, and Body in Human Feeling. Cambridge University Press, Cambridge.

Krauss, D. (Dir.), 2016. Extremis [Documentary]. Netflix.

Láinez Ramos-Bossini, A.J., Tercedor-Sánchez, M., 2021. Lenguaje y emoción. Estudio basado en corpus de foros de debate sobre salud mental. In: Flores Borjabad, S.A., Pérez Cabaña, R. (Eds.), Nuevos retos y perspectivas de la investigación en literatura, lingüística y traducción. Dykinson, Madrid, pp. 1712–1737.

Lang, P.J., 1995. The emotion probe. Studies of motivation and attention. Am Psychol 50 (5), 372–385. https://doi.org/10.1037//0003-066x.50.5.372.

López-Rodríguez, C.I., 2019. Verbal patterns to express DISEASE in English medical texts. In: Simonnaes, I., Andersen, O., Schubert, K. (Eds.), New Challenges for Research on Language for Special Purposes (Forum für Fachsprachen-Forschung). Frank & Timme, Berlin, pp. 219–233.
López-Rodríguez, C.I., Sánchez-Cárdenas, B., 2021. Theory and Digital Resources for the English-Spanish Medical Translation Industry. Peter Lang, Bern.. https://doi.org/10.3726/b18119.
Lynge Olsson, N., 2021. Subtitle Edit (version 3.6) [computer software]. http://nikse.dk.
Maiese, M., 2011. Embodiment, Emotion, and Cognition. Palgrave Macmillan, Basingstoke.
Martin, J.R., White, P.R.R., 2005. The Language of Evaluation: Appraisal in English. Palgrave Macmillan, Basingstoke.
Mohammad, S.M., Turney, P.D., 2013. NRC Word-Emotion Association lexicon. National Research Council, Canada. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.
Ohayon, M. (Dir.), 2016. Cristina [Documentary]. Netflix.
Ortony, A., Clore, G.L., Collins, A., 1988. The Cognitive Structure of Emotions. Cambridge University Press, Cambridge.
Peck, D. (Dir.), 2018. Ram Dass: Going Home [Documentary]. Netflix.
Plutchik, R., 2001. The Nature of Emotions. Am. Sci. 89, 344 http://www.jstor.org/stable/27857503.
Prentice, S., Knight, J., Rayson, P., El Haj, M., Rutherford, N., 2021. Problematising characteristicness: A biomedical association case study. Int. J. Corpus Linguist. 26 (3), 305–335. https://doi.org/10.1075/ijcl.19019.pre.
Prieto-Velasco, J.A., 2017. Recepción y percepción de las imágenes en textos médicos para pacientes: estudio experimental sobre la repulsión. Panace@: Revista de Medicina, Lenguaje y Traducción 18 (45), 50–60 https://www.tremedica.org/panacea/v18-n45-junio-2017/.
Rayson, P., 2008. From key words to key semantic domains. Int. J. Corpus Linguist. 13 (4), 519–549. https://doi.org/10.1075/ijcl.13.4.06ray.
Rayson, P., 2009. Wmatrix: a web-based corpus processing environment (Version 4 and version 5) [Computer software]. Lancaster University, Computing Department http://ucrel.lancs.ac.uk/wmatrix/.
Russell, J.A., 2003. Core affect and the psychological construction of emotion. Psychol. Rev. 110 (1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145.
Sacharin, V., Schlegel, K., Scherer, K.R., 2012. Geneva Emotion Wheel rating study (Report). University of Geneva, Swiss Center for Affective Sciences.
Scherer, K.R., 2005. What are emotions? And how can they be measured? Social Sci. Inf. 44 (4), 693–727. https://doi.org/10.1177/0539018405058216.
Scherer, K.R., 2009. Emotion theories and concepts (psychological perspectives). In: Sander, D., Scherer, K.R. (Eds.), Oxford Companion to Emotion and the Affective Sciences. Oxford University Press, Oxford, pp. 145–149.
Scott, M., 1997. PC analysis of key words—and key key words. System 25 (2), 233–245. https://doi.org/10.1016/S0346-251X(97)00011-0.
Semino, E., Demjén, Z., Hardie, A., Payne, S., Rayson, P., 2018. Metaphor, Cancer and the End of Life. A Corpus-Based Study. Routledge, London.
Sokolova, M., Bobicev, V., 2020. What Sentiments Can Be Found in Medical Forums? Proceedings of Recent Advances in Natural Language Processing (Hissar, Bulgaria, 7-13 September 2013), pp. 633–639. https://aclanthology.org/R13-1083.pdf.
Sosea, T., Caragea, C., 2020. CancerEmo: A Dataset for Fine-Grained Emotion Detection. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (Online). Association for Computational Linguistics, pp. 8892–8904. https://aclanthology.org/2020.emnlp-main.715.
Tiedemann, J., 2016. Finding Alternative Translations in a Large Corpus of Movie Subtitles. Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), pp. 3518–3522. http://www.lrec-conf.org/proceedings/lrec2016/pdf/62_Paper.pdf.
Tercedor-Sánchez, M., Láinez Ramos-Bossini, A., 2017. La comunicación en línea en el ámbito de la psiquiatría: caracterización terminológica de los foros de pacientes. Panace@: Revista de Medicina, Lenguaje y Traducción 18 (45), 30–41 https://www.tremedica.org/panacea/v18-n45-junio-2017/.
Tyrrell, P., Harberger, S., Siddiqui, W., 2021. Stages of Dying. StatPearls [Internet]. StatPearls Publishing.
Valitutti, A., Strapparava, C., Stock, O., 2004. Developing affective lexical resources. PsychNology J. 2 (1), 61–83. http://www.psychnology.org/File/PSYCHNOLOGY_JOURNAL_2_1_VALITUTTI.pdf.
Zad, S., Jimenez, J., Finlayson, M., 2021. Hell Hath No Fury? Correcting Bias in the NRC Emotion Lexicon. Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021), Online. Association for Computational Linguistics, pp. 102–113. https://aclanthology.org/2021.woah-1.11.pdf.