Adjectives expressing sadness in Swedish and English: A contrastive lexical analysis

Sandra Magni

Department of Linguistics
Bachelor's Thesis 15 ECTS credits
Linguistics – Bachelor's Course, LIN600
Bachelor's Programme in Linguistics 180 ECTS credits
Spring semester 2023
Supervisors: Bernhard Wälchli, Robert Östling
Swedish title: Adjektiv som uttrycker ledsenhet på svenska och engelska: En kontrastiv lexikal analys

Adjectives expressing sadness in Swedish and English: A contrastive lexical analysis

Sandra Magni

Abstract

This study examines how Swedish and English adjectives relating to the semantic domain of SADNESS are used and what some of their semantic and syntactic characteristics are. The primary objectives of this study are to determine whether the adjectives in question are used more often *attributively* or *predicatively*, whether they are used more frequently along with the semantic role of *stimulus* or *experiencer* and whether there are any syntactic, semantic or lexical disparities in how Swedish and English encode the domain of SADNESS. To investigate this, parallel corpus data retrieved from movie and TV subtitles were analyzed in combination with dictionary data. The results showed that predicative usage was more dominant than attributive. While the semantic role of stimulus was more prevalent in the dataset of this study, experiencer-oriented adjectives showed higher prevalence when considering the individual frequencies of each word in the corpus. Notably, the Swedish word *tråkig* combines the meanings of 'sad' and 'boring', illustrating differences in emotion-related concept encoding between the two languages.

Keywords

sadness, lexical typology, predicative, attributive, experiencer, stimulus

Sammanfattning

Denna studie undersöker hur svenska och engelska adjektiv som relaterar till den semantiska domänen LEDSENHET används och vad några av deras semantiska och syntaktiska egenskaper är. De primära målen med denna studie är att avgöra om adjektiven i fråga används mer ofta *attributivt* eller *predikativt*, om de används oftare i samband med den semantiska rollen *stimulus* eller *upplevare* och om det finns några syntaktiska, semantiska eller lexikala skillnader i hur svenska och engelska kodar domänen LEDSENHET. För att undersöka detta analyserades parallella korpusdata hämtade från film- och TV-undertexter i kombination med ordboksdata. Resultaten visade att predikativ användning var mer dominerande än attributiv. Medan den semantiska rollen stimulus var mer utbredd i datamaterialet för denna studie visade upplevare-orienterade adjektiv högre prevalens när de individuella frekvenserna för varje ord i korpusen beaktades. Det svenska ordet *tråkig* kombinerar betydelserna av 'ledsen' och '(lång)tråkig' vilket illustrerar skillnader i emotionsrelaterad begreppskodning mellan de två språken.

Nyckelord

ledsenhet, lexikal typologi, predikativ, attributiv, upplevare, stimulus

Contents

1 Introduction **1**

2 Background **2**
2.1 Introduction . . . 2
2.2 The semantic domain of SADNESS . . . 2
2.3 Lexical typology . . . 3
2.4 Attributive and predicative adjectives . . . 4
2.5 Semantic roles . . . 5
2.6 Previous linguistic studies in the domain of SADNESS . . . 6

3 Purpose and research questions **8**
3.1 Purpose . . . 8
3.2 Research questions . . . 8

4 Method **9**
4.1 Introduction . . . 9
4.2 Data . . . 9
4.2.1 Corpora . . . 9
4.2.2 OpenSubtitles . . . 9
4.2.3 Sample . . . 10
4.2.4 Dictionaries . . . 10
4.3 Procedure . . . 11
4.3.1 Choice of words to be examined . . . 11
4.3.2 Data extraction . . . 12
4.4 Analysis . . . 13

5 Results **14**
5.1 Introduction . . . 14
5.2 Frequency . . . 14
5.3 Predicative and attributive use . . . 15
5.4 Semantic roles . . . 16
5.5 The difference and relation between syntactic function and semantic role orientation . . . 18
5.6 Lexical comparison . . . 20
5.6.1 Ledsen . . . 20
5.6.2 Sad . . . 21
5.6.3 Tråkig . . . 22
5.6.4 *Unhappy*, *unfortunate*, *tragic* and *depressing* . . . 23
5.6.5 *Trist*, *sorglig* and *sorgsen* . . . 24
5.6.6 The meaning-frequency relationship of words . . . 25
5.6.7 Key observations from the lexical comparison . . . 26

6 Discussion **27**
6.1 Introduction . . . 27
6.2 Discussion of results . . . 27
6.2.1 Syntactic function . . . 27
6.2.2 Semantic role orientation . . . 28

6.2.3 Lexical differences: Etymology and semantic role orientation 29
6.3 Discussion of method 30
6.4 Future research 31

7 Conclusions **32**

References **33**

A A selection of the dataset **36**

B The Python code used for making Figure 1 **37**

1 Introduction

The meaning and significance of human emotions has been a subject of inquiry for many years and the fundamental characteristics of emotions are still being explored in linguistics, psychology, biology and many other fields (Jackson et al. 2019). While most human languages have words for describing emotions such as "sadness", "joy", "anger" and "fear" it is not clear whether these emotions have similar meanings across languages or why their meanings might vary.

Contrastive language studies primarily aim to identify, analyze, and elucidate similarities and differences between a pair of languages, often shedding light on their unique syntactic, semantic, and lexical characteristics. By juxtaposing languages, these studies deepen our understanding of linguistic features, cross-linguistic universals, and the cultural nuances that shape them. In line with these objectives, this study delves into the nuances of Swedish and English adjectives within the semantic domain of SADNESS. By comparing these two languages, the research seeks to highlight potential disparities in how sadness is encoded, thereby contributing to the broader framework of contrastive linguistic analysis. Specifically, this study aims to investigate how adjectives relating to the semantic domain of SADNESS are used in Swedish and English and what their semantic and syntactic characteristics are. Several previous studies have examined the semantic domain of SADNESS in English and Swedish as well as in many other languages, however, most of them have done so by looking at how sadness is conceptualized in metaphors and metonymies. The current study fills a gap by examining some of the syntactic, as well as semantic, aspects of adjectives denoting sadness. It seeks to provide a more comprehensive and nuanced understanding of the meaning of words relating to sadness in Swedish and English as well as the interplay between syntactic and semantic characteristics and does so by analyzing parallel corpus data retrieved from movie and TV subtitles. Such sources, often reflecting colloquial usage and cultural norms, can offer a genuine glimpse into the natural language usage of words related to emotions while the use of parallel corpora enables us to compare more precisely which concepts are encoded by what words in different languages. In order to ascertain whether there are any lexical differences in how Swedish and English encode the concept of sadness, dictionaries were consulted in addition to the corpus data as a complementary lexical resource.

The primary objectives of this study are to determine whether the adjectives in question are used more often *attributively* or *predicatively*, whether they more frequently align with the semantic roles of *stimulus* or *experiencer* and whether there are any syntactic, semantic or lexical disparities in how Swedish and English encode the concept of sadness. Through this exploration, the study reaffirms the understanding that while emotions are a shared human experience, the ways in which different languages interpret, experience and articulate these emotions might vary across languages.

2 Background

2.1 Introduction

This section provides an overview of the theoretical background for the current study. Section 2.2 begins by describing what is meant by "the semantic domain of SADNESS" in the context of this study. In section 2.3, a short introduction is given to the field of lexical typology, a subfield of linguistics highly relevant to the topics and aims of the current study. Section 2.4 details attributive and predicative adjectives as well as secondary predication and section 2.5 describes the semantic roles relevant to this study. These sections explain how the two aspects, one being syntactic and one semantic, can relate to one another and how each of them pertain to the aims and research questions of this study. A summary and review of previous research within the domain of SADNESS is presented in section 2.6.

2.2 The semantic domain of SADNESS

A semantic domain (also referred to as a semantic field) can be defined as consisting of a set of lexemes all relating to the same content domain (Lehrer and Kittay 1992, p. 3). The words in a semantic domain, though not synonymous, all share some common semantic property and the notion of a semantic domain can be extended to any set of lexemes with a close relation in meaning, all of which can be subsumed under the same general label (Akmajian et al. 2001, p. 240). In this study, adjectival lexemes belonging to the semantic domain of SADNESS, namely, those that in some way relate to or denote the concept of sadness, will be examined.

Sadness is an intricate emotion, deeply embedded in the human experience. As noted by Power (2010), sadness, much like other emotions, is easily felt or recalled but remains challenging to define or describe in clear terms (p. 1486). A leading perspective in psychology's study of emotions, including sadness, revolves around the "basic emotions" approach. This paradigm underscores the potential existence of a finite set of fundamental emotions, present universally across cultures, from which more nuanced emotions evolve. Paul Ekman, a chief proponent of this approach, has demonstrated that emotions like sadness are universally expressed and identified via facial cues across diverse cultures. In their seminal cross-cultural study, Ekman and his colleagues (1972) concluded that the six basic emotions are anger, disgust, fear, happiness, sadness and surprise. However, in a subsequent exploration, Ekman (1999) expanded this list to incorporate emotions such as amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame, highlighting the complex tapestry of human emotions not solely expressed through facial cues. Central to the experience of sadness is the contemplation of loss or failure, whether actual or perceived, set in the past, present or an anticipated future. Such reflections may revolve around various subjects or aspirations, ranging from individuals, places, unfulfilled ambitions, cherished possessions or even the collapse of a once-held moral or ideal value. What induces sadness can be diverse, influenced by individual and cultural interpretations of loss. Nevertheless, recurrent catalysts can include feelings of rejection, the illness or death of a loved one, or disillusionment from unmet expectations. As Ekman (1992) aptly encapsulated, sadness is an expansive term that embraces a spectrum of feelings, from mere disappointment to profound despair and torment.

In her semantic analysis of words relating to sadness in English and Russian, Wierzbicka (1999, p. 39) states that the full definition of "sadness" includes both a specific feeling and a sequence of thoughts or a cognitive scenario which includes recognizing that something undesirable has occurred, wishing such things didn't happen and understanding there's nothing one can do about it. Wierzbicka argues that this cognitive scenario embedded in the English word

"sadness" isn't necessarily universal and that other languages may encode different cognitive scenarios. She states that differentiating between core and peripheral aspects of a concept is vital. In Russian, the emotional landscape painted by the term "sadness" is not represented by a single word, but rather by a spectrum of words, each with its own nuances. The language offers *grustnyj* and *pečal'nyj* as translations for 'sad', and when expressing 'sadness', it uses terms like *grust*, *pečal'*, and at times, *toska*. Each of these Russian terms, while being in the same emotional vicinity as "sadness", carries its own distinct connotations. For instance, while *grust* might denote a melancholic type of sadness, *pečal'* could imply a deeper, more sorrowful sentiment, and toska often conveys a sense of deep longing or yearning. Simply equating them directly with the English "sad" and "sadness" overlooks these intricacies. It's crucial to approach such translations with sensitivity, understanding that emotional concepts might be spread across different terms in various languages, each reflecting unique cultural and linguistic sensibilities.

Psychology as a whole, and perception specifically, is closely related to language in several ways. On the most fundamental level, the acoustic stimulus provided by our voice must itself be perceived in order for linguistic communication to succeed. At a more subtle level, the way in which people perceive the world around them might be affected by the way they talk about it, and conversely (Miller and Johnson-Laird 1976, p. 2). Oatley and Johnson-Laird (1989) explain that "[...] language and its underlying conceptual apparatus is intimately related to the real nature of emotions, and the meanings of emotional terms are neither arbitrary nor analyzable but do indeed relate to experience." (pp. 105). Emotions have psychological causes and are created by cognitive evaluations, such as the perception of a predator making one fearful and its disappearance reducing the fear. Emotions serve a communicative function both within the brain and within the social group. External signals such as facial expressions, movements and gestures and tone of voice make it possible for people to communicate emotions to others without having to use words. The use and interpretation of bodily expressions and nonverbal cues often vary across cultures and they are of great significance when examining words related to emotion since they can complement and help to better understand the meaning of the words themselves (Wierzbicka 1999, p. 22).

2.3 Lexical typology

Lexical typology can be generally defined as the systematic study of cross-linguistic variation in words and vocabularies or, put more simply, the study of the ways in which different languages express meanings with words. This can be approached from a few slightly different perspectives, one of them being by examining how, and through which lexical items, a certain domain is expressed in different languages (Koptjevskaja-Tamm et al. 2015, p. 434). Previous lexical typological studies have investigated semantic and conceptual domains such as color terminology (Berlin and Kay 1969), body-part terminology (Brown 1976), kinship terminology (Nerlove and Romney 1967), motion verbs (Talmy 1985) and perception verbs (Viberg 1984), among others.

In lexical typological research, it is also important to consider various syntactic patterns and constructions since these can reveal important clues as to how different experiences, perceptions and sensations are conceptualized in different languages. This is evident in *The linguistics of temperature* (Koptjevskaja-Tamm 2015), a comprehensive typological study of the conceptualization of temperature in more than fifty genetically, areally and typologically diverse languages, edited by Koptjevskaja-Tamm. It showcases, among other things, how the semantics of temperature words depend on their syntactic functions to a large extent and it consists of several studies in which syntactic constructions are used for talking about temperature-related situations. Although lexical typology is a linguistic approach, it does not exclusively deal with entirely lin-

guistic phenomena. The field of lexical typology is interested in examining the meanings and implications of words from linguistic as well as psychological, anthropological and philosophical perspectives. The field of psychology, in particular, has long had a close relation to semantic and lexical typological research. This is evident especially in the area of color terminology, which is one of the most extensively investigated areas in lexical typology. Some examples are the studies and experiments conducted by psychologist Eleanor Rosch who further developed the theory originally put forward by anthropologist Brent Berlin and linguist Paul Kay on how named color categories in the world's languages are organized around universal focal colors (Berlin and Kay 1969; Heider 1972; Rosch 1973). Similarly to that of color terminology, the linguistic study of emotion is dependent upon, and may have significant implications for, the field of psychology, and vice versa. Drawing from Ekman's (1972) research on universal emotions and Wierzbicka's (1999) exploration of the intricate semantic differences in words like "sadness" between English and Russian, it becomes evident that the intertwining of psychology with lexical and semantic typology offers a richer understanding. This combination helps reveal how feelings and concepts are uniquely encoded across various languages and cultures, emphasizing the depth and variety in emotional expression.

In the field of lexical typology, the comprehensive and broad bibliography presented in this study reflects an understanding of the interdisciplinary nature of language, drawing on insights from linguistics, psychology and anthropology. Engaging with diverse sources and theoretical perspectives ensures a multifaceted examination of the conceptual or semantic domain in question, showcasing the evolution of thought and the interconnectedness of linguistic phenomena. This comprehensive approach mirrors the real-world complexity of language and strengthens the study's theoretical foundation in general linguistics.

2.4 Attributive and predicative adjectives

Compared to the lexical categories of nouns and verbs, adjectives have received less attention in the linguistic literature (Cabredo Hofherr and Matushansky 2010, p. 1). Although sadness and other emotions can be encoded through various word classes including verbs (e.g., *to mourn*) and nouns (e.g., *grief*), adjectives are the most common way of encoding the concept of sadness in the Germanic languages, of which English and Swedish are a part (Wierzbicka 1999, p. 18). Given this, the examination of how adjectives relate to the concept of sadness is particularly compelling. This study will follow the standard classification of adjective usage into two main types: attributive and predicative. Although this is a formal, syntactic, distinction which might not typically be of interest to lexical and semantic typology, the distribution of attributively and predicatively used adjectival lexemes should not be taken for granted. Investigating how attributive and predicative use relates to the semantics of adjectives might provide a better understanding of whether this distinction is a solely syntactic notion or whether it has semantic implications as well.

Attributive adjectives appear in the first position in a noun phrase, as demonstrated in (1), while predicative adjectives appear in the second position in a clause, as in examples (2a) through (4b).

(1) The silly rooster has taught us a ***sad*** truth.

Englebretson (1997) examines the distribution of predicative versus attributive adjectives in a corpus of 51,852 words of naturally occurring spontaneous English conversation and describes three types of predicative adjectives present in their data. The first type is one that includes linking verbs, i.e., a verb that describes the subject by connecting it to a predicate adjective or

predicate noun such as a copula or a verb of perception. The first type is demonstrated in (2a) and (2b).

(2) a. That sounds ***depressing***. b. It's ***sad***.

The second type of predicative adjective acts as a predicate complement without a linking verb, as illustrated in (3).

(3) I keep it really ***cool***. (Englebretson 1997)

The third type consists of assessments since they predicate a property of a referent that can only be discerned with awareness of the previous discourse or situational context. This is demonstrated in (4a) and (4b).

(4) a. How ***unfortunate***. b. ***Tragic***.

Secondary predication is a subtype of predicative adjectives and something which, although it is not a main focus, will be looked at in this study. They are generally subdivided into resultatives and depictives. Resultatives express a result of the state encoded by the main predicate, as in (5a). Depictives, on the other hand, express a state that holds during the reference time of the event encoded by the main predicate, as in (5b).

(5) a. The movie left me ***depressed***. (constructed)

b. He walked around ***sad***. (constructed)

Constructions containing secondary predicates provide a number of challenges for syntactic and semantic theory. Whether all languages have the same kinds of secondary predicate constructions and what kind of semantic content secondary predicates commonly treat are relevant questions particularly for typological analyzes on adjectival lexemes (Himmelmann and Schultze-Berndt 2005, p. 2).

2.5 Semantic roles

Semantic roles (also known as thematic roles, theta roles or thematic relations) such as agent, patient, goal, source, stimulus and experiencer, represent the underlying relationship that a participant has with the main verb in a clause (Payne 1997, p. 47). A concept relevant for this study, relating both to semantic roles and to the function of adjectives, is that of valence. It refers to the number of arguments that a head may take and can be described in both syntactic or semantic terms with the syntactic valence of a head being the number of arguments it may have and the semantic valence of a head being the number of semantic roles associated with it (Van Valin Jr 2001, p. 92). Adjectives such as the ones examined in this study, e.g., *sad* and *unhappy*, can take arguments when used predicatively. They can have one or two arguments, as demonstrated in (6a) and (6b). Since *unhappy* in (6b) takes two arguments, *Jessica* and *Paul*, and is associated with two semantic roles, *Jessica* having the role of experiencer and *Paul* the role of stimulus, it has the syntactic and semantic valence of 2. In (6a) on the other hand, *unhappy* has the syntactic and semantic valence of 1.

(6) a. Jessica is ***unhappy***. (constructed)

b. Jessica is ***unhappy*** with Paul. (constructed)

Adjectives can also have the syntactic and semantic valence of 0, as demonstrated in (7).

(7) ***Sad***!

Understanding the difference between syntactic and semantic valence is essential for studies examining the characteristics of adjectives, as it helps discern the structural requirements and meaning relationships within sentences.

The semantic roles pertinent to this study are experiencer and stimulus. The role experiencer typically constitutes someone or something having or undergoing a given experience while the role stimulus typically represents someone or something giving rise to a certain experience (Crawley and Kleinman 1994). The experiencer is an animate entity that in some way perceives or passively undergoes an event or activity typically created by a stimulus but involving neither volition nor a change of state. Correspondingly, the stimulus is an entity that causes emotional or sensory responses without agency (Payne 1997, p. 50).

Van Valin (2001) notes that "There is no agreement among syntacticians or semanticists as to the 'correct' set of semantic roles" (p. 23). He discusses semantic roles from the perspective of two levels of generality or specificity: in terms of verb-specific roles, such as speaker or entity broken, and in more general terms, such as agent and patient. Van Valin refers to roles like runner and hearer as 'verb-specific semantic roles' and roles like agent, theme, experiencer and stimulus as 'thematic relations' in order to distinguish between the two types of roles terminologically. In this interpretation, verb-specific semantic roles such as thinker and believer, which are further grouped into the role of 'cognizer', hearer and feeler, which are grouped into the role 'perceiver' and liker and hater, which are grouped into the role 'emoter' all make up the thematic relation of experiencer. Meanwhile, the verb-specific semantic roles of something which is seen, heard, liked etc. make up the thematic relation of stimulus.

2.6 Previous linguistic studies in the domain of SADNESS

There are several studies that examine the semantic domain of SADNESS in English, Swedish and many other languages. However, most of these do so by looking at how sadness is conceptualized in metaphors and metonymies. Some examples include: *On the concept of sadness: Looking at words in contexts derived from corpora* (Tissari 2008) and *A comparative and contrastive study of sadness conceptualization in Persian and English* (Moradi and Mashak 2013). There are a few exceptions, however, one of them being the previously mentioned analysis of words relating to sadness in English and Russian by Wierzbicka (1999). Another one is the study conducted by Jackson et al. (2019) in which emotion semantics across a sample of 2474 spoken languages were evaluated by examining cases of colexification, that is, instances in which multiple concepts are expressed by the same word form within a language. By employing a large comparative linguistic database to create vast networks of how people use words to name experiences, the study provides new knowledge about diversity as well as universality in how humans experience and understand emotion.

Sadness expressions in English and Chinese have been investigated by Zhang (2014) in a valuable large-scale corpus-based examination. In order to capture their subtle nuances of meaning, reveal the culture-specific conceptualizations encoded in them and uncover their respective cultural distinctiveness, Zhang conducted a contrastive semantic analysis of expressions relating

to the general concept of sadness in English and Chinese. The author describes how each language has its own set of emotion words to characterize emotional experience and how, therefore, the comparison of the emotional lexicon of a language with that of another can point to similarities and differences in their respective conceptualizations of emotion. Similar to what the current study is intending to do, Zhang compares and contrasts lexical items that do not match neatly between languages and both studies do so by examining corpus data in combination with dictionary data. While Zhang looks at sadness expressions in terms of the corpus-linguistic structural categories colligation, collocation, paraphrase and semantic association or preference the current study aims to explore the expressions in terms of which semantic roles they are more oriented towards and whether they are used more attributively or predicatively. In their book, Zhang emphasizes the significance of the way in which language, thought and culture influence and shape one another.

In conclusion, while many studies delve into metaphoric expressions of sadness, only a few, notably those by Zhang (2014), Wierzbicka (1999), and Jackson et al. (2019), deeply analyze the specific words and structures languages use to convey this emotion.

3 Purpose and research questions

3.1 Purpose

The purpose of this study is to explore how Swedish and English adjectives relating to the semantic domain of SADNESS are used, what some of their semantic and syntactic characteristics are and whether there are any syntactic, semantic or lexical disparities in how the two languages encode this concept. The languages Swedish and English were chosen in order to enable a deeper and more thorough understanding and analysis, something which might not be feasible with a large sample of languages that I as the author am not familiar with.

3.2 Research questions

1. Are the adjectives examined used more often attributively or predicatively?
2. Are they predominantly stimulus- or experiencer-oriented?
3. Are there any differences in how the closely related languages Swedish and English encode the concept of sadness?

4 Method

4.1 Introduction

In this section, the methodology of the study is presented. Section 4.2 contains a presentation of the data sources and section 4.3 provides a walk-through of the procedure pursued for deciding upon which words to include in the study together with the data extraction. Section 4.4 details the analysis of the data.

4.2 Data

4.2.1 Corpora

The data used in this study consists of Swedish and English parallel corpora made up of movie and TV subtitles. Part of the reason why movie and TV subtitles were chosen for this study is that they offer an informal and colloquial language sample. The language exhibited in movies and TV shows represents spoken discourse and is typically meant to reflect spontaneous conversations. Because of this, movie and TV subtitles constitute a valuable resource for the compilation of parallel corpora and can have several advantages over other existing parallel corpora such as legal documents, translations of the Bible or European Parliament proceedings (Levshina 2016; Lison and Tiedemann 2016). Another reason why movie and TV subtitles are suitable is that, apart from the language itself, movies and TV shows are accompanied by other visual and auditory contextual information (Ahn et al. 2012). When it comes to investigating words relating to emotion especially, external and bodily signals such as facial expression and tone of voice can serve as a useful complement to the analysis of language. An examination of these factors was not conducted within the current study but is something that might be instrumental for future research on the language of emotions.

The reason why a parallel corpus, as opposed to two separate monolingual corpora, was chosen for this study is because using parallel corpora enables one to compare more precisely what concepts are encoded in which words in different languages. Being able to examine how the Swedish words correspond to the English ones and vice versa is necessary in order to ascertain any potential differences in how the two languages encode the concept of sadness. As will be reported on further ahead in sections 5.6.1 to 5.6.5, the translations of individual words given in a bilingual dictionary might not cohere with the translations found in parallel corpora and the use of parallel corpora might therefore provide a broader picture of what words are actually used to express particular emotions in two different languages.

4.2.2 OpenSubtitles

The data used in this study consists of movie and TV subtitles in Swedish and English retrieved from the OpenSubtitles parallel corpora 2018, a collection of parallel corpora made up of translated movie and TV subtitles available at `https://www.opensubtitles.org/`. OpenSubtitles is part of the OPUS project which is maintained by Jörg Tiedemann. The OpenSubtitles 2018 parallel corpora of Swedish and English respectively are available on and were collected via the corpus manager and text analysis software Sketch Engine. The OpenSubtitles parallel 2018 - English corpus contains approximately 1 billion words in total and the OpenSubtitles parallel 2018 - Swedish corpus contains approximately 100 million words in total. It's uncertain how extensive the truly parallel portion is between the OpenSubtitles parallel 2018 English and Swedish corpora. While the English corpus comprises approximately 1 billion words and the

Swedish one has 100 million words, the parallel section can be estimated at most around 100 million words, derived from the Swedish content and its corresponding English translations. This approximation, however, can vary due to several factors. The alignment might not consistently be one-to-one, meaning a single Swedish sentence could align with multiple English ones or vice versa. Not every Swedish segment may have an English counterpart, potentially reducing the aligned content. Furthermore, the presence of metadata or other non-content elements can impact the total word count. So, even though the maximum parallel portion could be around 100 million words, the genuine aligned content might be somewhat less due to these variables.

4.2.3 Sample

The sample used in this study consists of data taken from the drama, crime, thriller, horror, sci-fi, war and mystery genres since these are the genres where the greatest quantity of sadness expressions are likely to be found. In order to avoid instances of the sadness expressions being used in a sarcastic or ironic manner, data from the comedy genre as well as a few others that often coincide with it, such as fantasy, romance, adventure, family and animation, were not included.

No delimitations were made regarding the release date or original language of the movie and TV shows. This decision was made because neither release date nor original language were deemed as factors that would affect or considerably change the outcome of the analysis. In the current study, 100 instances of five English and five Swedish words, i.e., a total of 1000 sentences, were collected. How the sample size was determined is described below in 4.3.1.

4.2.4 Dictionaries

In order to ascertain whether there are any lexical differences in how Swedish and English encode the concept of sadness, dictionaries were consulted in addition to the corpus data as a complementary lexical resource. As mentioned previously, the translations of individual words given in a bilingual dictionary might not correspond to the translations found in parallel corpora and comparing the two could therefore be of great interest in order to gain a complete understanding of the properties and use of each word.

For the English words, Collins Dictionary, available online at `collinsdictionary.com`, was used. Collins Dictionary consists partly of the Collins COBUILD Advanced Learner's Dictionary which is itself based on the Collins Corpus, a 4.5 billion-word language database of words from a vast range of sources of spoken and written English. For the Swedish words, the dictionary *Svensk ordbok utgiven av Svenska Akademien* (SO), available online at `svenska.se`, was used. SO contains descriptions of modern Swedish and emphasizes the meaning and use of words in combination with historical information. The English–Swedish dictionary *NE:s stora engelska ordbok: Engelsk–svensk/svensk–engelsk*, available online (with a paid or institutional subscription) at `ne.se`, was consulted for translations. The translations used for the Swedish words in the following sections are the one or two first ones given in the English-Swedish dictionary.

The definitions retrieved from SO presented in this study were translated to English by the author. As will be demonstrated further on, there are a lot of subtle nuances between the words examined in this study and their equivalents or translations in the other language. When translating the Swedish dictionary definitions to English, the words were chosen based on the foremost translations given in the English–Swedish dictionary *NE:s stora engelska ordbok: Engelsk–svensk/svensk–engelsk*. Because of this, the word *nedstämd* used in the Swedish dictionary definition of *ledsen* was translated to *depressed* even though many other translations, such as *dejected*, *gloomy*, *low* or *downhearted* to name a few, could have been appropriate as well.

4.3 Procedure

4.3.1 Choice of words to be examined

As previously mentioned, the data sample used in this study consists of 100 instances of five English and five Swedish words, i.e., a total of 1000 sentences. Because all of the sentences collected from the corpora were going to be individually and manually annotated, a limit needed to be made in regards to the number of sentences to include in the study. With the time frame of the study in mind together with the idea that at least 100 instances of each word would be required in order to have a representative data sample, it was decided that five words per language would be a suitable amount.

The choice of the words to be included in the study was based on a method which utilizes a vector space model. In this approach, a word's representation is the aggregate of its substrings' vector representations. Substrings are smaller sequences of characters within a word, and in vector space models, both words and substrings have their own semantic vector representations. Thus, the final vector of a word is derived by summing the vectors of all its substrings (Bojanowski et al. 2017). The method, which for the sake of simplicity will be referred to as the *distributional method*, provides a list of the distributionally most similar words to any given word. The distributional method is based upon the distributional hypothesis according to which words with similar distributional properties have similar meanings. It utilizes the distributional patterns of words, such as their neighboring words or grammatical environments, to represent semantic similarity between them (Sahlgren 2006, p. 17). The training data used for a model based on the distributional hypothesis is vital and might affect its outcome. Rich and diverse linguistic contexts in the data enable the model to discern nuanced semantic relationships. Incorporating subword information allows for a deeper understanding, especially in languages with complex morphology. Moreover, the specificity of contexts in the data impacts the model's ability to identify broad and nuanced semantic relationships between words. These models are typically trained on convenience samples from Wikipedia or similar repositories containing text from a variety of different websites. Because of the mixed nature of the data, this model can offer a broad, unrestricted capability but might also introduce noise and potential biases. The diverse contexts can assist in understanding polysemous terms, but the sheer variety of the data can also mean that the model might not be finely tuned for specific, niche tasks (Bojanowski et al. 2017; Jurafsky and Martin 2023).

The distributional method was chosen to ensure that the words included in this study were not selected arbitrarily. A potential option would have been to let word frequency serve as the basis for the choice of words, to choose the words which, in addition to belonging to the semantic domain of SADNESS, have the highest frequency in the corpus they are being retrieved from. However, doing so might have provided a very limited scope of the sadness domain as a whole. In order to come to results which as accurately as possible reflect the entirety of the emotional lexicon of each language, this method was not chosen.

There are, however, several potential problems with using the distributional method. One is that word distribution is unable to indicate the nature of the semantic similarity relations between words, and therefore does not distinguish between things such as synonyms, antonyms, and hyponyms (Sahlgren 2006, p. 24). Because of this, using homographs or polysemous words as search words might provide irrelevant or inaccurate results when concerned with a specific similarity relation. The method is also unable to distinguish between different word forms. This became evident in the current study where search results such as *saddening*, *sadder*, *saddened* and *saddest* for the search word *sad* had to be excluded.

The search words used for this study were *tråkig* 'sad' or 'boring' and *ledsen* 'sad' for

Swedish and *sad* for English. The reason why the word *tråkig* was chosen to form a basis for the examination of sadness expressions in this study is because of its polysemous nature, which will be detailed further in section 5.6.3. The words *ledsen* and *sad* were chosen because they are the adjectives that most prototypically relate to and represent the semantic domain of SADNESS in Swedish and English respectively. The distributionally most similar words to *tråkig* are *trist* 'sad' or 'dull' followed by *sorglig* 'sad' and the distributionally most similar word to *ledsen* is *sorgsen* 'sad' or 'sorrowful'. The distributionally most similar words to *sad*, aside from the previously mentioned word forms as well as a few other words which were excluded from the analysis due to low frequency in the corpus, are *depressing*, *tragic*, *unfortunate* and *unhappy*.

4.3.2 Data extraction

A search was conducted in the Sketch Engine parallel concordance tool for each of the English and Swedish adjectives. The concordance tool displays the search word in its context which consists of approximately one sentence. The query type 'lemma', which finds all word forms of the lemma contained within the search word was used along with 'adjective' as the part-of-speech search criteria. This search provides different word forms such as plural, comparative and superlative as well as the two Swedish grammatical genders common and neuter. When the noun that the property is ascribed to is neuter, adjectives take the neuter ending -t and when the noun has the common gender there is no inflection, as demonstrated in examples (8a) and (8b).

(8) a. Livet är ***sorgligt***. 'Life is sad.'

b. Du är för ***sorglig***. 'You're too sad.'

It is important to note that adverbs in Swedish are formed from adjectives with an ending that is either the same as, or happens to coincide with, the neuter -t (Hallonsten Halling 2018, p. 62). This means that the part-of-speech search criteria 'adjective' might not be able to differentiate between adjectives and adverbs in Swedish. This did not pose a problem for the current study, however, since none of the 500 Swedish sentences analyzed contained an adverb. For the English words, applying the part-of-speech search criteria 'adjective' made it so that the search provided comparative, such as *sadder*, and superlative, such as *saddest*, forms while filtering out nouns, such as *sadness*, and adverbs, such as *sadly*.

In the Sketch Engine parallel concordance tool, different 'text types' referring to the values, such as metadata, assigned to the data in the corpus can be chosen in order to limit the search. As mentioned previously, the text types were confined to the drama, crime, thriller, horror, sci-fi and mystery genres but no limitation was made regarding the release dates, original language of or languages spoken in the movies and TV shows. For each of the 10 adjective searches, a random sample of 100 concordance lines was generated using the random sampling tool provided in Sketch Engine. The total of 1000 sentences were then downloaded and transferred to Excel were the annotations and statistical tests were later on performed. A selection of this dataset can be found in Appendix A. Sources to every data example presented in this paper, such as the movie or TV show title, document ID or other metadata needed in order to locate the example in the corpus, will not be given since these details were not saved, and were not always available, when collecting the data sample. However, since the Sketch Engine sampling tool always produces exactly the same concordance lines as long as the search criteria and the sample number is the same, it would be possible for someone to find the same set of sentences that were used in this study.

4.4 Analysis

Firstly, each sentence retrieved from Sketch Engine was manually analyzed in order to infer whether its adjective was used attributively or predicatively and whether the subject of the clause in which the adjective appears had the semantic role of experiencer or stimulus. In this study, the subject was taken to be the word or phrase with which the verb in the clause agrees. Since some of the sentences containing an attributive adjective consisted solely of a noun phrase, these instances were annotated based on which semantic role their head noun more closely aligns with. Two examples of this are demonstrated in (9a) and (9b) where the head nouns *development* and *girl* are the parts to be interpreted as having the semantic role of either stimulus or experiencer.

(9) a. An ***unfortunate*** development.

b. A ***sad*** girl.

Henceforth, the term *semantic role orientation* will be used to refer to what semantic role the adjectives examined are more concurrent with. Likewise, the notion of predicative and attributive use is going to be referred to as *syntactic function*.

Instances of the third type of predicative adjective, described in section 2.4, were excluded from the semantic role analysis since they do not contain an explicit subject and therefore lack semantic roles. Very few examples of secondary predication, mentioned in section 2.4.1, were found in the data and it was therefore not something that was included in the presentation or analysis of the results. Instead, the occurrences of adjectives used in secondary predication were annotated simply as predicative. Whether or not the exclusion of these instances in some way affected or distorted the results will be discussed further on in section 6.3. While it is important to be cognizant of the notion of syntactic as opposed to semantic valence, as mentioned previously in section 2.6, it is not something that was explicated specifically for the data used in this study since it was not considered to be a deciding factor in fulfilling its particular aims.

After each of the sentences in the data was analyzed and annotated, the dictionary definitions as well as the bilingual dictionary translations of each word were examined in order to gain a comprehensive understanding of their meaning and in order to ascertain whether the translations given in the bilingual dictionary agree with the ones given in the parallel corpus data. The results of this analysis will be detailed in section 5.6.7.

A chi-square test of independence was conducted on the values of all of the 10 words combined in order to ascertain whether there was any significant association between syntactic function and semantic role orientation. It is important to note, however, that the chi-square test, since it is based on the combined values of all 10 words, is affected by those words and the particular values they happen to have. This means that the same test conducted on a different set of words might give entirely different results. Because of this, a Fisher's exact test was conducted for each of the 10 words independently. The decision of which statistical test to use on a certain dataset depends on the specific characteristics of the data, such as sample size and expected cell frequencies, as well as the requirements of the analysis. There are, however, some general guidelines which may help in making the decision. It is often suggested that the Fisher's exact test is particularly useful when the sample size is small, typically below 20 or 25, or when a substantial number of the expected cell counts are cells below five (McEnery et al. 2006, p. 56). While the total sample size of each adjective never fell short of 25, the majority of the words had either one or two expected cell counts below five. With this in mind, it was decided that the Fisher's exact test would be more appropriate for testing the significance of the relation between syntactic function and semantic role orientation for each adjective independently and the chi-square test of independence for the 10 words combined.

5 Results

5.1 Introduction

In this section, the results from the analysis are presented for each of the research questions separately. In section 5.2, the individual frequencies in the corpus of each of the words examined are presented. The results regarding the distribution of predicative and attributive use are presented in section 5.3 and this is followed by a presentation of the results regarding the distribution of semantic role orientation in section 5.4. Section 5.5 looks at the difference and relation between the observed distributions of syntactic function and semantic role orientation and provides a statistical analysis of the potential significance of the association between the two. Sections 5.6 through 5.6.7 deals with the different meanings, dicitonary definitions, bilingual dictionary translations, parallel corpus translations and frequencies of the adjectives examined.

5.2 Frequency

Table 1 below displays the frequency of each of the examined words in the corpus they were retrieved from. The table shows the overall frequency of the Swedish words in the OpenSubtitles 2018 parallel - Swedish corpus and the English words in the OpenSubtitles 2018 parallel - English corpus. Note that the frequencies in Table 1 represent those of the Swedish and the English corpus respectively as a whole and not their truly parallel portions between them which, as explained previously in section 4.2.2, has not been determined. This is why the frequencies of the English words are higher than those of the Swedish ones, with the entirety of the English corpus (consisting of approximately 1 billion words) being significantly larger than the Swedish one (consisting of approximately 100 million words).

Based on the concept put forward and investigated by Zipf (1945), among others, which states that the more frequently a word is used, the more potential meanings that word tends to have, one would expect the same principle to apply to the words in this study. This is examined further in section 5.6.6, keeping in mind that the amount of words, their frequency as well as the number of different meanings of the words examined in this study are of a considerably smaller scale. Therefore, the results of the current study regarding a potential meaning-frequency relationship of words might not be fully comparable to those of Zipf's study, but nevertheless something worth examining.

Table 1: The frequency of each word in the corpus

Word	Frequency
Ledsen	79.368
Tråkig	8.231
Sorglig	3.558
Trist	3.272
Sorgsen	583
Sad	94.181
Unhappy	18.775
Unfortunate	16.278
Tragic	14.979
Depressing	6.822

5.3 Predicative and attributive use

The results regarding the distribution of predicative and attributive use are compiled below in Table 2. For the annotation of syntactic function, each of the words had a total of 100 instances. Regarding the Swedish adjectives, the results indicate that *ledsen* is exclusively used predicatively, whereas *sorgsen*, *tråkig* and *trist* are predominantly used as predicative adjectives. *Sorgsen* was used predicatively 85 times out of 100, *tråkig* 84 times and trist 81 times. *Sorglig* had a more balanced frequency of usage with 51 instances of predicative use and 49 instances of attributive use.

Table 2: The distribution of predicative and attributive use for each adjective

Word	Predicative	Attributive
Ledsen	100%	0%
Sorgsen	85%	15%
Tråkig	84%	16%
Trist	81%	19%
Sorglig	51%	49%
Depressing	82%	18%
Unhappy	77%	23%
Sad	76%	24%
Unfortunate	44%	56%
Tragic	30%	70%

Examples of each of the Swedish adjectives used predicatively are given below in (10a) to (10e) and examples of the same adjectives used attributively (except for *ledsen*) are given in (11a) to (11d). Each Swedish sentence is accompanied by its English translation from the parallel corpus data. With the scope, time frame and aims of the study in mind, it was decided that giving the examples in the form of interlinear morpheme-by-morpheme glosses would be too time consuming and not entirely necessary. Additionally, since most of the adjectives have five or more translations, with both *tråkig* and *trist* being translated to both 'sad' and 'boring', it would be difficult to maintain consistency between the English words used in the glosses and the translations given in the data. Another thing that could have been cause for confusion in a glossed example is that the translations given in the parallel corpus data rarely cohere completely with the source sentence with, for instance, words frequently being added or left out. Therefore, each example simply consists of the sentence containing the word in question followed by its parallel corpus data translation.

(10) a. Jag är ***ledsen***. 'I'm sorry.'

b. Du har varit så ***sorgsen***. 'You've been so sad lately.'

c. Det var ***tråkigt***. 'That's really too bad.'

d. Att vara kung kan vara så ***trist***. 'Being king can be so dull.'

e. Ja, men den är ***sorglig***. 'Yes, it is, but it's rather sad.'

(11) a. Hon var en ***sorgsen*** flicka. 'She was kind of a sad little girl.'

b. Det var en väldigt kort och ***tråkig*** resa. 'It was a very short and boring trip.'

c. Jag gjorde en ***trist*** upptäckt idag. 'I made an unfortunate discovery today.'

d. En ***sorglig*** historia hur man än ser på det. 'A sad story by anyone's estimations.'

As for the English adjectives, the data shows that *depressing*, *unhappy* and *sad* are primarily used as predicative adjectives while *unfortunate* and *tragic* are most frequently used as attributive adjectives. *Depressing* was used predicatively 82 times out of 100, *unhappy* 77 times, *sad* 76 times, *unfortunate* 44 times and *tragic* 30 times.

Examples of each of the English adjectives used predicatively are given below in (12a) to (12e) and examples of the same adjectives used attributively are given in (13a) to (13e). Each English sentence is accompanied by its Swedish translation given in the parallel corpus data.

(12) a. God, this is ***depressing***. 'Gud, så deprimerande.'

b. I can't stand being so ***unhappy***. 'Jag är så ledsen.'

c. I didn't think it would be that ***sad***. 'Jag trodde inte att den skulle bli så ledsam.'

d. It's very ***unfortunate***. 'Det är väldigt beklagligt.'

e. It's ***tragic*** but it's true. Det är tragiskt, men så är det ju.'

(13) a. You kept singing that ***depressing*** song. 'Du sjöng den där sorgsna sången.'

b. Only ***unhappy*** people abhor dancing. 'Bara olyckliga hatar att dansa.'

c. They bring ***sad*** news. 'De kommer med sorgliga nyheter.'

d. Our first encounter was under such ***unfortunate*** circumstances. 'Vi träffades under olyckliga omständigheter.'

e. A ***tragic*** twist of fate. 'Ett tragiskt öde.'

Overall, the adjectives were used predicatively in 710 out of 1000 (71%) instances and attributively in 290 (29%) instances. If we take into account the frequencies of each word (displayed above in Table 1), predicative use remains more frequent. The total frequency of the adjectives predominantly used predicatively (by 50% or higher), that is, *ledsen*, *sorgsen*, *tråkig*, *trist*, *sorglig*, *depressing*, *unhappy* and *sad*, is 218.348. The total frequency of the words most commonly used attributively, i.e. *unfortunate* and *tragic*, is 31.257.

5.4 Semantic roles

The results regarding the distribution of semantic role orientation are compiled below in Table 3. Due to several sentences being exluded from the annotation of semantic role orientation as explained previously in section 4.4, the total number of instances varied for each word. The total number of instances annotated for semantic role orientation was 623. For the Swedish adjectives, the data showed that *ledsen* exclusivelt had the semantic role orientation of experiencer while *trist* only had that of stimulus. *Sorgsen* was more oriented towards the role of experiencer with 84 instances out of 96 (87.5%) being experiencer-oriented. *Sorglig* and *tråkig* were more oriented towards the role of stimulus with *sorglig* having an orientation towards stimulus 83 times out of 91 (91.2%) and *tråkig* 77 times out of 89 (86.5%).

Table 3: The distribution of semantic role orientation for each adjective

Word	Experiencer	Stimulus
Ledsen	100%	0%
Sorgsen	87.5%	12.5%
Tråkig	13.5%	86.5%
Sorglig	8.8%	91.2%
Trist	0%	100%
Unhappy	78%	22%
Sad	57.6%	42.4%
Unfortunate	12.8%	87.2%
Depressing	0%	100%
Tragic	0%	100%

In (14a) to (14d) below, examples are given of the words *ledsen*, *sorgsen* and *tråkig* occurring together with either a subject or head noun having the role of experiencer. Examples (15a) to (15d) demonstrate the Swedish words being used along with the semantic role of stimulus.

(14) a. Jag är ***ledsen***. 'I'm sorry.'

b. Jag är väldigt ***sorgsen***. 'I'm really sad.'

c. Jag hade ***tråkigt***. 'I was bored.'

d. Din ***sorgliga***, ömkliga lus. 'You sad, miserable louse.'

(15) a. Regniga dagar gör mig alltid ***sorgsen***. 'Rainy days always make me sad.'

b. Ja, det är för jädra ***tråkigt***. 'Yes, it's too damn boring.'

c. Visst är det ***sorgligt***? 'Isn't that sad?'

d. Engagemang är ***trist***. 'Commitment is dreary.'

When it comes to the English adjectives, *depressing* and *tragic* were entirely oriented towards the role of stimulus. The majority of the subjects for *unfortunate* had the role of stimulus with 82 out of 94 (87.2%). *Unhappy* was more oriented towards the role of experiencer with 78 instances out of 100 (78%). *Sad* had a slightly more balanced frequency with 53 instances out of 92 (57.6%) being experiencer-oriented and 39 (42.4%) being stimulus-oriented.

In (16a) to (16c), examples of the words *unhappy*, *sad* and *unfortunate* occuring alongside with a subject having the role of experiencer are demonstrated. In (17a) to (17e), the same is presented for each of the English adjectives with regards to the semantic role of stimulus.

(16) a. I can't stand being so ***unhappy***. 'Jag är så ledsen.'

b. You have to stop being so ***sad*** about this. 'Du måste sluta vara så ledsen över detta.'

c. Who were those ***unfortunate*** people? 'Vad var det för olyckliga människor?'

(17) a. I'm making everyone ***unhappy***. 'Jag gör alla olyckliga.'

b. Either way it's very ***sad***. 'Hur som helst är det väldigt sorgligt.'

c. This is an ***unfortunate*** loss to both the police department and the entire country. 'Det är en stor förlust för polisstyrkan och landet.'

d. There's nothing so ***depressing*** as a winning streak that won't stop streaking. 'Det finns inget som är så deprimerande som ett turserie som aldrig tar slut.'

e. Mr. Moody, you are a ***tragic*** disappointment. 'Du är en tragisk besvikelse, Mr. Moody.'

Overall, the total number of instances oriented towards the semantic role of stimulus was 584 out of 928 (62.93%) and the number of instances oriented towards the semantic role of experiencer was 344 (37.07%). However, taking the frequencies of each word into account, experiencer-oriented adjectives are more frequent than the stimulus-oriented ones. The total frequency of the adjectives that are predominantly experiencer-oriented (by 50% or higher), namely, *ledsen*, *sorgsen*, *unhappy* and *sad* is 192.907. Meanwhile, the total frequency of the adjectives that are predominantly stimulus-oriented, *tråkig*, *sorglig*, *trist*, *tragic*, *depressing* and *unfortunate*, is 53.140.

5.5 The difference and relation between syntactic function and semantic role orientation

By comparing the range of values for syntactic function and semantic role orientation, one is able to get a basic sense of which factor is associated with more variability in the use of the words. For each adjective except for *sad*, the difference between its experiencer and stimulus-orientation was greater than the difference between its attributive and predicative use. By looking at Tables 2 and 3 one can see that the range of variation in the values across all of the words spans 52% for the Swedish words and 49% for the English words when it comes to their syntactic function. When it comes to semantic role orientation, the range spans 100% for the Swedish words and 78% for the English words, indicating that semantic role orientation has a greater influence on the differences in usage than syntactic function.

Table 4 below displays the combined occurrences of syntactic function and semantic role orientation. It shows that occurrences of adjectives being used attributively while having the semantic role orientation of experiencer were uncommon, making up 5.06% out of a total of 928 occurrences. The combination of predicative use and stimulus-orientation was the most common with 37.07%. The combination of predicative use and experiencer-orientation was the second to most common (32.01%) followed by that of attributive use together with stimulus-orientation (25.86%).

Table 4: A contingency table displaying the combined occurrences of syntactic function and semantic role orientation

	Predicative	Attributive	Total
Stimulus	344	240	584
Experiencer	297	47	344
Total	641	287	928

Figure 1 below provides a visualization of the syntactic use and semantic role orientation of each adjective examined in this study. It was generated using the Python code provided in Appendix B. The red color represents the Swedish words while blue represents the English ones. It also displays each word's total frequency in the corpus with larger dots and fonts representing higher frequency and smaller ones representing lower frequency.

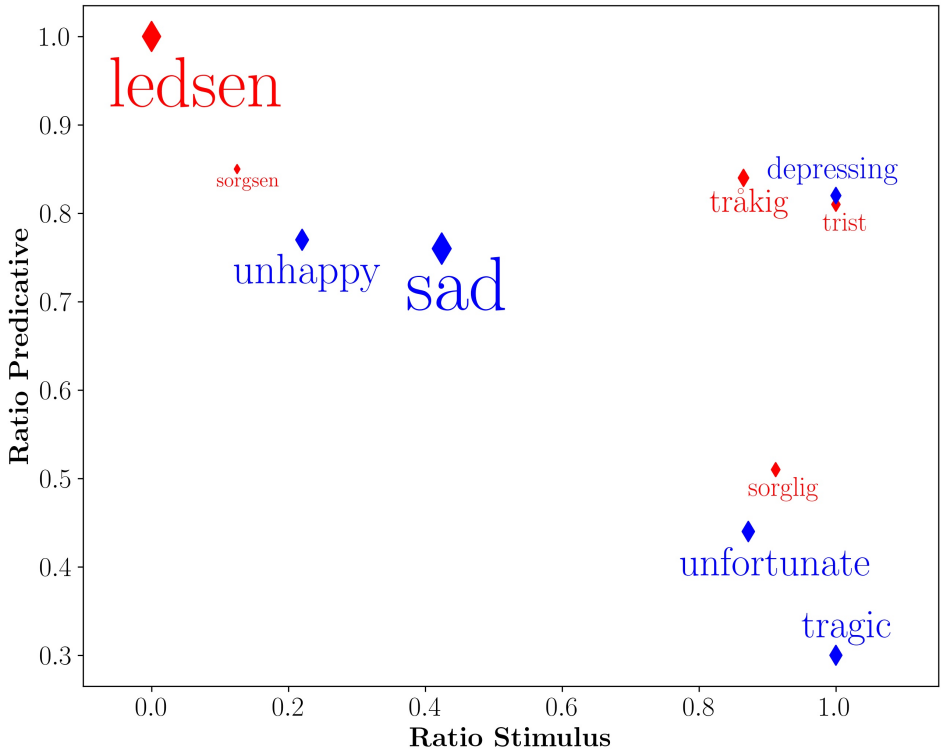

Figure 1: Visualization of syntactic use and semantic role orientation in combination with total frequency

Analyzing the frequency of the different combinations predicative–stimulus, predicative–experiencer, attributive–stimulus and attributive–experiencer might make it possible to identify potential patterns, rules and preferences in the usage of the words investigated. Finding out whether certain combinations are more prevalent than others, and if so, why, could provide evidence for the ways in which different semantic roles interact with syntactic structures. The ways in which an adjective is used syntactically and the semantic properties it possesses are not independent of each other; instead, they often interact in complex ways to shape meaning (Cabredo Hofherr and Matushansky 2010, p. 3). Understanding these types of relations could lead to insights into how emotional states are represented linguistically.

The significance level of 0.05 was selected for this study to balance the risk of Type I and Type II errors, in line with conventional statistical practices, and to determine the statistical significance of the observed results indicating that the probability of obtaining results as unlikely as, or more unlikely than, the observed ones by random chance alone is 5% or less. The chi-square test of independence, performed on the data displayed in Table 4 above, indicated that there is a significant association between syntactic function and semantic role orientation for the adjectives examined in this study, X^2 (1, N = 928) = 76.27, p < .001. The results of the Fisher's exact test for the words *sad* (p = 0), *sorglig* (p = 0.0067), *unfortunate* (p = 0.0011) and *sorgsen* (p = 0) individually indicate a significant association between syntactic function and

semantic role orientation. Meanwhile, the results of the Fisher's exact test for the words *tråkig* ($p = 0.114$), *unhappy* ($p = 0.149$), *ledsen* ($p = 1$), *trist* ($p = 1$), *tragic* ($p = 1$) and *depressing* ($p = 1$) do not.

5.6 Lexical comparison

The aim of this study is to examine whether there are any differences in how Swedish and English encode the concept of sadness and besides the potential syntactic and semantic differences, lexical ones might also be found. As Zhang (2014) demonstrates in their book investigating sadness expressions in English and Chinese, languages may vary in which emotions they lexicalize, how many different words there are to talk about an emotion or where the conceptual and practical line is drawn between similar emotions. In the following sections, the different meanings of each adjective, the different ways in which they might be used and their dictionary definitions will be explored. Additionally, the various translations of the words found in the parallel corpus data as well as the bilingual dictionary will be presented.

5.6.1 Ledsen

In the data used in this study, the word *ledsen* has been translated to *sad* only four times out of one hundred. In the English–Swedish dictionary, *ledsen* was translated to *sad*, *unhappy*, *distressed*, *worried*, *grieved*, *annoyed*, *disappointed*, *hurt* and *upset*, however, none of these translations, except for *sad*, were found in the current data sample. Besides the four instances of *ledsen* being translated to *sad*, *ledsen* was translated to *sorry* in all of the instances contained within the data sample. The word *sorry* can be used as an adjective describing someones feelings but in the sentences found in the current data it was exclusively used interjectionally as a conventional way of apologizing or expressing regret. The two meanings of the word *ledsen*, one being closely equivalent to *sad* and the other to *sorry*, are reflected in the entries given in SO:

1. [Someone] who feels depressed (usually because of some adversity).
2. [Someone] who apologizes or admits that one has committed an error.

When using the parallel concordance search tool, *sorry* was translated to *ledsen* (45.019 times) as well as *förlåt* (44.415 times) and *ursäkta* (13.786 times). *Förlåt* is an interjection closely equivalent to *sorry* and *ursäkta* can be used either as a transitive verb, to excuse, pardon or forgive someone for something, as a reflexive verb, to excuse oneself, or as an interjection similar to *förlåt*. The phrases in (18a) to (18c) below show *ledsen*, *förlåt* and *ursäkta* used as interjections and each of them were given in the corpus as translations of the English phrase *I'm sorry*.

(18) a. Jag är ***ledsen***. 'I'm sorry.'
b. ***Förlåt***. 'I'm sorry.'
c. ***Ursäkta*** mig. 'I'm sorry.' (alternatively: 'Excuse me.')

For the results presented in tables 2 and 3 above, *ledsen* stands out in that it is the only word that is used exclusively predicatively as well as the only one that is oriented solely towards the semantic role of experiencer. The word *ledsen* can, however, be used attributively. One can, for instance, talk about *en ledsen hundvalp* 'a sad puppy', *en ledsen smiley* 'a sad smiley' or *en ledsen min* 'a sad face'. *Ledsen* is typically only used to denote either an animate being, such as a puppy, or something that has the ability to convey subjective and internal emotions, such as

a smiley or a facial expression, but is rarely used to refer to inanimate objects or events. This might be why *ledsen* was exclusively experiencer-oriented. One may deduce that the reason why *ledsen* was used as an equivalent to *sorry*, in accordance with its second definition presented above, 96 times and as the equivalent to *sad*, as described in its first definition, lies in the discursive and pragmatic nature of the movie and TV subtitle-data. However, it might also be due to *sad* having a total frequency of 94.181 in the corpus while *sorry* has that of 1.521.669.

5.6.2 Sad

Table 5 below presents the frequencies of the translations found in the data for *sad*. It shows that *sad* is most often translated to *ledsen* but also to *sorglig*, *sorgsen*, *tråkig* and *trist* as well as *beklagligt* 'regrettable', *mörkt* 'dark', *tragisk* 'tragic', *dyster* 'gloomy', *olycklig* 'unhappy', *patetiskt* 'pathetic' and *synd* 'shame' (as in *vad synd* 'what a shame'). Seven out of the 12 translations found in the data were given also in the English-Swedish dictionary with the five words not included being *tragisk*, *synd*, *olycklig*, *beklagligt* and *patetiskt*.

Table 5: The frequency of the Swedish translations of *sad* found in the data

Translation of *sad*	Frequency
Ledsen	57
Sorglig	24
Sorgsen	6
Tråkig	3
Tragisk	2
Synd	2
Olycklig	1
Trist	1
Beklagligt	1
Mörkt	1
Dyster	1
Patetiskt	1

The definitions for *sad* given in Collins Dictionary are:

1. If you are sad, you feel unhappy, usually because something has happened that you do not like.
2. Sad stories and sad news make you feel sad.
3. A sad event or situation is unfortunate or undesirable.
4. If you describe someone as sad, you do not have any respect for them and think their behavior or ideas are ridiculous.

The word *sad* in the definitions above does not possess four different meanings in the way that *tråkig* (as explained further in section 5.6.3 below) can mean both *sad* and *boring* or in the way that the word *blue* can refer to a color as well as an emotional state. Rather, they illustrate the different grammatical patterns that the word can have and what type of things the word can

be used to refer to. In definition (1) above, the word is used predicatively and is referring to an abstract phenomenon, specifically a feeling. In definition (4), the word refers to a person and would typically be followed by a noun, i.e., be used attributively. The definitions concur with the data in which *sad* was used both predicatively and attributively. As Table 2 above shows, *sad* was used predicatively, in constructions similar to definition (1) and the latter part of definition (2), in 76 out of 100 instances and was used attributively, in sentences similar to the first half of definition (2) as well as definitions (3) and (4), in 24 instances. The definitions also display that *sad* can be used both in sentences where the subject or the head noun has the semantic role of stimulus, as in definition (3), as well as experiencer, as in definition (1). In the data used for this study, *sad* had the semantic role orientation of experiencer 53 times out of 92 (57.6%) and that of stimulus 39 times (42.4%) as can be seen in Table 3.

5.6.3 Tråkig

In the data, the word *tråkig* has been translated to both *sad*, as in (19a), and *boring*, as in (19b), meaning that the word can belong to more than one semantic domain. Similar to *ledsen*, *tråkig* can be used as a way to politely express sympathy or convey that you're feeling sad or sorry about something, as in (19c) where *tråkigt* has been translated to *too bad*. Example (19d) shows that *tråkig* can also be used as a translation of the word *sorry* used as a descriptive adjective.

(19) a. Jag har ***tråkiga*** nyheter. 'I have sad news.' (alternatively: 'I have bad news.')

b. En ***tråkig*** film. 'A boring movie.'

c. Det var ***tråkigt*** att du inte kunde komma. 'It was too bad that you couldn't come.' (alternatively: 'It was a shame that you couldn't come.' or 'I'm sorry that you couldn't come.')

d. Ja, det var ju en ***tråkig*** historia det här, för alla inblandade. 'Well, this is a rather sorry tale for everyone involved, you might say.'

The polysemous nature of the word *tråkig* was reflected also in its dictionary definitions:

1. [Something] which fails to arouse any joy or interest (due to monotony, long-windedness or the like; about person, object or abstract phenomenon).

2. [Something] which is sad or regrettable.

Besides *sad* and *boring*, *tråkig* was also translated to *tedious*, *drab*, *dreary*, *bored*, *bore*, *regrettable*, *unfortunate*, *dull*, *plain bad*, *too bad*, *shame* (as in *what a shame*) and *sorry*. All of the translations found in the data except for *plain* and *regrettable* are also given in the English–Swedish dictionary. Table 6 below displays the frequencies of the translations found in the data for *tråkig*.

Table 6: The frequency of the Swedish translations of *tråkig* found in the data

Translation of *tråkig*	Frequency
Boring	43
Sad	11
Bore	10
Too bad	8
Bored	7
Sorry	6
Bad	2
Unfortunate	2
Dreary	2
Tedious	2
Dull	2
Drab	1
Dry	1
Plain	1
Regrettable	1
Shame	1

Since the word *tråkig* showcases a difference in how Swedish and English encode the concept of sadness it makes for a particularly interesting object of study. This will be discussed further in section 5.6.7.

5.6.4 *Unhappy*, *unfortunate*, *tragic* and *depressing*

In the majority of the sentences, *unhappy* has been translated to *olycklig* which, according to the English–Swedish dictionary consulted in this study, is the closest translation of the word. Besides this, *unhappy* has also been translated to *missnöjd* 'dissatisfied' and, in a few instances, *ledsen*. Each of the translations found in the data were also given in the dictionary. Three definitions were specified for *unhappy* in Collins Dictionary:

1. If you are unhappy, you are sad and depressed.
2. If you are unhappy about something, you are not pleased about it or not satisfied with it.
3. An unhappy situation or choice is not satisfactory or desirable.

The word *unfortunate* was most often translated to *olycklig* which, as with *unhappy*, is the most fitting translation of the word according to the dictionary used in this study. *Unfortunate* was also translated to *tråkig*, *sorglig*, *beklaglig* 'regrettable', *otrevlig* 'unpleasant', *oturlig* 'unlucky', *stackars* 'poor' and *synd* 'shame' none of which, except for *beklaglig*, were given in the English–Swedish dictionary. Three definitions were given for *unfortunate* in Collins Dictionary:

1. If you describe someone as unfortunate, you mean that something unpleasant or unlucky has happened to them.
2. If you describe something that has happened as unfortunate, you think that it is inappropriate, embarrassing, awkward or wrong.

3. You can describe someone as unfortunate when they are poor or have a difficult life.

In the data used for this study, *tragic* was most often translated to *tragisk*, which is the only translation given in the English–Swedish dictionary, and a few times to *sorglig*. Only one definition is given of *tragic* in Collins Dictionary:

1. A tragic event or situation is extremely sad, usually because it involves death or suffering.

In the majority of the sentences, *depressing* was translated to *deprimerande* and, in a few, to *nedslående* 'disheartening', *dyster* 'gloomy' and *trist* 'sad' or 'dull', all of which, except for *trist*, were also given in the English–Swedish dictionary. *Depressing* has only one definition in Collins Dictionary:

1. Something that is depressing makes you feel sad and disappointed.

The definition already indicates that *depressing* is a word typically or entirely used to describe something having the role of stimulus rather than experiencer as the passive or patentive form *depressed* would be more likely to. Whether the exlusion of the word *depressed* might have had any affects on the results of this study will be discussed further in section 6.3.

5.6.5 *Trist*, *sorglig* and *sorgsen*

The definitions of *trist* are very similar to those of *tråkig*, as evidenced by its entries in SO:

1. [Something] which fails to arouse any joy or interest (due to monotony, mediocrity, grayness or the like; about person, object or abstract phenomenon).

2. [Something] which makes a disheartening, disappointing or unpleasant impression.

The only notable difference between the two, in regards to their dictionary definitions, is that *tråkig* is described as something which 'is sad or regrettable' while *trist* is describes as something which 'makes a disheartening, disappointing or unpleasant impression'. The similarity between *tråkig* and *trist* was also somewhat indicated in their respective translations. Both words were translated to *boring*, *sad*, *dull*, *dreary*, *drab*, *unfortunate*, *too bad* (as in *that's too bad*) and *shame* (as in *what a shame*). Unlike *tråkig*, however, *trist* was also translated to *disappointing* and *tiresome*. Out of the various translations of *trist* found in the corpus data, only *sad* and *dreary* were presented also in the bilingual dictionary. Unlike *trist* the word *tråkig* was, as displayed previously in Table 6, translated to *bore*, *bored*, *bad*, *tedious* and *regrettable*.

The word *sorglig* has only one entry in SO:

1. [Something] which causes (great and painful) depression or dejection.

In the current data sample, *sorglig* was translated to *sad* in almost all of the sentences with only a few exceptions including *tragic*, *deplorable*, *depressing*, *pitiful*, *pathetic* and *sorry*. In the dictionary, all of the translations just mentioned apart from *pathetic*, *depressing* and *sorry* were included. When translated from *sorglig*, *sorry* was used not interjectionally, as when translated from *ledsen*, but as a descriptive adjective, as demonstrated in (20) below.

(20) All right, listen up, you ***sorry*** bunch of leathernecks! 'Hör upp era sorgliga stackare!'

Similar to *sorglig*, the word *sorgsen* has only one entry in SO:

1. [Someone] who feels sadness, sorrow or grief.

In the data, *sorgsen* was most often translated to *sad* but also, in only one instance each, to *grievey*, *soulful*, *plaintive*, *heartsick*, *sorrowful*, *bereft* and *grief stricken*. The translations provided in the dictionary that did not appear in the data were *sorrow-stricken*, *grieved*, *melancholy*, *mournful*, *woeful* and *rueful*.

5.6.6 The meaning-frequency relationship of words

As mentioned previously, while the amount of words, their frequency and the number of different meanings of the words examined in this study are of a relatively small scale, it might be interesting to consider whether Zipf's (1945) meaning-frequency relationship theory applies to them as well. Table 7 presents the total frequencies of each word in the corpus together with the number of definitions that each of them have in the dictionary. The results presented in this table indicate that the theory stands, especially in regards to the English words.

Table 7: The frequency and number of dictionary entries of each word

Word	Frequency	No. of dictionary entries
Ledsen	79.368	2
Tråkig	8.231	2
Sorglig	3.558	1
Trist	3.272	2
Sorgsen	583	1
Sad	94.181	4
Unhappy	18.775	3
Unfortunate	16.278	3
Tragic	14.979	1
Depressing	6.822	1

The word with the highest frequency overall, *sad*, has four dictionary entries in Collins Dictionary and the word with the second to highest frequency overall, *ledsen*, has two in SO. It is important to note that the number of dictionary definitions for each word might vary depending on which dictionary is being consulted. In order to come to more definitive conclusions, one would need to look at and compare the entries of several different ones. It could be argued, however, that although *sad* has four definitions these are not as distinct, or do not define different meanings, in the way that the definitions of *ledsen* does. The two definitions of *ledsen* show that the word can belong to more than one word class and can be used for two entirely different purposes. The meanings of *sad* do not differ as much and it is not definitive whether the second and third definitions ("Sad stories and sad news make you feel sad." and "A sad event or situation is unfortunate or undesirable.") actually describe two different meanings or whether they simply describe different aspects or nuances of the word.

5.6.7 Key observations from the lexical comparison

This section highlights some particurarly salient issues regarding the lexical comparison including the complexities of lexical categorization across the two languages and the potential discrepancies between dictionary definitions and real-world language usage, emphasizing the need for supplementary sources like movie and TV subtitles to capture the full range of lexical variability.

The Swedish word *tråkig*, which can be translated to English as both *sad* and *boring*, showcases a difference in how Swedish and English encode the concept of sadness. In English, *sad* and *boring* are distinct adjectives, each representing a separate emotional state. In Swedish, on the other hand, these two concepts are encompassed by the single term *tråkig*, suggesting something that could be described as a type of shared semantic space or an overlap between two. This polysemy suggests that native speakers of Swedish may categorize or understand these emotional states differently, perhaps considering them as more similar or interrelated than they are in English. This difference in lexical categorization offers a clear example of how the encoding of emotional concepts like such as sadness or boredom can differ across languages. The lexical peculiarity of the Swedish word *tråkig* encompassing both sad and boring in its meaning draws parallels to Wierzbicka's observation of the distinct emotional categorizations in Russian. Just as Russian words *grust* and *pečal'* offer shades of sadness not entirely captured by the English term "sadness", *tråkig* presents a distinctive blend of two emotional states.

The word *sorry* is another example that shows how the demarcation of the semantic domain of SADNESS might be difficult to pinpoint. While the word often insinuates a sense of sadness it is not specifically about sadness but rather about acknowledging one's own actions and the effects that those actions might have on others. While *sorry* certainly has some relation to the notion of sadness, one could interpret the word as fitting more accurately into the domains of REGRET, REPENT, APOLOGY, HONORIFICS or something of the like.

Regarding whether the translations given in the bilingual dictionary agrees with the ones given in the parallel corpus data, the examination presented above did reveal somewhat of a discrepancy between the two. Among all of the translations found in the corpus data for each of the ten adjectives there were a total of 28 words that were not represented in the bilingual dictionary. This finding suggests that while dictionaries provide a valuable and robust foundation for understanding language translation, they may not entirely encompass the breadth of lexical variability seen in natural language use as showcased in movies and TV shows. Consequently, this highlights the importance of leveraging real-world data, such as the movie and TV subtitle corpora used in the current study, in supplementing dictionary-based approaches to research in lexical typology or, perhaps, language in general.

6 Discussion

6.1 Introduction

This section begins with a discussion of the results of the study. Section 6.2.1 consists of an interpretation of the observed distribution of syntactic function, section 6.2.2 deals with the observed distribution of semantic role orientation and section 6.2.3 examines lexical differences in terms of word origin and its potential influence on semantic role orientation. Section 6.3 discusses the methodological aspects of the study and, finally, section 6.4 provides some suggestions for future research.

6.2 Discussion of results

6.2.1 Syntactic function

One semantic factor that seems to be crucial when it comes to the division of adjectives into predicative and attributive is that of time-stability. The notion of time-stability, originally put forward by Givón (2001), refers to the rate of change that something has over time and is, together with complexity, concreteness and spatial compactness, one of the four most classificatory or generic semantic features of our conceptual lexicon. On Givón's scale of time-stability, which has been employed as an explanatory model in many typological accounts of parts of speech, nouns tend to occupy the most time-stable end while verbs tend to occupy the least time-stable end (Hallonsten Halling 2018, p. 185). This is because the properties of prototypical nouns tend to change only little or not at all over time while those of verbs tend to be more transitory with verbs such as *break* or *jump* encoding rapid changes or ones such as *read*, *walk*, *dream* or *know* encoding longer and more enduring, albeit not permanent, states. Adjectives, on the other hand, typically fall somewhere towards the middle of the scale depending on what type of properties they are denoting. That is, an adjective used to describe an object, such as the adjective *red* being used to describe the color of a chair, would fall closer towards the time-stable half of the scale where the property it is being used to describe, the chair, also falls. Adjectives describing emotions or feelings, such as any of the ones investigated in the current study, encode temporary, non-physical states that are mental or evaluative rather than physical which means that they would fit in better closer towards the least time-stable end of the scale than the ones describing objects.

In their investigation of the treatment of property concepts in languages with small and closed classes of adjectives, Stassen (1997) examined the semantic and contextual factors that determine or affect whether the property concepts are used in a verbal (e.g., predicative) or nominal (e.g., attributive) form. Stassen provides a characterization of the seven semantic types of adjectives originally identified by Dixon (1982) in terms of their underlying time-stability with AGE, DIMENSION, VALUE and COLOR representing more durable or time-stable properties and HUMAN PROPENSITY, PHYSICAL PROPERTY and SPEED representing more fleeting or time-unstable concepts. In accordance with Givón's notion of nouns being more time-stable and verbs less so, Stassen found that some languages tend to treat the more time-stable semantic types as nominal elements and the more time-unstable semantic types as verbal elements.

In order to find out whether there is any correspondence between the semantic notion of time-stability and the use of adjectives in either predicate or noun-modifying form, Saylor (2000) evaluated speech samples from natural interactions between three English-speaking children and their caretakers. The study revealed that the speakers appeared to be guided by the underlying conceptual time-stability of the concept encoded in the adjectives when deciding whether to use

it attributively as part of a noun phrase (e.g., "The big house") or predicatively as part of an adjectival predicate (e.g., "The house is big"). Similar to Stassen, Saylor found a tendency for more time-stable meanings encoded in AGE, DIMENSION and COLOR terms to be used in noun-modifying or attributive constructions and more time-unstable meanings encoded in PHYSICAL PROPERTY and HUMAN PROPENSITY terms to be used in predicative constructions.

The results of the current study can also be related to, and contextualized within, the existing knowledge of a relation between time-stability and the constructions of adjectives. Generally, attributive adjectives are considerably more frequently than predicative ones, both in spoken and written language (Croft 1991, p. 122). One might therefore predict that the adjectives included in the current study would be used attributively in the majority of the instances. Nevertheless, the results show that predicative use was far more common than attributive regardless of the frequency of each word in the corpus. Perhaps the reason for the results contradicting the expectations is that the adjectives examined belong to the semantic type of HUMAN PROPENSITY which, as shown by Stassen (1997) and Saylor (2000), most often represent less time-stable notions and, consequently, tend to be used in predicative constructions. It is important to note, however, that if the adjective is used to refer to an inanimate object it no longer falls under the semantic type of HUMAN PROPENSITY and can no longer be considered as time-unstable. In a sentence such as "That book is totally depressing", although the opinion that the book is depressing is subjective and might change over time, the book itself and whatever qualities or contents it has are very, if not entirely, time-stable.

When placing the findings on the usage of sadness-related terms in Swedish and English alongside Wierzbicka's (1999) comprehensive analysis of emotional terms in English and Russian, several parallels become evident. A key observation from the current study is the predominance of predicative usage of adjectives related to sadness, suggesting that in both Swedish and English, sadness is commonly depicted as a state or outcome. This resonates closely with Wierzbicka's cognitive scenario, where "sadness" arises as an emotional response to an undesirable event. For instance, the frequency with which adjectives like *ledsen* and *sad* are used predicatively underscores this pattern, reflecting the internal processing of a negative situation as outlined in Wierzbicka's cognitive scenario.

6.2.2 Semantic role orientation

As stated earlier, orientation towards the role of stimulus was predominant for the sample examined in this study but when taking the overall frequency of each adjective into account the adjectives that were predominantly experiencer-oriented in this study's sample (by 50% or higher) had a significantly greater frequency than those that were stimulus-oriented. This discrepancy could be attributed to several factors. It may be indicative of the inherent variability and complexity of natural language use, where context-specific factors significantly influence lexical choice. The higher prevalence of stimulus-oriented adjectives in the sampled data could be a result of the specific sentences chosen for analysis consisting of movie and TV subtitles which represent spoken discourse and are typically meant to reflect spontaneous conversations, potentially skewing towards contexts where sadness is provoked by external factors. It is also worth considering that the overall frequency might be influenced by a few highly frequent experiencer-oriented adjectives that are used across many different contexts, while the stimulus-oriented adjectives might be more varied but each less frequent individually. Consequently, experiencer-oriented adjectives might have a higher frequency overall even if stimulus-oriented adjectives are more prevalent in certain contexts. In order to ascertain whether these somewhat contrary results are due to the nature of the data sample or something else it might be necessary to conduct a similar

analysis on a sample derived from other contexts.

As mentioned in section 5.5, by simply comparing the range of values for syntactic function and semantic role orientation it seems that, overall, the difference between the adjectives semantic role orientations were greater than the difference between their attributive and predicative use. That is, it appears that semantic role orientation may contribute more significantly than syntactic function to the differentiation in the usage of the words. This may be due to some inherent flexibility of these words to be used alongside with either experiencer or stimulus, offering a wider range of expressive possibilities in communicating feelings of sadness. Due to established language usage patterns or grammatical rules, the words may be more constrained in their syntactic function which might be what leads to less differentiation in terms of attributive or predicative usage. The wider range observed in semantic role orientation than that of syntactic function might be a reflection of differing nuances and constraints within the language structure. It is important to note that this method of simply comparing the distribution of predicative and attributive use with that of semantic role orientation for the 10 adjectives examined in this study provides a rather simplistic measure and does not offer any insights into the statistical significance of these differences. Testing the significance of the range of variation between semantic role orientation and syntactic function would probably require more complex tests which would fall outside of the scope and time frame of this study and were therefore not conducted. This comparison does not account for the potential influences that word characteristics, context, and other possible linguistic factors can have on a word's usage within these categories.

Circling back to Wierzbicka's (1999) analysis of English and Russian,, the dataset of the current study reveals a tilt towards the semantic role of stimulus, aligning with Wierzbicka's assertion that sadness is often evoked by an external trigger or event. However, an intriguing divergence is observed when considering the overall word frequencies. When doing so, there is a more pronounced emphasis on the experiencer's internal emotional state, as demonstrated by the higher prevalence of experiencer-oriented adjectives like *ledsen* and *unhappy*. This subtlety echoes Wierzbicka's emphasis on the experiencer's internal recognition and processing of the adverse event.

6.2.3 Lexical differences: Etymology and semantic role orientation

Out of the 10 adjectives examined in this study, four of them are words that are derived from, or originate in, other languages. They are the following: *trist* (of French *triste*) (Hellquist 1922, p. 1007), *tragic* (modeled on Latin *tragicus*, from Greek *tragikos*), *unfortunate* (made up of *un-* (prefix of negation) *not* + *fortunate* (adj.) with *fortunate* coming from Latin *fortunatus*) and *depressing* (present-participle adjective from *depress* (v.) which itself comes from Old French *depresser* 'to press down, to lower', from Late Latin *depressare*) (Onions et al. 1966, pp. 935, 372, 258). Looking at Figure 1 presented previously in section 5.5, one can see that all four words appear to the far right in the plot, that is, all of them are strongly stimulus-oriented. In order to ascertain whether this pattern is random or significant, more data on a larger number of words would be required and this is something which might lend itself to future research. Perhaps, if not due to chance, the reason why words derived from other languages tend to be more stimulus-oriented is because the notions of stimulus and experiencer involve different degrees of objectivity. The stimulus typically constitutes an external, tangible force while the experiencer involves an animate entity that subjectively experiences, perceives or undergoes some emotion or state. Maybe there exists some subconscious tendency to associate words that are modeled on words from other languages with more external and concrete concepts such as those that typically have the role of stimulus.

6.3 Discussion of method

An important question to ask is whether any data excluded from the study might have significantly affected or distorted the results. In the current study, the only data points that were excluded from the analysis were predicative adjectives that consist of assessments, as described earlier in section 2.5, since they predicate a property of a referent that can only be discerned by being aware of the previous discourse or situational context. These types were excluded from the semantic role analysis since they do not contain an explicit subject and therefore lack semantic roles. No data entries were excluded from the analysis of predicative and attributive use. Overall, 72 sentences out of one thousand were excluded from one part of the analysis. The proportion of the excluded sentences together with the relatively low variability of the dataset suggest that the impact on the overall results might be minimal. Since the sentences are included and accounted for in one part of the analysis there is no risk of any potentially anomalous information held within them not being mirrored or represented elsewhere in the dataset.

Another questions worth asking is whether the use of parallel corpora, rather than two monolingual corpora, was necessary and whether any important information or particular aspects of the results would have gone lost without them. The use of parallel corpora made it possible to compare the translations given therein with those given in a bilingual dictionary. This, for instance, showed how out of the 10 translations of *ledsen* found in the dictionary, only two were used in the parallel corpus translations. Because of the interactive nature of the movie and TV subtitle data, the translation of *ledsen* to *sorry* was the most prevalent. The word *sorry*, in turn, shows that the demarcation of the semantic domain of SADNESS might be difficult to pinpoint.

The choice of which genres to include and which not to was not based on any exact science but rather an educated guess or estimation considering the themes and content typically present in the different genres. In order to avoid any uncertainty or subjectivity when gathering the data sample, perhaps it would have been better to simply include all genres and, if it would have become necessary, simply exclude any sarcastic or ironic expressions.

Although the release dates and original languages of the movies and TV shows were not considered to be factors that would likely affect or considerably change the outcome of the analysis, taking them into account might have had several benefits. In order to ascertain whether original language and release date are decisive factors in determining the syntactic and semantic characteristics of emotion terms one would have to actually include them in the analysis. Perhaps this is something that would lend itself to future research in this area. By not taking these factors into account it is possible that some aspect of how English and Swedish encode the concept of sadness might have gone unnoticed.

Another potential shortcoming of this study concerns the words that were examined in it. Based on *depressing* being an active adjective typically causing depression and *depressed* a passive one typically referring to someone experiencing depression, it is not surprising that *depressing* turned out to be exclusively stimulus-oriented (Fox and Hopper 1994, p. 99). For more substantial and definitive results it would probably have been better to include the passive form of the adjective as well, in order to find out whether the semantic differences between the two are as uncomplicated as they may seem.

6.4 Future research

As mentioned previously in section 4.2.1, external and bodily signals such as facial expression and tone of voice can serve as a useful complement to the analysis of language, particurarly when it comes to investigating words relating to emotion. Although an examination of these factors was not conducted within the current study it is something that, if examined in future research, might reveal more important information about the semantic domain of SADNESS specifically and the language of emotions in general.

As mentioned previously in section 4.4, secondary predication not something that was included in the presentation or analysis of the results of the current study, however, it is something that may be investigated in detail in future studies.

Given the results of the current study it would be interesting to see what a similar study on a larger sample of languages from multiple different language families might bring. Perhaps a study on languages with greater cultural and linguistic differences than that of English and Swedish would yield different results. The distinction between predicative and attributive adjectives is a somewhat simplified, albeit important, one. For future research it might be useful to broaden the scope of potential syntactic differences. Besides their predicative and attributive functions one could, for example, examine the gradability and comparability of adjectives and whether they are restrictive or non-restrictive (Martin 2014; Rusiecki 1985). Another possible direction for future research would be to extend the investigation to other emotion-related semantic domains like HAPPINESS, ANGER, FEAR, etc. This would allow for a more comprehensive understanding of how different emotions are represented linguistically and how they are used across languages.

Perhaps the patterns found in the results of this study regarding etymology and semantic role orientation might be linked to code-switching, that is, the juxtaposition of passages of speech belonging to two different grammatical systems or sub-systems (in this case, words originated in other languages and those not), within the same exchange (Gumperz 1977). If it is the case that the notions of stimulus and experiencer involve different degrees of objectivity, perhaps this has something to do with alienation which involves a separation of the speaker from the self (Haiman 2001). While intriguing, this particular aspect was beyond the scope of the current study and could be the focus of future research.

The current study may be perceived as a pilot study laying the foundation for a larger lexical typological inquiry of a number of different languages into the potential relation between etymology and semantic role orientation, the interplay between semantics and syntax or the semantic domain of SADNESS in general.

7 Conclusions

The aim of this study has been to examine and determine some of the syntactic, semantic and lexical characteristics of adjectives relating to sadness in Swedish and English. In this section, a brief answer is given to each of the research questions.

1. Are the adjectives examined used more attributively or predicatively?

   Overall, the adjectives were used predicatively in 710 out of 1000 (71%) instances and attributively in 290 (29%) instances. If we take into account the frequencies of each word, displayed previously in Table 1, section 5.2, predicative use remains more frequent. The total frequency of the adjectives mostly used predicatively (by 50% or higher), that is, *ledsen*, *sorgsen*, *tråkig*, *trist*, *sorglig*, *depressing*, *unhappy* and *sad*, is 218.348. The total frequency of the words most commonly used attributively, i.e. *unfortunate* and *tragic*, is 31.257. While the concrete data suggests a predominance of predicative usage, what this reflects at a deeper level is the innate interplay between semantics and syntax. In lexical typology, where the emphasis is usually on semantics, this study reveals the indispensability of syntactic structures in forming the full picture.

2. Are the adjectives more stimulus- or experiencer-oriented?

   Overall, the total number of instances oriented towards the semantic role of stimulus was 584 out of 928 (62.93%) and the number of instances oriented towards the semantic role of experiencer was 344 (37.07%). However, if one takes the frequencies of each word into account, experiencer-oriented adjectives are more frequent than the stimulus-oriented ones. The total frequency of the adjectives that are predominantly experiencer-oriented (by 50% or higher), namely, *ledsen*, *sorgsen*, *unhappy* and *sad* is 192.907. Meanwhile, the total frequency of the adjectives that are predominantly stimulus-oriented, *tråkig*, *sorglig*, *trist*, *tragic*, *depressing* and *unfortunate*, is 53.140. This distinction between stimulus and experiencer orientation, rather than being a mere classification, touches upon the more profound implications of how languages position emotional agency. The results indicate a subtle balance in how languages express external events or stimuli and the internal emotional responses of experiencers. The collective weight of experiencer-oriented adjectives, as demonstrated by their higher total frequencies, underscores their importance in linguistic representation.

3. Are there any differences in how the closely related languages Swedish and English encode the concept of sadness?

   The Swedish word *tråkig*, which can be translated to English as both *sad* and *boring*, showcases a difference in how Swedish and English encode the concept of sadness. In English, *sad* and *boring* are distinct adjectives, each representing a separate emotional state. In Swedish, on the other hand, these two concepts are encompassed by the single term *tråkig*, suggesting something that could be described as some sort of shared semantic space or an overlap between two. This difference in lexical categorization offers a clear example of how the encoding of emotional concepts like such as sadness or boredom can differ across languages. The lexical distinction between *tråkig* in Swedish and its English counterparts underlines the nuanced ways in which closely related languages can frame emotions. It's not simply about there being different words for the same feelings; it's about how languages can merge or differentiate emotional concepts based on their intrinsic syntactic and semantic patterns, without necessarily being tied down by vast cultural differences.

References

Ahn, Dohyun, Seung A. Annie Jin and Ute Ritterfeld (2012). Sad movies don't always make me cry: The cognitive and affective processes underpinning enjoyment of tragedy. In: *Journal of Media Psychology: Theories, Methods, and Applications* 24.1, pp. 9–18. DOI: 10.1027/1864-1105/A000058.

Akmajian, Adrian, Richard A. Demer, Ann K. Farmer and Robert M. Harnish (2001). *Linguistics: An introduction to language and communication*. 5th ed. Cambridge: MIT Press.

Berlin, Brent and Paul Kay (1969). *Basic color terms: Their universality and evolution*. Oakland: University of California Press.

Bojanowski, Piotr, Edouard Grave, Armand Joulin and Tomas Mikolov (2017). Enriching word vectors with subword information. In: *Transactions of the Association for Computational Linguistics* 5, pp. 135–146. DOI: 10.48550/arXiv.1607.04606.

Brown, Cecil H. (1976). General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. In: *American Ethnologist* 3.3, pp. 400–424. DOI: 10.1525/AE.1976.3.3.02A00020.

Cabredo Hofherr, Patricia and Ora Matushansky, eds. (2010). *Adjectives: Formal analyses in syntax and semantics*. Vol. 153. Amsterdam: John Benjamins Publishing Company.

Collins Online Dictionary (n.d.). Glasgow: HarperCollins. URL: https://www.collinsdictionary.com/.

Crawley, Rosalind A. and David Kleinman (1994). Thematic roles, focus and the representation of events. In: *Language and Cognitive Processes* 9.4, pp. 519–548. DOI: 10.1080/01690969408402130.

Croft, William (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press. ISBN: 9780226120904.

Dixon, R. M. W. (1982). *Where have all the adjectives gone?* Berlin: De Gruyter Mouton. DOI: 10.1515/9783110822939.

Ekman, Paul (1992). Facial expressions of emotion: New findings, new questions. In: *Psychological Science* 3.1, pp. 34–38. DOI: 10.1111/J.1467-9280.1992.TB00253.X.

— (1999). Basic emotions. In: *Handbook of Cognition and Emotion*. Ed. by T Dalgleish and M Power. Sussex: John Wiley & Sons. URL: https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=649485.

Ekman, Paul, Wallace V. Friesen, Maureen O'Sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E. Ricci-Bitti, Klaus Scherer, Masatoshi Tomita and Athanase Tzavaras (1972). Universals and cultural differences in the judgments of facial expressions of emotion. In: *Journal of Personality and Social Psychology* 53.4, pp. 712–717. DOI: 10.1037/0022-3514.53.4.712.

Englebretson, Robert (1997). Genre and grammar: Predicative and attributive adjectives in spoken english. In: *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure* 23.1, pp. 411–421. DOI: 10.3765/BLS.V23I1.1272.

Fox, Barbara A. and Paul J. Hopper, eds. (1994). *Voice: Form and function*. Vol. 27. Typological Studies in Language. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/TSL.27.

Givón, Talmy (2001). *Syntax: An introduction*. Vol. 1. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/Z.SYN1.

Gumperz, John J. (1977). The sociolinguistic significance of conversational code-switching. In: *RELC Journal* 8.2, pp. 1–34. DOI: 10.1177/003368827700800201.

Haiman, John (2001). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford: Oxford University Press. ISBN: 0195354060.

Hallonsten Halling, Pernilla (2018). Adverbs: A typological study of a disputed category. PhD thesis. Stockholm University. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-153421.

Heider, Eleanor R. (1972). Universals in color naming and memory. In: *Journal of Experimental Psychology* 93.1, pp. 10–20. DOI: 10.1037/h0032606.

Hellquist, Elof (1922). *Svensk etymologisk ordbok*. 1st ed. Lund: C. W. K. Gleerups förlag. URL: http://runeberg.org/svetym/.

Himmelmann, Nikolaus P. and Eva F. Schultze-Berndt, eds. (2005). *Secondary predication and adverbial modification: The typology of depictives*. Oxford: Oxford University Press. DOI: 10.1093/ACPROF:OSO/9780199272266.001.0001.

Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray and Kristen A. Lindquist (2019). Emotion semantics show both cultural variation and universal structure. In: *Science* 366.6472, pp. 1517–1522. DOI: 10.1126/science.aaw8160.

Johnson-laird, Philip N. and Keith Oatley (1989). The language of emotions: An analysis of a semantic field. In: *Cognition and Emotion* 3.2. DOI: 10.1080/02699938908408075.

Jurafsky, Dan and James H. Martin (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 3rd ed. (n.p.) ISBN: 9780131227989.

Koptjevskaja-Tamm, Maria, ed. (2015). *The linguistics of temperature*. Vol. 107. Typological Studies in Language. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/TSL.107.

Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina and Martine Vanhove (2015). The semantics of lexical typology. In: *The Routledge handbook of semantics*. Ed. by Nick Riemer. London & New York: Routledge, pp. 434–454. DOI: 10.4324/9781315685533.

Lehrer, Adrienne and Eva Kittay, eds. (1992). *Frames, fields, and contrasts: New essays in semantic and lexical organization*. London & New York: Routledge. DOI: 10.4324/9780203062821.

Levshina, Natalia (2016). Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. In: *Folia Linguistica* 50.2, pp. 507–542. DOI: 10.1515/FLIN-2016-0019.

Lison, Pierre and Jörg Tiedemann (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 923–929. URL: https://aclanthology.org/L16-1147.

Martin, Fabienne (2014). Restrictive vs. nonrestrictive modification and evaluative predicates. In: *Lingua* 149, pp. 34–54. DOI: 10.1016/j.lingua.2014.05.002.

McEnery, Tony, Richard Xiao and Yukio Tono (2006). *Corpus-based language studies: An advanced resource book*. London & New York: Routledge.

Miller, George A. and Philip N. Johnson-Laird (1976). *Language and perception*. Cambridge: Harvard University Press. ISBN: 9780674421288.

Moradi, Mohammad Reza and Shahrzad Pirzad Mashak (2013). A comparative and contrastive study of sadness conceptualization in Persian and English. In: *English Linguistics Research* 2.1. DOI: 10.5430/ELR.V2N1P107.

NE:s stora engelska ordbok: Engelsk-svensk/svensk-engelsk (n.d.). Malmö: NE Nationalencyklopedin AB. URL: https://www.ne.se/.

Nerlove, Sara and Antone Kimball Romney (1967). Sibling terminology and cross-sex behavior. In: *American Anthropologist* 69.2, pp. 179–187. DOI: 10.1525/AA.1967.69.2.02A00050.
Onions, Charles T., Robert W. Burchfield and George W. S. Friedrichsen, eds. (1966). *The Oxford Dictionary of English Etymology*. Oxford: Oxford University Press. ISBN: 9780198611127.
Payne, Thomas E. (1997). *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511805066.
Power, Michael J. (2010). Sadness. In: *The Corsini Encyclopedia of Psychology*. Ed. by Irving B. Weiner and W. Edward Craighead. Vol. 4. Hoboken: John Wiley & Sons, pp. 1486–1488. DOI: 10.1002/9780470479216.corpsy0810.
Rosch, Eleanor H. (1973). Natural categories. In: *Cognitive Psychology* 4.3, pp. 328–350. DOI: 10.1016/0010-0285(73)90017-0.
Rusiecki, Jan (1985). *Adjectives and comparison in English: A semantic study*. New York: Longman. ISBN: 0582291291.
Sahlgren, Magnus (2006). The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. PhD thesis. Stockholm University. URL: http://www.diva-portal.org/smash/get/diva2:1041204/FULLTEXT01.pdf.
Saylor, Megan M. (2000). Time-stability and adjective use by child and adult English speakers. In: *First Language* 20.58, pp. 091–120. ISSN: 01427237. DOI: 10.1177/014272370002005804.
Stassen, Leon (1997). *Intransitive predication*. Oxford: Clarendon Press. ISBN: 0199258937.
Svensk ordbok utgiven av Svenska Akademien (SO) (n.d.). Stockholm: Svenska Akademien. URL: https://svenska.se/so/.
Talmy, Leonard (1985). Lexicalisation patterns: Semantic structure in lexical forms. In: *Language typology and syntactic description*. Ed. by Timothy Shopen. Cambridge: Cambridge University Press, pp. 33–149. ISBN: 9780521257008.
Tissari, Heli (2008). On the concept of sadness: Looking at words in contexts derived from corpora. In: *Corpus linguistics, computer tools, and applications - State of the art*. Ed. by Barbara Lewandowska-Tomaszczyk. Frankfurt: Peter Lang, pp. 291–308. ISBN: 9783631583111.
Van Valin Jr, Robert D. (2001). *An introduction to syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139164320.
Viberg, Åke (1984). The verbs of perception: A typological study. In: *Explanations for language universals*. Ed. by Brian Butterworth, Bernard Comrie and Östen Dahl. De Gruyter, pp. 123–162. DOI: 10.1515/9783110868555.
Wierzbicka, Anna (1999). *Emotions across languages and cultures*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511521256.
Zhang, Ruihua (2014). *Sadness expressions in English and Chinese: Corpus linguistic contrastive semantic analysis*. London: Bloomsbury. DOI: 10.5040/9781472593719.
Zipf, George Kingsley (1945). The meaning-frequency relationship of words. In: *Journal of General Psychology* 33.2, pp. 251–256. DOI: 10.1080/00221309.1945.10544509.

Appendix A A selection of the dataset

This appendix presents a selection from the dataset crafted for and utilized in this study. Each image showcases five sentences with their corresponding translations sourced from the parallel corpora, accompanied by annotations indicating syntactic function and semantic role orientation.

Swedish	English	Syntactic function	Semantic role orientation
<s> Jag är ledsen . </s>	<s> I'm sorry. </s>	PRED	EXP
<s> Jag är ledsen . </s>	<s> I'm sorry. </s>	PRED	EXP
<s> Jag är ledsen ... </s>	<s> I'm sorry... </s>	PRED	EXP
<s> Jag är ledsen . </s>	<s> I'm sorry. </s>	PRED	EXP
<s> Jag är ledsen , fru Nihal. </s>	<s> I'm sorry, Ms. Nihal. </s>	PRED	EXP

Swedish	English	Syntactic fu	Semantic ro
<s> Äh, vad du är tråkig . </s>	<s> You're so boring. </s>	PRED	STIM
<s> Det var tråkigt . </s>	<s> That's really too bad. </s>	PRED	STIM
<s> Ja, det är för jädra tråkigt . </s>	<s> Yes, it's too damn boring. </s>	PRED	STIM
<s> Men det var inte så tråkigt som du sa. </s>	<s> But life there was not as dull as you pictured it. </s>	PRED	STIM
<s> Nej, men för att du är rå, grov och tråkig . </s>	<s> No. </s><s> Because you're crude, rude and a bore. </s>	PRED	STIM

Swedish	English	Syntactic func	Semantic ro
<s> - Det blir trist på hotellet nu. </s>	<s> Things will be sad now at the hotel. </s>	PRED	STIM
<s> - Trist ? </s>	<s> - Boring? </s>	PRED	x
<s> Livet är trist utan sprit. </s>	<s> Life gets pretty boring when you stop drinking. </s>	PRED	STIM
<s> -Kvantitativ analys är trist . </s>	<s> Quantitative analysis is... very boring. </s>	PRED	STIM
<s> Använd de trista orden igen och jag går. </s>	<s> Use these words again and I shall leave the room! </s>	ATTR	STIM

English	Swedish	Syntactic funct	Semantic role
<s> He seemed sad . </s>	<s> Han verkade ledsen. </s>	PRED	EXP
<s> Either way it's very sad . </s>	<s> Hur som helst är det väldigt sorgligt. </s>	PRED	STIM
<s> So, listen, Taylor's kind of sad tonight. </s>	<s> Du, Taylor är lite ledsen ikväll. </s>	PRED	EXP
<s> What a sad tone in your voice. </s>	<s> Du låter sa ledsen. </s>	ATTR	STIM
<s> It's sad . </s>	<s> Vad sorgligt </s>	PRED	STIM

English	Swedish	Syntactic fu	Semantic ro
<s> It was quite tragic . </s>	<s> Det var rätt så tragiskt. </s>	PRED	STIM
<s> My husband died in a tragic hunting accident. </s>	<s> -Min man dog i en jaktolycka. </s>	ATTR	STIM
<s> I do have the details about these tragic deaths. </s>	<s> Men jag har detaljerna om dessa tragiska dödsfallen. </s>	ATTR	STIM
<s> That's tragic . </s>	<s> - Det är sorgligt. </s>	PRED	STIM
<s> So many tragic endings. </s>	<s> Så många sorgliga slut. </s>	ATTR	STIM

English	Swedish	Syntactic fu	Semantic ro
<s> This won't be an unhappy occasion. </s>	<s> Det här ska inte vara en olycklig tillställning. </s>	ATTR	STIM
<s> I'm so unhappy , Pacco. </s>	<s> Jag är så olycklig, Pacco. </s>	PRED	EXP
<s> How you could be unhappy . </s>	<s> Hur kan du vara olycklig . </s>	PRED	EXP
<s> You've made me unhappy Dresden. </s>	<s> - Du har gjort mig olycklig, Dresden. </s>	PRED	STIM
<s> I've never had an unhappy customer. </s>	<s> - Jag har aldrig haft en missnöjd kund. </s>	ATTR	EXP

Appendix B The Python code used for making Figure 1

This appendix presents the Python code used for making Figure 1 found in section 5.5.

```
import matplotlib.pyplot as plt
import numpy as np
from matplotlib import rc
rc('font', family='serif', size=18)
rc('text', usetex=True)

words = ["ledsen", "sorgsen", "tråkig", "trist", "sorglig", "depressing",
↪  "unhappy", "sad", "unfortunate", "tragic"]
cl = ['r', 'r', 'r', 'r', 'r', 'b', 'b', 'b', 'b', 'b']

stim = np.array([0.0, 0.125, 0.865, 1.0, 0.912, 1.0, 0.22, 0.424, 0.872,
↪  1.0])
pred = np.array([1.0, 0.85, 0.84, 0.81, 0.51, 0.82, 0.77, 0.76, 0.44,
↪  0.3])
freq = np.array([79368, 583, 8231, 3272, 3558, 6822, 18775, 94181, 16278,
↪  14979])

x_loc = [-0.06, -0.03, -0.05, -0.02, -0.04, -0.1, -0.1, -0.05, -0.1,
↪  -0.05]
y_loc = [-0.08, -0.02, -0.039, -0.03, -0.03, 0.019, -0.05, -0.07, -0.05,
↪  0.02]

cfreq = np.array(np.sqrt(freq) / 150)
text_sizes = 10 + 40 * (cfreq / max(cfreq))

plt.figure(figsize=(10,8))
plt.scatter(stim, pred, c=cl, s=100*cfreq, marker="d")
plt.xlabel(r"\textbf{Ratio Stimulus}")
plt.ylabel(r"\textbf{Ratio Predicative}")

for i, x in enumerate(range(len(words))):
    plt.text(stim[i]+x_loc[x], pred[i]+y_loc[x], words[i], size= 10 + 40
    ↪  * (cfreq[x]/ max(cfreq)), color=cl[i])
plt.xlim(-0.1, 1.15)
plt.savefig('/media/sandra/T7/figure1.png', dpi = 300)
```

Stockholm University
SE-106 91 Stockholm, Sweden
Telephone +46 (0)8 16 20 00
https://www.su.se/