Explicit References to Social Values in Fairy Tales:
A Comparison between Three European Cultures

Alba Morollon Diaz-Faes

Institute for the Study of
Literature and Tradition
NOVA University of Lisbon
1069-061 Lisbon, Portugal
`albafaes@fcsh.unl.pt`

Carla Sofia Ribeiro Murteira

Institute for the Study of
Literature and Tradition
NOVA University of Lisbon
1069-061 Lisbon, Portugal
School of Communication
and Media Studies, IPL
Portugal
`carlamurteira@fcsh.unl.pt`

Martin Ruskov

University of Milan
20123 Milan, Italy
`martin.ruskov@unimi.it`

Abstract

The study of social values in fairy tales opens the possibility to learn about the communication of values across space and time. We propose to study the communication of values in fairy tales from Portugal, Italy and Germany using a technique called word embedding with a compass to quantify vocabulary differences and commonalities. We study how these three national traditions of fairy tales differ in their explicit references to values. To do this, we specify a list of value-charged tokens, consider their word stems and analyse the distance between these in a bespoke pre-trained Word2Vec model. We triangulate and critically discuss the validity of the resulting hypotheses emerging from this quantitative model. Our claim is that this is a reusable and reproducible method for the study of the values explicitly referenced in historical corpora. Finally, our preliminary findings hint at a shared cultural understanding and the expression of values such as Benevolence, Conformity, and Universalism across European societies, suggesting the existence of a pan-European cultural memory.

1 Introduction

Culture is defined "as a common heritage of a set of beliefs, norms, and values" (US DHHS 2001), that influences an individual's cognition and behaviour (Wong, 2013). Social values are understood as standards or criteria of the desirable, thus they guide the selection or evaluation of behaviours, policies, people, and events (Schwartz et al., 2020). Building on this understanding of values as a cornerstone of culture, we turn to literature as a mirror reflecting these values across different cultural contexts in the past. Developments in NLP, in particular word embeddings, have allowed for the quantitative analysis of historical corpora (Miaschi and Dell'Orletta, 2020; Rodriguez and Spirling, 2022).

With this work we want to test the limits of an approach for studying the social values present in fairy tales, one of the most widely spread forms of popular narratives and a privileged genre for the identification of patterns of cultural exchange, as they have historically migrated across different cultures and periods, creating a rich tapestry of storytelling traditions. In particular, we study the aggregated explicit tokens mapped on the values proposed by the Theory of Basic Human Values (Schwartz, 1992, 2012) across fairy tale corpora from three different European traditions – namely Portugal, Italy and Germany – in order to compare the quantitative representations and analyse the emerging patterns. We do this by first finding the stemmed matches of these tokens and enriching the text with the corresponding annotation. After that we employ a word embedding with a compass (Di Carlo et al., 2019) and clique percolations (Palla et al., 2005) to highlight the semantic variation between the three national corpora.

A critical investigation of the results of our method finds that its results correspond to findings of previous research. We also find indications that despite the differences on the expression of values in the three compared countries, it seems that Values of Benevolence (quality of interpersonal relationships), Conformity (respect for social norms and expectations) and Universalism (protection of the welfare of people and nature) have remained consistent in fairy tales across the three national traditions, which we also view as confirmation of the validity of our approach for the study of values embedded in historical, literary corpora.

2 Background

The study of explicit references of values in fairy tales is related to the accumulated social attitudes up to the historical period of codification of the

tales. To our knowledge, no systematic research of this wide topic exists. As such, we view it as being at the crossroads between the socio-historical, literary study of fairy tales, and the psychological study of social values which is shaped by contemporary research. On the other hand, such a study at scale would not be possible without the instruments and methods of computational humanities and word embeddings in particular.

2.1 Unpacking Fairy-Tale Studies from the Brothers Grimm to Digital Humanities

The late 18th century witnessed the rise of folklore studies as part of a quest for national and cultural identity, particularly in Europe (Schacker, 2003). Jakob and Wilhelm Grimm, riding the tide of renewed interest in popular culture among the upper-class intelligentsia, became pivotal figures in this domain. They first published their fairy-tale collection *Children's and Household Tales* in 1812, striving to present a pure German narrative tradition, untouched by foreign influence, particularly the French (Teverson, 2013). This publication sparked what would become the 19th century's golden age of fairy tales across Europe. This was a time of growing urbanisation, industrialization, and literacy. Scholars and nationalists, fearful of losing invaluable oral traditions due to these rapid societal changes, began the collection and preservation of folklore (Ostry, 2013). Among these custodians were collectors and writers such as Italy's Giuseppe Pitré and Portugal's Consiglieri Pedroso, whose texts feature prominently here alongside the Grimms'. Their work, heavily inspired by the Grimms, was driven by a desire to distil and dialectically construct their nations' cultural legacy.

Despite the nationalistic intentions of Brothers Grimm and others who embarked on preserving what they thought to be distinct national narratives, the study of fairy tales reveals as much about the interconnectedness of cultures as it does about their uniqueness. Fairy tales, at their core, are a blend of narratives that "migrate on soft feet" (Warner and Warner, 2016), indicating that they traverse and interweave across generic, geographical and temporal boundaries, sometimes in untraceable ways. Thus, while the Grimms and others sought to capture and enshrine a uniquely national heritage, their work also serves to underscore the similarities between narrative traditions.

Unpicking these similarities and differences, however, can prove to be quite a complex task. As scholars are frequently dependent on translations, the potential for misinterpretation or loss of nuanced meanings during this process is high. Translations, like the ones by Margaret Hunt, Thomas Crane and Henriqueta Monteiro used here, are enormously valuable artefacts, but must be recognised as acts of literary adaptation that might differ from the originals (Haase, 2016). Further complicating matters, the comparative analysis of several national traditions involves processing vast quantities of text to identify patterns. This challenge extends beyond the study of fairy tales and into the comparative study of literature as a whole.

In response to these challenges, digital humanities and computer-assisted literary studies offer innovative methodologies. Computational methods, in particular, aid in identifying and assessing literary patterns across scales, from individual texts to entire fields and systems of cultural production (Wilkens, 2015). These new approaches, to which our work is a contribution, help produce new types of evidence that enrich and expand humanities research. Indeed, computational approaches to fairy tales have already successfully been deployed in studies such as "Computational analysis of the body in European fairy tales" (Weingart and Jorgensen, 2013). In this study, the authors used digital humanities research methods to analyse the representations of gendered bodies in European fairy tales. They created a manually curated database listing every reference to a body or body part in a selection of 233 fairy tales, and its analysis revealed that the gender and age of fairy-tale protagonists correlate in ways that indicate societal biases, particularly against the ageing female body. A further exploration of gender bias in fairy tales is presented in "Are Fairy Tales Fair?" (Isaza et al., 2023). This study employs computational analysis to dissect the sequence of events in fairy tales, revealing that one in four event types exhibit gender bias when not considering temporal order, and that female characters are more likely to experience gender-biased events at the start of their narrative arcs. These studies underscore the potential of distant reading, data analysis and visualisation as powerful tools in the comparative study of fairy tales, particularly when used alongside subject expert close reading (Moretti, 2022). Nevertheless, perceptions and attitudes towards gender represent just a fraction of the broader societal values spectrum.

2.2 The expression of values across space and time in European Fairy Tales

Values are regarded as a shared societal understanding of what constitutes *good*, *wrong*, *fair*, *unfair*, *just*, *right* or *ethical* behaviour (Haidt, 2013; Kesebir and Haidt, 2010; Turiel, 2005). Values are cognitive representations of an individual's biological needs, an individual's requirements in interpersonal coordination, and the institutional demands focused on group welfare and survival (Schwartz and Bilsky, 1987). Nonetheless, it is crucial to acknowledge the significance of cultural and individual influence in the development and expression of values. Cultural Psychology postulates that human behaviours result from the reciprocal interaction between cultural and individual psyche (Shweder, 1991; Cohen, 2011; Schwartz et al., 2020). However, the manifestation of behaviors and values is contingent upon context and situation, implying that similar cultural processes might serve or facilitate different purposes based on cultural context (Rogoff, 2003; Schwartz et al., 2020). Therefore, one could examine variations in the expression of values across different regions and periods through the analysis of historical corpora. This stems from the expectation that literature can be used as a vehicle for the expression of cultural norms and values, thereby reflecting the distinct ideological attributes of the writers and the regions from which it emerges (Albrecht, 1956). Several Theories have been proposed to summarise values across different cultures (for a review of theories see Ellemers et al (2019)). In this paper we focus on the Theory of Basic Values (Schwartz, 2012), since it found validity expression across several cultures (e.g. (Spini, 2003; Schwartz et al., 2001, 2014; Davidov et al., 2008), and it has been applied in the study of European values (e.g., European Social Survey). The Theory of Basic Human Values (Schwartz, 2012) comprises 10 human values that are fuelled by four different and opposite motivations: Openness to Change vs. Conservation, Self-Transcendence vs. Self-Enhancement as observed in Figure 1.

Openness to Change relates to an individual's need for independence of thought, action, and feelings, and readiness for change, therefore comprises the values of Self-direction, Stimulation, and partly Hedonism. On the other hand, Conservation relates to the values of Security, Conformity and Tradition, as it emphasises the individual's needs for order,
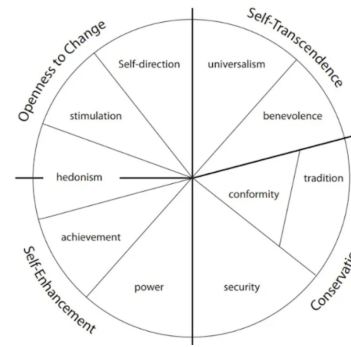

Figure 1: Theoretical model of relations among ten motivational types of values (Schwartz, 2012).

preservation of the past, and resistance to change. Self-enhancement considers the individual's needs to pursue their own interests, success, and dominance over others, therefore comprises the values of Power, Achievement, and partly Hedonism. On the other hand, Self-transcendence considers the values of Universalism and Benevolence, to focus on the welfare and better interests of others. For a definition of specific values, see Table 1.

Europeans can be regarded as having a common identity (Castano, 2004) that is expressed through their way of life, values and culture, and that has been building since ancient times (Pagden, 2002; Pinheiro et al., 2012) leading to the establishment of a broad set of European Values. Values such as human dignity, freedom, democracy, equality, rule of law, and human rights have been declared as the values of the European Union, to form "a society in which inclusion, tolerance, justice, solidarity and non-discrimination prevail" (EU, 2020). Based on several empirical studies and policy making guidelines, these values correspond to Schwartz's values of Universalism, Self-Direction, and Benevolence (for more information see: (Scharfbillig et al., 2021; Murteira, 2024). If these values are presumed to have been shared to some degree across the European territory since antiquity, it stands to reason that they could have been variously conveyed through fairy tales across the three regions under analysis. Constructs such as values can either be assessed by explicit or implicit measures. A psychological construct is implicitly assessed when the individual "is unaware that a psychological measurement is taking place, this type of measure is often used to assess values, attitudes, stereotypes, and emotions in social cognition research" (APA, 2023). On the other hand, a psychological construct is explicitly assessed when the "individual

Table 1: The definition of each of the ten motivational types of values (Schwartz, 2012).

Value	Definition
Security	Safety, harmony, and stability of society, of relationships, and of self.
Tradition	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide the self. Maintaining and preserving cultural, family or religious traditions.
Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.
Self-Direction	Independent thought and action-choosing, creating, exploring.
Stimulation	Excitement, novelty, and challenge in life.
Hedonism	Pleasure and sensuous gratification for oneself.
Achievement	Personal success through demonstrating competence according to social standards.
Power	Social status and prestige, control or dominance over people and resources.
Benevolence	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.
Universalism	Understanding, appreciation, tolerance, and protection of the welfare of all people and of nature.

is aware that a psychological measurement is taking place" (APA, 2023). Putting it simply, values can be measured explicitly when individuals are directly asked about values, and implicitly when the individuals are not aware of the measurement, because values are assessed using indirect questioning methods. Bearing in mind that art is a behavioural expression of culture that serves several purposes, including the *form of order*, which is the need for psychological and mental organisation of experiences (Dissanayake, 1980), we can hold the reasonable expectation that the historical corpora under analysis will reflect, to a degree, the explicit and implicit cultural ways and behaviours of societies in which these fairy tales were written. The presence of these values in our corpora was assessed by vocabulary quantification techniques through the development of Word Embeddings that communicate values in fairy tales.

2.3 Using Word Embeddings to Quantify Vocabulary Differences

Word embeddings have emerged as an important instrument for the quantitative analysis of textual corpora. These are mappings of vocabulary onto a multidimensional numerical space, based on their occurrences (Mikolov et al., 2013; Rodriguez and Spirling, 2022). Different techniques for creating word embeddings exist, but their common general principle is "a word is characterised by the company it keeps". It is useful to distinguishing between two categories of word embeddings: i) static (also called type-based) - those that feature a single representation for a word token, and ii) contextual (also called token-based) - those that distinguish between different representations of each word to capture potential differences in meanings, according to the surrounding context (Miaschi and Dell'Orletta,

2020; Lenci et al., 2022). Whereas contextual word embeddings better capture the richness of vocabulary, static word embeddings perform better on smaller corpora which do not have the volume that would allow for the semantic richness necessary for multiple meanings (Ehrmanntraut et al., 2021). Arguably, this is due to the fact that in a small thematic corpus, typically meanings are restricted by the context of its compilation.

A widespread approach that allows to overcome the challenge of small corpora and their lack of richness, is the combination of pre-training with a huge[1] generic corpus and the subsequent fine-tuning with the corpus of interest. However, corpora of these huge dimensions are inevitably contemporarily written, and due to cultural and linguistic change over time inevitably introduce unwanted biases. In confirmation of this consideration, in their particular context Manjavacas and Fonteyn (2022) observed that training from the ground up is more effective than fine-tuning of pre-existing models.

When it comes to comparing the word embeddings representing different corpora, a widespread approach is the so-called semantic change detection (Tahmasebi et al., 2021). Since for intercultural comparison, "change" might wrongly suggest a transition from one culture to the other, in the context here the phrase "semantic variation" would be more accurate. Still, whenever techniques for semantic change detection do not rely on any particular diachronic properties of the underlying corpora, we claim they could be reused also for synchronic linguistic analysis. More specifically we claim that

---

[1]Some widely established contextual models like BERT are trained on a corpus that includes the entire contents of Wikipedia which comprises of 2.5 billion word tokens (Devlin et al., 2019), others use training sets that are many orders of magnitude larger (Dodge et al., 2021)
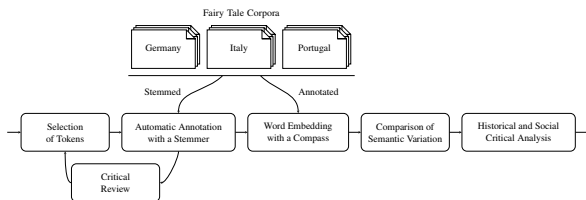
Figure 2: The outline of the process we followed.

an approach called temporal word embedding with a compass (Di Carlo et al., 2019) is applicable, for culture-specific rather than time-specific distinctive corpora. This approach consists of first creating an embedding on a cumulative corpus containing all texts from the different cultures to be considered. Then, from this baseline (compass) word embedding, further fine-tuning is performed on each of the corpora, to be compared so as to create culture-specific word embeddings. The result is a different (numerical) vector representations of each particular word token, which allows for quantitative comparisons between them, as done previously (Ferrara et al., 2022; Di Carlo et al., 2019).

3 Method

Our study of the explicit references to values in fairy tales follows the process illustrated in Figure 2. To provide an outline, it starts with the identification of tokens that represent values of interest. We group these tokens in groups that we consider to be synonyms. Then, we automatically annotate all occurrences in the text of the stems representing the considered tokens. Once this is done, we manually analyse the produced annotations to identify ambiguities and mistakes in our identification of tokens. The purpose of this analysis is to better understand the semantics behind their occurrences, in order to refine the selection of tokens and identify potential tokens representing multiple values. Finally, we apply a static word embedding with a compass and perform critical analysis on the differences from the resulting vector spaces.

Fairy Tales Corpora. The corpus selection had several stages. First we focused on the Grimms' *Children's and Household Tales*, using Margaret Hunt's 1884 English translation. We manually selected 30 tales that span well-known and beloved stories and lesser known ones, so as to provide a comprehensive representation of the entire collection. Then we selected 30 Portuguese and 30 Italian tales taken from two important contemporary

collections to the Grimms': *Portuguese Folktales* by Consiglieri Pedroso, translated to English in 1882 by Henriqueta Monteiro; and *Italian Popular Tales*, collected and translated to English in 1885 by Thomas Frederick Crane. These collections were chosen due to their cultural significance and their temporal proximity to the Grimms' collection, aiming to offer a comparative perspective on 19th century fairy tales across different European cultures.

Selection of Tokens. Assuming that the historical corpora are themselves mirrors of social behaviours and ways of living in societies in which the fairy tales were written, we are interested in the explicit expressions of values in the texts. Starting from Schwartz's model and the European core values, we initially compose a list of tokens that represent these values, based on three empirical studies regarding value-specific tokens. This list of tokens contains words that were selected from two dictionary studies about values, where each word is associated with a specific Schwartz's value. (Schwartz, 1992; Lindeman and Verkasalo, 2005; Murteira, 2024). For instance, the token "Peace" is associated with the value of Universalism, and the token "Cooperation" is associated with the value of Benevolence (see Table 4 in Appendix). Then we perform automatic identification of explicit references of values. We do this using stemming (Jabbar et al., 2020) on both the token lists and the fairy tale texts. This is because, in contrast e.g. to lemmatisation, stemming reduces different word forms the same originating token. We use the Snowball stemmer algorithm (Porter, 2001) to identify all occurrences of the stemmed tokens in the corpora and tag (i.e. annotate) them with a label corresponding to the group of synonym tokens.

Critical Review. We then critically analyse and refine by adapting tokens according to the desired annotation. This was done using a specifically de-

Table 2: Quantitative descriptors of the corpora. When we refer to tokens, we mean the ones that were identified by our automated annotation process. Complete list of included texts is available in Table 3 in Appendix.

Corpus	Texts	Symbols	Words	Tokens
Germany	30	306 475	59 500	1840
Italy	30	234 158	45 223	1808
Portugal	30	231 149	44 887	1439

Figure 3: Screenshot of a browsing page from the bespoke web instrument that reviews the produced annotations. Another view shows a clickable heatmap as Figure 7 in Appendix, which allows for a distant reading view.

veloped for the purpose web interface (Figure 3, accessible online at `https://tales.ko64eto.com`) that allows for a review of the texts in the corpora with the results of the automatic annotation highlighted in different colours. The outcome of this was a series of decisions to adapt the token selection as a way to refine it and guide subsequent iterations of this annotation process. Correspondingly, following this approach inspired by grounded theory (Rieger, 2019), the ultimately proposed list of tokens in this study emerges from exploration of the corpus and is not a result of deductive hypothesis research. We provide a statistical overview of the results of this annotation process in Table 2 and in Appendix we provide both the complete final version of our tokens and a Venn diagram of the occurrences of groups of synonym token across the three corpora.

Word Embedding with a Compass.   Due to the historical nature of the corpora we study and in order to avoid contaminating them with external biases from pre-training, we organise our analysis following the word embedding with a compass approach (Di Carlo et al., 2019). To do this, we create one generic culture-agnostic shared embedding from scratch containing all three corpora. Then, starting from this compass, we independently create three parallel fine-tunings for each of the cultures. For the creation of the compass, to avoid the possible introduction of biases, we chose not to include any further possible texts, neither from any of our three contexts, nor from others. Our approach to syntactic identification of references to values, is not contextual, i.e. we treat a reference to a

value-related stemmed token as the same for all its identified uses. In our critical review step we examine the validity of this generalisation. To represent the annotations in the word embedding algorithm, before and after each identified occurrence of a token we insert an indication of the corresponding group of synonym tokens (i.e. the first token in it).

Comparison of Semantic Variation.   The word embedding allows measuring contextual similarity between words, thus speaking of "change" and "variation". Once we have the three word embeddings for the cultural corpora, for each of them we consider only the distances between groups of tokens (represented by the annotation label, i.e. the first token in each synonym group) and experimentally define a similarity threshold above which we consider a pair of tokens to have a relating edge between them in a graph representation of tokens) in order to use clique percolations clustering with k=2 (Palla et al., 2005). In other words, for all similarities above that threshold we consider the corresponding tokens to be related in that embedding, and distances above the threshold mean the corresponding tokens are not. This results in a clustering that might assign one token to multiple clusters. It might also bring two tokens into the same cluster even if the distance between them is greater than the threshold, as long as there is a "bridge" of other tokens in between to connect them.

Historical and Social Critical Analysis.   At the end of our method we analyse the quantitative results using critical analysis from the perspectives of both literary studies and psychological research. This allows us to cross-validate (e.g. through triangulation (Noble and Heale, 2019)) our results with the established body of research and thus get an indication of their theoretical validity.

4 Results

An important part of the results of our approach is the reflective inspection of the produced automated annotation and possible corrections for these. An overall conclusion of this process is that, expectedly, the most impactful tokens capture the values they were intended to match well. The most important token that did not correspond to our initial interpretation was "faith". We originally ascribed the label "faith" to the value of "piety," indicating religious devotion. However, a careful examination of our corpus revealed an intriguing trend. The term

	mother	law	brother	love	know	justic	generos	cooper	pieti	kind	loyal	right	reward	smart	empathi	peac	pure	punish	curios	free	truth	correct	toler	equal	honest	solidar	dialogu	emancip	evid	**TOTAL**
Germany	446	398	108	169	71	107	99	93	78	57	57	35	25	20	15	13	8	6	7	11	5	3		6	6	2	2	1		**1840**
Italy	384	391	394	278	83	56	38	33	27	15	6	9	17	2	8	10	7	6	6	7	4	3	6	2	3	6		2		**1808**
Portugal	372	253	326	193	81	19	32	26	37	11		6	6	8	4	1	6	8	5		8	7	6	4	2	3	7	1	2	**1439**

Figure 4: Frequencies of identified occurrences of tokens across the three corpora. A more detailed heatmap between texts and labels is available on Figure 7 in Appendix.

"faith," contrary to our initial classification, typically expressed affiliation with "loyalty," mainly as per the usage patterns in various Grimm tales, particularly in "Faithful Johannes" (a German tale). As a consequence, we ascribe the token "faith" to the value associated to "loyalty".

Another token that provides an interesting example is "father," due to its potential multiple associations. On the one hand, it could represent "caring," similar to "mother," but on the other, it could be a symbol of authority (Hopp et al., 2021). When exploring the corpora, we found that "father" was predominantly associated with "caring," with a remarkable exception in "The Maiden and the Fish" (Portugal), where one out of four instances appeared associated with authoritative power.

A third, less impactful token we considered was "patient," which was initially associated with "kindness." However, an analysis of the corpus found that its usage related exclusively to an individual receiving medical treatment, and we consequently excluded it from our analysis.

Figure 4 shows the references to values by countries, according to the ascribed tokens. A more detailed mapping of occurrences of tokens in particular texts is provided in Figure 7 in the Appendix. From the resulting comparison of clusters across corpora, noteworthy is the one defined around tokens related to "mother." As the Venn diagram on Figure 5 shows, while in our German and Portuguese corpora it appears together with "brother", in the Italian and Portuguese corpora it also appears in relation to "know." Only in Germany does it relate to "generous."

4.1 Historical Analysis

Dolores Buttry elucidates on the usage of "faith" in Grimm tales to exclusively mean "loyalty," and not "piety." She writes that the related values of faithfulness and loyalty (which are "Treu" and "Treue" in German) have been foundational virtues in Ger-

many since ancient times (Buttry, 2011). Stories such as "Faithful Johannes," but also "The Frog King," exemplify extreme loyalty towards superiors, illustrating the importance of fidelity and respect for authority in their various manifestations. Buttry characterises the tale of the faithful/loyal servant as an enduring archetype, highlighting the recurring appearance of the words "Treu" (faithful) and "Treue" (loyalty, fidelity) in German tales (Buttry, 2011). She further suggests that, while respect for authority and the sanctity of oaths were nearly universal concepts before these stories were collected, they seem to have retained their vitality and cultural significance particularly in German-speaking traditions. This idea finds further support in one of the only non-German occurrences of "faith" in our corpus, as the label appears in "The Story Of Catherine and Her Fate," a Sicilian tale first collected by Swiss-German folklorist Laura Gozenbach.

It is also interesting to examine how values manifest in tales from different cultural contexts. In our results, we found that values of "piety" and "empathy" appeared clustered together in Italian and Portuguese tales, but not in German ones. This may be explained by the different religious traditions in all three countries, since both Italy and Portugal were majoritarily Catholic regions at the time the tales were collected, while there was a strong Protestant presence in the German territory. Indeed, Jack Zipes (2002) writes that the Grimms' tales portrayed the main values of Protestant ethics
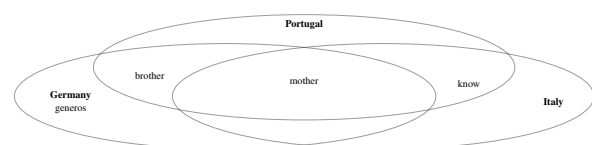

Figure 5: An illustration of the degree of overlap across the three national corpora for the token "mother"

and the bourgeois enlightenment. The heroes in their tales are predominantly concerned with self-preservation and the acquisition of wealth, and they assist others, including animals, only when they perceive a potential gain for themselves, demonstrating a calculated approach to empathy and compassion. This model of behavior, Zipes argues, exemplifies the general Protestant ethic of the time, and so empathy, although occasionally appearing in the Grimms' tales, is not a dominant theme (Zipes, 2002). We may advance the possibility that the differing religious ethos of Italy and Portugal would place more emphasis on empathy as it relates to Catholic piety.

4.2 Social Analysis

Frequency analysis shows that tokens such as "mother," "law," brother," and "love" have a strong presence (more than 100 appearances) across the three countries under analysis. Based on the elaborated correspondence between tokens and the Theory of Basic Values (see Appendix), the words "mother," "brother" and "love" are connected to Benevolence, and "law" is connected to Conformity. In Germany, the token "justice" has also a strong presence, and is connected with the value of Universalism which stands for the protection and welfare of all people and nature. Considering that the value Benevolence stands for the good quality of social connections between people, and Conformity stands for the preservation of social/cultural expectations and norms, then we could infer that these tales describe several social dynamics. The tales' plots are representative of social dynamics among fictional characters that may resemble society, in order to describe the quality of human relationships and social/cultural norms in place.

Interestingly, some differences across countries are expressed by the tokens' frequency related to Benevolence, Conformity and Universalism. For instance, in Germany, "mother" seems to be a stronger reference for communication of Benevolence than "brother" when compared to Portugal and Italy. Also, "love" seems to be a stronger reference for communication of Benevolence in Italy than in Germany and Portugal. However, in Germany, we may note that tokens such as "generous" and "cooperation" reinforce the communication and expression of Benevolence in those tales. When it concerns the need for rules and social welfare, it seems that in Germany and Italy the token

"law" is frequently used when compared to Portugal to express the value of Conformity. Finally, Germany shows a strong presence of token "justice" in their tales, which highlights the importance of Universalism in these tales and the need to convey the respect for human rights and dignity. In sum, while Portugal, Italy and Germany communicate strongly the values of Benevolence and Conformity, it seems that Germany also communicates the value of Universalism. Despite the differences between countries, it seems that European Values of Benevolence and Universalism are being communicated by the tales across countries.

5 Discussion and Conclusion

While the proposed approach is still in its infancy, and the emerging results would require more thorough examination, our preliminary analysis provides some concrete evidence that European Values have been a long-standing element in European cultural communication through fairy tales. The corpus analysis across different cultures revealed a significant variety in the representation of values. For example, the affiliation of the token "faith" with "loyalty" rather than "piety," particularly in German culture, illustrates the role of cultural and historical contexts in shaping value representations. Similarly, the differential clustering of "piety" and "empathy" in Italian and Portuguese tales compared to German tales further underscores the influence of religious and socio-cultural contexts in value representation. Interestingly, despite these differences, the analysis revealed a strong commonality across all three cultures, pointing at the communication of European Values through tales. Tokens associated with Benevolence, Conformity, and Universalism manifested frequently across fairy tales of all three countries. This finding is particularly noteworthy because it suggests a shared cultural understanding and expression of these values across European literary production, and, possibly and by extension, across European societies, thus hinting at the existence of a pan-European cultural memory.

We have identified clear limitations in our approach. Working at the syntactic level, both in terms of stemming and static word embeddings, limits the possibility to capture nuances, and with this some noise is introduced in the analysis. However, contrary to our expectations, our detailed analysis revealed that ambiguities are only isolated cases. This is valid to the extent that in none of

these cases a token bore semantic ambiguity that was a dichotomy rather than an outlier so that it could undermine the general results.

The focus on explicit references, unsurprisingly, resulted in an inability to annotate tokens such as "democracy" in the tales, as they were only implicitly referenced. Therefore, exploring methods to apply semantic word embeddings to historical texts could be a potential way to address not just explicit, but also implicit references to values (Ferrara et al., 2023). While such approaches already exist (Montanelli and Periti, 2023), we believe further attention should be paid to the possibility that the pre-trained embeddings may introduce biases unrelated to the corpus under study.

This work provides a foundational understanding of how European Values are represented in literary texts and highlights the potential of computational linguistics in cultural studies. This study encourages further interdisciplinary research in the field of literary studies, cultural analytics, and computational linguistics to expand our understanding of cultural values and their historical evolution.

Acknowledgments

References

Milton C. Albrecht. 1956. Does Literature Reflect Common Values? *American Sociological Review*, 21(6):722–729.

APA, 2023. 2023. APA Dictionary of Psychology.

Dolores Buttry. 2011. Treue in three tales by the Brothers Grimm. *Forum for World Literature Studies*, 3(2):166–173.

Emanuele Castano. 2004. *European identity: A social-psychological perspective*, Governance in Europe, chapter 3. Rowman & Littlefield, Lanham, MD.

Dov Cohen. 2011. Cultural Psychology. Technical report, Oxford University Press. Type: dataset.

Henriqueta Monteiro Consiglieri Pedroso. 1882. *Portuguese Folk-Tales*. WikiSource.

Thomas Frederick Crane. 2017. *Italian Popular Tales*. Project Gutenberg.

E. Davidov, P. Schmidt, and S. H. Schwartz. 2008. Bringing Values Back In: The Adequacy of the European Social Survey to Measure Values in 20 Countries. *Public Opinion Quarterly*, 72(3):420–445.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Valerio Di Carlo, Federico Bianchi, and Matteo Palmonari. 2019. Training Temporal Word Embeddings with a Compass. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6326–6334.

Ellen Dissanayake. 1980. Art as a Human Behavior: Toward an Ethological View of Art. *The Journal of Aesthetics and Art Criticism*, 38(4):397.

Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. ArXiv:2104.08758 [cs].

Anton Ehrmanntraut, Thora Hagen, Leonard Konle, and Fotis Jannidis. 2021. Type- and token-based word embeddings in the digital humanities. In *Proceedings of the Conference on Computational Humanities Research 2021*, number 2989 in CEUR Workshop Proceedings, pages 16–38, Aachen.

Naomi Ellemers, Jojanneke Van Der Toorn, Yavor Paunov, and Thed Van Leeuwen. 2019. The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, 23(4):332–366.

EU, 2020. 2020. The EU values. Last accessed 30 July 2023.

Alfio Ferrara, Stefano Montanelli, and Martin Ruskov. 2022. Detecting the semantic shift of values in cultural heritage document collections (short paper). In *Proceedings of the 1st Workshop on Artificial Intelligence for Cultural Heritage*, number 3286 in CEUR Workshop Proceedings, pages 35–43, Aachen.

Alfio Ferrara, Sergio Picascia, and Elisabetta Rocchetti. 2023. Augustine of Hippo at SemEval-2023 Task 4: An Explainable Knowledge Extraction Method to Identify Human Values in Arguments with SuperASKE. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1044–1053, Toronto, Canada. Association for Computational Linguistics.

J. Grimm, W. Grimm, and M. Hunt. 1884. *Household Tales by the Brothers Grimm*. George Bell and Sons.

Donald Haase. 2016. Challenges of Folktale and Fairy-Tale Studies in the Twenty-First Century. *Fabula*, 57(1-2):73–85.

Jonathan Haidt. 2013. *The righteous mind: why good people are divided by politics and religion*, 1st vintage books ed edition. Vintage Books, New York. OCLC: 900283765.

Frederic R. Hopp, Jacob T. Fisher, Devin Cornell, Richard Huskey, and René Weber. 2021. The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, 53(1):232–246.

Paulina Toro Isaza, Guangxuan Xu, Akintoye Oloko, Yufang Hou, Nanyun Peng, and Dakuo Wang. 2023. Are Fairy Tales Fair? Analyzing Gender Bias in Temporal Narrative Event Chains of Children's Fairy Tales.

Abdul Jabbar, Sajid Iqbal, Manzoor Ilahi Tamimy, Shafiq Hussain, and Adnan Akhunzada. 2020. Empirical evaluation and study of text stemming algorithms. *Artificial Intelligence Review*, 53(8):5559–5588.

Selin Kesebir and Jonathan Haidt. 2010. Morality (in Handbook of Social Psychology).

Alessandro Lenci, Magnus Sahlgren, Patrick Jeuniaux, Amaru Cuba Gyllensten, and Martina Miliani. 2022. A comparative evaluation and analysis of three generations of Distributional Semantic Models. *Language Resources and Evaluation*, 56(4):1269–1313.

Marjaana Lindeman and Markku Verkasalo. 2005. Measuring Values With the Short Schwartz's Value Survey. *Journal of Personality Assessment*, 85(2):170–178.

Enrique Manjavacas and Lauren Fonteyn. 2022. Adapting vs. Pre-training Language Models for Historical Languages. *Journal of Data Mining & Digital Humanities*, NLP4DH(Digital humanities in...):9152.

Alessio Miaschi and Felice Dell'Orletta. 2020. Contextual and Non-Contextual Word Embeddings: an in-depth Linguistic Investigation. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 110–119, Online. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space.

Stefano Montanelli and Francesco Periti. 2023. A Survey on Contextualised Semantic Shift Detection. ArXiv:2304.01666 [cs].

Franco Moretti. 2022. *Falso movimento: la svolta quantitativa nello studio della letteratura*. Extrema ratio. Nottetempo, Milano.

Carla Murteira. 2024. Towards an ontology of values: Elaboration of a dictionary of values' related words based on the theory of social values from schwartz. In preparation.

Helen Noble and Roberta Heale. 2019. Triangulation in research, with examples. *Evidence-Based Nursing*, 22(3):67–68.

Elaine Ostry. 2013. *Social Dreaming*, 0 edition. Routledge.

Anthony Pagden, editor. 2002. *The idea of Europe: from antiquity to the European Union*. Woodrow Wilson Center series. Woodrow Wilson Center Press ; Cambridge University Press, Washington, DC : Cambridge ; New York.

Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.

Teresa Pinheiro, Beata Cieszynska, and José Eduardo Franco, editors. 2012. *Ideas of | for Europe*. Peter Lang D.

M.F. Porter. 2001. Snowball: A language for stemming algorithms. Accessed 27 July 2023.

Kendra L. Rieger. 2019. Discriminating among grounded theory approaches. *Nursing Inquiry*, 26(1):e12261.

Pedro L. Rodriguez and Arthur Spirling. 2022. Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84(1):101–115.

Barbara Rogoff. 2003. *The cultural nature of human development*. Oxford University Press, Oxford [UK] ;a New York.

Jennifer Schacker. 2003. *National Dreams: The Remaking of Fairy Tales in Nineteenth-Century England*. University of Pennsylvania Press.

M. Scharfbillig, L. Smillie, D. Mair, M. Sienkiewicz, J. Keimer, R. Pinho Dos Santos, H. Vinagreiro Alves, E. Vecchione, and Scheunemann L. 2021. *Values and identities: a policymaker's guide*. Publications Office of the European Union, Luxembourg. OCLC: 1289305231.

Seth J. Schwartz, Ágnes Szabó, Alan Meca, Colleen Ward, Charles R. Martinez, Cory L. Cobb, Verónica Benet-Martínez, Jennifer B. Unger, and Nadina Pantea. 2020. The Convergence Between Cultural Psychology and Developmental Science: Acculturation as an Exemplar. *Frontiers in Psychology*, 11:887.

Shalom H Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier.

Shalom H. Schwartz. 2012. An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2(1).

Shalom H. Schwartz and Wolfgang Bilsky. 1987. Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3):550–562.

Shalom H. Schwartz, Gian Vittorio Caprara, Michele Vecchione, Paul Bain, Gabriel Bianchi, Maria Giovanna Caprara, Jan Cieciuch, Hasan Kirmanoglu, Cem Baslevent, Jan-Erik Lönnqvist, Catalin Mamali, Jorge Manzi, Vassilis Pavlopoulos, Tetyana Posnova, Harald Schoen, Jo Silvester, Carmen Tabernero, Claudio Torres, Markku Verkasalo, Eva Vondráková, Christian Welzel, and Zbigniew Zaleski. 2014. Basic Personal Values Underlie and Give Coherence to Political Values: A Cross National Study in 15 Countries. *Political Behavior*, 36(4):899–930.

Shalom H. Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. 2001. Extending the Cross-Cultural Validity of the Theory of Basic Human Values with a Different Method of Measurement. *Journal of Cross-Cultural Psychology*, 32(5):519–542.

Richard A. Shweder. 1991. *Thinking through cultures: expeditions in cultural psychology*. Harvard University Press, Cambridge, Mass.

Dario Spini. 2003. Measurement Equivalence Of 10 Value Types From The Schwartz Value Survey Across 21 Countries. *Journal of Cross-Cultural Psychology*, 34(1):3–23.

Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen. 2021. *Computational approaches to semantic change*. Language Science Press, Berlin.

Andrew Teverson. 2013. *Fairy Tale*, 0 edition. Routledge.

Elliot Turiel. 2005. Thought, Emotions, and Social Interactional Processes in Moral Development. In *Handbook of Moral Development*, pages 7–35. Psychology Press.

US DHHS 2001. 2001. *Mental health: culture, race, and ethnicity: a supplement to Mental health, a report of the Surgeon General*. Report of the Surgeon General Series. Department of Health and Human Services, U. S. Public Health Service, Washington, D.C. OCLC: 47911898.

Marina Warner and Marina Warner. 2016. *Once Upon a Time: A Short History of Fairy Tale*. Oxford University Press, Oxford, New York.

S. Weingart and J. Jorgensen. 2013. Computational analysis of the body in European fairy tales. *Literary and Linguistic Computing*, 28(3):404–416.

Matthew Wilkens. 2015. Digital Humanities and Its Application in the Study of Literature and Culture. *Comparative Literature*, 67(1):11–20.

Paul T. P. Wong. 2013. Positive Psychology. In Kenneth D Keith, editor, *The Encyclopedia of Cross-Cultural Psychology*, 1 edition, pages 1021–1027. Wiley.

Jack Zipes. 2002. *The Brothers Grimm: From Enchanted Forests to the Modern World*. Palgrave Macmillan US, New York.

A Appendix

The source code of the annotation analysis tool is available at:

`https://github.com/umilISLab/moreever/`.

Continued with figures and tables on the following pages.

Table 3: The Fairy Tales included in the corpora. The Italian corpus includes several collectors. When not indicated, collected by Giuseppe Pitré. Otherwise, 1. Vittorio Imbriani; 2. Domenico Comparetti; 3. Laura Gozenbach; and 4. Carolina Coronedi-Berti.

Germany (Grimm et al., 1884)	Italy (Crane, 2017)	Portugal (Consiglieri Pedroso, 1882)
• Allerleirauh • Briar Rose • Cinderella • Faithful John • Fitcher's Bird • Frau Trude • Godfather Death • Hansel And Grethel • King Thrushbeard • Little Red Cap • Little Snow White • Old Sultan • One Eye Two Eyes And Three Eyes • Our Lady's Child • Rapunzel • Rumpelstiltskin • Snow White And Rose Red • Strong Hans • The Frog King Or Iron Henry • The Giant And The Tailor • The Girl Without Hands • The Jew Among Thorns • The Juniper Tree • The King Of The Golden Mountain • The Lazy Spinner • The Robber Bridegroom • The Six Servants • The Three Spinners • The Two Kings Children • The Valiant Little Tailor	• Brother Giovannone • Cinderella[3] • Don Firriulieddu • Godmother Fox • King Bean Giuseppe Bernoni • Little Chick Pea Tuscan variant • Pitidda • Snow White Fire Red • The Cat And The Mouse • The Cistern • The Cloud[3] • The Crumb In The Beard[3] • The Crystal Casket • The Dancing Water The Singing Apple And The Speaking Bird • The Doctor's Apprentice • The Fair Angiola[3] • The Fair Fiorita[3] • The King Of Love • The King Who Wanted A Beautiful Wife[3] • The Lord St Peter And The Apostles • The Parrot Which Tells Three Stories • The Sexton's Nose • The Shepherd Who Made The King's Daughter Laugh[3] • The Stepmother • The Story Of Catherine And Her Fate[3] • The Story Of Crivoliu[3] • The Story Of St James Of Galicia[3] • The Three Admonitions • Thirteenth • Water And Salt	• May You Vanish Like The Wind • Pedro And The Prince • Saint Peter's Goddaughter • The Aunts • The Baker's Idle Son • The Cabbage Stalk • The Daughter Of The Witch • The Enchanted Maiden • The Hearth-cat • The Hind Of The Golden Apple • The Little Tick • The Maid And The Negress • The Maiden And The Beast • The Maiden And The Fish • The Maiden From Whose Head Pearls Fell On Combing Herself • The Maiden With The Rose On Her Forehead • The Prince Who Had The Head Of A Horse • The Princess Who Would Not Marry Her Father • The Rabbit • The Seven Iron Slippers • The Slices Of Fish • The Spell Bound Giant • The Spider • The Step Mother • The Three Citrons Of Love • The Three Little Blue Stones • The Three Princes And The Maiden • The Tower Of Ill Luck • The Two Children And The Witch • The Vain Queen

Table 4: List of tokens mapped with the values proposed in the Theory of Basic Values from Schwartz (Schwartz, 1992; Lindeman and Verkasalo, 2005; Murteira, 2024).

Token	Synonyms	Value
dialogu	conversation	Universalism
equality	equality, equal	Universalism
free	free	Self-Direction
right	right, claim	Universalism
justic	justice, judge, trial, fairness, just	Universalism
peace	peace	Universalism
cooper	help, together	Benevolence
curios	curiosity, curious	Self-Direction
empathi	compassion, pity	Conformity
evid	evidence	Universalism
emancip	liberty	Self-Direction
generous	hospitality, goodness	Benevolence
honest	honest, confidence	Benevolence
smart	clever, cleverness, wise	Achievement
kind	kind, kindness, graciousness, gentleness	Conformity
know	know, able, knowledge	Achievement
brother	brother, sister, brotherly, sisterly	Benevolence
love	love, married, wife, husband, marriage, wedding	Benevolence
loyal	honor, faith	Benevolence
pieti	piety, pious, god, virgin, saint, angel, pray	Tradition
mother	mother, father, motherly, fatherly	Benevolence
punish	punish, punishment	Conformity
pure	pure, innocent, innocence	Tradition
correct	correct, reason, correctness	Universalism
reward	reward, prize, pay, treasure, jewels	Power
law	lawful, king, queen	Power
solidar	harmony, support	Benevolence
toler	acceptance, permissiveness	Universalism
truth	truth	Universalism

Figure 6: A Venn diagram showing the occurrences of tokens across the national corpora.

Figure 7: Counts of identified occurrences of tokens across the texts of the three corpora. An interactive version of this heatmap is available at `https://tales.ko64eto.com`. In it clicking on a number takes you to the corresponding text for easier review.