

Computational analysis of the body in European fairy tales

Scott Weingart

Indiana University, School of Library and Information Science

Jeana Jorgensen

Indiana University, Folklore and Ethnomusicology

Abstract

This article explores how digital humanities research methods can be used to analyze the representations of gendered bodies in European fairy tales, a flexible and pervasive genre that has influenced Western children's education and acquisition of gender identity for centuries. By blending the theoretical and methodological concerns of folkloristics, gender studies, and large-scale scientific research, this article demonstrates the utility of cross-disciplinary collaboration in asking traditional questions of traditional materials with new methods. To facilitate this research, a hand-coded database listing every reference to a body or body part in the 233 fairy tales was created. Analysis revealed strong indications that the gender and age of fairy-tale protagonists correlate in ways that indicate societal value being placed on certain perspectives, with youthful and masculine perspectives being validated as universal, whereas feminine and aged bodies are often marked as 'other'. Feminist scholars have articulated some of these ideas in the past, and this approach allows for a more empirical exploration of such claims.

Correspondence:

Scott Weingart
School of Library and
Information Science
Indiana University
Wells Library 024
1320 E. Tenth Street
Bloomington,
IN 47401-3907 USA.
E-mail:
scbweing@indiana.edu

This article explores the possibilities of using computational analysis to understand the representations and constructions of gender and the body in European fairy tales. We hope to show how data analysis and visualization can be powerful tools in the humanist's toolbox. Like all power tools, however, not wielding them carefully and expertly could be disastrous. In this paper, we outline the parameters of the project, the methods we chose, and some of the more interesting results we discovered. We will also present our research process as a successful example of interdisciplinary collaboration, touching on the implications for digital humanities research as well as for gender and folklore studies.

Our collaboration began as one between two researchers from very different schools; Scott Weingart focuses on historical scientific research through

the lens of large-scale data analysis, while Jeana Jorgensen's background is in folklore, gender studies, and narrative. For this project, itself an offshoot of Jorgensen's dissertation, we found that both areas of expertise were necessary for our results to be both theoretically thorough and feasible within the short time span of Jorgensen's dissertation. Further, we found that the initial unfamiliarity with each other's expertise led to fruitful discussions, paving the way for both new sorts of questions to be raised about the content of fairy tales, and novel methods of data collection and representation for the digital humanities.

In order to address the basic question—do fairy tales construct and represent bodies differently according to gender—we surveyed the quantitative and digital methods that have already been

employed in folkloristics. Before the widespread use of computers, many folklorists simply made lists of the phenomena they were studying, counting instances of occurrences, and doing simple calculations such as tallying percentages.¹ By the 1960s, some folklorists were using computers to examine thematic clusters across cultural narratives. Like the founding father of digital humanities, Father Robert Busa, these scholars employed punch cards to compute large amounts of data such as regional superstition collections and supernatural folk narratives.² More recently, folklorists such as Kathleen Ragan and Tim Tangherlini have used varied quantitative methods including statistical analysis of the components of texts, applying network analysis to characters in texts from one genre, and creating maps of storyteller networks.³

1 Data Choice and Collection—Theory and Practice

Drawing inspiration from these varied approaches, we applied computational approaches to canonical fairy tales while utilizing tried and tested theoretical frameworks and critical analyses from feminist and folklore studies. We follow Jack Zipes in asserting that canonical fairy tales are those that have been both formative and enduring, while demonstrating conformity to traditional plots and tale elements. Zipes writes: ‘the literary origins of all our canonical fairy tales in the West mark the culmination of an oral tradition in history but certainly not its end’ (1993, p. 7). At the intersection of oral folkloric narratives and literary styles and contexts, canonical fairy tales exhibit great culture significance.

Our basis for selecting canonical tales was to evaluate the impact of both the collection and the individual tales therein; as such, there are multiple versions of tales like ‘Cinderella’, including the early and well-known texts by Charles Perrault and the Brothers Grimm, as well as later versions that have resonated enough with people to have been reabsorbed into folklore. We analyzed tales from six tale collections, some literary and classical, others collected using contemporary ethnographic standards.⁴ The sample included 233 tales, and as

we wished to list and compare every time a body or body part was mentioned or described in the tales, we needed to devise a method for handling this amount of data.

Our solution was to create an expertly hand-coded database that included every noun or adjective used in reference to bodies within all the tales, in all about 11,000 entries. We included thirteen points of data about tales themselves⁵ and another fourteen points of data about every word used in reference to bodies⁶. Some of these data required expert interpretation, which was inspired by previous theoretical frameworks; for example, was this body reference dealing with the grotesque? What structural section of the tale was it found in? It is important to note that this hand-coding was not done as it often is, by teams of undergraduate assistants who have been tested for intercoder reliability, but by one researcher: Jorgensen, the author of this article providing its theoretical framework and the experience in folklore.

Rather than a detriment, we feel this is an important step toward clarifying the humanistic importance and subjectivity of the observer, answering the recent call of Johanna Drucker (2011): ‘Rendering *observation* (the act of creating a statistical, empirical, or subjective account or image) as if it were *the same as the phenomena observed* collapses the critical distance between the phenomenal world and its interpretation, undoing the basis of interpretation on which humanistic knowledge production is based.’ While much of the data collection was straightforward, much of it required some level of interpretation, and it was important that this process was done with full knowledge of the coder’s *subjectivity* rather than their *objectivity*. This emphasis on being aware of one’s subjectivity in the coding process parallels the reflexive turn in anthropology, folklore, and other social sciences occurring in the 1970s and 1980s. As Nancy Scheper-Hughes states: ‘While the anthropologist is always a necessarily flawed and biased instrument of cultural translation, like every other craftsman we do the best we can with the limited resources we have at hand’ (1995, pp. 417–18). Working with an appreciation of one’s role in the process of

interpretation has thus become a recognized component of writing culture within social sciences disciplines.

This hand-coding method allowed us to move beyond traditional ‘count the word’ computational studies. While that approach can be powerful on a huge corpus of data, as has been shown with Google Ngrams (Michel *et al.*, 2011), it often comes at the expense of being able to apply a researcher’s own theoretical lens to the data. When the researcher is not able to control for what variables are presented, and how they were collected, a research project becomes enslaved to a limited dataset.

Our data also allowed us to find correlations and trends in these body representations across gender, nationality, status, age, and so forth. Importantly, it allowed us to create tables and visualizations which could be used as *tools for discovery*. None of these statistics by themselves are meaningful, but they allow us to find anomalies and unexpected trends easier and faster, and they suggest further areas of inquiry into which a domain expert can then look more thoroughly.⁷

These questions—how disciplinary perspectives influence methodology, and how to handle exploratory tools—are crucial not only for scholars working in, or peripherally too, the digital humanities, but also for scholars in not-so-digital disciplines. This point was recently driven home for folklorists when literary Darwinist Jonathan Gottschall chose to examine the evidence for universal sex differences in mate selection using folktales as his data, citing the reason that they are simple narratives that provide researchers ‘relatively direct access to the aspects of lives and ways of the traditional populations they are most interested in studying and upon whom they have the most difficulty gathering data’ (Gottschall *et al.*, 2003, p. 377). Treating folktales as universal and direct lines of access to what the ‘folk’ really think and believe, however, is problematic. According to criticism from folklorist Donald Haase, Gottschall’s methodology relies ‘on the assumption that all the published texts have oral origins, that orality is pure and natural, and that this natural original essentially survives intact and defines each text’s unequivocal and primary level of significance’ regardless of ‘the language of that

text and despite whatever mediation, alteration, or appropriation might occur at the hands of collectors, editors, and translators’ (2010, p. 21). Thus, while Gottschall’s attempt to apply computational methods to folktales is admirable, his results are based upon faulty assumptions and thus lose their utility.

The potential meanings of beauty in folk and fairy tales are myriad, from metaphorical expressions of inner worth and messages about gendered socialization to evaluations of good mating material. This latter perspective is summarized by Gottschall (which he hopes to prove as universally true by digitally examining beauty trends in folktales and fairy tales): ‘Men place great value on female physical attractiveness because it is a trustworthy indicator of relative fertility’ (2008, p. 176). Men’s physical attractiveness is less important than women’s because ‘male fertility is much less variable than female, and those variances are much more difficult to detect’ and because of the difficulties inherent in raising human young: ‘women—unlike most female mammals—must balance preferences for physically attractive mates with preferences for parentally investing mates’ (Gottschall *et al.*, 2008, p. 176).⁸ While our analysis has corroborated some of Gottschall’s findings—such as the higher incidence of young characters than old, and the importance of beauty to female characters—his explanation is biologically reductionist, and does not take into account either the constructed nature of gender or the fact that the tales, like any form of expressive culture, are filtered through multiple perspectives. As Indian folklorist A. K. Ramanujan notes, ‘literature refracts as much as it reflects’, and to use literary materials ‘in a literal straightforward fashion is to misuse them, or use them only to illustrate what we know already through other means’ (1999, p. 52). That seems to be the case with Gottschall’s analysis: while his application of digital methods to folkloric materials is innovative, he lacks the cultural sensitivity to properly frame his study or interpret his findings. The tales do not necessarily depict men’s and women’s gender roles as so entirely different because they *are* that way in reality, but rather because the tales are an important socializing force in ensuring that gender is properly

performed, and this is done by inscribing cultural values upon women's bodies.

The importance of bodies—especially gendered ones—in fairy tales is monumental. Feminist scholars often consider the female characters in fairy tales to be too passive, pretty, and domestic (if protagonists), or alternately too wicked, ugly, and vicious (if antagonists). Laurence Talairach-Vielmas writes of Victorian fairy tales and the earlier fairy tales they draw on, 'what we generally learn as children is that all princesses are beautiful and may even try to improve their beauty. In fact, their beauty is their wealth—quite literally, since being beautiful enables them to win a prince and a fortune' (2007, p. 5). Beauty is linked with not only success in fairy tales, but also with character. Marcia Lieberman, in one of the earliest feminist criticisms of fairy tales, writes: 'Good-temper and meekness are so regularly associated with beauty, and ill-temper with ugliness, that this in itself must influence children's expectations' (1972, p. 385). Thus, beauty is a problem in and of itself in fairy tales, but its associations with other supposedly female traits are thought to be problematic as well. Kay Stone concludes, based on her fieldwork on the reception of fairy tales among contemporary American audiences: 'Thus the message of the Cinderella story that seems most relevant for modern girls and women concerns the rewards one is supposed to receive for being pretty, polite, and passive; the primary reward, of course, is marriage, and marriage not just to anyone but to a "prince"' (1985, p. 136). Lori Baker-Sperry and Liz Grauerholz conclude that 'that messages in the Grimms' fairy tales, especially those that have been reproduced often, are consistent with other messages women and girls receive about the importance of feminine beauty' (2003, p. 724). Thus, fairy tales are but one outlet of Western society insisting on the importance and imbrication of traits that are supposedly inherently feminine: beauty, goodness, passivity, dependence, and an affinity for the domestic sphere. And with the prevalence of fairy tales in Western popular culture, especially due to the American mass media productions of Disney, the importance of fairy tales' potential effects upon children's notions of gender roles and body image cannot be underestimated.

2 Data Analysis

Once the database had been coded, the first step was to calculate basic statistics. These are simple but important raw numbers of how many references there were to each gender, age, and so forth; and also counts of the use of each adjective or noun associated with a body. Initially, we wondered which body parts were used in fairy tales most frequently, or at all. Which body parts are important enough to mention in connection with a fairy-tale plot, or as a narrator aside? Which body parts even existed in the popular imagination around the time when fairy tales were first being recorded? In our count of distinct, explicitly mentioned body parts, there were 139 different body parts mentioned (this count separates singular and plural instances, so *eye* and *eyes* are listed in separate rows). The top ten body parts listed are: *heart* (295 mentions), *eyes* (294), *head* (259), *hand* (249), *hair* (204), *hands* (194), *face* (139), *feet* (119), *tears* (115), and *blood* (110) (Fig. 1). Note that besides *heart* and *blood* all the other top-ten body parts are external: they are visible where they reside on the body, and can be observed visually; that is, vision is the main sense used to perceive them. This trend continues with the next ten most commonly described body parts: *arms* (95), *foot* (78), *mouth* (78), *finger* (77), *beard* (56), *form* (54), *shoulders* (52), *neck* (50), *back* (49), and *body* (47).

Some patterns immediately stand out: most of these body parts are external, but they convey features of identity and enable motion. A person's eyes, face, and hair are important identifying features, both in our world and in the fairy-tale world: heroes and heroines are the fairest in the land, which one assumes includes beautiful facial features, and hair is often described as golden or otherwise remarkable. A person's foot or feet help them move throughout the world, while their arms, hand or hands, shoulders, and fingers allow them to interact with the world tactilely, touching and manipulating the objects around them. The most frequently mentioned body parts in fairy tales, then, give us a sense of characters that have the limbs to explore the world, and the facial features to express information about their identities. It is

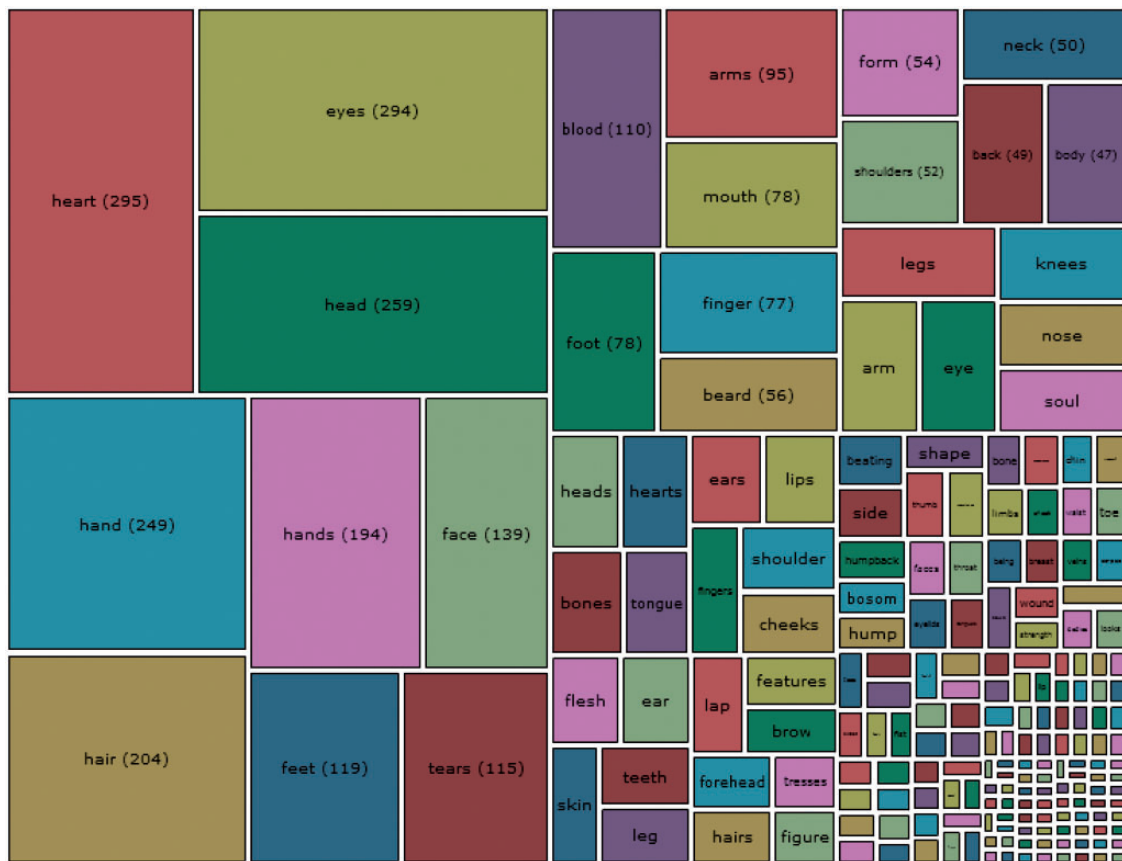


Fig. 1 List of body parts used in tales, with the areas of each box corresponding to the ratio of word usage to the total.

also worth noting that the top twenty body part nouns make up the majority of the references to body parts (fewer than 10% of the nouns are used approximately 65% of the time, exhibiting common Zipf-like properties), indicating that different parts of the body are not given equal weight in fairy-tale descriptions.

Adjectival descriptions serve to deepen our understandings of the physical appearances of characters, in order to paint a more detailed visual picture to help audiences imagine and sympathize with them, but also to elaborate on their inner states. Additionally, adjectives offer information about the values ascribed to bodies. The ten most commonly used adjectives in the sample are: *old* (689 mentions), *young* (543), *little* (438), *beautiful* (433), *poor* (342), *youngest* (148), *handsome* (125),

dead (114), *dear* (110), and *good* (103) (Fig. 2). As with nouns, adjectives were not lemmatized, so *fair* and *fairest* are counted separately, resulting in 297 distinct adjectives describing body parts. It is significant that three of the top ten deal with age (*old*, *young*, and *youngest*), three with appearance (*little*, *beautiful*, and *handsome*), and three with perceived state of being (*poor*, *dear*, and *good*). The last one, *dead*, deals with existence and its lack. These are significant themes of the fairy tale: the maturation of the protagonist, the obtaining and retaining of attractiveness and wealth, and the struggle to survive, though one's family members and helpers might not.

The next ten most common adjectives reinforce these themes. They are: *oldest* (94 mentions), *second* (77), *golden* (73), *eldest* (67), *ugly* (67), *fair* (61),

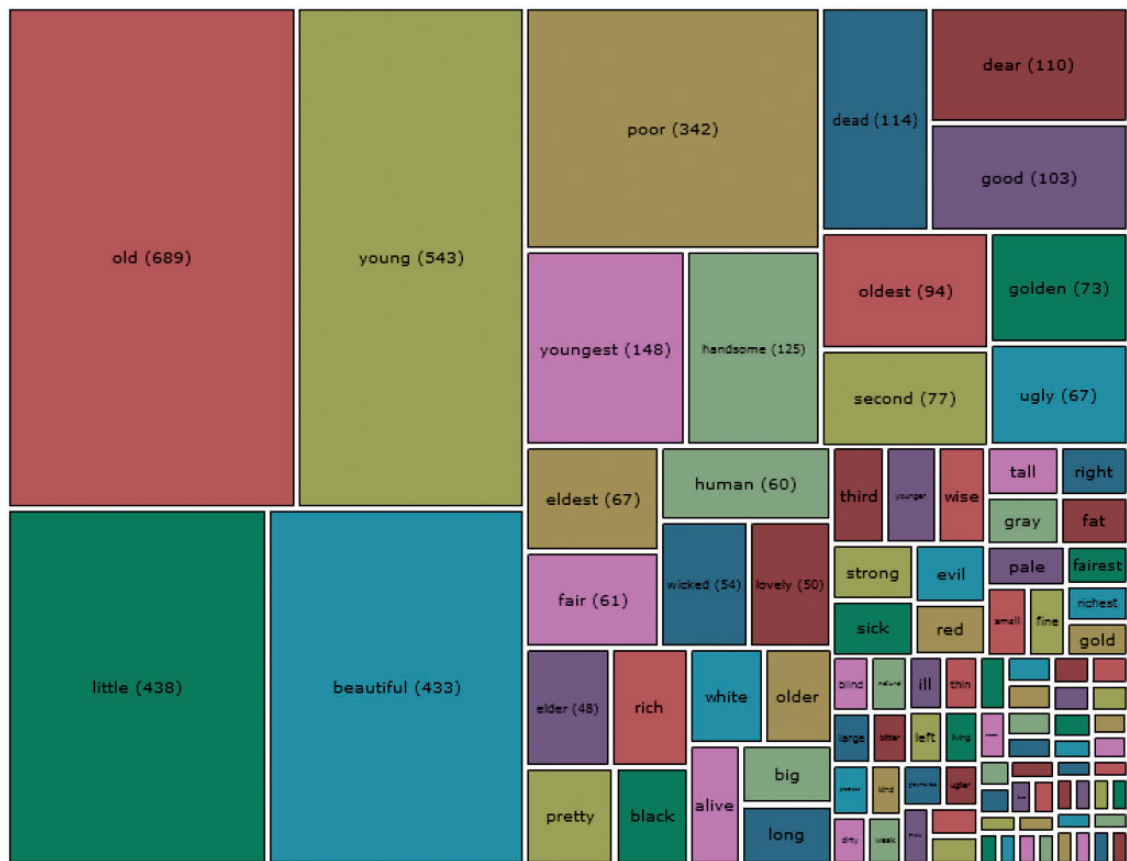


Fig. 2 List of adjectives used in tales, with the areas of each box corresponding to the ratio of word usage to the total.

human (60), *wicked* (54), *lovely* (50), and *elder* (48). Again, there is a preoccupation with age (*oldest*, *eldest*, and *elder*, plus *second*, which usually refers to a second son or daughter), appearance (*golden*, *ugly*, *fair*, and *lovely*), and state of being (*wicked* and *human*). The list of most common nouns also includes many words that correspond to easily visible surfaces, such as facial features. This corresponds to Max Lüthi's assertion that fairy tales are concerned with surfaces rather than depths. He also states, 'the folktale tends to render things and animate beings in metallic or mineral terms' (1986, p. 27), which correlates to this information, as the only color mentioned thus-far has been gold.

Adjectives also serve the function of invisibly ladling out social meaning. Some adjectives that appear fairly straight forward on the surface turn out to

convey additional meanings when analyzed in aggregate, and alongside other textual data. Age is an example of this phenomenon. The adjectives *old* and *young* are the top two in the adjective count, listed at 689 and 543 occurrences, respectively. This indicates that age is an important concern in fairy tales; indeed, age is one of the three oppositional axes that Bengt Holbek uses to describe the identity shifts that occur in fairy tales (tracking the progress from youth to maturity, as well as from low-status to high-status, and the connections between men and women). However, the numbers become even more meaningful when viewed in light of the fact that more bodies are implied to be young than old in fairy tales.

Besides listing explicit uses of the words *young* and *old*, the database also tagged bodies or

body parts that were implicitly described as belonging to someone young or old. In total, there were 2,767 instances of old/mature bodies described, and 7,762 instances of young/youthful bodies described.⁹ This means that nearly three-quarters of the time we see bodies in fairy tales, those bodies are youthful. Returning to the adjective count, in which the word *young* explicitly appeared 543 times, and *old* appeared 689 times, it is clear that there is a discrepancy in how many times old adjectives are used, considering that there are so many fewer old bodies in the tales (Fig. 3). While nearly 75% of references to bodies describe young people, the adjective *young* comprises fewer than half of the age-related adjectives. Compare that against only 25% of body references being to old people, while *old* comprises more than half of age-related adjectives. We interpret this to mean that there's something important about describing fairy-tale characters as old, something noteworthy in the fact that it's worth mentioning at all.

After examining and interpreting simple statistics, we then looked at basic correlations in the data; for example how often abstract nouns like *death* or *beauty* was used in reference to each gender. As expected, *beauty* was used more frequently with women than men, but this allowed us to compare precisely how much more frequently *beautiful* was used with women than *handsome* was used with men (Table 1).¹⁰ These correlations sometimes yielded surprising results, and were used as

launching points to dive further into the data. Looking at these correlations occasionally reinforced (and added to!) previous theories of gender representation. If different types of bodies (old/young, male/female) were described to an equal degree, we would expect the percentage of references to those bodies to be equal to the percentage of adjectives describing them. That is, if 10% of the bodies were old female bodies, we would expect an equal 10% of the adjectives used to describe old females. This, apparently, was not the case. For example, while only 10.7% of references to bodies referred to old females, 14.5% of all adjectives used were attached to old females. While old females were described in more detail than expected, young males were described in less than expected (29.6% of references were to young males, but only 28.2% of adjectives were used to describe them). Female bodies in general were described 2.2% more than expected, and old bodies were described 5.8% more than expected.

Our analysis then led us to word co-occurrence statistics using network representations. What adjectives and nouns were used together most frequently? In Fig. 4, adjectives are shown in relation to certain body categories. Beyond that, were certain clusters or groups of adjectives and nouns more likely to be used with certain categories? That is, would old and young people be described by two very different sets of words? Males and females? How significant was the variation? These questions

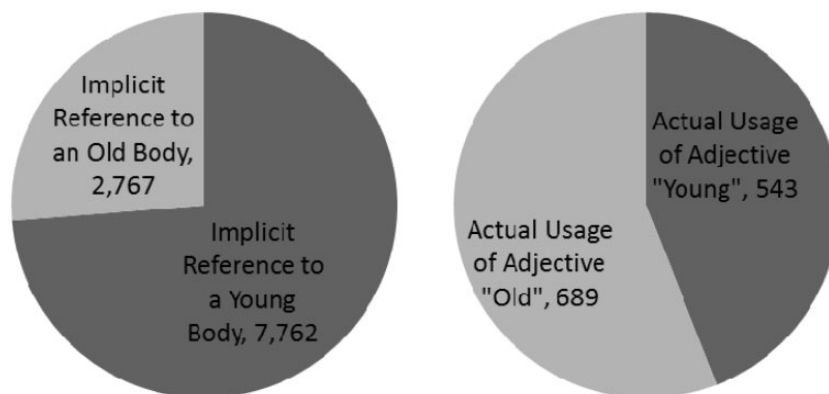


Fig. 3 Comparison of references that implicitly describe bodies as old or young compared to explicit statements that a body is *old* or *young*.

Table 1 Representative sampling of aesthetic words evaluating physical appearances

| Word | High percent (%) | Old percent (%) | Female percent (%) |
|--------------|------------------|-----------------|--------------------|
| Beautiful | 61.24 | 8.18 | 98.35 |
| Beauty | 27.79 | 3.53 | 97.83 |
| Handsome | 68.10 | 2.50 | 5.79 |
| Ugly | 31.25 | 24.19 | 79.37 |
| Fair | 61.67 | 1.67 | 86.89 |
| Lovely | 61.22 | 4.08 | 93.88 |
| Pretty | 48.78 | 2.44 | 90.00 |
| Ugliness | 80.00 | 7.14 | 73.33 |
| Fairest | 80.00 | 40.00 | 100.00 |
| Prettiest | 28.57 | 0.00 | 100.00 |
| Appearance | 87.50 | 12.50 | 87.50 |
| Hideous | 57.14 | 28.57 | 66.67 |
| Uglier | 57.14 | 0.00 | 85.71 |
| Homely | 0.00 | 0.00 | 100.00 |
| Handsomest | 83.33 | 0.00 | 0.00 |
| Loveliest | 100.00 | 0.00 | 100.00 |
| Good-looking | 50.00 | 25.00 | 25.00 |
| Prettier | 0.00 | 0.00 | 100.00 |

For each word, the percentage is shown of how often that word is associated with bodies of high status (versus those of low), old bodies (versus young), and female bodies (versus male). Percentages are shaded if they are over 50%. See Fig. 2 for prevalence of adjectives compared to the entire set.

are answered in Fig. 5, which shows all nouns and adjectives in relation to one another.

The Gender graph in Fig. 5 shows that there isn't a significant clustering of words by gender; for the most part, the same groups of words are used to describe both males and females (even if a few choice words are used solely to describe one gender or the other). Analysis of the database revealed that most words are in fact shared between men and women, with the gender-specific exceptions not being very significant due to their infrequent occurrences. For example, the top two words used solely with women are 'tresses' and 'bosom', used fourteen and ten times, respectively, but in a database consisting of 11,141 entries, these are rather small numbers. In contrast, the age graph in Fig. 5 shows a prominent clustering of young words on the bottom left and old words on the top right. This implies that old people are described significantly differently than young people, even more so than the difference between males and females.

The relationship between age and gender is also visible in Table 1, depicting the use of words in our data set that evaluate physical appearance in terms of aesthetics. Notice that with a few exceptions, almost all of the words correlate more with women than with men, and that the majority of the words have low correlations with old age, meaning that they apply more to the bodies of younger characters. Co-occurrences support this interpretation, and can give more information about how women's bodies are described in terms of traditionally 'beautiful' attributes. For instance, the phrase *golden hair*, reputed to be stock in fairy tales, occurs sixty-six times, out of seventy-three total mentions of the adjective *golden*. All but fourteen of those *golden hair* phrases refer to women; of the fourteen, five apply to the supernaturally enhanced golden hair of the boy in 'Iron Hans' (ATU 502 in the Grimms) and nine apply to the Devil's golden hairs, the object of a quest, in 'The Devil with the Three Golden Hairs' (ATU 461 in the Grimms).¹¹ Thus, males with golden hair get it via supernatural means (either by being supernatural creatures, as with the Devil, or by getting mixed up with supernatural creatures, as with the boy protagonist in 'Iron Hans'). The women with golden hair, it is implied, are naturally that way, and their golden hair is a special marker of beauty.

3 Preliminary Conclusions

Data analysis yielded several interesting trends, as reported in the previous section. Most references to body parts are to ones that are external and visible. The most-used adjectives tend to describe age, appearance, or state of being. A relatively small number of nouns and adjectives make up more than half of the total words used in describing body parts. It is far more likely for an old person to be described as *old* than for a young person to be described as *young*. Old people and females were both described more than expected, given the distribution of nouns and adjectives. The most descriptors were attached to old females, and the least to young males. Male and females are generally described in similar ways, though there are certain

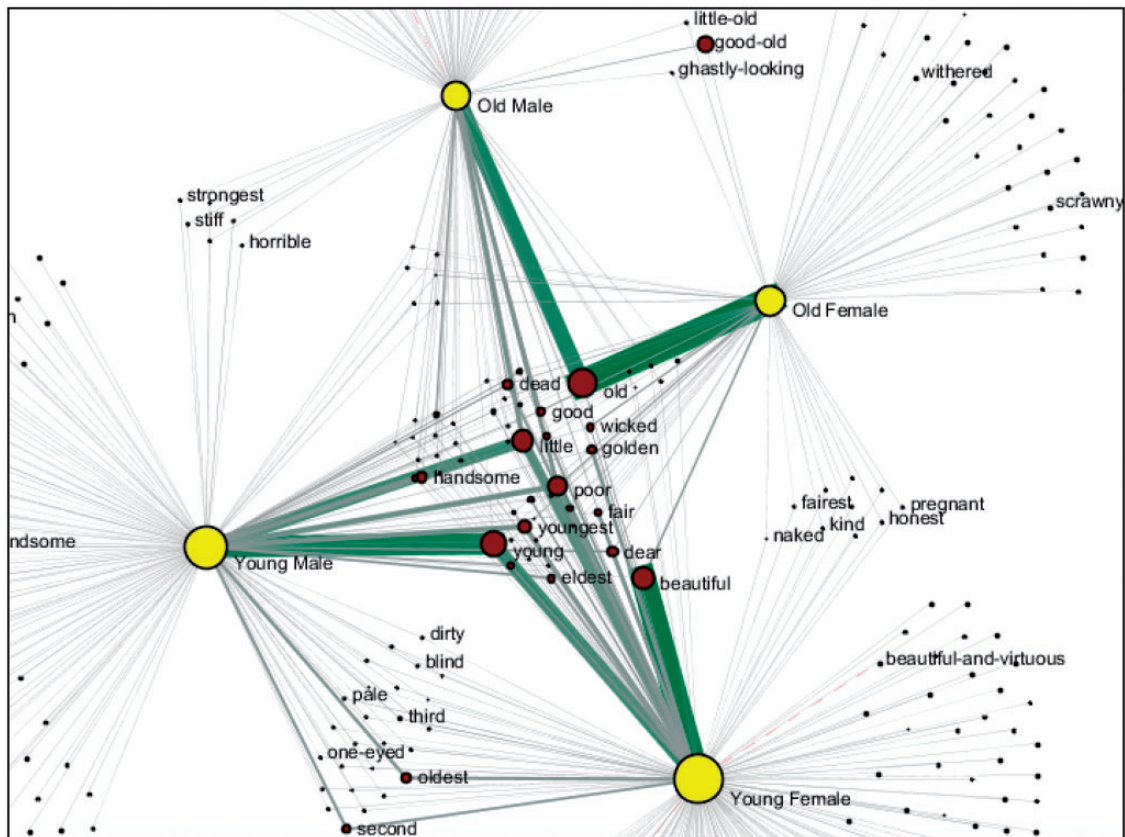


Fig. 4 The four light (yellow) dots represent *old male*, *young male*, *old female*, and *young female*, while each of the dark (red) dots represents an adjective. A line is drawn between an adjective (the dark/red dots) and a body type (the light/yellow dots) if one is used to describe the other. That line gets thicker if it is used more frequently. Thus, a thick line connects *beautiful* to *young female*. Because *beautiful* is in the middle, we can also see it is occasionally used to describe *old females* and *young males*—although never with *old males*.

terms that are used almost exclusively for men or women. In contrast, old and young bodies tend to be very polarized in their descriptions across the board. Aesthetic descriptions are generally attached to women rather than men.

To explain some of these differences, we turned to feminist theory. Second-wave feminists such as Simone de Beauvoir developed the notion of the universal masculine perspective, the idea that in Western culture, the public, unmarked, assumed universal position is in fact specifically male. As Judith Butler (1990) summarizes this argument: 'Beauvoir contends that the female body is marked within masculinist discourse, whereby the masculine body, in its conflation with the universal, remains

unmarked' (1990, p. 17). Our data support this assertion not only in terms of female bodies being marked within fairy tales, but we also believe that the same principle applies to young and old bodies.

Youthful bodies are assumed to be the unmarked universal category in fairy tales. This is not only because bodies are described using the adjective *old* more frequently than *young* (Fig. 3), but also because a number of adjectives cluster in groups by age. We interpret these findings to mean that old bodies must be differentiated in fairy tales, because they are no longer in the supposedly universal category of youth. Old bodies are qualified with more descriptions in order to give audiences a sense of who these characters are, since they don't fall into

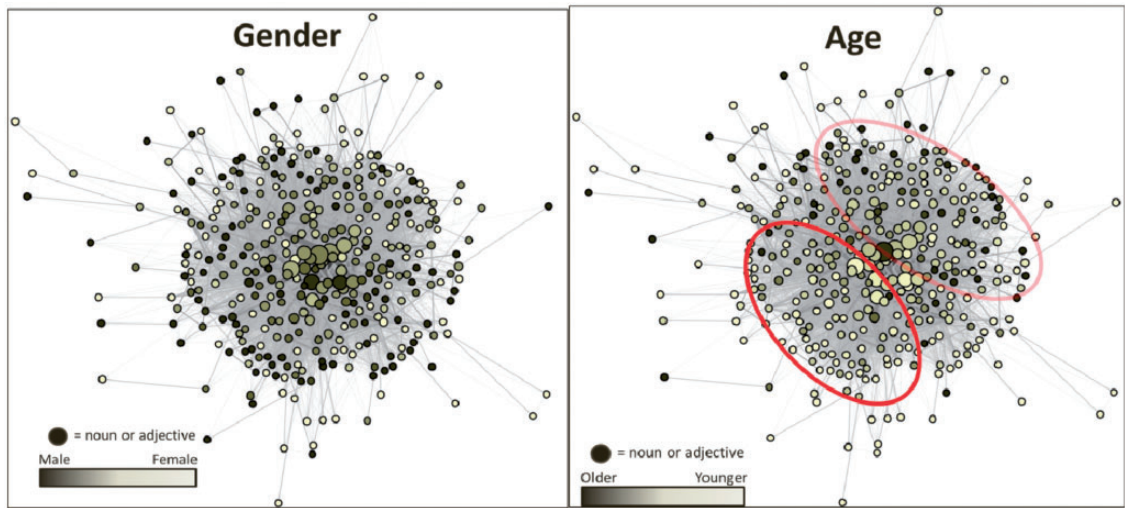


Fig. 5 In these graphs, each dot is a noun or adjective. Lines are drawn between words if they are used together, so if *pretty* and *face* are used together frequently, a thick line connects them. The bigger a dot, the more frequently that word is mentioned. In the gender graph (left), dots are shaded by whether the word is most frequently associated with men (darker) or with women (lighter). In the age graph (right), dots are shaded by whether the word is most frequently associated with old (darker) or young (lighter) bodies.

the category of the youthful protagonist, with whom listeners are supposed to easily identify. This information also supports previous theories in fairy-tale scholarship, such as the notion that the tales are from the child's perspective (Dundes, 1989).

4 Discussion

This study presents a successful model for collaboration in the digital humanities. Combining researchers from two very different backgrounds, but both grounded in humanistic training and ideals, we were able to nurture a continuing process of exploration and discovery. In an environment where both approaches have not been thoroughly explored in tandem, it became essential to bounce a myriad of ideas back and forth, which led to the development of novel questions and methods. We feel this combined approach was important in satisfying both the difficulties Drucker and others have expressed regarding data objectivity and the rigorous expertise requirements of data analytics.

The results we present here are still in preliminary form; they are explored in greater depth in Jorgensen's in-progress dissertation, with interpretations informed by feminist and folkloristic theories. In Jorgensen's write-up, there are strong indications that the gender and age of fairy-tale protagonists correlate in ways that indicate societal value being placed on certain perspectives. These results might help explain important questions in folkloristics, such as the gender bias noted by fairy-tale collectors, wherein the repertoires of male tellers contain mostly tales with male protagonists yet the repertoires of female tellers are evenly split down the middle (Holbek, 1998). In other words, if the presence of normalized masculine bodies indicates a more pervasive masculine perspective in fairy tales, it would make sense that male tellers gravitate more toward tales that feature their own point of view. Another topic to explore is mind/body dualism, wherein men are associated with the mind and women with the body. Thus far, the fact that men seem to undergo more transformations than women, along with the fact that women are described more than men with

appearance-evaluative words, leads us to think the fairy tales express rather than subvert the often-sexist mind/body dualism common to other aspects of Western culture.

Interdisciplinary scholars, led by folklorists and feminists, have studied the configuration of gender roles in fairy tales since the 1970s, trying to ascertain their influence on children, and society more generally. Folklorists have been slow to adopt digital methodologies for interpreting expressive culture, and no folklorist to our knowledge has used this kind of visual data modeling to study the linguistic content of fairy tales before, so this study is uniquely positioned to help us understand the dynamics of gender in the pervasive and influential genre of fairy tales.

Acknowledgements

We wish to thank the two anonymous reviewers for their constructive feedback and excellent suggestions, as well as the Alliance of Digital Humanities Organizations for awarding this paper the Paul Fortier Prize for the best young scholar paper at the 2011 Digital Humanities Conference in Stanford, California.

Funding

This work was supported by the University of Tartu [DoRa Scholarship to J.J.] and the National Science Foundation [Graduate Research Fellowship to S.W.].

References

- Baker-Sperry, L. and Grauerholz, L.** (2003). The pervasiveness and persistence of the feminine beauty ideal in children's fairy tales. *Gender and Society*, 17(5): 711–26.
- Bottigheimer, R.** (1986). Silenced women in the Grimms' tales: the 'fit' between fairy tales and society in their historical context. In Bottigheimer, R. (ed.), *Fairy Tales and Society: Illusion, Allusion, and Paradigm*. Philadelphia: University of Pennsylvania Press, pp. 115–31.
- Butler, J.** (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Crane, T.** (2001). *Italian Popular Tales*. Ed. and with an introduction by Jack Zipes. Santa Barbara, Denver, and Oxford: ABC-CLIO, Inc.
- Danielson, L.** (1979). Toward the analysis of vernacular texts: the supernatural narrative in oral and popular print sources. *Journal of the Folklore Institute*, 16(3): 130–54.
- Drucker, J.** (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(001). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (accessed 15 September 2011).
- Dundes, A.** (1989). The psychoanalytic study of the Grimms' tales: 'The Maiden Without Hands' (AT 706). In Dundes, A. (ed.), *Folklore Matters*. Knoxville: The University of Tennessee Press, pp. 112–50.
- Falassi, A.** (1980). *Folklore by the Fireside: Text and Context of the Tuscan*. Veglia, Austin: University of Texas Press.
- Gottschall, J., Berkey, R., Cawson, M. et al.** (2003). Patterns of characterization in folktales across geographic regions and levels of cultural complexity: literature as a neglected source of quantitative data. *Human Nature*, 14(4): 365–82.
- Gottschall, J., Anderson, K., Burbank, C. et al.** (2008). The 'beauty myth' is no myth: emphasis on male-female attractiveness in world folktales. *Human Nature*, 19(2): 174–88.
- Grimm, W. and Grimm, J.** (2003). *The Collected Fairy Tales of the Brothers Grimm. Third Edition*. Trans. Jack Zipes. New York: Bantam Books.
- Haase, D.** (2010). Decolonizing fairy-tale studies. *Marvels & Tales*, 24(1): 17–38.
- Holbek, B.** (1998). *Interpretation of Fairy Tales, Folklore Fellows Communications*. Helsinki: Academia Scientiarum Fennica.
- Jones, L. C.** (1944). The ghosts of New York: an analytic study. *Journal of American Folklore*, 57(226): 237–54.
- Lieberman, M.** (1972). 'Some Day My Prince Will Come': female acculturation through the fairy tale. *College English*, 34(3): 383–95.
- Lüthi, M.** (1986). *The European Folktale: Form and Nature 1982*. Trans. John D. Niles. Bloomington & Indianapolis: Indiana University Press.
- Massignon, G.** (1968). *Folktales of France*. Trans. Jacqueline Hyland. Chicago: The University of Chicago Press.

- Michel, J., Shen, Y. K., Aiden, A. P. *et al.* (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014): 176–82.
- Ragan, K. (2009). What happened to the heroines in folktales? An analysis by gender of a multicultural sample of published folktales collected from storytellers. *Marvels & Tales*, 23(2): 227–47.
- Ramanujan, A. K. (1999). *The Collected Essays of A.K. Ramanujan*. In Dharwadker, V. (ed.), Oxford: Oxford University Press.
- Ranke, K. (1966). *Folktales of Germany*. Trans. Lotte Baumann. Chicago: University of Chicago Press.
- Sackett, S. J. (1970). Using a computer on a belief collection. *Western Folklore*, 29(2): 105–10.
- Scheper-Hughes, N. (1995). The primacy of the ethical: propositions for a militant anthropology. *Current Anthropology*, 36(3): 409–20.
- Stone, K. (1985). The misuses of enchantment: controversies on the significance of fairy tales. In Rosan, J. A. and Kalčík, S. J. (eds), *Women's Folklore, Women's Culture*. Philadelphia: University of Pennsylvania Press, pp. 125–45.
- Talairach-Vielmas, L. (2007). *Moulding the Female Body in Victorian Fairy Tales and Sensation Novels*. Aldershot, England: Ashgate Publishing, Ltd.
- Tangherlini, T. (1998). 'Who Ya Gonna Call?': Ministers and the mediation of ghostly threat in Danish legend tradition. *Western Folklore*, 57(2–3): 153–78.
- Tangherlini, T. (2010). Legendary performances: folklore, repertoire, and mapping. *Ethnographia Europaea*, 40: 103–15.
- Zipes, J. (1993). *The Trials and Tribulations of Little Red Riding Hood*. New York: Routledge.
- Zipes, J. (2009 [1989]). *Beauties, Beasts, and Enchantment: Classic French Fairy Tales*. Kent, UK: Crescent Moon Publishing.
- quantitative approach in her essay 'Silenced Women in the Grimms' Tales,' listing tales in which female characters lose their voices, and counting the distribution of the word 'speak,' noting that it 'appears more often in conjunction with authority figures' (1986, p. 126).
2. This can be seen in the following two examples from the 1970s. Samuel J. Sackett encoded content and context information from a superstition collection onto punch cards, and although preparing the cards prior to running computer analysis on them was quite time-consuming, he noted: 'We will still have plenty of work to do, and plenty of decisions to make, but our load will be inestimably lightened by the computer' (1970, p. 106). Larry Danielson used punch cards to encode information about the content, structures, and styles of 143 supernatural narratives from both folk and print sources, aiming to obtain statistical data about the 'influences of topic on structural, stylistic, and content variables' (1979, p. 132).
 3. For successful examples, see Tangherlini (1998); Tangherlini (2010); and Ragan (2009).
 4. Because fairy tales exist on a spectrum from oral to written, it is important to strike a balance between folkloric and literary sources. This strategy ensures that Jorgensen's expertise as a folklorist will lead to proper handling of disciplinary materials, as it has long been lamented that scholars unfamiliar with folkloristics ignore non-literary versions of tales (Dundes, 1989, 119). The collections used are: *Beauties, Beasts, and Enchantment: Classic French Fairy Tales* translated by Jack Zipes (covering the classical French tradition from the 1690s onward), *The Collected Fairy Tales of the Brothers Grimm* translated by Jack Zipes (based on the 1857 edition of the Grimms' tales with additions from their notes), *Italian Popular Tales* compiled and translated by Thomas Crane and edited by Jack Zipes (spanning Italian literary and folk tales from the 1800s), *Folktales of France* edited and translated by Geneviève Massignon (tales that she and others collected in the 1950s), *Folktales of Germany* edited and translated by Kurt Ranke (compiled from fieldwork collections from approximately 1850–1950), and *Folklore by the Fireside* by Alessandro Falassi (tales that he collected in Tuscany in the 1970s and translated). In each collection, Jorgensen selected only proper fairy tales or *zaubermärchen* to analyze (those numbering 300–749 in the ATU tale type system, which is described in note 11, with a few exceptions for tales that fall outside that category but are still clearly tales of magic), discarding the legends and animal tales that appear in some of these collections.

Notes

1. For instance, Louis C. Jones analyzed ghost tales from New York City statistically, finding 'that revenants are either indifferent (fifty-eight percent) malevolent (thirteen percent) or benevolent (twenty-nine percent)' (1944, p. 246). He notes, however, 'These categories are entirely arbitrary, based upon my judgment of their actions' (Jones, 1944). Another good example is Ruth Bottigheimer's work on gender roles and speech acts in the Grimms' tales. She hints at a

We should also note that while some scholars use the terms ‘fairy tale’ and ‘folktale’ interchangeably, we follow the folkloristic convention of using folktale as a more general category, referring to (usually) oral narratives that are fictional and formulaic, while fairy tales, or *zaubermärchen* in German, have closer ties to literary culture and tend to focus upon magical aid and quests. As we are particularly concerned with the forms the fairy tale has taken in contemporary American culture (and scholarship) while at the same time reflecting general Western ideas about gender and the body, we work with tales in translation.

Where possible, we use translations done by scholars who are folklorists, guaranteeing some amount of cultural sensitivity to context. Additionally, it is possible to treat translations like any other folkloric text: an equally valid variant. Folklore is always in flux, always being translated between languages and cultures, so working with tales in translation is acceptable so long as the translation is trusted.

5. Tale, Collection, Author, Teller, Collector, Year of Writing/Collecting, Year of Publication, Tale Type, Region, Original Language, Gender of Teller/Writer, Gender of Collector, Gender of Editor, Gender of Protagonist.
6. Some of these points were fairly evident in the texts and required little interpretation (Noun, Adjective, Surrounding Text, Page Number, Gender, Young/Old, High/Low, Quoted Speech, Skin Tone), and some required more interpretation (Positive/Negative value, Grotesque, Violence, Nudity, Move). ‘High’ and ‘Low’ refer to social status, while ‘Young’ and ‘Old’ refer to the character’s age: youthful characters are of child-bearing age or younger, while old characters are beyond child-bearing years, or are the parents of children that are already mature. The ‘Move’ category is drawn from Holbek’s modification of Vladimir Propp’s 31 tale functions. Holbek condensed the thirty-one functions into five structural groupings, or moves. This is one of the unique aspects of this project, and one of the reasons folklorists should be doing digital work on fairy tales, because we can draw on the insights from our discipline to sift through the data better than an outsider.
7. For this reason, we avoid common statistical methods such as calculating *P*-values. This is preliminary research, and we hope to use these results to drive further inquiry and exploration into the data; as such, we do not want to limit ourselves by arbitrary *P*-value cutoffs. Apparent evidence of a trend will inspire deeper statistical exploration in future projects.
8. Despite the fact that this narrative is founded on a number of faulty assumptions—such as the supposed universality of marriage in human societies, preoccupation with paternal investment, and selfishness lying at the heart of human interactions—those utilizing an evolutionary psychology or sociobiology paradigm interpret everything through these lenses. For a refreshing counter-argument, see *Sex at Dawn*, a book committed to demonstrating that many of the foundations of evolutionary psychology are outdated and mired in the social constructs of the time periods when they were thought up: Ryan, C. and Jethá, C. (2010). *Sex at Dawn: The Prehistoric Origins of Modern Sexuality*. New York: Harper-Collins.
9. We have excluded a number of bodies that were either ambiguously aged or had no age (578), bodies that were described in the plural as both young and old (32) and bodies that were young but were being acted upon by older bodies (2).
10. Although many unique words in the dataset stand for very similar or overlapping body parts or concepts, we believe keeping them separate will reveal subtle variations in usage. Oftentimes statistical research aggregates data to the highest meaningful level; because we are not yet certain which effects subtle changes in word usage might prove important to, that highest meaningful level is the exact word itself.
11. ATU refers to the Aarne–Thompson–Uther tale type system, a method of cataloguing folktale plots as they travel across linguistic and ethnic borders (for instance, ‘Cinderella’, one of the most widely told folktales, does not bear the same title in every culture where it is told, so it is useful to have a numerical indexing system). For the latest revisions of the tale type index, see Uther, H.-J. (2004). *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson*. Helsinki: Academia Scientiarum Fennica. 3 volumes.