THEO MEDER, FOLGERT KARSDORP, DONG NGUYEN,
MARIËT THEUNE, DOLF TRIESCHNIGG, AND
IWE EVERHARDUS CHRISTIAAN MUISER

Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database

This paper describes the development of the Dutch Folktale Database as a digital archive of intangible heritage and a sophisticated research instrument. Current research focuses on automating the assignment of metadata to folktales and on obtaining a better understanding of classifications of folktales into types, as well as an understanding of motif sequences. These studies are useful for future fundamental research into patterns and structures, such as historical and geographical variation in oral tradition, on an international scale.

Keywords

AFS ETHNOGRAPHIC THESAURUS: Folktales, digital archiving, classification, variants, motifs

THERE IS AN INCREASING INTEREST in creating digital collections of folk narratives because of the potential to make large numbers of narratives accessible to scholars and—if properly annotated—to allow investigators to search on dimensions of interest such as theme and location (La Barre and Tilley 2012). In addition, by applying computational analysis and visualization methods to such collections, previously unperceived patterns may be revealed, leading to new insights (Abello, Broadwell, and Tangherlini 2012).

In this paper, we focus on one such digital collection: the Dutch Folktale Database, an online collection of folktales maintained by the Meertens Instituut in Amsterdam. The Dutch Folktale Database serves both as a digital archive and as a digital research instrument. To enhance its usefulness as a resource for scholars, the folktales in the collection have been annotated with metadata, ranging from the name of the narrator

THEO MEDER is a senior researcher, Meertens Instituut, Amsterdam
FOLGERT KARSDORP is a PhD candidate, Meertens Instituut, Amsterdam
DONG NGUYEN is a PhD candidate, University of Twente
MARIËT THEUNE is Assistant Professor, University of Twente
DOLF TRIESCHNIGG is a postdoc researcher, Mydatafactory

IWE EVERHARDUS CHRISTIAAN MUISER is a scientific programmer, Database Group of the University of Twente

All are involved in folk narrative research at the Meertens Instituut in Amsterdam

Journal of American Folklore 129(511):76–94

Copyright © 2016 by the Board of Trustees of the University of Illinois

to the international tale type. So far, these metadata have been assigned manually, a process that is both time-consuming and prone to errors. Recently, we have started to investigate the possibility of automating this process using various machine learning techniques that enable the computer to automatically learn from data to, for example, generate metadata on its own. In addition, we have experimented with the computational modeling of motifs in our collection. In the following sections, we present the primary results of these efforts, and discuss the implications for future research. First, we give an overview of the Dutch Folktale Database and its development over the years.

The Dutch Folktale Database 1.0–3.0

The Dutch Folktale Database (DFD) was first developed as a stand-alone database in 1994; it was mainly used by researchers at and visitors to the Meertens Instituut in Amsterdam. The database was designed to offer insight into folktale¹ traditions in the Netherlands, and to provide access to extant synchronic and diachronic variations. To meet this goal, complete story texts gathered from written historical sources or transcribed from audio recordings were included in the database. Relevant metadata were added to each individual folktale: Who was the narrator? Where was the story told? When was the story told? In what language or dialect was the story told? Who collected the story, and how was it recorded? Other metadata used to enrich the tales include a list of personal names, place names, keywords, a brief summary of the tale in modern Dutch, the genre and tale type numbers where appropriate, and motifs in the story according to Stith Thompson's *Motif-Index*.

The most important international catalog used initially to identify stories in the DFD was *The Types of the Folktale* by Antti Aarne and Stith Thompson (1964). This index works appropriately for classifying widespread, traditional fairy tales and anecdotes, but not for local tales, anecdotes, traditional legends, modern jokes, and urban legends. For identification of folktales within these genres, it was necessary to use additional catalogs such as (in order of importance): the Frisian folktales compiled by Jurjen van der Kooi (1984), the traditional legends and saints' legends compiled by Jacques Rudolf Willem Sinninghe (1943), the urban legends compiled by Jan Harold Brunvand (1994), and an additional general index compiled by Theo Meder.² The coverage of these additional catalogs is too limited to identify all stories; in many cases, it is impossible to classify unique stories, in particular family stories or personal narratives with a tale-type designation. In 2012, our group estimated that it was possible to identify 60 percent of the stories in the database with the folktale indices, leaving 40 percent of the tales unclassified.³ As the DFD 2.0 was being developed, the new index *Types of International Folktales* by Hans-Jörg Uther (2004), an expanded and updated index based on Aarne-Thompson, appeared. Rather than discarding the earlier AT-numbering, the database moved to ATU as the new tale type indexing standard. With this new index, new story types were added (such as ATU 777 "The Flying Dutchman"), while other types and subtypes were deleted or combined.

Originally, the folktales that were entered into the database were also labeled with motifs—narrative building blocks—based on the *Motif-Index* developed by Thompson

(1955–1958). However, those familiar with this resource understand that finding and assigning motifs is both tedious and cost-intensive, and is no longer common editorial practice for folktale collections. The search for motifs is time-consuming and requires great precision; it is much easier to find folktales through searches on keywords, names, or tale types. Consequently, the DFD group discontinued assigning motifs in 1996.

As the database grew, the *Contes de ma mère l'Oye* by Charles Perrault and the *Kinder- und Hausmärchen* by the Brothers Grimm were added to facilitate comparative study. Sound recordings, pictures (cartoons, “Photoshop-lore”), and videos were also added. In addition, a Lexicon of Folktales jointly written by researchers and students was added as an interpretative component providing background detail on the folktales in the database. The first entries in this lexicon targeted traditional fairy tales,⁴ and resulted in the publication of the encyclopedic book *Van Aladdin tot Zwaan Kleef Aan* (Dekker, Kooi, and Meder 1997). The Lexicon was subsequently expanded to include entries on Dutch traditional legends, saints’ legends, jokes, and contemporary legends, and now contains over 250 entries.

In 2004, the database was made accessible online for the general public (www.verhalenbank.nl). Since then, the number of pageviews by visitors has fluctuated between half a million and 3 million pageviews per year, depending on the attention the database has received in the media. From questions and responses by e-mail, we have developed an overview of visitors with several large categories comprising the

Figures 1–3. The Dutch Folktale Database 1.0, 2.0, and 3.0: the stand-alone version of FileMaker Pro (1994–2004), the first online version (2004–2013), and the second online version using the Omeka platform since 2013. The example is always a detail of a Little Red Riding Hood version (AT/ATU 333) as told by Anders Bijma in 1971.

Details ABDMA22

© 2004

Meertens Instituut

zoeken

geavanceerd zoeken

Meer informatie:

tekst

Grimm

Perrault

beeld

geluid

Aarne Thompson

AT Uther

lexicon

editen

hier

dupliceer

hier

alle verhalen van verteller: Bijma, Anders

alle verhalen van dit type: AT 0333

alle verhalen van het ATU-type: ATU 0333

verspreidingskaart van dit type: AT 0333

uniek id

ABDMA22

ATU type

ATU 0333

ATU omschrijving

Little Red Riding Hood

AT type

AT 0333

AT omschrijving

The Glutton (Red Riding Hood)

titel

Roodkapje

notulist

Y. Poortinga

taal

Fries (Woudfries)

schriftbron

Transcriptie Y. Poortinga, 28-9-1971; hs. Poortinga, archief Fryske Akademy, in bruikleen bij Frysk Letterkundich Museum en Dokumintaeisintrum, Ljouwert

plaats van vertellen

Boelenslaan (Friesland)

verteller

Bijma, Anders

plaats van handelen

datering

28 september 1971

literair

nee

subgenre

sprokje

motieven

J21.5 & K2011 & Z18.1 & F911.3 & F913 & Q426

samenvatting

Een meisje, Roodkapje geheten, moet eten naar haar zieke grootmoeder brengen in het bos. Als zij aan het touwtje trekt en de deur opent, ziet zij grootmoeder in bed liggen. Zij maakt opmerkingen over de grootte van haar ogen, neus en mond. Het is de grootmoeder niet, maar een wolf die het meisje in één hap opeet. Zodra grootmoeder thuiskomt ziet ze de wolf in haar bed liggen. Zij gaat de jager erbij halen die de buik van de wolf opent. Het meisje komt levend weer naar buiten. De buik wordt gevuld met stenen. Als de dorstige wolf uit de rivier wil drinken, valt hij erin door het gewicht van de stenen en verdrikt. Het meisje keert terug bij haar ouders.

trefwoorden

Vader, moeder, ouders, meisje, dochter, grootmoeder, ziekte, bos, huis, touwtje, trekken, deur, openen, bed, grootte, omvang, ogen, zien, neus, ruken, mond, eten, geweld, wolf, jager, buik, leven, redden, stenen, straf, dorst, rivier, verdrikt, dood, vertellen, fout, achtervolgen, vermommen, school

namen

Roodkapje, Anders Bijma, vrouw Bijma [Attje Bijma-Tuinstra]

Figure 2

Bron

Transcriptie Y. Poortinga, 28-9-1971; hs. Poortinga, archief Fryske Akademy, in bruikleen bij Frysk Letterkundich Museum en Dokumintaeisintrum, Ljouwert

Datum

1971-09-28 1971-09-28

Rechten

ja

Bereik

Boelenslaan (Friesland)

Tekst

Roodkapje
Der wiene in man en in frou dy hiene in famke. En dat famke hjitte Roodkapje. Mar sy hie in beppe en dy siik. En doe moast Roodkapje mei iten dan deis nei beppe ta. Sy kaem by beppe en dêr siet de doar fêst. Doe sei se: Trek maar aan het touwtje. Dan gaat de deur vanzelf open. Sy der yn en sy komt by 't bêd, doe sei se: Wat heeft grootmoeder grote ogen! - Nammensto beter kan 'k sien. Doe sei se: Wat heeft grootmoeder een grote neus. - Nog beter kan ik ruiken. Doe sei se: Wat heeft grootmoeder een grote mond. - Daar kan 'k nog beter mee happe. En dêr docht se in hap en dêr hapt se Roodkapje op. En doe wie 't in hiele dikke greute wolf en dy lei yn beppe har bêd. De echte beppe moest oer de heide en sy komt thús en dy hearde it en dêr wie in jager en dy siet dy wolf en de buk iepen en dêr komt Roodkapje wer út. En hwat de jager hat te dwaen? Dy dy wolf fol mei stienen en dy naet him wer dicht. En de wolf hie sa'n dorst en dy nei de rivier ta en dêr de stienen nei loaren en dêr fonsjpt de wolf net yn de rivier Mar men en hat hiene Roodkapje wer.
Frou Bijma: Dú haast it net goed forteld. Dy wolf waard it gewaar. Dy gong har achternei.
Anders Bijma: O ja.
Frou Bijma: En doe hearde er hwat sy sei.
Anders: Ja.
Frou Bijma: Dat grootmoeder sei: Trek maar aan het touwtje, en doe gong hy letter ek binne en doe die hy itselde. Dú haast it net goed forteld.

Bestanden

00:00

-03:17

Figure 3

majority of these visitors: researchers from the humanities carrying out specialized studies, students and pupils writing papers or reports, journalists checking whether a story is an urban legend or not, relatives of deceased narrators interested in their forebear’s repertoire, storytellers looking for specific themes or new stories to augment their repertoires, people searching for local stories, and people generally interested

in folktales.⁵ These impressions based on e-mail contacts were confirmed by the results from an opt-in questionnaire posted on the homepage of the website, which was completed by 88 visitors between June 2012 and June 2013. Most of these visitors (56 percent) indicated that they were interested in the folktales for personal use. Another large group of respondents (30 percent) used the folktales for the purpose of storytelling, while 17 percent of the visitors reported to have scholarly purposes, and 14 percent of the respondents reported to have educational purposes (Trieschnigg, Nguyen, and Meder 2013).

In its early years, the Folktale Database was akin to a digital museum of Dutch folktales displaying variants of stories. With the addition of rich metadata, the archive developed an encyclopedic value and offered the opportunity for scientific research. Since the early twenty-first century, the new discipline of e-humanities (Digital Humanities) has been gaining considerable ground. At the Meertens Instituut, two computational humanities projects received grants in 2012 from both the Netherlands Organization for Scientific Research (NWO) and the Royal Netherlands Academy of Arts and Sciences (KNAW); the projects were called Folktales As Classifiable Texts (FACT) and “Tunes & Tales.”⁶ Within the FACT project, the database is now transitioning to DFD 3.0, in which the digital archive and the “encyclopedia” are evolving into tools for fundamental research—for instance, research into variation in oral transmission and motif sequences. The database currently includes over 42,000 folktales and associated metadata.⁷

In the remainder of this paper, we discuss two aspects of computational folktale research. First, we describe our efforts to enrich the database through the automatic generation of metadata. We then describe research on computational modeling of traditional types and motifs using the DFD. Finally, we theorize how, in the future, “big folktale data” can help us find general patterns and structures in storytelling, and how the transcriptions of orally circulating stories in the database can give us a better understanding of variation in oral transmission.

Automatic Enrichment with Metadata

Even though the Meertens Instituut is not actively collecting new folktales, an increasing reservoir of digitized and digitally born narrative material is waiting to be processed and added to the online collection. Currently, the Meertens Instituut has about 35,000 folktale manuscripts in the archive, and probably as many tales in editions in the library. The institute does not have the human resources to feed all this material into the Folktale Database, with the main bottleneck being the assignment of the relevant metadata. Consequently, there is a need for automating this task. In our experience, not all input into the database can be automated. Sources, the names of narrators, and the dates and places of narration must be entered manually. We are, however, currently investigating ways to automate six other tasks: language or dialect identification, named entity recognition (NER), keyword extraction and assignment, automated story summarization, sub-genre recognition (fairy tale, joke, legend, etc.), and tale type identification. Given that the automated methods used to accomplish these tasks do not always work flawlessly, we will still need to rely on experts to

supervise the automatic annotation. Nevertheless, we expect that the annotation process will be speeded up considerably, as a good result can be approved in a matter of seconds.

Below, we describe our results so far on implementing (1) language and dialect identification, (2) sub-genre recognition, and (3) tale type identification. For all these tasks, we make use of supervised machine learning methods. These require a set of manually annotated examples to learn the relevant features that can be used to classify a text as belonging to, for instance, a certain genre. For this process, it is important that the example texts have been annotated in a correct and consistent manner.

Since the database is the result of 18 years of manual input by more than 60 researchers, information specialists, typists, interns, and volunteers, it has required significant cleanup. Errors that have accumulated in the database include incorrect or ambiguous information, typographical errors, data entry errors, and missing values. Entry standardization had not been a significant issue during the period that the database was used primarily as a digital archive or encyclopedia of texts. For example, it was largely irrelevant whether historical jokes were referred to as *boerde* or *klucht* (jest), or if the language was referred to as “General Civilized Dutch” or “Standard Dutch,” or whether a story was dated in 1660 or in the “third quarter of the seventeenth century.” Such ambiguous and fuzzy metadata values, however, emerged as stumbling blocks for computational approaches. Therefore, an intensive round of spring-cleaning took place in 2012–2013 to correct mistakes and inconsistencies. Along with the correction of errors and the addition of missing information, the entries were standardized across the database (Muiser et al. 2012). Effort was expended developing automatic procedures to make the data more consistent. Once cleaned, the database was transferred to a new platform, Omeka 2.0.⁸ We chose Omeka for several reasons, not least its promise of interoperability, since it is Dublin Core Metadata compliant. Dublin Core is a very basic and widely accepted metadata standard that contains 15 loosely interpretable pre-defined metadata terms. In addition, it allows for new terms to be added where necessary. The original database field terms were mapped to these pre-defined metadata terms as closely as possible, after being expanded upon with new terms that were very specific to the Dutch Folktale Database. The Omeka platform also offers several specialized features for search and display. Moreover, the Omeka platform can be easily extended with plug-ins, such as automated geolocation with the Google Map API (Application Programming Interface). Currently, we are working on the development of plug-ins that allow for the automatic annotation of various metadata, starting with the automatic identification of language, genre, and tale type, which have been the main topics of the FACT project’s research so far and are discussed below.

Research on automatic language and dialect identification was performed by Dolf Trieschnigg et al. (2012). The story material used was a selection of 39,003 folktales, written in a total of 16 different languages and dialects.⁹ The largest language groups in this selection were Frisian (17,347 stories) and Standard Dutch (13,632 stories). Furthermore, two historical linguistic groups could be discerned: Middle Dutch (656 tales) and seventeenth-century Dutch (2,361 tales). The remaining folktales were in 12 languages or dialects, from Groningen (854 stories, provincial dialect) and Waterland

Figure 4. Detail from the Omeka version of the Folktale Database, with the location of storytelling shown in Google Maps. The place of action (which is not necessarily the same) is recorded as well, and opens perspectives for a future tourist app for folktale safaris.

(153 stories, regional dialect) to Dordrecht (64 stories, urban dialect). Since the stories were taken from the DFD, they had all been manually annotated for language and dialect. Different supervised machine learning techniques were tested to automatically determine (classify) the story language. In supervised learning, the computer is given a number of examples, in this case, a number of stories and their manually annotated language. Based on these examples, the computer builds a model that is subsequently used to classify a test set, all of which are unseen stories. To evaluate different automatic language-identification approaches, *n*-fold cross-validation was used. In *n*-fold cross-validation the collection is split into *n* (for instance 10) parts (folds). The collection is split into sets of documents called folds. Each fold of documents is used to test a model trained on the documents in the other folds combined. In this way, all documents in the collection were used for both training and testing. Only the written transcripts of the stories were used as a basis for language identification; metadata such as storyteller location were not used. Experiments were carried out with different representations of the story texts (using words and variations of character *n*-grams, i.e., fixed-length sequences of characters in a text) and different types of machine learning models (nearest neighbor and nearest prototype).¹⁰ The performance of the techniques was measured in *precision*, the proportion of correct predictions in that language; *recall*, the proportion of documents in that language that is correctly predicted; and *F-measure*, the harmonic mean of precision and recall. A higher value (between 0 and 1) indicates a better performance. Table 1 shows the results for the different languages for the best performing machine learning model. The computer performance was nearly flawless for stories in Frisian, Middle Dutch, and seventeenth-century Dutch (F-measures above 0.96). On recognizing Standard Dutch, the program scored lower (an F-measure of 0.74) because confusion occurred

with (Dutch) dialects that are closely related to Standard Dutch, and with mixed languages such as Standard Dutch mixed with the dialect of the province of Overijssel. The performance for different dialects varied strongly (F-measures between 0.09 and 0.88). We hypothesize that this can be attributed to the limited number of training documents for these languages, the varying style in which the stories have been transcribed, and the similarity of some of the dialects. The experiments also revealed errors in the manual annotation process. Manual inspection of some of the misclassified folktales showed that their manual classification was sometimes incorrect. It shows an additional strength of automatic metadata assignment: it can be used to check manual annotation work, and signal possible errors. A better-defined subset of dialects, more training data, use of available geographic information, and possibly support from dialect dictionaries will further enhance the effectiveness of language recognition.

Dong Nguyen and associates performed experiments on the automatic recognition of sub-genres (2012). Support Vector Machines (a type of supervised machine learning model) were trained to classify folktales according to nine sub-genres: fairy tale, traditional legend,¹¹ saint's legend, urban legend, joke, ballad, riddle, situation puzzle, and personal narrative.¹² We restricted our experiments to 14,963 folktales in Standard Dutch. The folktales were divided into training, development, and test sets. A variety of features was tested: lexical features (unigrams and character n-grams), stylistic and structural features (part of speech, punctuation, white space, sentence length, etc.), named entities, and manually annotated metadata (keywords, named entities, summary, date).

Our experiments revealed that character n-grams (sequences of 2–5 characters) were the most effective, and that the system performed best on recognizing jokes, riddles, and legends. The system had more difficulty with recognizing ballads, personal narratives, and fairy tales (see table 2). The number of ballads included in the corpus

Table 1. Machine learning results: language identification

Language	Precision	Recall	F-measure
Frisian	0.999	0.976	0.987
17th-century Dutch	0.983	0.978	0.980
Middle Dutch	0.952	0.974	0.963
Liemers (d)	0.861	0.909	0.884
Gronings (d)	0.882	0.785	0.830
Standard Dutch	0.879	0.633	0.736
Gendts (d)	0.942	0.560	0.703
Noord-Brabants (d)	0.331	0.558	0.415
Zeeuws (d)	0.692	0.265	0.383
Flemish	0.229	0.810	0.357
Dordts (d)	0.207	0.609	0.310
Drents (d)	0.196	0.707	0.307
English	0.112	0.887	0.199
Waterlands (d)	0.091	0.824	0.163
Standard Dutch mixed	0.259	0.088	0.131
Overijssels (d)	0.055	0.250	0.090
<i>Macro average</i>	0.542	0.676	0.527
<i>Micro average</i>	0.799	0.799	0.799

Table 2. Machine learning results: genre identification

Genre	Precision	Recall	F-measure
Fairy tale	0.69	0.53	0.60
Traditional legend	0.82	0.92	0.86
Saint's legend	0.81	0.40	0.53
Urban legend	0.59	0.91	0.71
Joke	0.93	0.71	0.81
Ballad	0.00	0.00	0.00
Situation puzzle	0.70	0.58	0.64
Personal narrative	0.69	0.52	0.59
<i>Macro average</i>	0.66	0.59	0.61

was small and did not provide the algorithm with sufficient training material, while personal narratives constitute a residual group that appears to have too few defining characteristics to be recognized as a specific sub-genre. Fairy tales were regularly classified with legends. This failure in classification may be attributable to the difficulty of classifiers to distinguish between supernatural phenomena in legends (which may or may not be believed by tellers and audiences) and the magic in fairy tales (which is understood as fictional by tellers and audiences).

The automatic classification of folktales in terms of tale types is a complicated task: folktales can be assigned to tale types from at least six different type indices. Nguyen, Treischnigg, and Theune (2013) carried out a study to see if tale types could be automatically assigned to fairy tales and urban legends using the work of Hans-Jörg Uther (2004) and Jan Harold Brunvand (2012). In these experiments, the folktale material was limited to Standard Dutch.

The task was framed as a ranking problem: Given a folktale, the goal was to rank as highest the most appropriate tale types. Such a ranking could be presented to annotators, or the top-ranked tale type could be assigned directly. As in previous experiments, the folktales were divided into training, development, and test sets. The system identifies the tales that are most similar to the current given folktale based on features capturing lexical and semantic similarity between the folktales. The corresponding tale types are then used to generate a ranking of tale types. Learning to Rank techniques (a type of supervised machine learning) were used to learn a ranking model that uses various types of features. The similarity between individual tales, and the similarity between a folktale and all tales of a candidate tale type were most effective. The best system was able to rank the most appropriate tale types at the highest position 82 percent (ATU) and 76 percent (Brunvand) of the time. In a follow-up study, Nguyen, Trieschnigg, and Theune (2014) investigated perception of narrative similarity. Similarity ratings by both experts and non-experts were compared with the assignment of tale types. Their results indicated that a more nuanced view is needed of narrative similarity than is captured by tale types alone. Future work includes extending the experiments to other type indices (e.g., Sinninghe 1943 and Kooi 1984) and other languages.

The results we have obtained so far on the automatic detection of the language, genre, and tale type of folktales are encouraging. Other tasks we intend to automate

include the extraction of personal names, place names, and keywords from folktales, as well as the creation of summaries of folktales. Even though the automatic methods are not perfect, it is clear that they will facilitate and speed up the addition of new folktales to the DFD so that more folktales will become available online for humanities researchers to study, with more complete and consistent metadata than is currently available. We intend to make the DFD accessible in English as well.¹³ The translation of the contents of the fields—both data and metadata—still remains a challenge for Natural Language Processing techniques. For the time being, we use Google Translate. Eventually, we would like the DFD to connect with other folktale databases in the world to facilitate international research.

We also plan to investigate whether new interdependencies between stories can be discovered using computational methods: Can one distinguish new formal or conceptual relationships and clusters of tales within a large corpus of texts algorithmically?¹⁴ It is not in any way certain that the human classification of tales, as it has developed over the course of a century, is the best. Perhaps one can distinguish clusters of folktales, based on mutual thematic or formal features that have not been previously recognized. Another challenge involves data visualization: What types of visualizations are the most useful to researchers, and can the computer “learn” from expert feedback?

With an automated, improved, and enriched folktale database, and especially with internationally interlinked databases, one can imagine the diachronic and synchronous research possibilities that lie ahead. A few possibilities for research in the Netherlands include the analysis of rich data collected in the 1960s and 1970s, when 24 folktale collectors collected tales, primarily legends, from storytellers throughout the country (Dekker 1978:24–5). They all carried the same list of questions (“Do you know stories about the nightmare? . . . the werewolf[?] . . . gnomes?” etc.). Once that material, over 32,000 stories, is incorporated into the database, the geographical distribution of stories and knowledge of traditional supernatural phenomena can be charted: gnomes, werewolves, witches, wizards, magic, ghosts, nightmares, giants, omens, freemasons, changelings, fortune-tellers, faith healers, magic trees, and satanic rituals (see Broadwell and Tangherlini, in this issue). One of the most successful collectors in the mid-twentieth century was the Frisian collector Dam Jaarsma (1914–1991), who collected over 16,000 legends in a Frisian area. Incorporating his collections would enable us to research family repertoires and to investigate how repertoire overlaps between siblings and other family members. Another possibility would be gendered research: In what way do male folktale repertoires differ from female repertoires?¹⁵

Computational Modeling of Motifs in Oral Tradition

In the project “Tunes & Tales,” we focus on variation and motif sequences in folktale texts and folk song melodies. The methodological starting point is that both melodies and stories are made up of a succession of motifs: the basic building blocks that together determine the “DNA” of a story or song (metaphorically speaking). It is known that, in oral tradition, stories and songs vary from performance to performance on many dimensions (Rubin 1997), both in style, structure, and content; in

fact, every new representation of a story will generate new variations (Bartlett 1932; Owens and Bower 1979). Although narrative plots and melodies vary, most of the time, they remain recognizable. Some motifs disappear, and some change places or are replaced, yet other motifs tend to remain stable. For example, it is important that Cinderella (ATU 510A) goes to the ball to dance with the prince, but that episode need not happen three consecutive days in a row. Similarly, as a partner test, Cinderella's foot must fit the slipper, but the stepsisters do not necessarily have to cut off their toes or heels to fit the shoe. These details are often left out by contemporary narrators as being too cruel; the more fairy tales became considered a children's genre over the course of the twentieth century, the more violence, cruelty, and sexuality disappeared from the tales, while, for instance, kindness to animals and love for the environment were newly introduced.¹⁶ The project "Tunes & Tales" investigates this variation in motif sequences in oral tradition, and attempts to develop a model that explains the variation in both folktale and melody. From that perspective, it might even be possible to predict variation: in what ways will a story start to vary, and which narrative elements will remain unchanged?

In folktale research, the concept "motif" has been recognized as the smallest possible informational unit. In their indices, Aarne and Thompson (1964) and Uther (2004) distinguish complete tale types by only presenting a selective number of the more prominent motifs, while Thompson (1955–1958) provides an impressive index of these folktale motifs.

Thompson defines the motif as "the smallest element in a tale having a power to persist in tradition" (1951:415). Every story, even the shortest imaginable, consists of at least one motif. Even the famous "six-word story" allegedly composed by Ernest Hemingway—"For sale: Baby shoes, never worn"—contains two or three motifs.¹⁷ More elaborate stories consist of a string of motifs. Karsdorp and colleagues (2012a) have shown that, in the ATU catalog by Uther (2004), only two sub-genres clearly demonstrate how folktales draw from a common pool of motifs: the more elaborate "tales of magic" and "realistic tales" (which are, in a sense, "tales of magic" with the magic removed) share a large number of motifs. In contrast, tale types that belong to the remaining sub-genres in the ATU index have been assigned motifs that are not shared by other tale types. The degree of specificity and consequently the lack of co-occurring motifs across tale types in the ATU index make it hard to form generalizations over different tales with respect to their motif sequences. Although the index for folktales provides a selection of motifs for each tale most of the time, it fails to provide an exhaustive list of possible narrative building blocks, since the resulting number of possible variants would, in some cases, be essentially infinite. For many short tales, the index frequently provides a single unique and distinctive motif, which cannot be found elsewhere in the index. The negative consequence of this is that it blurs the distinction between tale types and motifs. Although extensive motif-searching is possible using the *Motif-Index* of Thompson (1955–1958) and the *Type and Motif-Index* of Ernest W. Baughman (1966), the folktales need to be annotated by hand to develop a complete sequence of motifs for each tale. These motif sequences serve two purposes: first, they identify a story as a version of a certain tale type, and, second, they help reveal how stories of the same type differ from each other. This identification needs to happen on a very elementary level with small (i.e., less

complex) motifs, but also on a more comprehensive level with abstract motifs, akin to Propp’s functions (1968), such as, for instance, the “Bride test” (Motif H360). An example of a smaller motif might be the “Slipper test” (Motif H36.1), while a very large and abstract motif such as “Better be content with what you have, than try to get more and lose everything” (Motif J346) fits the entire tale of “The Fisher and His Wife” (ATU 555).¹⁸

Figures 5–6. Top: Bar graph of the sub-genres in the ATU catalog that mutually exchange motifs (only based on the data the ATU catalog itself provides). Bottom: Visualization of folktales that share one or more motifs with one or more other folktales, as specified by the ATU catalog. Big and dark knots represent many shared motifs (esp. in the Tales of Magic and the Realistic Tales), while the smallest dots only share one motif with another folktale. In this visualization, all loners, the unique motifs, were omitted; otherwise, the visualization would be swarming with small islets of single dots.

Ideally, the identification of motifs should be automated. Karsdorp and colleagues (2012b) offer a partial solution to this problem, proposing a method that isolates the triplets *subject-verb-object* in each sentence, providing strong indications of which motifs are in play through the identification of the cast of actors who do something. This is not as easy as it sounds. Detecting the cast is related, but not identical to, named entity recognition. Many folktale characters not only lack a name, but many are not human: to wit, talking animals, animated pancakes, and the gingerbread man. To detect such actors, Karsdorp et al. (2012b) propose a method that relies on the intentionality of actors. A clear reflection of intentionality is found in direct and indirect speech. When an entity uses direct or indirect speech, we are more certain that we are dealing with an actor.¹⁹ By looking at the dispersion of actors in the text (rather than just making a simple frequency count), a ranking of major and minor actors can be made. The assumption is that actors who are more important are expected to occur more often in a given story and to be more evenly distributed over the story than are less important actors. Figure 7 visualizes this intuition. It plots the occurrences of all actors in a variant of the tale type “The Fleeing Pancake” (ATU 2025). As can be seen clearly from the graph, the pancake (being the protagonist) is present throughout the entire text. Other characters, such as the fox, are only briefly present. The expectation following these different distributions is that the fox plays a less prominent role than does the pancake (cf. Lüthi 1979:25–32). Both the farmer and the fox occur four times in the story. However, because the farmer is more evenly distributed over the text, the expectation is that he is more important.

Two folktale research experts from the Meertens Instituut were asked to provide a list of all actors in a collection of 78 Dutch folktales. In addition to this list, they were asked to rank the actors on a scale of importance. Karsdorp et al. compared the dispersion approach to a simple frequency count baseline. In this baseline approach, all nouns in subject position were ranked according to their document frequency in decreasing order (2012b). Evaluation was done using Mean Average Precision (MAP), which computes on a scale of 0 to 1 how well the system is able to allocate the highest ranks to the actual actors in the stories. The proposed system outperformed the

Figure 7. Dispersion of actors in a version of ATU 2025.

baseline by a great margin, and thus proved to be successful, with a MAP of 0.74 for the baseline and 0.91 for the dispersion method. The MAP of 0.91 means that, on average, nearly all actors are ranked on top.

The identification of actors is an important step in the process of detecting motifs and motif sequences in folktales. In addition to the actors, the settings, events, and actions need to be identified, thus providing all the narrative ingredients that are essential for a scenario: the cast, the action, and the scenery. To enable inter-operability with other folktale databases, we would like to provide a connection between the observed triplets in a folktale and the triplets comprising motifs in the *Motif-Index*. This still remains a big challenge. Fortunately, digital versions of the motif indices made by Thompson (1955–1958) and Baughman (1966; with additional motifs from English and North American folktales) exist; taken together, they account for ±55,000 motifs. Dutch developer Dirk Kramer built a hierarchically structured database with numerous search options and with links from motifs to folktales containing that motif in the DFD.²⁰ This motif database already functions as a useful tool in the manual assignment of motifs to folktales. Another Thompson motif database was built by Folger Karsdorp, called MOMFER, and this search engine can deal with synonyms and abstractions: when looking for “glass footwear” or “poisoned fruit,” it will also find “glass shoe” and “poisoned apple” (see www.momfer.ml; and Karsdorp et al. 2015).

In a pilot study, Karsdorp and van den Bosch investigated how well existing retrieval and machine learning techniques performed the task of identifying motifs from Thompson’s *Motif-Index* in Dutch and Frisian folktales (2013). They compared two systems: a standard retrieval model with a document-ranking function commonly used in information retrieval, and the supervised-topic model Labeled Latent Dirichlet Allocation (L-LDA) (Ramage et al. 2009). Topic models (Blei, Ng, and Jordan 2003) are particularly interesting for motif identification because of the analogous function of topics and motifs. In topic models, topics are represented by groups of semantically related words. Each document is represented by a distribution over these topics.

Figure 8. Screenshot from the digital *Motif-Index* by Thompson (enlarged with Baughman’s index), a hierarchical database and search engine built by Dirk Kramer (under construction). The yellow markup of the Slipper Test indicates that the motif can be found in the DFD.

Figure 9. The gray (small) motifs are recognized in a version of the Little Red Riding Hood fairy tale without any problems. Because the hierarchy of the *Motif-Index* is known, the higher motifs can be represented too.

Similarly, we can approach motifs as being represented by a group of words that are typically associated with a particular motif. Folktales can then be represented as distributions over motifs. Karsdorp and van den Bosch show that L-LDA functions as a competitive method for the identification of motifs in folktales (MAP for Dutch tales of 0.72 and for Frisian of 0.88) (2013). However, the relatively simple retrieval method outperforms the topic model by a relatively great margin (MAP for Dutch of 0.78 and 0.88 for Frisian).

A Future for Computational Folkloristics

The automated methods for folktale classification and analysis described in this paper make it possible to efficiently enter many more folktales than was previously possible, along with unambiguous metadata, into the DFD. Our current focus is on folktale typologies and motif sequences: to what extent can we cluster tales to complement the current system of types and subtypes? To what extent can computational methods improve upon the analysis of structural elements in folktales? Can these methods contribute to a more precise definition of “motif”? How big or small can a motif be? Does a sequence of combined motifs really constitute the (metaphorical) DNA of a folktale?²¹ Is the way in which Thompson and Baughman have distinguished motifs the most obvious and logical method? Or are there other linguistic or structuralist methods for approaching the elementary building blocks of stories that would make the matter much more transparent?²² After all, the algorithmic detection of a sequence of elementary building blocks is one thing, but understanding variation in oral transmission is another.

The computational tools developed for the DFD can be made available for other folktale databases as well. A possible solution to the lack of inter-operability across digital folklore archives is a “harvester” that pulls data from other collections into a research corpus.²³ This approach would enable more detailed research into geographical distribution or synchronic and diachronic variation, for example, without

requiring different systems to inter-operate on the systems level. Diachronic researchers could look into all kinds of historical shifts; for instance, in humor and taboos, based on jokes from the fourteenth century through today, can we discern certain topics becoming less or more prominent over time? From an international perspective, once data can be easily exchanged or can inter-operate, we can begin to chart the international occurrence and distribution of contemporary legends. One could, for example, trace certain narratives and their relationships to the shifting politics of ethnicity and cultural identity.

Not all questions within folklore research can be answered using computational methods. Many of the questions concerning context and meaning, identity, and perception are, and will remain, almost completely dependent on human interpretation. Practicing computational humanities only makes sense if we develop questions that computers are better able to answer than are researchers. With questions related to patterns and structures, and with tasks related to visualization, computers can be extremely helpful. Computational methods can help to organize and interpret vast amounts of data in situations in which the human brain quickly loses all overview.²⁴

Notes

1. The term “folktale” is taken in the broadest sense of the word, as in “tales folk tell”; that means not only fables, fairy tales, and anecdotes, but also legends, riddles, tall tales, and jokes.

2. This is not a book, but an online index as part of the DFD that is still growing. The Theo Meder index numbers are created ad hoc, whenever multiple stories of the same type enter the DFD that cannot be identified with any existing catalog number.

3. It can be expected that this percentage will not change dramatically if more catalogs are used, like the (mainly) medieval fable catalog by Gerd Dicke and Klaus Grubmüller (1987), the catalog of medieval exempla by Frederic C. Tubach (1969), or the catalog of migratory legends by Reidar Th. Christiansen (1958). Many of the fables are already covered in the ATU, while the DFD only contains a few exempla. Jacques Rudolf Willem Sinninghe (1943) covers most of the well known Dutch legends, so there are perhaps only a few percentages to gain here. It would be better altogether if an international group of folk narrative researchers would develop a new type index for international legends that would improve identification and enhance comparative research.

4. “Traditional fairy tales” refers to the animal tales, tales of magic, and anecdotes from the AT index.

5. Anthropologist Eric Venbrux (2009), for instance, wrote an article on twentieth-century folk belief and funerary culture in Friesland, solely based on data from the DFD.

6. The projects were carried out in close cooperation with the University of Twente and Radboud University Nijmegen. See www.elab-oralculture.nl.

7. When Jan Harold Brunvand for the first time compiled his “Type-Index of Urban Legends” (1994), he only listed titles and did not add tale type numbers. In order to use Brunvand’s index in DFD, it was necessary to add them. Brunvand has subsequently adopted the numbering for the new version of his type-index in his two-volume *Encyclopedia of Urban Legends* (2012).

8. See <http://omeka.org/>.

9. All texts were without restrictions on copyright, privacy, or offensiveness—this way, the material could also be offered to other scientists for review or own research.

10. More details may be found in Trieschnigg et al. (2012).

11. A traditional legend is a belief legend in magical and supernatural phenomena and beings; they are called *Sage* in both Dutch and German. A saint’s legend is a Catholic belief legend dealing with holy people, places, and objects; In Dutch and German, they are called “*Legende*.” An Urban Legend is a modern belief legend concerning modern technology, crime, accidents, embarrassments, and so on.

12. For this genre, in Dutch called “*kwispel*,” see Meder and Burger (2006).

13. Localizing Omeka is fairly straightforward, as the fields, buttons, and links can be translated easily.
14. See Abello, Broadwell, and Tangherlini (2012).
15. See, for instance, Dégh (1989); Tangherlini (1994); and Venbrux and Meder (1999).
16. Cf. Joosen (2012).
17. See http://en.wikipedia.org/wiki/For_sale:_baby_shoes,_never_worn.
18. Thompson (1951:413–27).
19. Our definition of indirect speech does not only include constructions like “he said” or “she answered,” but also constructions with verbs like “to promise, to beg, to understand, to regret, to wish, to discover,” and so on.
20. The database is still under construction, but can be found at <http://www.dinor.demon.nl/motif/>.
21. Cf. Irwin (2004) about selfish word strings.
22. Cf. Dundes (1964); Greimas (1966); Bal (2009); Meder (2012).
23. As proposed in Meder (2010, 2013, 2014).
24. Cf. Tangherlini (2013) and Finlayson (2011).

References Cited

- Aarne, Antti, and Stith Thompson. 1964. *The Types of the Folktale: A Classification and Bibliography*. Revision No. 2. FF Communications No. 184. Helsinki: Suomalainen Tiedekatemia.
- Abello, James, Peter M. Broadwell, and Timothy R. Tangherlini. 2012. Computational Folkloristics. *Communications of the ACM* 55(7):60–70.
- Bal, Mieke. 2009. *Narratology: Introduction to the Theory of Narrative*. Toronto: University of Toronto Press.
- Bartlett, Frederic. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Baughman, Ernest W. 1966. *Type and Motif-Index of the Folktales of England and North America*. The Hague: Mouton & Co.
- Blei, David, Andrew Ng, and Michael Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3:993–1022.
- Brunvand, Jan Harold. 1994. A Type-Index of Urban Legends. In *The Baby Train and Other Lusty Urban Legends*, by Jan Harold Brunvand, pp. 325–47. New York: W.W. Norton.
- . 2012. A Type-Index of Urban Legends. In *Encyclopedia of Urban Legends*, ed. Jan Harold Brunvand, pp. 741–65. 2 vols. Santa Barbara, CA: ABC-CLIO.
- Christiansen, Reidar Th. 1958. *The Migratory Legends: A Proposed List of Types with a Systematic Catalogue of the Norwegian Variants*. FF Communications No. 175. Helsinki: Suomalainen Tiedekatemia.
- Dégh, Linda. 1989. *Folktales and Society: Story-Telling in a Hungarian Peasant Community*, trans. Emily M. Schossberger. Bloomington: Indiana University Press.
- Dekker, Anton J. 1978. 150 Jaar Nederlands Volksverhaalonderzoek. *Volkskundig Bulletin* 4:1–28.
- Dekker, Ton, Jurjen van der Kooi, and Theo Meder. 1997. *Van Aladdin tot Zwaan Kleef Aan: Lexicon van sprookjes: ontstaan, ontwikkeling, variaties*. Nijmegen: SUN.
- Dicke, Gerd, and Klaus Grubmüller. 1987. *Die Fabeln des Mittelalters und der frühen Neuzeit, Ein Katalog der deutschen Versionen und ihrer lateinischen Entsprechungen*. Munich: Wilhelm Fink.
- Dundes, Alan. 1964. *The Morphology of North American Indian Folktales*. FF Communications No. 195. Helsinki: Suomalainen Tiedekatemia.
- Finlayson, Mark. 2011. Learning Narrative Structure from Annotated Folktales. PhD diss., Massachusetts Institute of Technology.
- Greimas, Algirdas Julien. 1966. Réflexions sur les modèles actantiels. In *Sémantique structurale: Recherche de méthode*, ed. Algirdas Julien Greimas, pp. 172–91. Paris: Librairie Larousse.
- Irwin, Robert. 2004. *The Arabian Nights: A Companion*. London: Tauris Parke Paperbacks.
- Joosen, Vanessa. 2012. *Wit als sneeuw, zwart als inkt: De sprookjes van Grimm in de Nederlandstalige literatuur*. Tiel: Lannoo.
- Karsdorp, Folgert, and Antal van den Bosch. 2013. Identifying Motifs in Folktales using Topic Models. In *Proceedings of the 22nd Annual Belgian-Dutch Conference on Machine Learning, BENELEARN-2013*, pp. 41–9, Nijmegen.

- Karsdorp, Folgert, Marten van der Meulen, Theo Meder, and Antal van den Bosch. 2015. MOMFER: A Search Engine of Thompson's *Motif-Index of Folk Literature*. *Folklore* 126(1):37–52.
- Karsdorp, Folgert, Peter van Kranenburg, Theo Meder, Dolf Trieschnigg, and Antal van den Bosch. 2012a. In Search of an Appropriate Abstraction Level for Motif Annotations. In *Proceedings of the 2012 Computational Models of Narrative Workshop*, ed. Mark Finlayson, pp. 22–6, Istanbul.
- Karsdorp, Folgert, Peter van Kranenburg, Theo Meder, and Antal van den Bosch. 2012b. Casting a Spell: Identification and Ranking of Actors in Folktales. In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, ed. Francesco Mambrini, Marco Passarotti, and Caroline Sporleder, pp. 39–50, Lisbon.
- Kooi, Jurjen van der. 1984. *Volksverhalen in Friesland; Lector en mondelinge overlevering. Een typen-catalogus*. Groningen: Stifting FFYRUG.
- La Barre, Kathryn, and Carol L. Tilley. 2012. The Elusive Tale: Leveraging the Study of Information Seeking and Knowledge Organization to Improve Access to and Discovery of Folktales. *Journal of the American Society for Information Science and Technology* 63(4):687–701.
- Lüthi, Max. 1979. *Märchen*. Stuttgart: Metzler.
- Meder, Theo. 2010. From a Dutch Folktale Database towards an International Folktale Database. *Fabula* 51(1–2):6–22.
- . 2012. *Avonturen en structuren. Op zoek Naar de bouwstenen van volksverhalen*. Amsterdam: Meertens Instituut.
- . 2013. Review of *Corpora Ethnographica Online*, 26–28 September 2012, Rostock. *Fabula* 54(1–2):117–20.
- . 2014. The Folktale Database as a Digital Heritage Archive and as a Research Instrument. In *Corpora Ethnographica Online. Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*, ed. Holger Meyer, Christoph Schmitt, Stefanie Janssen, and Alf-Christian Schering, pp. 120–8. Münster: Waxmann.
- Meder, Theo, and Peter Burger. 2006. “A Rope Breaks. A Bell Chimes. A Man Dies”: The *Kwispel*: A Neglected International Narrative Riddle Genre. In *Toplore: Stories and Songs*, ed. Paul Catteeuw, Marc Jacobs, Sigrid Rieuwerdt, Eddy Tieleman, and Katrien Van Effelterre, pp. 28–38. Trier: WVT-Verlag.
- Muiser, Iwe, Everhardus Christiaan, Mariët Theune, and Theo Meder. 2012. Cleaning Up and Standardizing a Folktale Corpus for Humanities Research. In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, ed. Francesco Mambrini, Marco Passarotti, Caroline Sporleder, pp. 63–74, Lisbon.
- Nguyen, Dong, Dolf Trieschnigg, Theo Meder, and Mariët Theune. 2012. Automatic Classification of Folk Narrative Genres. In *Proceedings of the Workshop on Language Technology for Historical Text(s) at KONVENS 2012*, pp. 378–82, Vienna.
- Nguyen, Dong, Dolf Trieschnigg, and Mariët Theune. 2013. Folktale Classification using Learning to Rank. In *Proceedings of the European Conference on Information Retrieval*, pp. 195–206, Moscow.
- . 2014. Using Crowdsourcing to Investigate Perception of Narrative Similarity. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 321–30, Shanghai.
- Owens, Justine, and Gordon Bower. 1979. The “Soap Opera” Effect in Story Recall. *Memory Cognition* 7(3):185–91.
- Propp, Vladimir. 1968. *Morphology of the Folktale* (2nd edition). Austin: University of Texas Press.
- Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D. Manning. 2009. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, ed. Philipp Koehn and Rada Mihalcea, pp. 248–56, Singapore.
- Rubin, David C. 1997. *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads and Counting-Out Rhymes*. Oxford: Oxford University Press.
- Sinninghe, Jacques Rudolf Willem. 1943. *Katalog der Niederländischen Märchen-, Ursprungssagen-, Sagen-, und Legendenvarianten*. FF Communications No. 132. Helsinki: Suomalainen Tiedeakatemia.
- Tangherlini, Timothy R. 1994. *Interpreting Legend: Danish Storytellers and Their Repertoires*. New York: Garland.

- . 2013. The Folklore Macroscope: Challenges for a Computational Folkloristics. *Western Folklore* 72(1):7–27.
- Thompson, Stith. 1951. *The Folktale*. New York: Dryden Press.
- . 1955–1958. *Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Medieval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends*. 6 vols. Copenhagen: Rosenkilde and Bagger.
- Trieschnigg, Dolf, Djoerd Hiemstra, Mariët Theune, Franciska de Jong, and Theo Meder. 2012. An Exploration of Language Identification Techniques in the Dutch Folktale Database. In *Proceedings of the Workshop on Adaptation of Language Resources and Tools for Processing Cultural Heritage (LREC 2012)*, pp. 47–51, Istanbul.
- Trieschnigg, Dolf, Dong Nguyen, and Theo Meder. 2013. In Search of Cinderella: A Transaction Log Analysis of Folktale Searchers. In *Proceedings of the First ACM SIGIR Workshop on the Exploration, Navigation and Retrieval of Information in Cultural Heritage*, Dublin.
- Tubach, Frederic C. 1969. *Index Exemplorum: A Handbook of Medieval Religious Tales*. FF Communications No. 86. Helsinki: Suomalainen Tiedeakatemia.
- Uther, Hans-Jörg. 2004. *The Types of International Folktales. A Classification and Bibliography*. FF Communications Nos. 284–296. Helsinki: Suomalainen Tiedeakatemia.
- Venbrux, Eric. 2009. Spoken op het kerkhof. Verhalen rond de dood uit de Friese Wouden. In *Spoken op het kerkhof: Verkenningen van protestantse vertelcultuur*, ed. John Exalto and Fred van Lieburg, pp. 121–34. Zoetermeer: Meinema.
- Venbrux, Eric, and Theo Meder. 1999. Anders Bijma's Folktale Repertoire and Its Collectors. *Fabula* 40(3–4):259–77.