

ISEBEL

Searching in English for Legends
A test run (2 weeks old)



Theo Meder



Intelligent (English) search in ISEBEL

- In search we try to get our data as accurate and complete as possible
- **GOOD NEWS:** ISEBEL is excellent in avoiding false results, so: **precision looks good**
- **Recall is less good:** in some cases ISEBEL misses out on legends that should have been found

The Danish devil

A man is very ill, and his wife warns that they may only open the door after there has been knocked 3 times. If the door gets opened before that...

- “...for det er **Fanden**, der vil ind og gøre min mand fortræd.”
- “...for it is the **Fuck** that will come in and make my husband upset.”

Still, this Danish devil will be found

- The “devil” is not mentioned as such in the text
- The metadata about the tale type contain crucial information:
 - ETK 27: About the Devil and being in league with him
 - ETK 27.540: Called the Devil, advice for visits by him

Tale types are part of the search in ISEBEL (Unfortunately the Dutch tale types are in German, not in English).

Another Danish devil

A minister is calling out names, but no one answers.

- “Jeg tror, **Fanden** har taget dem alle sammen!”
- “I think **the hell** has taken them all!”

Getting closer, but this time no “devil” in the data at all
 (“Fanden” will be found, “devil” will not)

Tale type = ETK 18: Ministers; ETK 18.358: Named
priests

Fanden = devil is in the Domain Specific Word List.

Werewolf AND handkerchief (1)

Werewolf bites in girl's handkerchief – later his girlfriend sees threads between his teeth.

Legends in the Dutch Folktale Database

(weerwolf AND zakdoek):

- 23 legends:
 - Standard Dutch: 14
 - Frisian: 7
 - Dialect of Brabant: 2



Werewolf AND handkerchief (2)

- German search in ISEBEL (Werwolf AND [Taschen]tuch): 0
- Danish search in ISEBEL (Varulv AND lommetørklæde): 0
- Dutch search in ISEBEL (weerwolf AND zakdoek): 23 (exact match)
- English search in ISEBEL (werewolf AND handkerchief): 10 (9 standard Dutch, 1 dialect of Brabant)

Missed: 13 in total: 5 in standard Dutch, all 7 in Frisian, 1 in the dialect of Brabant.

Taking along all the standard Dutch keywords (or summaries) in the translation would solve quite some misses.

Werewolf AND handkerchief (3)

Positive: search did not identify any completely wrong legends: 0%

Less positive: search found 43% of the legends

Less positive: search missed out on 57% of the legends

Especially Frisian remains a problem.

Frisian nightmare and werewolf

Frisian text: “Ien fan sawn opienfolgjende dochters is in nachtmerje. Ien fan sawn opienfolgjende soannen is in wearwolf.”

[Every 7th daughter is a nightmare, every 7th son is a werewolf]

Machine translation: “fan fan . . .’s daughters is in . . . fan fan . . . is in . . .”

[The only correct word is “daughters”, probably because it is spelled the same way as in standard Dutch]

Werewolf AND handkerchief (4)

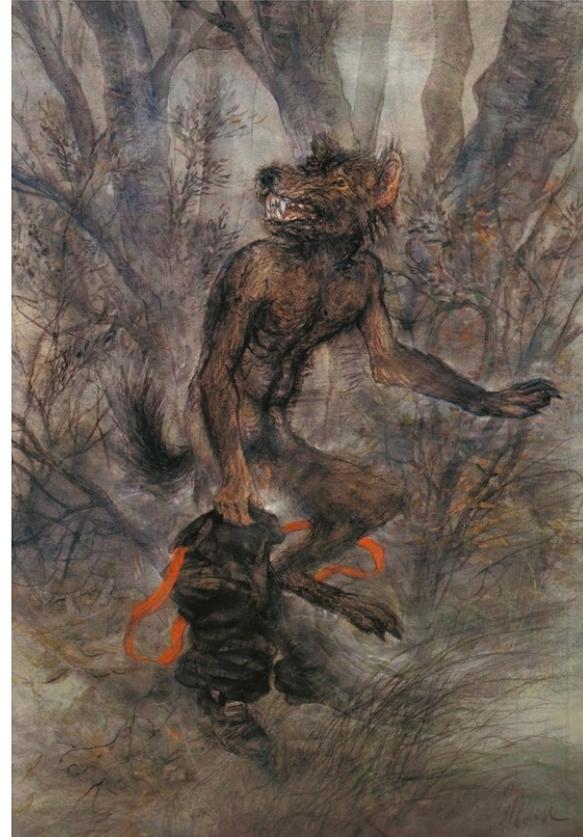
Why did ISEBEL miss 5 legends in Standard Dutch?

- Machine translation of longer legends breaks off too early (3x), before we get to the handkerchief scene
- Translation of title failed once (“The Broeker Werewolf”)
- Werewolf in 2 texts is just called a “big (black) dog”
- One text uses slightly deviant word “werwolve”

Again: solution for almost all of these translation problems is – translate the Dutch keywords as well.

Werewolf AND handkerchief (5)

- As far as the legend of the werewolf and the handkerchief is concerned, are we really dealing with a local Dutch legend here? Are there no German or Danish variants at all?



How about places? (1)

How can we have 10 werewolf legends, yet 12 dots on the map?



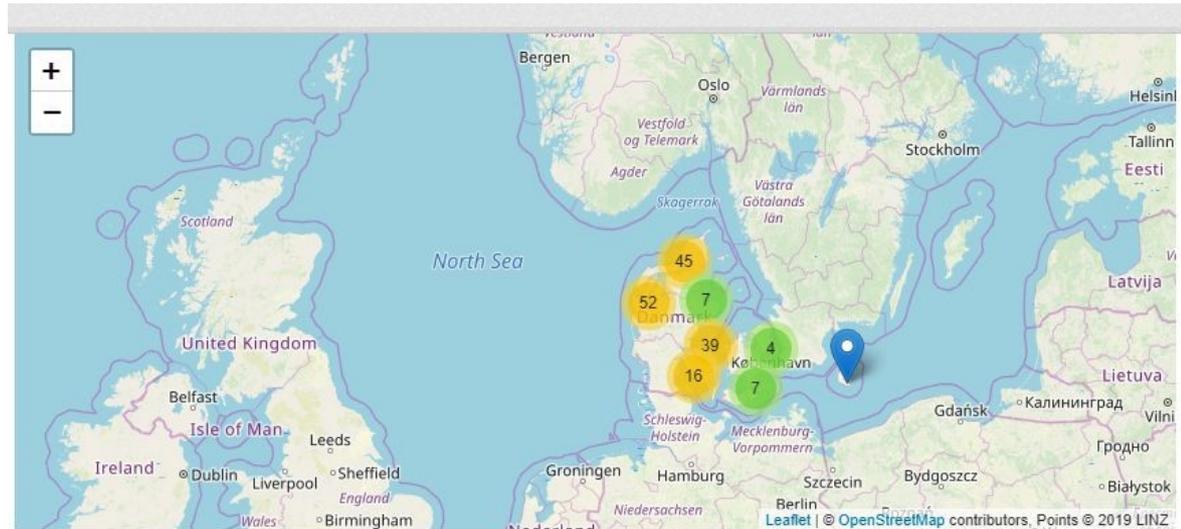
werewolf AND handkerchief

10 datasets found for "werewolf AND handkerchief"

Order by: ▼

How about places? (2)

We get 83 Danish legends about mermaids, but there are 171 dots on the map. Not only places of narration but also all places of action?



mermaid

83 datasets found for "mermaid"

Order by: Relevance

Organizations: [UCLA](#)

How about places? (3)

Legends about mermaids spread out over a long coastal area. Adding more legend material will show that even better.



Starting the witch hunt (1)

Search in ISEBEL (no wildcard possible?):

- German “hexe”: 192
- Danish “heks”: 456
- Dutch “heks”: 3496 (in DFD: 3079)
- Frisian “tsjoenster”: 1404 (in DFD: 1057)
- English “witch” total results: 3274

(sum = 5548; about 2274 legends lost in translation?)

Starting the witch hunt (2)

- ISEBEL finds 192 German legends on “hexe”
- ISEBEL finds 20 German legends on “witch”

Explanation:

- It looks like 172 legends don't have an English translation yet, neither manually nor automatically generated

Looking for the butterwitch

Search in ISEBEL:

- German “Hexe” and “Butter”: 2
- Danish “heks” and “smør”: 39
- Dutch “heks” and “boter”: 71
- Frisian “tsjoenster” and “bûter”: 7
- English “witch” and “butter”: 238 (while the sum above is 119)

Many more Danish legends were found (208) but less Dutch (28). In the search for Danish legends, **tale types** are being taken into account too, like: ETK 29: Witches and their sport; ETK 29.613: Witches take cream, butter, grain, etc.)

The trouble with languages: ambiguity (1)

A witch in standard Dutch is not only a “heks”, but also a “kol” and a “toverkol”. Most of the time “toverkol” is recognized and translated through the Domain Specific Word **Longlist**. In most cases, “kol” is not recognized and translated.

The trouble with languages: ambiguity (2)

The Little People are in Dutch mostly called “kabouter”, but sometimes “alvermanneke”, “aardmannetje”, “ierdmantsje”. In the other languages it is: nisse, ligtemand, ellemand, bjærgmand (D), Gnom, Kobold, Puk, Heinzelmännchen, Gnoom, Unnerierdsche, Erdmännchen, Ierdmännken (G)...

What’s the English equivalent? Gnome, brownie, leprechaun, elf, goblin, hobgoblin, imp, dwarf, puck, hidden folk, hidden people, fairy.

TECHNICAL CONCLUSIONS

- ISEBEL looks great (even better today)
- The legends are imported and the search engine works
- Search gives very few irrelevant results
- Some automatic translation issues need to be solved (Frisian)
- Should the Domain Specific Word List be a first filter (Danish “Fanden” = devil)?
- Translations of long legends were sometimes cut off
- Automatic translation of tale types, keywords (and summaries) will improve the search results

FOLKLORISTIC CONCLUSIONS

- The legend of the werewolf and the handkerchief seems to be a local Dutch legend
- Legends of mermaids show despersion over a long coastal area
- Witches sabotaging the making of butter, and stealing butter, milk, and cream is a belief legend that can be found all over the researched areas of the Netherlands, Denmark and North-eastern Germany.
- Of course, these results should be researched more in detail.
- TO DO: adding gender to narrators

AND SOME FINAL REMARKS...

- Can researchers get a **secret button** (or a secret link) so they can keep looking under the hood?
- Let's continue the ISEBEL project after September, even if we don't get funded (and let's continue the Zoom sessions)
- Let's invite other folktale databases to join us

THANKS EVERYBODY

for investing so much time and effort in this project