



GWC 2016, Bucharest, 27-30 January 2016

# Towards a WordNet based Classification of Actors in Folktales

Thierry Declerck, Tyler Klement,  
Antonia Kostova

Saarland University, Germany



# Background

- A student project on the topic „ Classification of Folktales: Building and Querying an Ontology with Folktales Classifications“
  - Goal: To design a piece of software that could take any given folktale, and display a list of categories to which this folktale belonged.
- Prerequisite: Formalisation of classification schemes used by folklorists
  - Additionally, investigating how WordNet can be used for identifying similar elements in different (formalized) classification schemes (topic of the current presentation).

# 2 Classification schemes

- Two well-known classification systems used by folklorists:
  - TMI - Thompson-Motif-Index of Folk-Literature
  - ATU - Aarne-Thompson-Uther classification
- Both of them are available as printed sources, or as online resources in html or pdf format. Since the two systems are related to each other (from ATU to TMI) , our aims were to:
  1. organize them in one ontology with appropriate references,
  2. make the resulting ontology available online,
  3. implement a web interface for SPARQL querying, and
  4. implement an automatic classifier of texts based on statistical approach.

# TMI

<http://www.ruthenia.ru/folklore/thompson/>

## S. Thompson. Motif-index of folk-literature : a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books, and local legends.

Revised and enlarged edition. Bloomington : Indiana University Press, 1955-1958.

[Restore frame](#)

Grant support: INTAS project 05-1000008-7922, ĐÓÔÊ #06-06-80-420a, ĐÓÔÊ #07-06-00441-à

- [A. Mythological Motifs](#)
- [B. Animal Motifs](#)
- [C. Motifs of Tabu](#)
- [D. Magic](#)
- [E. the Dead](#)
- [F. Marvels](#)
- [G. Ogres](#)
- [H. Tests](#)
- [J. the Wise and the Foolish](#)
- [K. Deceptions](#)
- [L. Reversals of Fortune](#)
- [M. Ordaining the Future](#)
- [N. Chance and Fate](#)
- [P. Society](#)
- [Q. Rewards and Punishments](#)
- [R. Captives and Fugitives](#)

Anankian 20; African: Werner African 127ff.; †Probenius and Fox, (Loango): Fechter-Loesche 207; Hindu: Penzel 110; Buddhist myth: Malalasekera II 338; Icel.: Boberg, MacCulloch Eddic 326; Irish myth: Cross.

A1. †A1. *Identity of creator.*

A1.1. †A1.1. *Sun-god as creator.*--Egyptian: Müller 69; Persian: Carnoy 260.

A1.2. †A1.2. *Grandfather as creator.*--S. Am. Indian (Paressi): Métraux BBAE CXLIII (3) 359, (Guarayú): Métraux RMLP XXXIII 147.

A1.3. †A1.3. *Stone-woman as creator.*--Paressi: Métraux BBAE CXLIII (3) 359.

A1.4. †A1.4. *Brahma as creator.*--Buddhist myth: Malalasekera II 338.

A2. †A2. *Multiple creators.*

A2.1. †A2.1. *Three creators.*--Icel.: Boberg, MacCulloch Eddic 327.--Oceanic: Dixon 24; Hawaii: Beckwith Myth 42.

A2.2. †A2.2. *First human pair as creators.* (Cf. †A1270.) Chinese: Eberhard FFC CXX 115 No. 70.

A3. †A3. *Creative mother source of everything.*--India: Thompson-Balys.

A5. †A5. *Reason for creation.*

A5.1. †A5.1. *Gods make earth to have place to rest their feet.*--Hawaiian: Beckwith Myth 43.

A7. †A7. *Creator's descendants.* (Cf. †A32.)

A7.1. †A7.1. *Creator has two sons.*--Guarayú: Métraux RMLP XXXIII 147.

A10. †A10. *Nature of the creator.*--India: Thompson-Balys.

A11. †A11. *Invisible creator.*--Jewish: Neuman.--Ackawoi: Alexander Lat. Am. 269.

A11.1. †A11.1. *Invisibility of creator learned from the impossibility of staring at the sun, his servant.*--Jewish: Neuman.

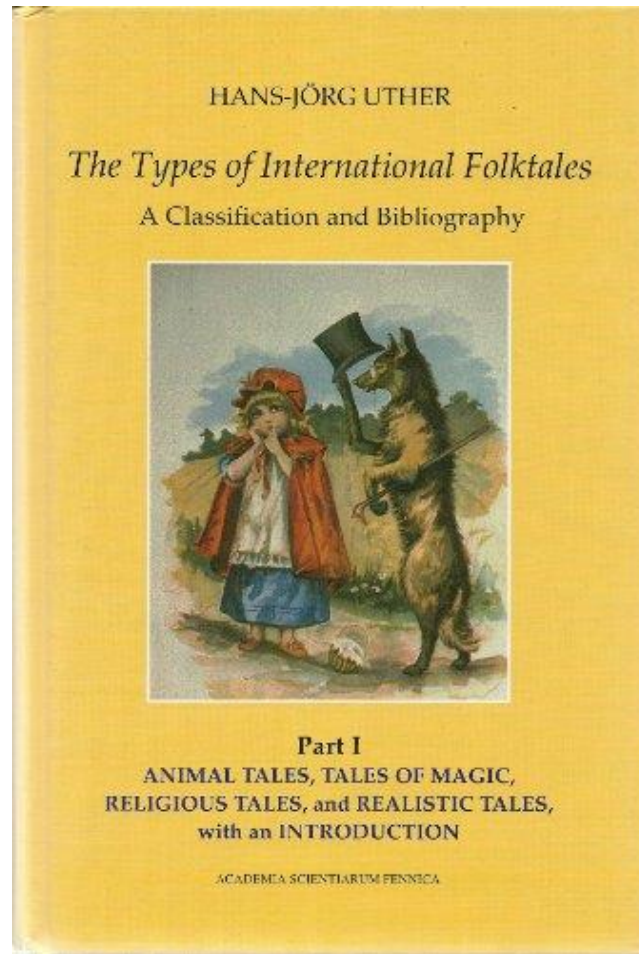
A12. †A12. *Hermaphroditic creator.* The creator is half man and half woman or is thought of as both male and female.--\*Lang Myth I 200f., 299; Güntert 324.--Greek: Eisler 396; Egyptian: Maspéro Histoire ancienne des peuples de l'Orient classique 141; Indian (Hindu): Keith 75.--Aztec: Alexander Lat. Am. 88.

A12.1. †A12.1. *Male and female creators.* Japanese: Anesaki 222; Hawaii: Henry Ancient Tahiti 345.

A13. †A13. *Animal as creator.*

A13.1. †A13.1. *Beast as creator.*

# ATU



# ATU Textfile

- **1**     ***The Theft of Fish***. (Including the previous Types 1\* and 1\*\*.) A fox (hare, rabbit, coyote, jackal) lies in the road pretending to be dead. A fisherman throws him on his wagon which is full of fish (cheese, butter, meat, bread, money). The fox throws the fish out of the wagon [K371.1] and jumps down after them [K341.2, K341.2.1].
- A wolf (bear, fox, coyote, hyena) tries to imitate this and pretends to be dead, too. The fisherman catches him and beats him [K1026]. Cf. Types 56A, 56B, and 56A\*.
- In some variants one animal (rabbit, fox) pretends to be dead in order to distract a man who is carrying a basket of food. Another animal (fox, wolf) steals the basket. (Previously Type 1\*, cf. Type 223.) Or an animal makes a hole in the basket so that the contents fall out. (Previously Type 1\*\*.)

# TMI vs ATU

## Thompson-Motif-Index

Motif is a repeated story element, e.g., a character, An object, an action, or an event.

- smaller units
- organized in hierarchical structure

## Aarne-Thompson-Uther Types

Type is a main story line that can be found in several cultures.

- bigger units
- parts of type descriptions refer to motifs

# Transforming TMI and ATU into an integrated Ontology

- Preprocessing TMI and ATU Text

HTML format

RE

line database

text format

RE

line database

Output:

Motif-id	Motif name
A	Mythological motifs
A1	Identity of creator
A1.1	Sun-god as creator
A1.2	Grandfather as creator
A1.3	Stone-woman as creator
A1.4	Brahma as creator
A2	Multiple creators

Output format:

[ATU number]~[ATU Title]~[ATU Description]~[List of references to TMIs, separated by commas]

Example:

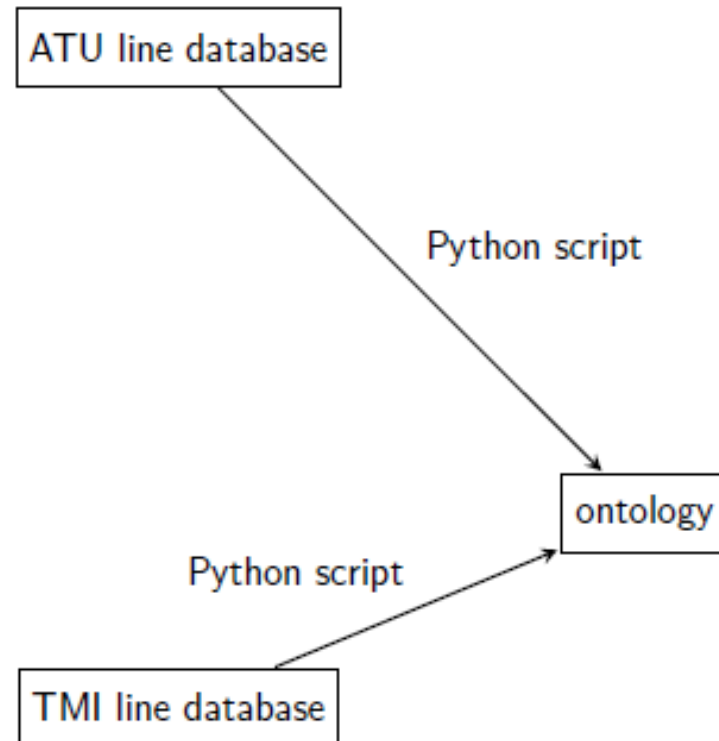
130~The Animals in Night Quarters~The Animals in Night Quarters. (Bremen Town Musicians.) Donkey, dog, cat and rooster are ill-treated by their owners because they...~[B296,N776|K335.1.4,K1161]



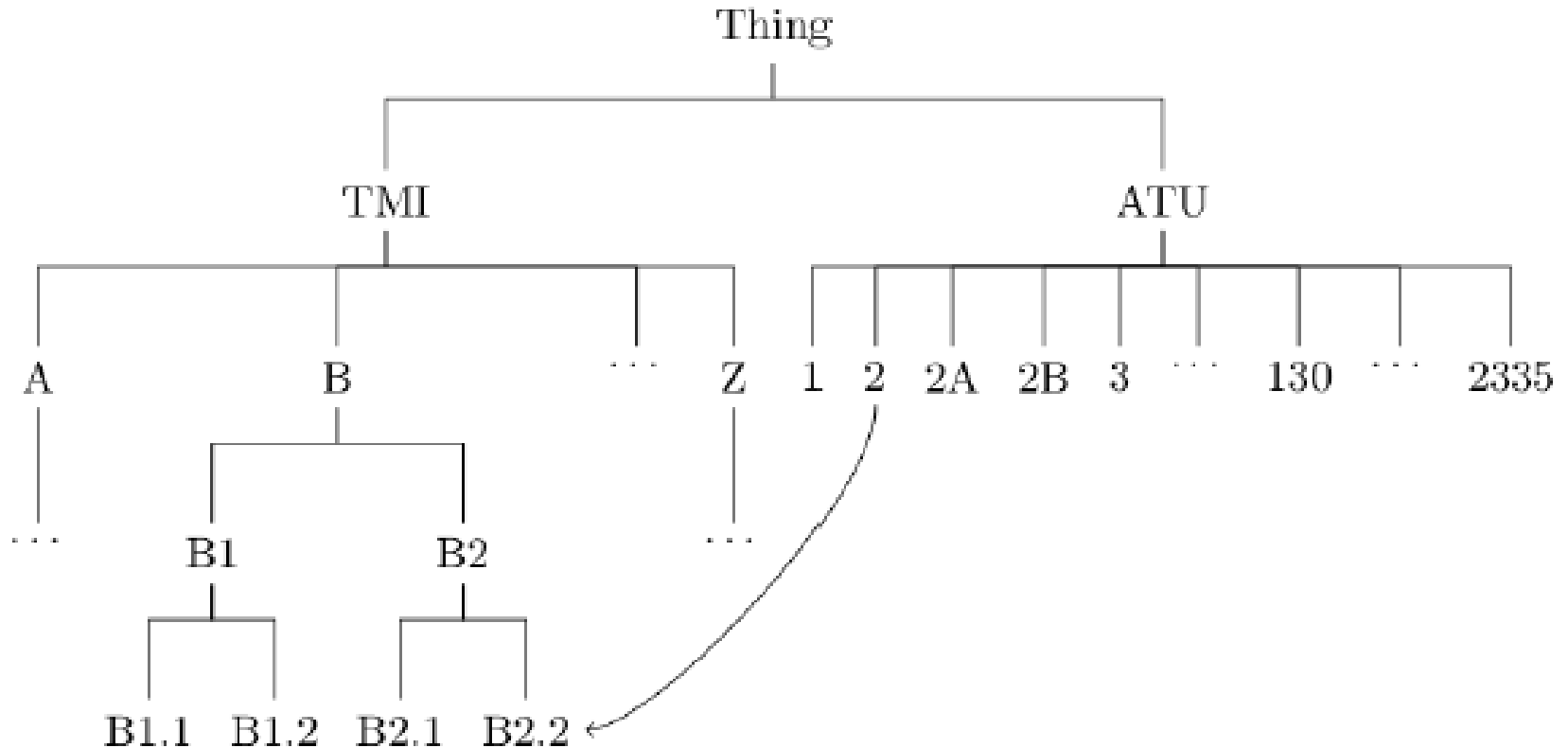
# Creating the Ontology for TMI and ATU

[ATU\_number]~[ATU\_Title]~[ATU\_Description]  
~[List of references to TMIs]

Motif-id	Motif name
A	Mythological motifs
A1	Identity of creator
A1.1	Sun-god as creator
A1.2	Grandfather as creator



# Ontology for TMI and ATU -- Structure



# Example of two ontology class entries in RDF(s) Syntax

```
File Edit Options Buffers Tools XML Text Help
Save Undo
tmi_atu_rdf.owl

<!-- http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D1213 -->
<owl:Class rdf:about="http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D1213">
  <rdfs:label xml:lang="en">&quot;Magic bell.&quot;</rdfs:label>
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D"/>
  <rdfs:comment xml:lang="en">&quot;Index D1213 of TMI&quot;</rdfs:comment>
</owl:Class>

<!-- http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D1213.1 -->
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D1213.1">
  <rdf:type rdf:resource="http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#Motif"/>
  <rdf:type rdf:resource="http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#D1213"/>
  <rdfs:label xml:lang="en">&quot;Magic gong.&quot;</rdfs:label>
  <rdfs:comment xml:lang="en">&quot;Terminal motif D1213.1&quot;</rdfs:comment>
</owl:NamedIndividual>

U:--- tmi_atu_rdf.owl 42% (250931,104) Git-master (nXML Validated:78% +2)
```

# Ontology Visualization (2)

tmi-ontology (http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#) : [/home/tonka/Classification\_Folktales/classification-of-folktales/ontology/tmi\_atu\_rdf.owl]

File Edit View Reasoner Tools Refactor Window Help

tmi-ontology (http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#) J2461.2

Active Ontology x Entities x Individuals by class x

Class hierarchy Class hierarchy (inferred)

Class hierarchy: "Literal following of instructions about greetings."

Annotations: "Literal following of instructions about greetings."

Annotations +

label [language: en]  
"Literal following of instructions about greetings."

comment [language: en]  
"Index J2461.2 of TMI"

Asserted in: <http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology#>

Description: "Literal following of instructions about greetings."

Equivalent To +

SubClass Of +

• "What should I have done (said)?"

General class axioms +

SubClass Of (Anonymous Ancestor)

- "Unprofitable association of unequals."
- "Unscrupulous conduct of business learned from observation of usurer's ow"
- "Useful and ugly preferred to expensive and beautiful."
- "Useful wins contest over beautiful."
- "Usefulness better than speed."
- "Useless surgical operation from misunderstanding."
- "Valuables given away or sold for trifle."
- "Value depends upon real use."
- "Value of silence."
- "Varieties of wisdom."
- "Weak fear company of strong."
- "Wealth and glory sacrificed for freedom and virtue."
- "Weight of bodily member chosen rather than its loss."
- "Well man made to believe that he is sick."
- "What one has is neglected in search for other things."
- "What should I have done (said)?"
- "Literal following of instructions about actions."
- "Literal following of instructions about greetings."
- "White sheep-skin used as a source of light."
- "Wife outwits her husband."
- "Wife persuades husband that she has returned immediately."
- "Wisdom (knowledge) from dream."
- "Wisdom (knowledge) from sage (teacher)."
- "Wisdom (knowledge) from women."
- "Wisdom (knowledge) taught by parable."
- "Wisdom (knowledge) taught by parable."
- "Wisdom chosen above all else."
- "Wisdom from angel."
- "Wisdom from books."
- "Wisdom from fools."
- "Wisdom from holy man."
- "Wisdom from old person."
- "Wisdom of child decides to run away."

# Ontology Visualized in Protégé

The screenshot displays the Protégé ontology editor interface. The top menu bar includes File, Edit, View, Reasoner, Tools, Refactor, Window, and Help. The address bar shows the URL: `http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-ontology/`. The main window is divided into several panes:

- Class hierarchy:** Shows a tree view of classes. The selected class is `""What Should I Have Said (Done)?" Miscellaneous type"`. Other visible classes include `""Today for Money, Tomorrow for None"`, `""Wait till I Am Fat Enough"`, `""Where Have you Been, Goose?"`, `""Where is the Warehouse?"`, `""Where Was Christ when he Was Neither in Heaven nor on Earth?"`, `""Who Gives his Own Goods Shall Receive it Back Tenfold"`, `""Who Has Lost This?"`, `""Who Was the Father of Noah's Sons?"`, `""You Don't Know what you Are Missing"`, `""You Shall See me a Little While Longer"`, `""You, or your Brother?"`, `""Thompson Motif Index of Folk-literature"`, `"ANIMALS."`, `"CAPTIVES AND FUGITIVES."`, `"CHANCE AND FATE."`, `"DECEPTIONS."`, `"HUMOR."`, `"MAGIC."`, `"MARVELS."`, `"MISCELLANEOUS GROUPS OF MOTIFS."`, `"MYTHOLOGICAL MOTIFS."`, `"OGRES."`, `"ORDAINING THE FUTURE."`, `"RELIGION."`, `"REVERSAL OF FORTUNE."`, and `"REWARDS AND PUNISHMENTS."`
- Class Annotations:** Shows annotations for the selected class:
  - `label [language: en]`: `""What Should I Have Said (Done)?" Miscellaneous type"`
  - `comment [language: en]`: `"Type 1696 of ATU"`
  - `seeAlso`: `""Literal following of instructions about greetings.""`
  - `isDefinedBy [language: en]`: `""What Should I Have Said (Done)?" Miscellaneous type. (Including the previous Type 1696A*.) A mother tells her stupid son (man tells his wife) what he should have said (done) in a particular situation. The son follows the advice at the next opportunity, where it turns out to be inappropriate. He is punished (is told again what he should have done or said, and he follows that advice in the wrong circumstances, etc.) For example, the fool congratulates mourners and offers sympathy to a bridal couple . Cf. Types 1681A, 1681B, and 1691B.`
- Description:** Shows the description of the selected class, which is identical to the `isDefinedBy` annotation.
- Equivalent To:** Shows no equivalent classes.
- SubClass Of:** Shows the class `""The Types of International Folktales Aarne-Thompson-Uther""`.
- General class axioms:** Shows no axioms.
- SubClass Of (Anonymous Ancestor):** Shows no axioms.

# Goal of the use of WordNet

- Detect similar characters/actors within and across the tale classification systems.

# Input Data for WordNet Analysis

- **2** *The Tail-Fisher*. **A bear (wolf)** meets a fox who has caught a big load of fish. He asks him where he caught them, and the fox replies that he was fishing with his tail through a hole in the ice. He advises the bear to do likewise and the bear does. When the bear tries to pull his tail out of the ice (because men or dogs are attacking him), it is frozen in place. He runs away but leaves his tail behind [K1021]. Cf. Type 1891.
- **Combinations:** This type is usually combined with episodes of one or more other types, esp. 1, 3, 4, 5, 8, 15, 41, 158, and 1910.

# Input Data for WordNet Analysis – Pre-processed for machine reading

- 6~Animal Captor Persuaded to Talk.~ A fox (jackal, wolf) catches a chicken (crow, bird, hyena, sheep, etc. ) and is about to eat it. The weak animal asks a question and the fox answers. Thus he releases the prey and it escapes. ~K561.1



# Use NLTK for accessing WN

- Searching for the least common hypernym (LCH) for the two words used in the pattern “A/An Noun (Noun):
  - `Synset(man.n.01), Synset(fox.n.05) => LCH(Synset(person.n.01))`
  - `Synset(fox.n.01), Synset(jackal.n.01) => LCH(Synset(canine.n.02))`
  - `Synset(fox.n.01), Synset(cat.n.01) => LCH(Synset(carnivore.n.01))`
  - `Synset(raven.n.01), Synset(crow.n.01) => LCH(Synset(corvine_bird.n.01))`

# Filtering out LCH results?

- Is “Synset(man.n.01), Synset(fox.n.05) => LCH(Synset(person.n.01))” not delivering a too generic synset?
- Testing the the NLTK function “path\_similarity” for filtering out:
  - “man.n.01” and “fox.n.05: ‘0.2’
  - “fox.n.01” and “jackal.n.01” : ‘0.33’
  - 0.33 as a threshold for selecting a hypernym?

# Flitering (2)

- filtering out the selected hypernym on the basis of the length of the path leading from it to the root node. The LCH “canine.n.02” has a much longer path to “entity” as does the LCH “person.n.01”.
  - Is “canine” then more appropriate for a precise detection of character similarities across classification systems?

# Extending the term base via the NLTK hyponym search

- synset “overlord.n.01”
  - hyponyms “feudal\_lord”, “seigneur” and “seignior”,
- But “fox.n.01”
  - “Urocyon\_cinereoargenteus” or “Vulpes\_fulva”
  - Are such scientific names useful for the task at hand? Still: it allows to link to another type of literature.

# NLTK function for generating multilingual equivalents (for example: FR)

- `Synset('fox.n.01') :: Synset('wolf.n.01') :: ['renard'] and ['loup', 'louve']`
- `Synset('dragon.n.02') :: Synset('monster.n.04') :: ['dragon'] and ['démon', 'monstre', 'diable', 'Diable']`
- `"Synset('enchantress.n.02') :: Synset('sorceress.n.01') :: ['sorcière'] and ['enchanteur', 'ensorceleur', 'sorcière']"`

# Future Work

- Evaluation of the results presented in this talk.
- Generation of multilingual classification systems, with the help of other sources
- Extending the work to other classification systems (for example Grimm)