

Toward Sequencing Multiple Motif Co-Occurrences

Sándor Darányi[†] and László Forró[‡]

[†] University of Borås, Swedish School of Library and Information Science, 50190 Borås, Sweden.
sandor.daranyi@hb.se

[‡] Vörösmarty u. 14, 5421 Abádszalók, Hungary.
salmonix@gmail.com

Abstract: *Catalogs project subject field experience onto a multidimensional map which is then converted to a hierarchical list. In the case of the Aarne-Thompson-Uther Tale Type Catalog (ATU), this subject field is the global pattern of tale content defining tale types as canonical motif sequences. To extract and visualize such a map, we considered ATU as a corpus and analysed two segments of it, "Supernatural adversaries" (types 300-399) in particular and "Tales of magic" (types 300-749) in general. The two corpora were scrutinized for multiple motif co-occurrences and visualized by two-mode clustering of a bag-of-motif co-occurrences matrix. Findings indicate the presence of canonical content units above motif level as well. The organization scheme of folk narratives utilizing motif sequences is reminiscent of nucleotid sequences in the genetic code.*

Keywords: *tale type, motif space, motif co-occurrence, 2-mode clustering, visualization.*

DOI: It would be provided by publication house

Összefoglalás. – A szellemi kulturális örökség megőrzésének egyik fontos feladata a hagyományok átörökítését végző szövegek feldolgozása és szolgáltatása. Ezeknek a szövegeknek csak töredéke található meg az Interneten, java részük az archívumokban vagy múzeumi gyűjtések anyagában lapul. Hogy komoly állományokról lehet szó, azt sejteti, hogy már egy évtizede is két és félmillió cédulát tartottak nyilván a világ legnagyobb folklórarchívumában, a Finn Irodalmi Társaság katalógusában. Ezzel párhuzamosan például a mesék motívumkatalógusai tízezerszámra sorolták fel a világ mesekincsének főbb motívumait, és a mesetípusok száma is ezekre rúg. Vagyis csak a mesék mint folklór szöveges műfaj esetében nyelvek százain átörökített, jellemző vonások tízezreit variáló, típusok ezreiben összefoglalható szövegvariánsok millióira lehet számítani.

Erre a számvetésre természetesen nem volt szükség az információs társadalmak létrejötte előtt, hiszen ha nincs eszköz a feladat elvégzésére, a probléma sem érdekes, noha valódi. Viszont a fordulópontot e tekintetben az jelenti, hogy ezek a szövegállományok vagy már digitalizált, vagy digitalizálható állapotban vannak, és a számítógépes feltáráshoz, feldolgozáshoz, információkereséshez szükséges módszerek is mind rendelkezésre állnak. Tehát „csak” használni kell őket, hogy a hagyományok egy tekintélyes részét, s vele

együtt a szöveges folklór műfajokat visszacsatoljuk az információs társadalom tudatalattijából a tudatába.

Ennek módjára nézve a szakmai közmegegyezés ma világszerte az, hogy a digitalizált szövegeket automatikusan indexelve és osztályozva, azok alkalmassá válnak információkeresésre és az eredmények vizualizálására (sőt a szövegfejlődés modellezésére is). Ugyanakkor ez a közmegegyezés a mesék automatikus osztályozása szempontjából abból indul ki, hogy a szövegeket a bennük szereplő motívumokkal lenne a legcélszerűbb indexelni, hiszen a hagyományos módszertan is ezt teszi. Most eltekintve attól a nem érdektelen részletkérdéstől, hogyan lehetne a motívumokat számítógéppel felismerni, amikor még a mibenlétükre sincs egyértelmű meghatározás, további bonyoldalmakat rejt az a mozzanat, hogy a meseszövegekben vajon a motívumok-e a legmagasabb szintű tartalmi indikátorok, vagy e fölött a szint fölött is vannak-e még típusképzésre alkalmas jegyek? Más szóval, felmerül a kérdés, vajon csakis egyedi üvegyöngyökkel (motívumokkal) rendelkezünk, vagy a mesetípusok inkább jellegzetes gyöngyfüzerekhez hasonlíthatók, amelyekben az egyes elemek sorrendje is fontos. Ezt a kérdést a hagyományos módszerekkel dolgozó mesekutatás eddigi tudtunkkal nem tette fel.

Az alábbiakban ez a feltevés a munkahitopézisünk. A hipotézist nem eredeti szövegeken vizsgáltuk, hanem az Aarne-Thompson-Uther mesekatalóguson (ATU) mint szöveggörpuzson. Álláspontunk ugyanis az volt, hogy – első lépésként – a típusok leírása mint tartalmi kivonat is hitelesen tartalmazza azokat az elemeket, amelyeket részletesebb vizsgálatok a jövőben nyilván az eredeti szövegekben is meg fognak találni.

Az ATU formalizmusa nagyobb fejezetekbe rendezi anyagát, majd az Aarne-Thompson motívumkatalógus (AaTh) jelzeteit használja az egyes mesetípusok rövid leírására. Újrafogalmazva a munkahitopézist, arra voltunk kíváncsiak, hogy ezekben a leírásokban ismétlődnek-e nagyobb darabok, azaz a motívum vagy motívumok közös előfordulása-e az a specifikus tartalmi egység, amely két mesetípusot megkülönböztet. Vizsgálatunkban az ATU „Varázsmesék” (300-749-es típusok) fejezetét, illetve ennek „Természetfeletti ellenfelek” című alfejezetét (300-399-es típusok) elemeztük. A görpuszban talált motívumok kilistázása után arra voltunk kíváncsiak, hogy (1) mely motívumok fordulnak elő közösen, (2) hány 2-, 3-, ..., n -elemű

„motívumcsomó” van az anyagban, illetve (3) az eredményeket térképpé alakítva, látszanak-e a korpuszban olyan tartalmi „csomósodások”, amelyek szerint néhány típus nagyobb egységgé állt össze, és ezek közösen használt motívumokra vezethetők vissza. A fenti elemzésben a motívumok sorrendjét, azaz kollokált mivoltát nem vizsgáltuk.

A motívumcsomók térképezéséhez párhuzamos klaszterálást használtunk (*block clustering, two-mode clustering*). Ez egyetlen munkamenetben két, egymásra merőleges tartalmi hierarchiát állapít meg és ábrázol faszervezettel (dendogrammal): a motívumok és a típusok összefüggésrendszerét. A két hierarchia egymásra vetítve megmutatja az anyagban rejlő homogén „tartalmi szigeteket”, esetünkben pedig igazolta a munkahipotézist. Eszerint csakugyan vannak „motívumcsomók” az anyagban, illetve több típust is jellemez ezek közös használata.

Mivel a fenti térképezési módszert a bioinformatikából vettük át – amely a művészetek mellett szintén használja a motívum fogalmát –, a cikket azzal a felvetéssel zárjuk, hogy a nyelvi és a genetikai kommunikáció párhuzamainak tanulmányozásával esetleg egy általánosabb modellhez jutunk, amelynek az említett két tartalomátviteli (közlési, kommunikációs) mód részben hasonló, részben különböző realizációja.

Kulcsszavak. – mesetípusok, motívumtér, motívumok közös előfordulása, párhuzamos klaszterálás, vizualizálás

I. INTRODUCTION

In cultural heritage objects, digitized or not, content indicators occurring on higher than word level are often called *motifs* or their equivalent. A motif is an element that keeps recurring in an artifact – e.g. in film, music, but also in folklore or scientific texts – by means of which often a narrative theme is conveyed. For example, the victory of the youngest son against all odds is a motif in folktales. In bioinformatics, the motif of a gene array study forms the mold of countless articles.

In the newly developed area of web sciences, a common rhetorical motif is to refer to the threats of information overload on people. In all of these different fields, insiders are familiar with these motifs, while outsiders are not; motifs constitute a kind of high-level jargon. The modeling of motifs (and especially the automatic detection of motifs) is an important, currently missing aspect of the analysis of cultural heritage and scientific communication texts beyond the sentence level (Darányi and Lendvai 2010:5).

Motif recognition for document classification and retrieval is largely unresolved. Work on identifying rhetorical, narrative and persuasive elements in scientific texts has been progressing, in several, but largely

unconnected tracks. The AMICUS project¹ (running between 2009 and 2012) set out to test a possible way to resolve these issues, starting with the identification of Proppian functions in folk tale corpora and adapting the solution to the identification of tale motifs or their functional counterparts.

Our current work goes back to (Darányi 2010) who suggested that prior to trying to map out a possible conceptual overlap between text-based and biological communication, it is necessary to show the theoretical underpinnings between them. He argued that evidence from different disciplines amounts to fragmented pieces of a bigger picture. By compiling them like pieces of a puzzle, one can see how the concept of *formulaity* applies to folklore texts and scholarly communication alike. Regardless of the actual name of the concept (e.g. motif, function, canonical form), what matters is that document parts and whole documents can be characterized by standard sequences of content elements, such as formulaic expressions enabling higher-level document indexing and classification by machine learning, plus document retrieval. For a comparison of folklore motifs and similar indexing constructs using Harris’ sublanguage model (2002), the interested reader is advised to the above paper.

Further, it is a research opportunity to be explored that one can use different domain-specific classification schemes, fragmentary or complete, to test the idea of motif extraction in general. E.g. it would be important to explore the relationship between concepts describing the life and accomplishments of the hero, unifying different typologies in one overarching concept (Campbell 2004); or to study the folkloristic underpinnings of classical Greek mythology and their geographic distribution as contrasted with archaeological evidence (Burkert 1979, Kirk 1970, Leach 1973, Nilsson 1964, Nilsson 1972). Other areas where formulaic text components such as Proppian functions can be applied to plot analysis include creative writing (Polti 1922, Polti 1924), collaborative narrative generation (Gervás *et al.* 2005, Lönneker *et al.* 2005, Peinado and Gervás 2006), or drama typology (Tomaszewski and Binsted 2007), among others.

This paper is organized as follows: Section 2 lists background considerations leading to the line of thought followed here. Section 3 discusses experiment design, Section 4 and 5 present and discuss the results. Finally, Section 6 offers our conclusions.

II. BACKGROUND CONSIDERATIONS

“As the history of type and motif indexes shows, the search for principles serving the classification of folk narratives has not yet produced a satisfying system, but indexes have provided scholars with ‘many valuable and practical research instruments, many methodical and theoretical by-products’, as Vilmos Voigt (1977:570) asserts.” (Uther 2009:11). Uther also states, based on *Acta Ethnographica*, that outlines of a new international

¹ <http://amicus.uvt.nl>

classification are now emerging (2009:10). Here we continue to show the relevance of automatic text classification for folklore archives (Voigt *et al.* 1999), with or without machine learning, to such studies.

We depart from the assumption that in catalogs, one meets domain-specific knowledge mapped onto a hierarchical structure. However, by nature such knowledge is also multivariate, i.e. describes many objects of the subject field by many characteristic features, and can be expressed by multivariate classification methods, with or without information visualization.

The case we want to test this hypothesis on is the Aarne-Thompson-Uther Tale Type Catalog (ATU), a classification and bibliography of international folk tales (2004). In the ATU, tale types are defined as canonical motif sequences such that motif string A constitutes Type X, string B stands for Type Y, etc. Also, it is important to note that types were not conceived in the void, rather they extract the essential characteristic features of a body of tales from all corners of the world, i.e. they are quasi-formal expressions of typical narrative content, mapped from many to one.

Together with the Aarne-Thompson Motif Index (AaTh) (Thompson 1955-58), ATU is the standard reference tool for librarians and digital curators alike, although other manuals such as Jason (2000) also come handy as means of orientation. However, when using ATU, it is regarded as a matter of fact that its descriptive units, motifs, constitute the highest level of abstraction, and there are no units of content above this. Therefore our research question was, does this assumption hold? If one regards the ATU type descriptions as text, and their entirety as a corpus, is its content evenly distributed as in the case of a divisive classification with no overlapping categories, or is there granularity (heterogeneity) to it?

III. EXPERIMENT DESIGN

A. Materials and methods

To extract and visualize a map of the respective segments, we considered ATU as a corpus and analysed sub-section “Supernatural adversaries” (types 300-399) in particular and section “Tales of magic” (types 300-749) in general. The two corpora were scrutinized for multiple motif co-occurrences and visualized by the two-mode clustering of a bag-of-motif co-occurrences (BOMC) matrix. After having excluded types not indexed by motifs at all, the first part of the experiment (300-399) worked with 52 tale types defined on the basis of 281 motifs, and the second part (300-745A) with 219 types and 1202 motifs, respectively.

After some early structural exploration by multidimensional scaling (MDS, PROXSCAL algorithm) which yielded inconclusive results (Fig 1), we turned to motif co-occurrence extraction. We augmented a standard lexicographic combination algorithm to compute combinations of tokens with a posting list indexing to calculate frequencies for each token. We applied a fre-

quency threshold to filter out valid but infrequent co-occurrences, assuming that significant occurrences are the more frequent ones (see Fig 2 for the pseudocode). This reduced the combinatorial results to manageable size. We will refer to multiple co-occurrences as *multi-plets* below (i.e. duplets, triplets, etc.). Respectively, multiplet-type matrices were constructed for two-mode clustering and visualisation.

For the latter, we used HCE3 (Seo and Shneiderman 2004). This is a program developed for genome sequence analysis but can be used to text structure

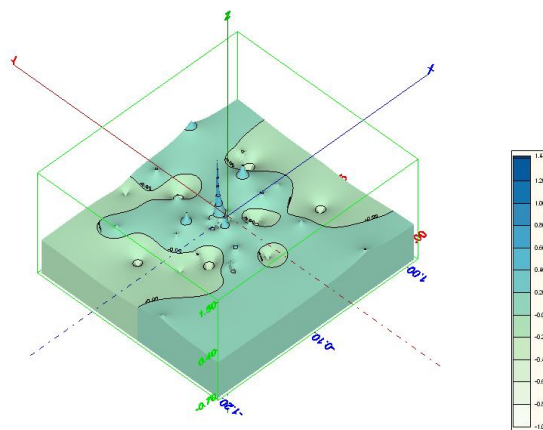


Figure 1. Type clusters as sinks in motif space (Supernatural adversaries segment of ATU)

analysis as well. For the results presented here, we applied row by row normalisation of data with single linkage (nearest neighbour) clustering using Euclidean distance as a similarity measure.

As an example for items in these corpora, tale type 725 (*Prophecy of Future Sovereignty*) reads as follows: “A clever boy refuses to tell his dream (about his future sovereignty) [M312.0.1, D1812.3.3] to his father and to the king. He is punished and endures various adventures (imprisonment) [L425]. A princess nourishes him in prison. War is to be declared on the emperor if he is not able to solve two riddles and a task. The clever boy solves the riddles and the task, tells the answers to the princess, and is freed from prison. So the boy averts war, marries the princess [H551], and finally receives two kingdoms.”

Clearly, the backbone of this type is the motif sequence [M312.0.1/D1812.3.3][L425][H551] where / refers to a forking alternative.

B. Background considerations

It is known from narrative studies that only canonical sequences of tale functions (a limited set of action types used by another limited set of actors) result in “valid”, i.e. acceptable Russian fairy tales (Propp 1968). However, due to the limited number of Propp’s examples, the

role of structural units combined with grammar in type creation is not at all clear. Our preliminary study took a first step to remedy for this situation.

III. RESULTS

Early on, for Part A of the experiment MDS indicated granularity in motif space (Fig 1), but type clusters – sinks in the motif landscape – were constructed on purely formal grounds, i.e. how many motifs had indexed groups of types. This result was inconclusive to decide

```

fun co-occurrence ( pos, n, post, S )
  if n == 0
    return iterate_last_element( pos, post, S )

  for c ( p..len(L)-n )
    if no ( post )
      post = make_post ( L[c] ) # initial state
      next
    if post = validate( post, L[c], unit ) # if positions are shared
      push S, L[c]
      push Result, co-occurrence( c+1, n-1, post, S )
      pop S, L[c]
  return Result

```

Figure 2. The co-occurrence extraction algorithm

about the null hypothesis.

At the same time we found that multi-plets occurred among the motifs. Their list for the *Supernatural adversaries* segment of ATU with the respective type numbers is given in Table 1. Such motif strings are displayed by two-mode clustering as horizontal band lines per type which, if the motif co-occurs in several types, form blocks (Figs 3-4).

Further, where they occurred in more than one tale type, all the triplets and quadruplets in Table 1 were also collocated, following the same sequential arrangement (story line).

	Motif numbers	Types
1	E341-M241-M241.1	505, 507
2	H1210.1-H1242-K1932	550, 551
3	R155.1-D231-F171.3-F171.1	471
4	B211.1.8-B422-B435.1-F771.4.1	545A, 545B
5	L161-C611-K1933-T68.1	301D
6	Z16-H621.2-H504-F660.1	653
7	Q2-S31-G466-H935	480
8	S31-K1911-K1911.1.2-D688	403, 450

Table 1. Motif triplets (1-2) and quadruplets (3-8) in the “*Supernatural adversaries*” ATU segment.

In Part B of the experiment, dealing with the “Tales of magic” section of ATU, we repeated motif duplet and triplet detection but did not visualize the results. To manage combinatorial explosion, we applied a detection

threshold which filtered out co-occurrences below specified levels. Statistics are displayed in Table 2.

Based on the above, the working (null) hypothesis could be rejected because we have found granularity in ATU on two levels, in the pattern of motif co-occurrences and in collocated motif co-occurrences.

Threshold	Duplets	Triplets
1	4293	618980
2	66	1408
3	4	16

Table 2. Motif duplet and triplet statistics for the “*Tales of magic*” ATU segment.



Figure 3. The red block in the middle indicates three co-occurring motifs in tale types 300 and 303.

IV. DISCUSSION

By two-mode clustering, the general view one gains is that the structure of ATU is mostly non-overlapping: types defined as motif strings are *almost* unique, their length depends on the number of motifs characteristic for a type, and to compare two tale types equals the matching of their respective motif strings. Whereas on word and sentence level this results in expressions of content similarity, e.g. by mapping tales as clusters of locations into some space and recognising types as their centroids, on merely formal grounds – as was the case with texts being indexed by motif numbers only – the result will reflect formal similarity, e.g. type clusters based on the number of motifs in them. For motifs and types, the striking novelty of multiple co-occurrence analysis was that the motif strings are not *entirely* unique, i.e. some of them have been persistent enough to be reused in different plots.

Apart from being eye-openers as well, these results are interesting for two major reasons. The first broad context is the perception of text variation as an evolutionary process, and the task of mapping evolving semantic content onto structures with both hierarchical and multivariate access. In this frame of thought, the reason why some motif strings have evolved and survived relates to a kind of selection pressure in a cultural

historical setting, yet to be modelled. To this end, ATU and AaTh as tools have pioneered and mastered the hierarchical approach to content description but are wanting in terms of being understood as multivariate products at the same time. This is a current deficiency that cannot be overlooked or neglected when it comes to any kind of their overhaul in and for a digital environment.

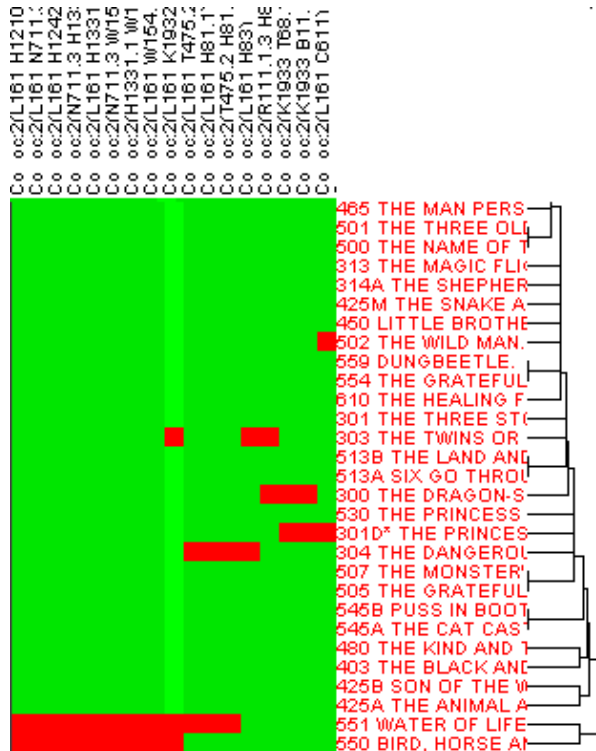


Figure 4. Blocks of motif duplets over tale types.

In other words we need descriptive units of content which can index the source material in its entirety, are both multivariate by nature and fit the hierarchical classification structure, plus are flexible enough to evolve, that is, become more and more enriched variants of the original standard classifications. Indexing by single text words or phrases plus by motifs is clearly not enough to meet this goal – the existence of persistent motif strings in multiple copies underlying several types proves that more than one level of semantic metadata may pertain to the body of tales we want to index.

The other broad context is the parallel between the linguistic and the genetic code as vehicles of information transfer over time. Both use coded transfer mechanisms to transmit their messages, capture instructions to reproduce meaning from form (we regard context as form here); and in both, sequence plays an important role in the coding and decoding process.

Tale types as motif sequences follow the sublanguage approach to content representation, pioneered by Harris (2002). As pointed out by Darányi (2010), this domain-specific practice from the life sciences can be recognized in formal descriptions of narrative content, too. A few similarities between their communication

patterns can be considered for methodology import between the two domains:

(1) Content is sequential, coded by an alphabet and compiled based on the combinations of its elements, i.e. irrespective of their order on a basic observation level. This holds for nucleotides – the building blocks of nucleic acids such as DNA and RNA – and motifs, the building blocks of tale types alike.

(2) On a next level, adding grammar and moving over to permutations, sequences start to play a role. Canonical nucleotide sequences generate secondary and tertiary – in fact spatial – structures such as the famed double helix; canonical motif sequences may contribute to the evolution of tale types, themselves representatives of tale variants in the plenty. Moreover, function sequences develop into fairy tale subtypes as shown by plot analysis (Propp 1968), and canonical mytheme sequences constitute myths and mythologies (Lévi-Strauss 1964-71, Maranda 2001). In a sense, reading and understanding the genetic code and narratives alike demands the mastering of abstract grammars with their equally abstract vocabularies.

(3) The concept of motifs is widely used in bioinformatics as well. Motifs in this sense mean primary nucleotide sequences of functional importance for structure generation. Sequential motifs include structural and regulatory motifs, with different functionalities pertaining to them; there may be methodological undercurrents linking the two knowledge domains which would need to be explored in more detail.

(4) Chromosome and story mutations may be more similar than thought previously. Chromosomal mutations produce changes in whole chromosomes (more than one gene), or in the number of chromosomes present, with the major types being (a) deletion – loss of part of a chromosome; (b) duplication – extra copies of a part of a chromosome; (c) inversion – reverse the direction of a part of a chromosome; and (d) translocation – part of a chromosome breaks off and attaches to another one.

Whereas most mutations are neutral and have little or no impact on the functionality of the product, their adding up can dramatically affect the survival rate of the outcome, leading to new genotypes and phenotypes in the course of evolution. In the same vein, deletion and translocation could be standard tools in the narrative building toolkit; inversion is suggested to play a central role in the Bible (Christensen, 2003), and duplication is evident e.g. in the case of the Proppian narrative scheme where complete tale moves may be repeated several times or combined with one another by different embeddings (Propp 1968). This indicates the need for a theory of text evolution as a series of narrative element recombinations, forming from simple to more complex structures by “mutation mechanisms”.

V. CONCLUSIONS

For a proof-of-concept investigation, we analysed two segments of ATU to find out whether the catalog con-

tained any internal structure as a reflection on overlapping narrative content in the real world. Tale types were indexed by their motifs and the resulting matrices were exposed to two-mode clustering and multiple co-occurrence analysis, respectively. Visualised results were used to highlight those motif combinations which occurred above a frequency threshold and thereby could be regarded as emerging structures in solidification.

Preliminary findings suggested that our line of thought worked because the null hypothesis could be rejected. Due to this, one can consider tale types as strings of single motifs and their multiplots, sort of “motif phrases”, which is new evidence. In our eyes, the popping up of the latter is proof for text evolution.

It is our understanding that two-mode clustering isolated the raw material (i.e. non-located sequences) of motif strings acting in their located variants as “narrative nucleotides”. However the nature of motif location will demand more detailed investigation.

We are looking forward to applying this technique with cautious optimism. As the AaTh contains about 40.000 motifs (Thompson 1955-58), this would allow for the prevalence of motif sequences as a new kind of metadata, and enable the use of both single and chained motifs as tags for semantic markup.

REFERENCES

- Burkert, W., *Structure and history in Greek mythology and ritual*. University of California Press, Berkeley. (1979).
- Campbell, J., *The hero with a thousand faces*. Princeton University Press, Princeton. (2004).
- Christensen, D. L., *The unity of the Bible: exploring the beauty and structure of the Bible*, Paulist Press, Mahwah, N.J. (2003).
- Darányi, S., “Examples of Formulaity in Narratives and Scientific Communication”, in Darányi, S. and P. Lendvai, eds., *Proc. 1st Intl. AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, Vienna, Austria, 29-35 (2010).
- Darányi, S. and Lendvai, P., eds., *Proc. 1st Intl. AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, Vienna, Austria, 29-35 (2010).
- Gervás, P., Díaz-Agudo, B., Peinado, F. and Hervás, R., Story plot generation based on CBR. *Knowledge-Based Systems* 18, 4-5, 235-242. (2005).
- Harris, Z. S., “The structure of science information”, *J. of Biomed. Informatics* 35, 215–221 (2002).
- Jason, H., *Motif, Type and Genre. A Manual for Compilation of Indices & A Bibliography of Indices and Indexing*, Academia Scientiarum Fennica, Helsinki. (2000).
- Kirk, G.S., *Myth: its meaning and functions in ancient and other cultures*. University of California Press, Berkeley. (1970).
- Leach, E., *Claude Lévi-Strauss*. The Viking Press, New York. (1973).
- Lévi-Strauss, C., *Mythologiques I-IV*, Plon, Paris. (1964-71).
- Lönneker, B., Meister, J.C., Gervás, P., Peinado, F. and Mateas, M., Story generators: models and approaches for the generation of literary artefacts. In *Conference Abstracts of the 17th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing* (Victoria, BC, Canada, June 2005). Humanities Computing and Media Centre, University of Victoria, 126-133. (2005).
- Maranda, P., ed., *The double twist: from ethnography to morphodynamic*, University of Toronto Press, Toronto. (2001).
- Nilsson, M.P., *A history of Greek religion*. W.W. Norton and Company, New York. (1964).
- Nilsson, M.P., *The Mycenaean origin of Greek mythology*. University of California Press, Berkeley. (1972).
- Peinado, F. and Gervás, P., Evaluation of automatic generation of basic stories. *New Generation Computing* 24, 3, 289-302. (2006).
- Polti, G., *The art of inventing characters*. James Knapp Reeve, Franklin, Oh. (1922).
- Polti, G., *The thirty-six dramatic situations*. James Knapp Reeve, Franklin, Oh. (1924).
- Propp, V.J., *Morphology of the folktale*, University of Texas Press, Austin. (1968).
- Seo, J. and B. Shneiderman, "A Rank-by-Feature Framework for Unsupervised Multidimensional Data Exploration Using Low Dimensional Projections," *Proc. IEEE InfoVis2004*, Austin, USA, 65-72 (2004).
- Thompson, S., *Motif-Index of Folk-Literature 1–6*, Indiana University Press, Bloomington. (1955-58).
- Tomaszewski, Z. and Binsted, K., The limitations of a Propp-based approach to interactive drama. In *Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies* (Westin Arlington Gateway, Arlington, Virginia, November 9-11, 2007). (2007).
- Uther, H.J., *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson, Part I*, Academia Scientiarum Fennica, Helsinki. (2004).
- Uther, H.J., “Classifying tales: Remarks to indexes and systems of ordering” *Nar. umjet.* 46, 1, 15-32 (2009).
- Voigt, V., M. Preminger, L. Ládi, and S. Darányi, “Automated motif identification in folklore text corpora”, *Folklore* 12, 126-141 (1999).