SEFAD, 2017 (37): 225-238
e-ISSN: 2458-908X
DOI Number: http://dx.doi.org/10.21497/sefad.328469

UNITS OF MEASUREMENT: ORAL TRADITION, TRANSLATION STUDIES AND CORPUS LINGUISTICS

Prof. Dr. John ZEMKE
University of Missouri
Department of Romance Languages & Literatures
ZemkeJ@missouri.edu
ORCID ID: *http://orcid.org/0000-0002-3038-0812*

Abstract

The study of the world's verbal arts offers an opportunity to consider ways that computational analysis and modeling of narratives may lead to new understandings of how they are constructed, their dynamics and relationships. Similarly, as corpus linguistics operations must define metrics, it offers an occasion to review basic interpretive concepts such as "units of analysis, context, and genre." My essay begins with an admittedly cursory overview from a novice perspective of what capabilities corpus linguistics currently possesses for the analysis and modeling of narratives. Consideration is given to the epistemological issue in the social sciences with the positivistic prescription or empiricist description of units of analysis and the potential pitfalls or advantages corpus linguistics encounters in searching for adequate equivalent terms. This review leads naturally to reflection on the crucial determinative action of context on meaning and the extent to which current computational interfaces are able to account for and integrate into global analysis of linguistic and performance dimensions such as performer, intonation, gesture, diction, idioms and figurative language, setting, audience, time, and occasion. As a tentative conclusion from this review, it can be stated that artificial intelligence for modeling narratives or devising narrative algorithms must develop capacities to account for performance dimensions in order to fulfill their analytical potential.

Keywords: Narrative, interpretation, prosody, context, units of analysis.

ÖLÇÜ BİRİMLERİ: SÖZLÜ GELENEK, ÇEVİRİ ÇALIŞMALARI VE BÜTÜNCE DİLBİLİMİ

Öz

Dünya söz sanatları çalışması anlatıların sayısal analiz ve örneklemelerinin oluşturulması, dinamiklerinin ve birbirleriyle bağlantılarının yeniden anlamlandırılmaları için imkân sağlar. Aynı şekilde, bütünce dilbilim çalışmaları ölçü bilimi tanımlarken, ''birim analizi, bağlam ve tür'' gibi temel yorumsal kavramları gözden geçirmeye imkan sunar. Makalem anlatıların analizi ve örneklemleri için bütünce dilbilimin hâlihazırda sahip olduğu yeni bir bakış açısıyla aslen yüzeysel bir incelemeyle başlamaktadır. Bütünce dilbilimin yeterli ve eşdeğer terimleri aramayla uğraşırken karşılaştığı birimlerin analizi ve olası güçlükler ya da avantajları pozitivist kurallar ya da ampirisizm tanımlamalarıyla birlikte sosyal bilimlerdeki epistemolojik (bilgikuramsal) mesele göz önüne alınmıştır. Bu inceleme doğal olarak bağlamın anlam üzerindeki önemli belirleyici etkisine ve halihazırdaki sayısal arabilimleri hesaplama, küresel dilbilim analiziyle ve konuşmacı, tonlama, el hareketleri, diksiyon, deyimler, mecazi söyleyişler, yer, seyirci, zaman ve olay gibi uygulama boyutlarıyla birleşmeye yol açmaktadır. Bu incelemeden çıkan kesin olmayan sonuç şudur: örnekleme anlatılar için olan yapay zeka ya da düzenleyici anlatı algoritmaları kapasitelerini analitik yeterliliklerini yerine getirebilmek amacıyla performans boyutlarını hesaplamak için geliştirmelidir.

Anahtar Kelimeler: Anlatı,yorum, ölçü, bağlam, analiz birimleri.

Gönderim Tarihi / Sending Date**:** 02-05-2016
Kabul Tarihi / Acceptance Date**:** 13-03-2017

INTRODUCTION

The study of the world verbal arts offers an opportunity to consider ways that computational analysis and modeling of narratives may lead to new understandings of they are constructed, their dynamics and relationships. Similarly, because corpus linguistics operations must define metrics, it offers an occasion to review basic interpretive concepts such as "units of analysis, context, and genre. My essay begins with an admittedly cursory overview from a novice perspective of what capabilities corpus linguistics currently possesses for the analysis and modeling of narratives. Consideration is given to the epistemological issue in the social sciences with the positivistic prescription or empiricist description of units of analysis and the potential pitfalls or advantages corpus linguistics encounters in searching for adequate equivalent terms. This review leads naturally to reflection on the crucial determinative action of context on meaning and the extent to which current computational interfaces are able to account for and integrate into global analysis linguistic and performance dimensions such as performer, intonation, gesture, diction, idioms and figurative language, setting, audience, time, and occasion.

The translation of a tale telling into a tale told necessitates appraisal of how translation decodes and recodes narratives. One insight into verbal arts that John Miles Foley expounded detailed their highly idiomatic essence, organized by a narrative logic of metonymy into larger systems of cultural meaning, Traditional Referentiality. Is corpus linguistics equipped with the necessary tools to code for idiomaticity? If, or when, the answer is in the affirmative, then the analysis of verbal arts by means of the methodologies of Corpus Linguistics and Artificial Intelligence promise to chart new understandings of culture specific narrative traditions, and, even, cross-cultural comparisons for the analysis of oral and written epic (and affine verbal arts).

Corpus Linguistics

The undertaking requires both large sets of digitized narrative as well as a set of cross-culturally viable terms, "units of analysis," in machine-readable form for marking up or "tagging" identifiable lexical, semantic, syntactic, and pragmatic features, as well performance features. Creation of such digital corpora of verbal arts is a Herculean task that poses interesting challenges to folkloristics and to computer science—researchers report drawbacks in the schemes of marking-up of texts manually by discipline experts or automatically by means of an Artificial Intelligence interface. Such difficulties notwithstanding, once specific language corpora are prepared, comparative study of cognate traditions could be undertaken. How are appropriate and adequate terms of comparison for discriminating similarities and differences to be devised? What criteria determine them? Should researchers recur to already existing convenient comparanda such as AaTh motifs, Thompson's Tale-Types, Propp's thirty-two functions, Bremond's fifteen or six narrative process? Could each scheme be computed sequentially or concurrently to parse texts into relevant units characterizing the various units' relations in the hierarchy of the narrative? Critics of such analytic approaches have identified deficiencies in them, and the prudence of extending those systems into the exploration the vast oceans of text is uncertain. Yet the sheer volume of materials that corpus linguistics must process in order to realize its analytic advantage makes serious consideration of whether a top-down or a bottom-up approach will prove most useful an imperative. A ground-up, empirical inference from the primary sources that would make explicit both overt and covert units, categories and relationships could lead to compelling new insights into narrativity. The Linnean classificatory system has not become totally obsolete, but rather, supplemented and revised, it remains useful for scientists. Üther's supplemented revision of AaTh suggests that corpus linguistics could play a role in reassessing folkloristic units of analysis.

Challenges for corpus linguistics include how it would preserve contexts that span the gamut from ritual to entertainment. Is it necessary to leave such parameters irrevocably

unrecoverable? I think Lauri Honko's entextualization of the Siri epic suggests that they can be represented.

The key strength of corpus linguistics is its ability to find, sort, and count items as a basis for linguistic description (Munday 11). It is an empirical methodology that makes use of computers for analyzing patterns either automatically or manually detected in large collections of texts using quantitative and qualitative techniques (Munday 12).

Specialists report that monolingual corpus based lexicography is well-suited to addressing six research questions: word meanings, word frequencies, association and connotation patterns, collocation, word sense and use, and use and distribution of synonyms (Munday 12). The discipline claims that the quality of its analysis of colloctaion and use norms is far superior to the human analyst's intuition (Munday 16). Yet, the field appears to remain anchored in word-forms, since they are the easiest to find, sort, and count. This limitation excludes from consideration speech pattern, intonation and other pragmatic features. When corpus linguistics incorporates tools that represent social and cultural features of narratives in communication, it may help to bridge the gap between linguistics and cultural studies (Munday 9). Corpus linguistics endeavors to derive "ontological structures" from text by applying statistical rules to linguistic annotations and by proposing a set of heuristic rules to derive ontology classes from the linguistic annotation (Mihaela and Declerck 2012, no pagination). Ontology generation detects lexical and syntactic properties that "offer a pre-structuring for the detection of relevant semantic properties" (*id*.).

A Digital Humanities research group working with a corpus of German Business Language recognizes that with its present tools, the ability to extract an ontology is constrained to the word level, leaving phrases, sentences, and the relationships between them, such as succession of states, or states in temporal or locative contexts is unrealized. The ambitious AMICUS research project aimed to develop a resource to "enable multilingual, content level indexing of folktale texts" (3). The AMICUS network focuses analysis on motifs, defined as "recurring conceptual, textual, audio or visual units [...] cognitively complex notions [...] expressed by lexically and syntactically variable units" (DeClerck, Lendavi, Daranyi 2012, 1). AMICUS considers the motif to be "a means to explain certain cross-cultural, cross linguistics, diachronic constancies in folktale" (*id*.). Semantically labeled versions of Üther's *Types of International Folktales* and Thompson's 1995 *AaTh* Motif Index were combined and linked together into domain specific ontology classes "that created an automatized index of folktale corpora" [...] that allow for graphing plots, connecting motifs at different levels of hierarchy; the plot graphs reveal "concatenation patterns of motif categories" together with "probabilities for network constructions" (*ibid.*, 3).

A research team comprised of James Abello, Peter Boradwell, and Timothy R. Tangherlini make the following claim for hypergraphic maps of folktale narrative schemes: "A multimodal network representation of a folklore corpus liberates folklore exploration from the limitations of existing classification terms" (60). A hierarchy tree of story elements presents, in essence, an inter-semiotic translation. The hierarchy tree is developed by a grasping technique that involves four major tasks: (1) Define similarity measures in stories; (2) construct a weighted graph among stories, where the weight of stories' connections is obtained through order-preserving transformation of their similarity norms; (3) decompose the graphs; and (4) compute a "hierarchy tree" for the obtained graphs (Abello et al., 68). Motifs are modeled as 23 "shallow ontologies" which are extensible at the level of specific manifestations as story attributes (69). Ideally, the shallow ontologies would be derived automatically, with expert correction as needed. Their current system precludes narrative structure as story attributes (69). Their optimistic conclusion announces: "computational folkloristics offers an opportunity to read and interpret culture in a more holistic fashion than ever before" (69). They expect such hypergraphs to map people in time and space, connecting stories in social networks. Nevertheless, major hurdles remain which must be overcome. Texts, together with audio and visual recordings, housed in folklore archives around the

world must first be digitized. Natural Language Processing technologies that can recognize and classify non-standard linguistic registers must be developed as well as ology tools that can ”automatically recognize and code narrative features and narrative structures’ (69).

Regarding meaning systems, Linda Garro avers that cross-cultural research is exponentially more difficult to carry out than intra-cultural research. On the basis of his work with digitalized corpora of German legends, Christoph Schmitt confirms that available techniques lack “a practical and reliable type of index system” for the comparison and analysis of international legends. Any analytical process, he observes, must connect the performed text to its performers and locate the narrative in its broader context (16th World Congress, Vilnius, 2013, *Catalogue*, 180).

In order to carry out extensive comparisons Muiser, Theune, and Meder address standardizing folktale corpora. Their goal is to *automate* metadata extraction, classification, and the clustering of folktales (2). They report similar constraints: Natural Language Processors cannot presently handle recorded oral performances and non-standard linguistic registers. For international cooperation they advocate standard metadata and infrastructure. Into the Dublin Core of standardized metadata terms the team incorporates additional specific terms: motif, subgenre, keyword, named entities, definition/descriptions, and geo-references, several of which admit interpretive “uncertainty” that leaves searching and separating values “problematic” (7). They conclude that standardization of cultural heritage databases entails significant problems that are best resolved in anticipation by a carefully designed and prepared process.

A semi-automatic annotation program developed by Mark Finlayson, “The Story Workbench,” allows untrained subjects to make semantic annotations on corpora of digitalized narratives. Finlayson’s interest lies in cognitive processes involving analogy and Artificial Intelligence routines for automatic labeling: “Most computational models of analogy require semantic input, supplied as semantically annotated texts. Greater complexity and sophistication requires more and more complex annotations making the manual process of assembling annotation sets inefficient” (1). Automatic systems that can assign word senses to words and arguments to verbs, he reports, are fast and consistent but not without flaws, and do not allow investigation of “the variation of human understanding of the same text” (1).

Finlayson reports a representation scheme for Discourse Coherence that posits 23 types of relationships (3-4). At current levels of knowledge it is, however, impossible to simultaneously code all the different aspects of meaning in a particular text (5). Understanding narratives and narrative processes is a key to understanding human cognition. Narratives transmit values, norms, and beliefs that ground cultural groups. Different cultures operate with specific assumptions about the world and these condition and individual’s interpretation and understanding of a narrative. Stories communicate such assumptions, and according to Finlayson, plot enfolds specific cultural information into narrative. Cross-cultural comparison of tale morphologies indicates that they share significant overlap in rough identity of common high-level structures (preparation, motivation, main actions, conclusion) but vary considerably in specific function sequence among other details (128). Finlayson’s algorithm for computational identification of morphologies, the Analogical Story Merger, is applied to semantically coded input with computer readability. An analogical mapper, the Structure Mapping Engine, determines similarities between portions of stories; and, finally, measures the fit of the derived morphology to the story corpus (a Bayesian model merging (128)).

Let us consider an example that Finlayson offers of the merging of two simple stories. In story X, a boy and girl are at play, the boy chases the girl, the girl runs away, and the girl thinks the boy is unpleasant. In story Y, a man stalks a woman, the woman is frightened, she flees, and the woman thinks the man is crazy (131). At one level the stories are similar, each consists of four sequences or states: pursuit, fear, flight, judgment. The homologation illustrates a flaw in the general methodology, similarities in the general abstract are mistaken for experiential similarities

where the, ostensibly, same states articulate very different psychological tensions and power relationships.

David Elson and Kathleen McKeown propose a methodology for discourse annotation, a story graph that is "tuned to the analysis of a text as narrative" (2010, 1) and is more a model of the received story than an absolute record of story content (*ibid*., 7). This sophisticated methodology encodes "underlying events" and views narrative holistically, revealing the interplay of "word sense, disambiguation, co-reference, interpretation of tense and aspect into a symbolic view of time and modality, and the encoding of actions, statives, and modifiers into propositional structures " (*ibid*. 1).

To semantically code an Aesop fable "The Fox and the Crow" passive voice phrases must be recast in active voice and figures of speech and idiomatic phrases must be replaced by simple synonyms (7). This method endeavors to account for the interpretive connections that give a story its cohesion (7). It is not similarities in actions between two stories that are of interest, but rather similarities manifested as goals, plans, and conflicts that constitute useable *comparanda*.

If the current abilities of Corpus Linguistics require that idiomatic and figurative speech be replaced by the substitution of synonyms, then this tool of the digital humanities is hardly equipped to reveal the dynamic word power and interplay between metonymic systems, what John Miles Foley called "Traditional Referentiality." Its usefulness lies elsewhere in the discovery and mapping of minimal motifs, allo-motifs, moti-phemes, and narremes, rather than the scope of meanings created with such devices by the singer of tales.

Lauri Harvilahti carefully titles his project the *Semantic Kalevala* not the *Experiential Kalevala*. Lauri Honko made a signal contribution with his work on entextualization.

Computational folkloristics, it may be expected, will one day enable the interpretation of culture in a more holistic fashion (Abello et al., 69) And therein liesgreat promise, yet to accomplish this desideratum will necessitate marking up into machine readable language all of the contextual features at present unaccountable for by the methodology's constraints and limitations. Corpus linguistics offers the performed text in an inter-semiotic translation, transforming a source language into a machine readable language, this implies certain costs.

Translation, Context, Genres

Commenting on the long tradition of translating sacred writings, James Kugel notes that any translation involves interpretation: "ambiguities in the original can rarely be duplicated in the translation, and as a result, the translator must take a stand and render the ambiguity one way or another (38). A wholly neutral translation, a so-called perfect translation is "a utopian dream" (Rubel and Rosman, 16). In fact, such a translation would necessarily require an axiom that meaning is universal and translatable between languages without, or with only minimal, loss. Despite much concentrated effort by the linguistic community, the evidence for universals is slight, and by one estimate, trivial. An "exact" translation would code "a universal cognitive structure [requiring] a kind of natural metalanguage" (Rubel and Rosman, 15). Roman Jakobson argued there a triadic system of translation: intra-lingual, rewording; inter-lingual, translation proper; and inter-semiotic, transmutation (Snelling-Hornby, 19).

Human languages operate with systems of phonology, lexicon, syntax, and prosodic features, in other words, diction, content, and context. Diction matters. Susan Jamison's illustration of how it matters is codified in the Goldilocks Principle. For English speakers the word "Porridge" immediately transports one to a particular narrative space that the synonymous "Oatmeal" does not. So too, the unremarkable question, "Who has been sleeping in my bed?" gives instant access to the same narrative space.

As a general proposition, translation may be understood as a way of relating the specific and particular to the universal or semi-universal by relating the local to a set of analytical concepts,

ideal-types. The subject's intervention to reformulate the specific as a generalization is "unavoidable" (Bremond, "A Critique of the Motif," 134), and raises the question of "units of analysis." The formulation of a "motif" involves a translation, a recoding. Identification of a singular motif requires, as Bremond observed that more weight be given to certain analogies that "we judge to be essential, giving less weight to differences that we judge to be accidental [...] what is abstracted is not a mere quotation" (*ibid*., 133).

The matter of translation of social phenomena sends us back, or ahead, to the units of analysis question. Unlike phenomena in nature, social facts are not things, they are representations, *comparison* of them entails translation, and translation raises the issue of incommensurability—phenomena that are impossible to measure or compare with the same metric. Clifford Geertz observed that translation entails "a question of the commensurability of conceptual structures from one discourse community to the next [...] puzzles of translation, with how meaning in one system of expression is expressed in another—cultural hermeneutics, not conceptive mechanics" (151). Here, then, is the underlying epistemological dilemma between universals and particulars, pitting positivist generalities against descriptivist particulars, which remains in contention despite several millennia of serious philosophical attention and thought. Ruth Benedict cogently demonstrated that comparison reveals differences, as well as similarities. To her way of thinking, to claim that there are neutral terms of translation for cross-cultural comparisons—that every term, trait or unit in one culture could be matched by a corresponding equivalent one in another—is naïveté (cited in Handler, 634).

Cultural comparison requires sophisticated translation practices, engaging in what Richard Handler characterizes as the "imperfect work of translation that can render difference meaningful and indispensable" (644). Wilhelm von Humboldt, who first connected culture, language, and behavior, claimed that language is an activity (*energeia*) based on a consensus among its speakers and dismissed characterizations of it as a static inventory of items, products of an activity (*ergon).* It is a truism in Translation Studies that discourse is situated in a time and a place, and this judgment is not an *a priori* decision but rather an inescapable conclusion from the standpoint of observation.

Discourse does not exist as such, but is always relative to an immediate situation. The recently deceased translator of modern Italian literature, William Weaver, remarked on the difficulty of translating even a simple greeting: "You need to know", he said," not only the time of day the scene is taking place, but where in Italy [...] [because] "in some places they start saying *buona sera*—'good evening'—at 1:00 p.m. [...] but you can't translate it as 'good evening' because the scene takes place at 3:00 p.m." (New York Times, November 16, 2013, Obituaries). And there is the case of Buddhist scriptures that contain a vast array of technical terms for which there are no English equivalents (Long, 7). The translation theorist Franz Paepcke emphasized that translators work not with words or languages but with texts: "every text is embedded in a situation which is not itself language" but rather a cultural, social or economic space in and from which the text speaks to us (Snell-Hornby, 33).

A recognized deficiency of multilingual corpora is the elimination of context, background situation, and cultural component (*ibid*., 158). This matters because it is context that disambiguates, determines propositional forms, and determines if a proposition is *explicature* or *implicature* (Gutt, 73). One attribute of metaphorical language is its capacity to convey a wide range of propositions: "the wider the range of implicatures, the greater the hearer's responsibility for constructing them, the more poetic the effect" (Sperber and Wilson, 1986a, p. 236 quoted in Snell-Hornby, 84). It is explicatures and implicatures that convey the intended interpretation of an utterance (Snell-Hornby, 94). This dimension of language corresponds with what John Foley termed "Traditional Referentiality" the recognition that a metonym references metonymous networks or systems of meaning in a particular class of discourse.

The linguist and anthropologist A. L. Becker formulated an elegantly simple statement of the fundamental problem encountered in translation: "The paradox of cross cultural communication is but a magnification of the one we encounter in every day conversation: I am like you/I am not like you" ("Attunement: An Essay on Philology and Logophilia," 370). Becker refers to a thought experiment that José Ortega y Gasset devised to compare languages. Superimpose two languages respective templates and their points of coincidences and divergences, what each declares or silences, as well as their respective preferences for vocabulary, morphology, syntax, phrase, and so on will be revealed (*ibid*., 371). Becker claims the task of the philologist is to bring texts that are distant in space or time into the here and now; to do this the philologist first installs him- or herself inside the author's language "the unquestioned frame for experience, that makes experience seem coherent" (*ibid*., 371). Philology is, then, the study of other frames (*ibid*., 372), frames that may be remote and radically different.

William Frawley observes that "translation" means "recodification," though there can be no exactness in the recoding except "in all but rare or trivial cases" ("Prolegomena to a Theory of Translation,"253). He assesses three separate arguments for identity across linguistic codes, the referential claim (absolute or relative synonymy), the conceptual/biological claim (all humans cognize the world in essentially the same manner), and the universals claim. Frawley dismisses the referential argument since "there is no meaning apart from code" (*ibid*., 254). He rejects the conceptual argument on the grounds that there is no satisfactory demonstration of a correlation between grammar and cognition. Finally, he observes that identified language universals such as negation, agency, conjoining, and embedding are necessary and thus trivial (*ibid*., 255-56).

Frawley understands inter-lingual translation to be the recoding of one set of semiotic constraints into a second different set of constraints, "the notion of identity is actually antithetical to the notion of translation. There is no meaningful (meaningful ≠ semantic) information except that it is coded, and the very fact of differential coding militates against 'exact translation'" (i*bid*., 257). He juxtaposes the standard measure of translation accuracy, "the fidelity of a new linguistic text to its 'original'" (*ibid*., 260), against his understanding that translation creates a third independent code, a unique sign-production which is not derivative (*ibid*., 261). In his view, there is no transferable "semantic essence," translation is a process of twin recoding: "The translation itself, as a matter of fact, is a third code which arises out of the bilateral consideration of the matrix and target codes; it is, in a sense, a sub-code of each of the codes involved" […] "a code in its own right" […]; this he avers is "both the bane and soul of the translator's existence" (*ibid*., 257).

Charles Fillmore distinguishes formal propositional knowledge from experiential knowledge, which exists as memories. He asserted: "comprehending a text involves mentally creating a world whose properties depend on the individual interpreter's private experience" ("Scenes-and-Frames Semantics," 61). As for the question of what constitutes meaning, whether meaning is the fulfillment of a checklist of attributes, that is, a series of conditions to be satisfied in order for a linguistic expression to be appropriate or truthful or if, rather, meaning is an appeal to a prototype, Fillmore weighs in soundly on the side of the prototype model: "There is no way of understanding the meaning of a word without having a notion of the whole experiential setting" (*ibid*., 73). Fillmore proposed a "Scenes-and-Frames Theory" of semantics, which is reminiscent of Friederich Schleiermacher's hermeneutic circle, by which understanding is viewed as a conceptual movement of repeated returns from the whole to the part and the part to the whole (*ibid*., 57). Just as setting is a crucial feature of meaning so too are affect and intention: "Affect and intentionality are culturally specific" (Ausbel and Rosmman, "Introduction," 13).

Appealing to Kenneth Burke's theory of the four "realms to which words may refer": nature, social relations, other words, and the supernatural ("Attunement,"372-73), A. L. Becker writes: "Meaning is to be observed in the relations of an organism with its particular environment" (*ibid*., 380). Becker maintains that linguistic relations with the environment are five in number —

interpersonal (pragmatics, tone, the cline of person), referential (the what, Burke's realm of nature), intertextual (X reports to Y, speaking the past), intratextual (requests for clarification, the need for the illusion of coherence), and intermedial (diagrams, syllabaries, alphabets), all five being located on a ground of silence. By this account, misinterpretation arises from mistaking or omitting from consideration one of the five realms (*ibid*., 381-86).

Lynne Long places authorship in relation to theoretical approaches to translation: "The question of text ownership and authorship, its position in the source and target polysystem, the motive for translating, the translator's ideology, the way the text is marketed and its intended readership, all have a bearing on the way in which translation is approached" ("Introduction: Translating Holy Texts," 7).

One is left, then, with an objective dilemma. On the one hand there is the claim that meaning is radically contextual and cannot be transferred or conveyed intact between semiotic systems. The Sapir-Whorf hypothesis asserts that human beings speaking different languages do not live in the same 'real' world because language shapes perception. On the other, meaning is held to be universal and translatable across languages without loss of meaning or only minimal loss. And for this to take place would, Paula Rubel and Abraham Rosman note, entail "a universal cognitive structure [requiring] a kind of natural meta-language" ("Introduction," 15). Even in this case a translation is mediated by the translator, whose own language constitutes what Wittgenstein described as a frame: "One thinks one is tracing the outline of the thing's nature over and over again, and one is merely tracing round the frame through which we look at it" (*Philosophical Investigations*, 1958; quoted in Becker, "Attunement," 372).

Clifford Geertz cautions against placing signifiers in stripped down abstract rule systems that purportedly "generate them":

> "the social contextualization of such "signifiers" [Yoruba line, Abelam color, Renaissance composition, Moroccan rhetoric] is a more useful way to comprehend how they signify, and what, than is forcing them into schematic paradigms or stripping them down to abstract rule systems that supposedly "generate" them. What enables us to talk about them usefully together is that they all inscribe a communal sensibility, present locally to locals a local turn of mind" (12).

Context

Heda Jason identifies context as the agent that transforms the motif as –etic unit into an –emic one: "[the] use of a motif in a text, its "meaning" and "function" will depend on the place the respective motif has in the framework of the "meaning" and "function" of the whole literary piece" (Jason 2000: 22). Noting the myriad examples of historic narratives, folkloric and otherwise, whose context is irrevocably unrecoverable, except from what can be gleaned of it from a textual document, Carl Lindhal contends that attending to performance leads to the understanding that the meaning of folklore is discernible in its performance (264) and to the appreciation that every folklore utterance unfolds in its immediate context: "in continuity with traditional knowledge and understanding" (265). As for narrative meaning of a constant plot shape, Lindhal cautions that different tellings may mean very different things, pattern stability does not vouchsafe stable meaning yet "a change in pattern almost certainly signals a major shift in signification (267).

Karl Reichel observed that epic that is sung rather than spoken, performed as song rather than spoken as poetry (163). He writes perceptively on the experiential differences between witnessing the performance of a ballad or epic and reading the text (177-78). A performance requires a performer, an audience, rhythm, diction, content, and circumstances. Recognizing that there are limits to documentation, Reichel queries whether it is even possible to adequately document a communicative event such as the performance of an epic (156-59). As Joshua Katz notes concerning Indo-European epics when diction is eliminated "what appears to be the same

stories show up again and again" (47). Once a performance is transformed into an edited text, Reichel observes, the music and all performance phenomena tend to be eliminated in the process of entextualization. Yet one need hardly remind oneself that folklore research has concerned itself with understanding the "context" of an expressive form (Abello et al., 69). Indeed, setting and performance are part of the message being conveyed (Rubel and Rosman, 12). Noting that Propp's studies of structure isolated the narrative from its social and cultural context, Alan Dundes granted Lévi-Strauss' attempt to relate the mythic paradigms to the cultures that operate with them in his model formulation. If epic is, as Richard Martin understands it, a total social event and not just the text that is culled from it, and if all discourse is situated in a time and place (Snell-Hornby, 165), and the time of a performance can affect meaning (Ben-Amos 1983: 287), and is embedded in a situation that is itself not language but cultural, economic, and social, then Üther's declaration that the description of a tale-type can reveal its variable elements, yet paradoxically, not its meaning or functions comes as no surprise. Context determines whether a proposition is an *explicature* or an *implicature,* a figurative extension, and this determination is always culturally specific (Gutt, 73). Mina Skafte Jensen has argued convincingly that common traits link epic traditions together, and she cites the special register of style and language, respect for facts, feeling of shared values, the entertainment power of narrative, and, she concludes: "these characteristics are all based on the event of performance" (49). Since no cultural expression can be performed outside a frame-of-reference, then according to Dan Ben-Amos' account "it is subject to generic conventions of a specific society. The a posteriori taxonomic conceptualization of an expression depends on the set of cultural conventions that marked out the expression to begin with (1992, 23). "Tellers always tell tales" observed Ulrich Marzolph, and these take place in the reality of a narrative performance that embraces the improvisation, variability, and creativity sanctioned by the art of story telling, which the tale-type as an analytical construct does not contemplate ("'Folktale', 'Tale-Type'" 221). Consequently, from Marzolph's perspective, the analysis and classification of "pure" tale type fails to account for the fact that "tales always convey ideas related to specific cultural contexts" (*id.*).

Myths and epics, observed Heda Jason, are considered to be highly specific to particular culture or cultural areas, and she reported that no indices that would encompass more than one tradition have been made (2000). The master indexer of motifs advocated separate investigations of the specific culture's oral traditions with an eye to identifying their endemic genres "the genres of a culture [...], she wrote, "form a system ordered along several axes " (2000, 36), and genre, she maintained is specific to particular cultural areas and are is a not universal (2000, 137). If, in Jason's view, the dragooning of AaTh units of analysis for the study and classification of non EuroAfroAsian cultures would be a mistake, the preferred alternative would be the development of schemes keyed to specific cultures, even if the scheme is, ultimately, arbitrary. The assertion that genre cross-culturally is not substantive but, rather, only nominal, has been forcefully articulated by Dan Ben-Amos (1992, 5) who insists that it is ethnic taxonomy that is meaningful, not least because performance type is a genre distinguishing feature. He locates the crux of the matter as characteristic of the difference between the natural and the social sciences (1992, 21). Tracing the nomothetic ideal-type units of analysis embodied in Propp's functions through successor systems, Ben-Amos notes their common methodology: "morphological analysis [that] follows the principles of ideal type construction" (1992, 11), and bolsters the claim that what is a "fact" in history and literature is determined by the interests of the researcher, citing Max Webber's observation: "There is no absolutely "objective" scientific analysis of culture" (Webber 1949, 72; quoted *ibid*., 12). The point is that narrative categories are scholarly inventions devised and introduced by European intellectual discourse—Myth, according to Edmund Leach, is not a phenomenon ethnographers encounter (1982, 3; quoted in Dundes 1968, 17). For a corrective to the positivist position, Ben-Amos again invokes Max Webber for whom the ideal-type "is a conceptual construct (*Gedankenbild*) which is neither historical reality or a "true" reality" (1992, 13).

The vexing nature of classification in folklore arises from arrays of native genres that do not correlate with scholarly categories, and here corpus folkloristics could clarify matters. If ethnic genres are cultural models for the organization of text (Dundes 1969: 275), how can corpus linguistics synchronize different systems, each possessing its singular internal logic (cf. Ben-Amos 1969: 275-76)? A decade ago, Chao Gejin and John Foley undertook the comparison of four different epic traditions, Mongolian, South Slavic, Old English, and Homeric, on five fundamental questions. Their work led them to several important insights. First, what an epic poem is depends for its structure and meaning on the specific tradition. Second, a *typical scene* is built from narrative increments yet each tradition has its own language of such scenes. Similarly, a *poetic line* may be measured by syllables, stress, melody, or rhyme, each must be understood in its own terms. Since *formulaic* phraseology is defined by metrical components, it, too, varies across traditions. Finally, comparison of the registers in oral epic revealed that all share metonymic phraseology and narrative, traditional referentiality, the epics operate with a language of *implicature*. Ben-Amos avers that to construct cross-cultural models on the basis of any one particular cultural system is to confuse the deductive model for the empirical ethnic taxonomies (1969: 283).

That narrative genre, as well as typology, are intellectual concepts is recognized by Uther who considers the first expectations that the description of exact narrative systems comparable with those of natural sciences and biological classification as no longer tenable: "a description of a tale type can show its various and changing structural elements, but not its meaning or functions" (I, 10). Distinguishing between Tale Type and motif cannot be done, according to Ûther, because their "boundaries are not distinct" (*id.*) Ulrich Marzolph notes that AaTh classifications of folktales in India, Pakistan, and Sri Lanka have been made, and identifies certain difficulties in the existing materials. No large-scale narrative collections are available, and the classifications rely on translation, which makes them questionable. The texts date to the colonial period and are approached from a Western point of view. Nevertheless, performance studies have uncovered indigenous classification systems, and the analysis of a large group of narratives could allow for classification based on empirical study ("Narrative Classification," 434).

Units of Analysis

What units of analysis are suitable for the study of verbal arts? It may be noncontroversial to state that three key features figure in all verbal art: diction, content, and circumstances. These are three legs on which the narrative stool stands. Recalling S. Jamison's Goldilocks principle—"porridge" instantly implies what "oatmeal" does not—diction can be ignored only at the analyst's peril. When two legs, diction and circumstance are removed, the narrative must balance on one leg, which is a trick, or it will certainly fall down. If in its natural *oikos* the performance of a narrative also conveys meaning, by decontextualizing the content one intellectually transforms the event into an entity, translating diction into ideal units of analysis and parsing the entity by applying the metric, motif, episode, type, genre, texture, image, formulae, figure of speech, and ordering them systematically into logical forms, propositions, and content patterns (Jason 1997: 221). For Jason, meaning is embedded in content not in form (*ibid*., 223). Yet it is the performer's conscious decision to will the narrative into instantiation for the pleasure or edification of others that makes the narrative socially meaningful (*ibid*., 223-24).

Stith Thompson described the motif as "the smallest element of a tale having the power to persist" (1948: 415-16, quoted in Jason 2000: 22), while for Alan Dundes the *allo*-motif is content that can fill a structural slot, a motif-*eme* (*id.)*. Jason holds both to be "context-free and free-floating" (*id.*).

On the other hand, motifs *qua* motifs acquire significance. In Claude Bremond's estimation, the gain meaning only in relationship to other motifs, carrying out its function as an antecedent cause, a consequent effect, acting as a means, or functioning as an end (1982: 133). In his criticism of the motif as a unit of analysis, Bremond asserts that sequences of actions, functions "are the true structural unit not arbitrary creations of the analyst" (139), and offer a surer method to understanding narrative operations. He proposes devising an anthropologically valid system of events, actions, and roles. Whatever unit of analysis is to be applied, its usefulness, Ben-Amos contends, is premised on the idea that "delimiting minimal units and analyzing their combinatory possibilities will reveal the nature of narration" (1983: 279). Here then is a level on which corpus fokloristics initially operates.

Janos Honti proposed three different ways the Tale-Type could be a viable unit of analysis: as a specific binding of motifs; any one type stands as a unique entity to any other; and as a Platonic form that is manifested by multiple instances (Dundes 1997: 195). Though Honti discounted the utility of the concept of tale-type, he asserted scholarship's ability to "construct a conceptual unit, considering the variants as constantly changing phenomena of an unchanging process" (*ibid*, 195). Reviewing major and minor criticism of the motif, its alleged independence, Eurocentrism, overlapping of motif with tale-type, Dundes concluded that the better unit of analysis is the narrative plot (197). Imperfect as they are, they are keystones in the comparative method (200).

Could corpus linguistics provide a method for address inadequacies in the AaTh motifs or the Thompson Tale-Type? Thompson himself noted that his system could not be extended to areas such as central Africa, indigenous North American groups or Oceania: Each one would need an index based strictly on its own traditions (1987, 8). Ben-Amos underscores particularity, that folktales and their performances have their own cultural subjectivity: "Their patterns, regularities, and rules are culturally regulated not scholarly constructs" (1992: 21). Like Jason, Ben-Amos views meaning as being in the texts, not in the constructed concepts, and its rules are discoverable, not constructed (1992: 22). Two epistemological traditions, positivist and interpretive, oppose each other at the juncture of the dilemma of universality and particularity. To Ruth Benedict's credit, she noticed that comparison reveals differences as well as similarities (quoted in Handler, 632), and such differences can be made meaningful and indispensable (*ibid*., 644).

Positivists assume the viability of common standards of measure, cultural units derivative of theory, that can identify and analyze patterns and allow for the stating of hypotheses and the testing of theories (de Munck, 278). Karl Popper's dictum on similarity is instructive: "Two things (e.g., societies), which are similar, are always similar in certain aspects. Similarity [...] always presupposes the adoption of a point of view. This point of view is, of course, one of comparison" (Popper 1959: 420-22; quoted in de Munck, 281). Popper's statement cuts both ways: there is a point of view. The empiricist social constructionists argue that reliable cultural units cannot be constructed, all such units are arbitrary (Gatewood, quoted *ibid*., 285), and no claim to objectivity that fails to take cultural meaning seriously is worthy of attention as it simply imposes one subjectivity and calls it objective (Garro, quoted *ibid*., 287). Though the positivist claim that "a strictly logical argument leads to an extreme relativist view that no two units are comparable" as de Munck observes, it is the case rather that "examining meaning-systems cross culturally [is] exponentially more difficult" and that it is far easier to "compare local understandings" (Garro *ibid*., 287). Finally, Linda Garro points to the advantages that coordinating top-down and bottom-up analysis would confer, controlling for -etic units that distort indigenous meanings (*ibid*., 228).

Max Louwerse ground his theory for developing a "standard world wide database" in an understanding of the reciprocity of AaTh motif and Propp's function. This methodology parses narrative propositions, identifying and classifying predicates and arguments by means of compiling lemmata and retrieving them in salient collocation patterns (245-47). The narration is

parsed, to some extent, on its own terms. Exploring the possibilities of creating databases that would make belief narratives "searchable for a worldwide audience" (179), Christoph Schmitt notes the need for a reliable type of index system that would allow for international comparison and computational tools that can "reveal connections between the performance [and place the] narrative in a broader context" (180).

Claude Bremond proposed a logic of narrative possibilities, elementary sequences that open, achieve, or close an attained result, as legitimate units of analysis. He initially described a system consisting of the operation of fifteen such processes but subsequently simplified the scheme to two processes: one that modifies a situation and another that preserves it, and whether the outcomes are favorable or unfavorable to the characters. He distributes these on the Aristotelian Square. Every action is coupled with a reaction. Modificatory processes are hindered by counter processes so that the "dynamics of the plot rest on the constant tension between these contrary forces" (411).

Is corpus lnuistics sufficiently developed to identify, classify, and map Bremond's processes? Abello *et al* conclude that single-classifier schemes are unable "to describe the complexity of tradition" (63). Commenting on the complexities of Indo-European historical linguistics, Joshua Katz avers that the larger and more unusual a unit of analysis is the harder it is to deal with, which he notes, explains historical linguistics' predilection for lexicon and semantics rather than syntax and discourse or what he asserts is "in some ways hardest of all—the artful devices of poetry" (47).

BIBLIOGRAPHY

ABELLO, James & BROADWELL, Peter et al. (2012). "Computational Folkloristics." *Communications of the ACM* 55 (7): 60-70.

BECKER, A. L. (1995). *Beyond Translation: Essays toward a Modern Philology*. Ann Arbor: University of Michigan Press.

BEN-AMOS, Dan (1969). "Analytical Categories and Ethnic Genres." *Genre* 2: 275-301.

BEN-AMOS, Dan (1983). "Introduction." *Research in African Literatures* 14: 277-82.

BEN-AMOS, Dan (1992). "Do we Need Ideal Types (in Folklore)? An Address to Lauri Honko." NIF Papers No. 2. Turku: Nordic Institute of Folklore.

BLACKBURN, Stuart H. & CLAUS, Peter J. et al. (1989). *Oral Epics in India*. Berkeley: University of California Press.

BREMOND, Claude (1980). "The Logic of Narrative Possibilities." *New Literary History* 11 (3): 387-411.

BREMOND, Claude (1982). "A Critique of the Motif." In *French Literary Criticism Today*. 125-46. ed. Tsevtan Todorov. trans. R. Carter. Cambridge: Cambridge University Press.

BRUNER, Jerome (1991). "The Narrative Construction of Reality." *Critical Inquiry* 18 (1): 1-21.

DARÁNYI, Sándor & WITTEK, Peter et al. (2012). "Toward Sequencing 'Narrative DNA': Tale Types, Motif Strings, and Memetic Pathways. *Proceedings of the Workshop on Computational Models of Narrative.* Istanbul.

DE MUNCK, Victor (2000). "Introduction: Units for Describing and Analyzing Culture and Society." *Ethnology* 39: 279-92.

DECLERCK, Thierr & LENDVAI, Piroska et al. (2012). "Multilingual and Semantic Extension of Folk Tale Catalogues". *Digital Humanities Conference, July 16-22.* University of Hamburg. http://www.dh2012.uni-hamburg.de/?s=Declerck.

DUNDES, Alan (1968). "Introduction." In Vladimir Propp, *Morphology of he Folktale*, 2d ed., xi-xvii. Austin: University of Texas Press.

DUNDES, Alan (1997). "The Motif-Index and the Tale Type Index: A Critique." *Journal of Folklore Research* 34: 195-202.

ELSON, David K. & MCKEOWN, Kathleen R. (2009). "A Tool for Deep Semantic Encoding of Narrative Texts." *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, 9-12. Suntec, Singapore.

ELSON, David K. & MCKEOWN, Kathleen R. "Building a Bank of Semantically Encoded Narratives." http://www.cs.columbia.edu/~delson/pubs/LREC2010-ElsonMcKeown.pdf

FINLAYSON, Mark A. "Collecting Semantics in the Wild: The Story Workbench." Association for the Advancement of Artificial Intelligence. www.mit.edu/~markaf/doc/finlayson.aaaifs.6.2008.46.pdf

FINLAYSON, Mark A. "Deriving Narrative Morphologies via Analogical Story Merging." www.mit.edu/~markaf/doc/finlayson.kokinov.nbupress.2009.137.pdf

FOLEY, John Miles and GEJIN, Chao (2012). "Challenges in Comparative Oral Epic." *Oral Tradition* 27 (2): 381-418.

FRAWLEY, William (1984). "Prolegomena to a Theory of Translation." In *The Translation Studies Reader*, 250-63. ed. Lawrence Venuti. London and New York: Routledge.

GEERTZ, Clifford (1983). *Local Knowledge: Further Essays in Interpretive Anthropology*. New York: Basic Books.

HANDLER, Richard (2009). "The Uses of Incommensurability in Anthropology." *New Literary History* 40: 627-47.

HANSEN, William (1997). "Mythology and Folktale Typology: Chronicle of a Failed Scholarly Revolution." *Journal of Folklore Research* 34: 274-280.

HARING, Lee (2001). "A Vade Mecum of Indexing." *FF Communications*, 2: 20-22.

HARVILAHTI, Lauri (2014). "The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala." *Oral Tradition* 28 (2). Forthcoming. pp. 1-9.

JASON, Heda (1997). "Texture, Text, and Context of the Folklore Text vs. Indexing." *Journal of Folklore Research*, 34: 221-25.

JASON, Heda (2000). *Motif, Type and Genre. A Manual for Compilation of Indices & A Bibliography of Indices and Indexing*. Helsinki: Suomalainen Tideakatemia; Academia Scientarum Fennica.

JENSEN, Mina Skafte (2005). "Performance." in *A Companion to Ancient Epic*, 45-54. Ed. John Miles Foley. Cambridge, Mass: Blackwell Publishing.

KATZ, Joshua T. (2005). "The Indo-European Context." *A Companion to Ancient Epic*, 20-30. ed. John Miles Foley. Cambridge, Mass: Blackwell Publishing.

KUGEL, James L. (1997). *The Bible As It Was*. Cambridge, Mass: The Belknap Press of Harvard University Press.

LINDHAL, Carl (1997). "Some Uses of Numbers." *Journal of Folklore Research* 34: 263-73.

LONG, Lynne (2005). *Translation and Religion: Holy Untranslatable*? Clevedon, Buffalo, Toronto: Multilingual Matters, Ltd.

LOUWERSE, Max (1997). "Bits and Pieces: Toward an Interactive Classification of Folktales." *Journal of Folklore Research*, 34: 245-49.

MARZOLPH, Ulrich (2003). "Narrative Classification." In *South Asian Folklore: An Encyclopedia*. Ed. by Margaret A. Mills, Peter J. Claus, and Sarah Diamond. New York and London: Routledge. pp. 433-34.

MARZOLPH, Ulrich (2003). "'Folktale,' 'Tale Type'." In *South Asian Folklore, An Encylcopedia*. Ed. by Margaret A. Mills, Peter J. Claus, and Sarah Diamond. New York and London: Routledge. pp. 220-22.

MUİSER, Iwe Everhardus Christiaan and THEUNE, Mariët et al. (2012). "Cleaning up and Standardizing a Folktale Corpus for Humanities Research." pp. 1-12. https://pure.knaw.nl/portal/files/477360/ARCH.2012.muiser.pdf.

MUNDAY, Jeremy (2002). "Translation Studies and Corpus Linguistics: An Interface for Disciplinary Co-operation." *Logos and Language* 3 (1): 11-20.

REICHL, Karl, ed. and trans. (2007). *Edige. A Karakalpak Oral Epic as performed by Jumabay Bazarov*. Helsinki: Academia Scientiarum Fennica.

RUBEL, Paula and ROSMAN, Abraham (eds.) (2003). "Introduction: Translation and Anthropology." In *Translating Cultures. Perspectives on Translation and Anthropology*, 1-22. Oxford: Berg.

SCHMITT, Christoph. "Belief Narratives in Online Databases: Retrieval Scenarios and the Problem of Internationally Valid Indexing, Based on the Example of the Wossidia Project." *Catalogure of the 16th Congress of the International Society for Folk Narrative Research: Folk Narrative in the Modern World: Unity and Diversity*, 179-80.

SNELLING-HORNBY, Mary (2006). *The Turns of Translation Studies*. Amsterdam, Philadelphia, PA: J. Benjamin.

UTHER, Hans-Jörg (2004). *The Types of International Folktales. A Classification and Bibliography*. 3 vols. Helsinki: Suomaleainen Tiedeakatemia.

VELA, Mihaela & DECLERCK, Thierry. "Heuristics for Automated Text-Based Shallow Ontology Generation." http://ceur-ws.org/Vol-401/iswc2008pd_submission_80.pdf.

WYATT, N. (2005). "Epic in Ugaritic Literature." in *A Companion to Ancient Epic*, 247-54. ed. John Miles Foley. Malden, MA: Blackwell Publishing.