

CONSTRUCTING DIGITAL DATABASES OF THE PERIPHERY OF ESTONIAN RIDDLES. DATABASE *ESTONIAN DROODLES*

Piret Voolaid

Our lives are in constant change, and so is our intellectual heritage. One of the genres of folklore that has undergone changes is riddles. The Estonian riddle material suggests that classical riddles in their traditional core form have been pushed to the state of static folklore. But since nature hates a vacuum, contemporary riddle types have already started to replace the old riddles.

The riddle material of the Estonian Folklore Archives may roughly be divided in two:

(1) the nearly 130.000 older traditional riddles (*Üks hani, neli nina? – Padi*. ‘One goose, four noses? – A pillow’) constituting the main, central body of the total riddle corpus of 175.000 variants; and

(2) the rest (45.000 variants) consists of peripheral riddle forms, such as: **conundrums** (*Kuidas sa tead, et elevant on voodi all? – Voodi on lae all*. ‘How do you know the elephant is under the bed? – The bed is under the ceiling’) with approx. 25.000 variants; **(compound) puns** (*Missugune suu ei räägi? – Kotisuu*. ‘Which mouth can not speak? – A bag’s mouth.’) approx. 5.000 variants; **initial letter puzzles** (*J-ga ilmas olen toiduks, R-iga mässu järel võiduks? – Jahu / Rahu*.) approx. 3.000 variants; **abbreviation formulae** (*SAA 1x5A2? – Saa ükskord viisakaks!*) approx. 1.000 variants; **acronyms** (*ETA? – Eeslid tulevad appi*.) approx. 2.000 variants; **attention tests** (*Te olete lendur ja lendate kaks korda nädalas liinil Moskva-Jerevan, vahemaandumisega Taškendis. Kui vana on lendur? – Teievanune*. ‘You are a pilot and fly twice a week en route Moscow-Yerevan with a stop in Tashkent. How old is the pilot? – Your age.’) approx. 200 variants; **puzzles of relatives** (*Ta on sinu vanemate laps, aga pole sinu õde ega vend? – Sina ise*. ‘Who is your parents’ child, who is neither your sister nor your brother? – You.’) approx. 500 variants; **calculation trick questions** (*Kahe poisi vahel jaotati neli õuna*.)

Kuidas sai kumbki poiss kaks õuna, kui korvi jäi üks õun? – Üks poiss sai õuna korviga. ‘Four apples were divided between two boys. How could both boys get two apples, if one apple was left in the basket? – One boy got an apple together with the basket.’) approx. 300 variants; **droodles** (*Neeger pimedas toas.* A black man in a black room) approx. 7,200 variants.

Traditional riddles have been entered into an extensive digital database and have been released in print in two academic volumes. The research group of short forms at the Department of Folklore Studies at the Estonian Literary Museum aims to provide a digital access to the newer peripheral riddle material for everyone interested, and to digitalise all the newer riddle forms listed above into electronic publications. The first in the line was the electronic database of 7.200 Estonian droodles, compiled by Piret Voolaid in May 2002 with the support of the Cultural Endowment of Estonia, <http://www.folklore.ee/Reebus>, ISBN 9985-867-36-X; owing to the growing international interest towards droodles the Estonian version has been translated into English and is available on page <http://www.folklore.ee/Droodles>. The e-database of nearly 5.000 Estonian (compound) puns (<http://www.folklore.ee/Sonamang>) and the digital database of 25,000 Estonian conundrums (<http://www.folklore.ee/Keerdkys>) of analogous outline, technical structure and form are under construction. The same is planned for all the peripheral subcategories of riddles listed above.

All the databases are based on separate paper files compiled mostly by the researchers of the short forms group, who have copied peripheral riddle material from the Estonian folklore collections and printed sources to the files since the 1930s.

The systematisation of the paper files has reached the point where it requires digitalisation into a database to continue and facilitate the work. The digital online database is a free e-publication, designed as an illustrative educational material for Estonian language classes in secondary schools and courses on folklore for university students majoring in language and literature, aiming to introduce one of the most productive short form genres in modern folklore. The constructed database will be especially useful for the members



of the research group: it should facilitate the obtaining of results considerably.

TECHNICAL STRUCTURE

The technical structure of databases enables to (1) reproduce peripheral riddle texts with a search engine, taking into consideration the textual data important from the folkloric perspective (question, answer, archive reference, collector, place and time of collection, subject word) and (2) reflects the potential of the genre's typological taxonomy. Let us have a closer look at this on the example of the database of Estonian doodles in English.

Search Engine of the *Estonian Doodles* digital database

Attribute **question**. Regarding the question and the answer as the two form elements of a doodle, we will realise that the question may be divided in two: the required picture may be accompanied by the verbal part, the 'What can you see in the picture?' 'What is it?'

question. The only category this does not apply to is narrative rebuses, where the question is a long story, requiring a one-word answer. The long question part generally contains a lot of unnecessary information. In performance situation the verbal part of the question is required, at least while presenting the first droodle. But since most of the droodles and rebuses of schoolchildren have reached the archives from written questionnaires, the textual question is missing altogether.

Attribute **answer** contains most of the information required by the user, as it presents the figurative explanation of the droodle itself.

e.g. We know that bear is the most productive character of droodles, but how productive is panda? The conducted search provides us the answer: panda occurs in two texts.

Attribute **archive reference** enables to search for texts by the unique archive reference, and provides an overview of specific material within a volume, which would be extremely time-consuming in paper files.

Attribute **collector** enables to reproduce all the texts by one collector, which in paper files has proven impossible.

Attribute **topography** marks the collection topography and includes the home parish of the informant and the specified known location. Since most of the droodles have been collected from schoolchildren, it is important to indicate the school: local variants of droodles are often submitted by pupils of the same school. A school generally gathers together children of the same parish, though some schools are attended by children from all over Estonia (e.g. the Haapsalu Sanatorium School, the Nõo Secondary School of Sciences). The home parish and the location of the school are separated by a semicolon.

Attribute **date** enables to search droodles by the year of collection. Originally, i.e. before the search engine was completed the texts were sorted by date: the active link on the main page opens the sorted lists.

The chart in the Estonian version of the database introduces the general collection work of droodle files. The oldest text was submit-

ted to the files in 1963. While browsing the database material by the years of collection, we cannot make generalisations claiming that droodles were not known in Estonia before 1963. They were, but before this date they were not submitted to the archives. The material collected later than 1996 have not been copied to the paper files, therefore they are not included in the database. In the nearest future, though, the droodles collected later than 1996 will be added to the digital database as well.

The differentiation of attribute **subject** was required during the systematisation of the paper files and was necessitated by the two facets of droodles (visual and textual) and the need to arrange them in the paper files. While pictures could not be sorted alphabetically, a prominent subject had to be selected from the text, often by intuition: only then it was possible to find a specific text by a specific subject word.

Database users may be surprised to find that some droodles, which traditionally contain the word *kass* 'cat' are placed under the subject word *elefant* 'elephant'. The number of subjects is rather limited, and this enables to find droodle texts from the paper files.

TYPOLOGICAL TAXONOMY IN THE DATABASE ESTONIAN DROODLES

The search engine providing all the identification data of a droodle is certainly an indispensable helper, though it has its limits.

That the digitalised data was already systematised for the paper files, it was important to show this in the electronic database. That is the reason why droodles have also been sorted by typological taxonomy; a link on the database home page opens droodle types. More than 600 groups or types of droodles, formed of variants sharing certain characteristic features, have been determined. Variants of the same type are grouped under the most characteristic, prototypical text, i.e. the text head. On the type row there is an archive reference to the text variant, known as the text head, and the number of droodles in a given type. Type redactions, i.e. variants overlapping within the same type are formed of more or less identical droodles.

While systematising any folkloric genre, categorisation into types is always subjective. The typology of the database of droodles has also specific problems. A database, therefore, reflects the typological decisions of its compiler(s) and can never be fully objective.

The formation of droodle redactions is multidimensional: during systematisation the compilers had to decide whether to proceed from (a) the varying question part, or (b) the varying answer. Both could not be considered at the same time: one had to be favoured over the other. Currently, the redactions clearly appear in the question part, i.e. the picture. Each similarly sketched picture within the corpus of the material constitutes a separate redaction. One and the same picture may have texts of widely different answer variants.

Information compiled into the database can generally be divided in three:

- (1) primary or real data, which reflects reality or a real system;
- (2) secondary or metadata, which is based on primary data and is the result of various modifications;
- (3) organisational or additional data, which enables to structure data base information in a certain way.

The digitalisation of the periphery of Estonian riddles has reached the point where the processing of primary or real data will soon be completed and the structure and technical potential of the database enables to add and complete it with new material. The use of different methods and software enables to modify primary data and achieve results that would be hard to achieve, if not unachievable, while working with the paper files.