

Ladies and Gentlemen, Fellow Colleagues!

Allow me to welcome you personally. I am extremely happy to see this conference finally materialize as a first operative step in hopefully a far leading journey at the end of which we will have a typological database of the Uralic languages.

I do remember how this idea first arose in my mind. For the last few years, I have been working on a piece of research that focused on the Uralic languages and displayed a typological character from the outset, or rather a historical typological one, to be precise. My aim was to unearth prenominal relics in the Uralic languages. In order for me to do this, I had to compare several typologically relevant grammatical phenomena in these languages, drawing naturally on an existing range of descriptive grammar in relation to them. It might have been a matter of inexperience that I only realized in the process to what extent such a seemingly average task of comparing given languages in typological terms could be prone to such difficulties in the case of Uralic languages. First of all, grammatical descriptions available for the relevant languages to be compared can be extremely different in regards to their scope and depth. Similarly, they differ in employing methods and patterns: some grammatically relevant phenomena are dealt with under totally different headings. To give you just one example, it is not easy at all to gather the data for patterns of possession in a given language if, say, the adnominal possession – somebody's something – and the habitive construction – somebody has something – are displayed in two completely different places in the description. Nevertheless, even this circumstance would not have set me back if I hadn't have to realize that the main cause for me not to be able to find the answer to some typologically relevant questions was that those questions had not been asked at all. That is what I call a really depressing experience as a background for typological comparison, given that the typologist's chances to discover a phenomenon in a language first in order to typologically analyse it later are virtually zero. Let us assume he or she is still capable of doing that in a particular Uralic language. Even so, there remain some eighteen of them that he or she cannot do anything about.

I have not mentioned yet that whilst working on my research, over several years I returned to the library of the Helsinki institute for linguistics, Kotus, which seems to house the largest collection of materials in relation to Uralic languages and uralistics as a whole in the world. This institute is in the process of moving right now or has already moved to a new place in Helsinki, but I guess most of us remember it as it used to be. The library in itself was a very long room occupied in its greater part by a range of mobile bookshelves beside which there was a considerably narrow path to move along. So I remember

walking up and down that path along the room, irritated by the fact that even though I had access to almost all the descriptive grammar available literally within reach I could not find out how a particular grammatical phenomenon looks for some of the Uralic languages. It was then that this thought came to my mind, that there is a definite necessity for having a unified, transparent, operable typological database of the Uralic languages. Of course, I became aware as well of the fact that such a device can only be achieved by a continuous collective work lasting for several years.

Since I diligently went to work at Kotus over a period of several years, I cannot precisely recall whether this happened in 2003 or 2004. Be as it may, by 2005, the idea had matured to the extent that it could be presented to uralists, namely at the Tenth International Congress of Finno-Ugrists held in Yoshkar-Ola, Mari El. Those of you who were present may remember that event as well. Around 20 or 30 colleagues gathered to listen to my proposition with interest, some of them were perhaps even inspired, and many of those present subscribed to the list of would-be participants of the project. It was of vital importance that the idea of creating a website for the project arose and this was later carried out by Professor Johanna Laakso to whom we are very grateful – as you know this website has been operating since that time and I hope its role will only improve in the years to come.

I was speaking about inspiration. Of course, the most enthusiastic about the idea was myself. I even went as far as saying back then that the database could possibly be realized in five years for which I was later called a utopian by one of the colleagues in his contribution to the debate on the website. In hindsight, I must admit he was right, providing – as I hope – the label of utopian related to the time frame and not the project itself. Nevertheless, it cannot be denied that enthusiasm decreased in the subsequent years, even though among the remarks and contributions appearing on the website there were quite a few excellent ideas which I am thankful for. But fresh impetus was needed anyway, and that was one of the components of motivation by putting forward the idea of our present conference. As I say, it was just one of the components, maybe not even the most important one, given that there are major concrete tasks to achieve before our conference which I will be talking about somewhat later on.

This is an outline of what I wanted to say about the past of the project. Now allow me to invite you on a journey into the future. Imagine a kind of time-machine, launching us into a year in the future, say, 2025, when the typological database of the Uralic languages is already operating at full capacity and will be available for any linguist's service, whether an expert in the general typology, in the Uralic languages or in any other field in linguistics. In what follows I will try to outline a vision about that operating device.

On first approach, the typological database of Uralic languages is nothing other than a virtual, online table or grid, i.e. an arrangement of data in rows and columns. The columns of the table represent the Uralic (Finno-Ugric and Samoyedic) languages, such as *Hungarian*, *Mari*, *Nganasan* etc. whereas the rows of the grid relate to typological parameters like *basic word order*, *possessive construction*, *conditional clause* etc. The number of columns should be no fewer than 18 as we are keeping count of no fewer than this number of Uralic languages, of course it is nearly a question of decision for us to divide certain languages into several dialects, thus arriving at a greater number of columns as a consequence. The number of rows, in turn, is unlimited, providing we have arrived at a full list of the typological parameters, it could amount to hundreds of them. I anticipate that the final layout of rows and columns of the grid will be considered as closed by 2025.

Thus, our grid consists of cells, each of which is located at a given crossing point of a row and a column as would be the case in any table. Each concrete cell displays the value of the typological parameter or criterion accounted for by the relevant row in the relevant language or dialect depicted by the given column. Supposing row 48 displays the type of numeral phrases in all those languages represented in the grid, and column 7 relates to the Erzya-Mordvin language, this then shows the relevant information about the typological characteristics of Erzya-Mordvin numeral phrases in the cell labelled as row 48 column 7. In the left and right cells next to it – i.e. its neighbouring cells in the same row – the numeral structures of Mari and Moksha-Mordvin, respectively, could be found for example. The proximate cells above and below this cell would display other typological features like e.g. possessive phrases or adpositional phrases, respectively, naturally still in Erzya-Mordvin, given that all information depicted in that column specifically relates to this language. If necessary, one could take into consideration not only any specific row or column of the grid but any subset of them as well, in the same way as any partial string of some rows and/or columns could be separated. Take a specified row of interest, numeral phrases for example, but a case when you are only interested in Finnic languages. Then, you can arrange the table to display the relevant data in the columns of all respective Finnic languages but not those of other languages. Similarly, we can specify two columns at any distance apart in the table, for instance we can make a comparison between the typological features of the numeral phrases in Finnish and Erzya-Mordvin without taking into consideration any further languages. Specification refers of course to columns as well. Therefore, we can separate the description of all kind of attributive phrases in a language – which practically means picking out five to ten cells of a given column – without considering any other data from that column. Filtering of both kinds could be carried out simultaneously as well if we are going to consider

more than one, but a limited number of, typological parameters in more than one, but limited number of, languages, say, all the attributive phrases in just the Permic languages.

Clicking on either of the cells in the first, farthest left column, i.e. that of the typological parameters, we could find a general description of the relevant parameter as well as its variation in the world's languages (somewhat in the same way as the World Atlas of Language Description puts it in its *description* section). In the same place, basic literature references concerning the question could be given as well as links to some databases already present on the web including the just cited WALS for that matter. This would be of use for specialists of Uralic languages in the main who might happen not to be acquainted to the necessary degree with the general typological wisdom about the phenomenon considered. Of course, such description and/or references and hints could be attached to the cells of the uppermost row as well, i.e. to the denominations of languages and dialects, which might be useful for the non-uralist typologists seeking some details about those languages they are just taking their examples from or checking data offered by others from those languages.

You will have noticed that what I have said about the cells so far is no more than a suggestion of ways to specify or group them together in line with our purposes. All that we can know in advance about any given cell is its theme, in other words, if it was like an encyclopaedia displaying a usual explanatory article on the specified topic, we would only be able to name the title of this article. Indeed, I have not said anything about what this cell should contain in terms of concrete data. First of all: should the content of a cell be conceived of as containing a normal text or rather a piece of information formalized in some way? In view of the diversity of the prospective users, I would like to answer: let this information be given both ways. This is because in order for our typological database to fit in the body of already existing ones (i.e. for a free transition between them in either direction to be granted), it is of great use if it is coded in some common, preestablished way; on the other hand, if we want it to be accessible for a wide range of linguists, not only for typologists themselves, the information in them should be displayed in a traditional textual form as well. This question is of a theoretical–methodological character, the answer to which should be discussed and selected later on, in the same way as the desirable degree of detail and depth of any information required when displayed in a typological database.

Let me return to a point mentioned earlier, namely the fact that our table is a virtual one. Of course, it could be real as well in the sense that, if necessary, it could be displayed on the screen of a computer (or whatever will be its

descendant by 2025) in the shape of a usual grid or a part of it – I imagine those responsible for the table might need such a display. But from the point of view of gaining information from, or drawing on, the database it doesn't seem to be a practical option. Take, for instance, a researcher interested in the typological characteristics of objective conjugation in the Uralic languages. While dealing with this question, he or she wants to consider only those languages which in fact display such a phenomenon, at the same time they need the information about these conjugations to be offered in a unique and systematic way. In other words, what the user wants to have is specification and a readable or even printable display of the relevant data. The means to achieve this goal could be operating a smart searching device. Normally, the user doesn't need to see a grid or table in its rows-and-columns shape. Given a searching device, the user types in a question and expects an answer to it in a formalized or textual way whichever he or she prefers. At the same time, the user doesn't even need to know which cell or cells are underlying this answer. In other words, the database should operate in a user friendly way. By 2025, of course, this problem is bound to have been long resolved by linguists and IT-experts.

So that is how the database would work in outline – but how does (or seen from 2025 backwards, how has) it come into being? By what means is setting the content of the cells and maintaining the table as a whole going on? Here, I would like to remind you that we keep speaking about a vision, so what I am offering you is simply a possible model of things evolving.

Our database is being shaped by a continuous collective work. This, of course, doesn't mean that everyone puts in it whatever they want at their own risk as e.g. I understand Wikipedia is managed. First of all, we should have a constant supervisory board. What I have in mind is a staff consisting of linguists without the permission of which neither new material may occur in the database nor can existing ones be modified. Let us take for the sake of simplicity an example with a void cell i.e. one not containing thus far any data for the specified parameter in the given language. The supervisory board could select a linguist they consider to be the best in relation to the topic in question and then offer that individual the task of collecting material and writing it up into a final form. Otherwise, the board could invite linguists in general to submit their offer (of course following the rules and specifications given in advance as general requirements concerning content and form). The submitted material should undergo a peer-review process in a similar way as is customary for articles submitted to a journal. Should the material prove acceptable, the board gives permission to feed it into the table (in the respective cell). The content of the relevant cell could be modified then at any time, either at the board's discretion or on the basis of an initiative coming from any user. Any modification offered

in either way should follow the same method of preliminary evaluation. *Mutatis mutandis*, the process is similar when an invited or voluntary author offers to fill more than one cell, e.g. supplying material concerning some concrete parameter for several languages or working out several parameters for the same language.

The database should keep being developed on a continuous basis. It could happen, for instance, that somebody offers a new parameter (typological feature) to be attached to the grid. Then a new row appears with void cells at the outset. It is much less probable for a new column i.e. a new dialect to be offered but even that would not be a problem, while, of course a range of empty cells would emerge as well, waiting to be filled in with the relevant data. Thus, the table is constantly developing in the course of time while being simultaneously operable at any time on the grounds of its already existing content. It keeps being both developed and made use of by generations of linguists.

Let me note in passing that we should arrive at a wide range consensus about the original creation or a considerable modification of a cell's content being considered as a normal piece of publication. With a click of the mouse, the author or contributor should be identifiable as has been the case with articles and their authors in some printed encyclopaedias. Given that we are imagining such a scenario, it might even be the case that the authors and possibly the referees or even the members of the supervisory board received some kind of financial reward for their contribution. It possibly is not an inevitable precondition for their work but would certainly contribute to the enthusiasm and increase the speed of developing the database.

So that is how the state of affairs appears in perspective, as I imagine it for 2025. Now we can set out on our journey back to the present, performing an inverted countdown, so to speak.

Let us first take a short stop in 2020, at the 13th international congress of finno-ugrists. (Our time machine is far from perfect in the sense that it proves incapable of predicting the venue of the congress, whereas it can provide us with an approximate timing, given that such occasions are held every fifth year.) Imagine being a part of the audience and listening to the speech of the chair of the main plenary session. He or she is commenting on the advantages of the Uralic typological database which by that time has already been operating for ten years. The first point the chair stresses is to what a huge extent the work on the typological database has contributed to the improvement of the grammatical description of the Uralic languages. This is due to the fact that the questions emerging in connection with the task of filling in the cells of the database often highlighted unforeseen gaps in the research on certain languages, in some cases they even revealed the total lack of inquiries within the relevant topic. In fact,

the requirements presented by the work on the database will have enforced Uralists to overcome deficiencies in uncovering the properties of some languages, which both served the interests of grammatical description of the relevant languages and contributed to the improvement of the typological database itself. Another asset of the database, the plenary session is hearing, is that challenging void cells provide an almost inexhaustible supply of topics for research, on a wide range of work from graduate students' assignments to PhD dissertations. Thus, our database has been playing the role of a catalyst, as it were, in Uralistics as a whole.

Let us now take another ten year's jump back to 2010. We are again attending the finno-ugrists' congress, but this congress is only the eleventh in the series. This in turn is close enough in time in order for us to know the venue as well. It is Piliscsaba, Hungary. According to the schedule of this conference, work is carried out in sections on a particular day. In one of the sections or symposia, as the organisers put it, it is precisely questions of typology that are discussed, with a special regard to the typological database of Uralic languages. It is during this section's work that an official announcement is made that the frames of the project have been set and therefore the concrete work on it can be started. We can imagine that participants should be given a booklet containing the core information about the ways of operating the project and this content should appear on the website of the project as well.

In order for this announcement to be made, the working groups of the project should carry out serious preliminary work in the period between our present conference and the 11th congress. Thus, here are the tasks for the important period between the autumn of 2008 and the summer of 2010. The *theoretical* workgroup – albeit tentatively – defines the set of languages and the set of parameters (i.e. the core of the table which we know is liable to be later extended in both directions); further, it produces an elaborate scheme or pattern for the prospective authors giving guidelines as to how to fill in a cell. The *operative* workgroup carries out the necessary improvements and changes for the website, compiles and continuously takes care of the list of the contributors, developing a system of contact with the authors and prospective users. As far as the *financial* workgroup is concerned, its main task is, of course, fund raising which means seeking out financial sources as well as preparing and submitting the relevant applications. (I dare not set it as a task for them to succeed with every application, even though it would be of course highly desirable.)

All this considered, we are back to our present conference, more precisely its closing phase, tomorrow. What do I expect for our conference to achieve in terms of organisation? The main point is that the three working groups just mentioned should be created, if not wholly than at least their leaders and some

of their members need to be chosen (with the possibility for their staff to be extended later). The members of the respective groups are expected to get in contact and keep in touch with each other so they can carry out collective work that, in my opinion, could in reality be performed via the internet and by email, from time to time perhaps even conducting online meetings.

The essence of the task for the next two days ahead of us is theoretical and organisational preparation for the planning of our project. As it has been worded on the website about the project, as well as put by our chair a few minutes ago, our present meeting has been planned as a kick-off conference, so to speak, which should set in motion at least the work on the preparation of the project. Given that the aim of our project in preparation is to create a typological database for the Uralic languages, the rationale behind organising this conference has been to close the existing gap between the two relevant disciplines i.e. linguistic typology per se and the study of Uralic languages. For the time being, it is exactly this gap which seems to be the main obstacle in such an undertaking. For the last two centuries, specialists of Uralic languages have admittedly made an enormous effort in describing them. At the same time, it cannot be denied that, especially for the first part of this period, the focus has mostly been on proving the genetic connectedness between these languages and stating the extent of their relationship, in other words, description was rather conceived of as preparation of the soil for the genetic analysis. As far as interest in typology is concerned, it does not seem to have had a very long past in uralistics, and even then, the attention has to a greater extent been paid to morphological rather more than syntactic typology. Although undeniably efforts to address typological characteristics of particular Uralic languages have been made in the last few decades, this kind of approach has not often been applied to the language family as a whole. Nor can the interest in typological topics be considered as overwhelming in present day uralistics, albeit the situation is considerably better than it was earlier.

On the other hand, and perhaps not quite so independently from what has just been said about the status of typology within uralistics, general linguistic typology, which has exhibited notable evolution, if not revolution during the last fifty years, seems to have paid somewhat limited attention to the Uralic languages. It often turns out to be the case that one encounters precipitous statements about Uralic languages in articles and books written by non-uralist typologists, sometimes what we can find about a relevant Uralic language in some typological surveys of the world's languages is quite surprising for a specialist of that language. So if I said earlier that in the descriptions of the Uralic languages we can often encounter a lack of answers to the questions which would be of special interest for the typologist, now I can add that in

typological work, in turn, there sometimes seem to be answers given to questions that haven't actually been asked.

This kind of “keeping a distance”, so to speak, between uralistics and typology is certainly undesirable anyway, but it is particularly unacceptable as a background when creating a typological database of Uralic languages. Among other things, I consider our planned work very important because it can be capable of filling in this void between these two disciplines. While becoming involved, uralists can learn to approach their languages from the perspective of typology whilst typologists can get a real picture about the typological shape of those languages.

Having said that, we must admit that such a mutual learning process cannot be simultaneous for the time being. In order to create a typological database of Uralic languages, it is the uralists who need to learn about typology first. We are certainly not inventing the concept of a typological database right now. Typological databases do exist, and even though we cannot presume that they can be directly adapted to the Uralic languages without any corrections, it remains to be the case that there is absolutely no sense in shaping our own database without having studied the existing ones. Therefore, in the first part of our conference we are going to listen to presentations from invited lecturers providing us with the most important information on existing typological databases.

At the same time, it would be of great importance to have a closer understanding of the differences between a typological approach and a more or less traditional uralistic one. What do we – I mean typologists from the one side and uralists from the other – have to say to each other, what can we offer and what can we expect from each other? Concerned with this aspect of things I invented the genre of twin talks, though the name for them, as far as I can remember, was given by Professor Bakró-Nagy. The point is that we have invited typologists and uralists in pairs to address the same subject presenting it from their respective point of view. There is a double advantage to be enjoyed from this kind of presentation. On the one hand, we are given a picture as to how typology, on the basis of its knowledge about the world's languages, approaches the relevant topic, what typology knows about it in general, and to what types of questions it expects a proper typological description of any language, a Uralic one for that matter, to answer. On the other hand, we can become informed as to what uralistics has to offer concerning the same topic, what uralists know about their languages in the given respect and what they expect typologists should pay attention to.

In turn, in the last phase of our conference, as previously mentioned, we have to deal with questions of the actual preparation for our project, creating the relevant working groups.

Of course, over and above all of what has been mentioned earlier, one of the most important aims of our conference is to establish personal contact with colleagues interested in furthering the cause of this project to create a typological database of the Uralic languages. What may seem to be a bit out of the ordinary is that all of our contributors have been personally invited and mostly with preliminarily specified topics. But we hope that those of you present are aware of the fact that it really could not have happened in any other way in regards to this kick-off characteristic of our conference, let alone the “genre” of the twin-talks. I must apologise to those who are of course not present but who displayed willingness to deliver interesting papers based on their own research in some other fields of linguistics and who have nevertheless been regretfully declined by the organisers on the basis of the specific character of this conference. In contrast, we are very pleased to greet from our audience those colleagues who were kind enough to accept the invitation to contribute to our efforts at launching the project giving us theoretical and practical advice on the basis of all their professional knowledge, despite the fact that they will not be delivering a paper.

As is often the case, some contributors indicated in the preliminary programme regrettably were unable to attend the conference after all. Some of the “twins” turned out to be an “only child” instead which certainly doesn’t make their task easier, but, in view of the high professional level of the audience present, I think we will be able to compensate for the lack of the other twin’s paper with our comments on an ad hoc basis. Of course, the overwhelming majority of our speakers are present which we are very glad to see.

So let us proceed without further delay to the presentations. I wish every success for all of us. Thank you for listening to me.