Well-being in Information Society 2014

Conference proceedings

Well-being in Information Society 2014
Conference proceedings

13-14 November 2014
University of Pécs

Editors:
Gábor Rappai - Csilla Filó

Layout and Cover Design:
z_

This is an open access book.

ISBN:
This book is available from
http://jollet.pii.pte.hu/menu/32/27

published by:
University of Pécs

supported by:
TÁMOP-4.2.2.C-11/1/KONV-2012-0005
European Union
European Social Fund

Content

Bipolar Influence-Chain Families Instead of Lists of Argument Frames

Alberti, G. - Kilián, I.
University of Pécs (Hungary)

1. ℜEALIS: Our theoretical linguistic framework

ℜeALIS, *REciprocal And Lifelong Interpretation System*, is a new "post-Montagovian" theory concerning the formal interpretation of sentences constituting coherent discourses [15–18], with a *lifelong* model of lexical, interpersonal and cultural / encyclopedic knowledge of interpreters in its center including their *reciprocal* knowledge on each other [6–8].

The decisive theoretical feature of ℜeALIS lies in a peculiar reconciliation of three objectives which are all worth accomplishing in formal semantics but could not be reconciled so far. The first aim concerns the exact *formal basis* itself, which is often mentioned as Montague's Thesis: human languages can be described as interpreted *formal* systems [18]. The second aim concerns *compositionality*, practically postulating the existence of a homomorphism from syntax to semantics. In Montague's interpretation systems a traditional logical representation played the role of an intermediate level between the syntactic representation and the world model, but Montague argued that this intermediate level of representation can, and should, be eliminated. The post-Montagovian history of formal semantics, however, seems to have proven the opposite, some principle of "discourse representationalism": "some level of [intermediate] representation is indispensable in modeling the interpretation of natural language" [19].

The Thesis of ℜeALIS is that the two fundamental Montagovian objectives *can* be reconciled with the principle of "discourse representationalism"—by embedding discourse representations in the world model, getting rid of an intermediate level of representation in this way while preserving its content and relevant structural characteristics. This idea can be carried out in the larger-scale framework of embedding discourse representations in the world model *not directly* but as parts of the representations of interpreters' minds, i.e. that of their (permanently changing) information states:

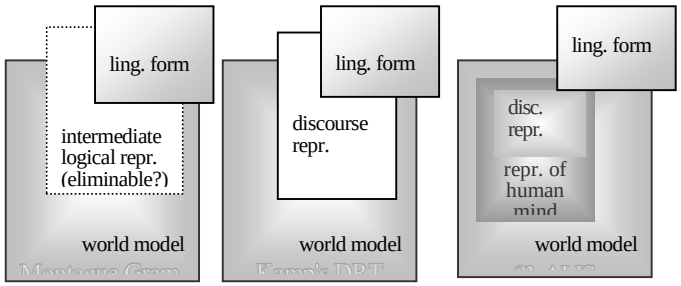


Fig. 1 Three models of the relation between linguistic form and world model

The ℜeALIS model means the formalized elaboration and linguistic application of this approach. Its 40 page long mathematical definition is available in [20]; but now we do not enter into details. Instead, we are arguing in favor of this model's relevance in language technology. Our general argument lies in the fact that intelligent tasks of language technology (e.g. machine translation, producing semantic representation) require the most sophisticated modeling of communicating human minds.

For example, a Hungarian text like the one in (2a) can only be translated into English (2b) or German (3c) by someone who knows the actual world around him/her: namely the given president's identity and sex.

a. Megjelen-t *az elnök*, de nem válaszol-t a kérdés-ek-re.
appear-Past the president but ~~he/she~~ not answer-Past the question-Pl-onto
b. *The President* appeared but *he/she* answered no questions.
c. *Der Präsident/Die Präsident-in*
the president / the president-Female
stell-te sich ein, aber *er/sie* hat keine Frage-n be-antwort-et.
place-Impf Refl in, but he/she has no question-s prefix-answer-Perf

Fig. 2 Interlocutors' knowledge about the world

The interpreter has numerous "famous" referents in his/her cultural/encyclopedic knowledge (stored in an appropriate hierarchy in his/her information state), which can be invoked by a *name* (3a). The same name, however, can refer to another entity (3b), and the interpreter should identify the appropriate one. A rich *ontology* concerning the world is also available for him/her (3c).

a. *Mozart* had a powerful influence on the work of *Beethoven*.
Beethoven knew much of *Mozart's* work.
b. J. G. Leopold Mozart (November 14, 1719 – May 28, 1787) was a composer, conductor, teacher, and violinist. *Mozart* is best known today as the father and teacher of Wolfgang Amadeus Mozart.
c. I have a half-St. Bernard and half-Scottish Shepherd, a Dalmatian and a parrot. The two *dogs* often frighten the poor *bird*.
d. Peter married yesterday. *The priest* spoke very harshly.
e. Péter el-es-ett. Jancsi csúnyán meglök-te.
Peter away-fall-Past. John brutally perf-push-Past
e'. Peter fell. John brutally pushed him. (Narr.)
e". Peter fell. John had brutally pushed him. (Reason)
f. Péter-nek van egy unokahúg-a. ∅ / *Az* kedvel-i őt.
Peter-DAT is a niece-PossSg3 ∅ / *That* like-Sg3def him/her
'Peter has a niece. He likes her. / She likes him.'

Fig. 3 Different sorts of knowledge at the interpreter's disposal

The interpreter also stores non-logical relations (3d), to be applied while building stories again and again—typically based on already-built similar stories (3d-e)), and while searching for contacts between *temporal*, *eventual* and normal referents. Text (3e) is ambiguous, for instance: if event e" of John's pushing Peter is taken to stand in a

Narrative relation with event e' of Peter's falling, the temporal referent t" belonging to e" chronologically *follows* that of e' (t'); whilst if e" is construed as the Reason of e' then t" *precedes* t' [17]. The interpreter's information state should be modeled so that such relations could be taken into account. A translator should choose between (3e') and (3e") above on the basis of the pragmatic difference concerning the rhetorical/discourse relation of the events described by the two sentences in (3e).

The model of the interpreter's information state should also contain a "topic cursor", because in the course of building discourses from clause to clause, interlocutors are aware of the topic changing permanently. The Hungarian example (3f) above illustrates this phenomenon: it is made explicit by the absence or presence of a pronoun which participant is taken to be the *topic* of a sentence relative to the preceding sentence. This factor is also relevant to the translator (as pronouns show no gender in Hungarian).
The set of examples in Figure 4 below concerns the rich and explicit Hungarian system of operators to be interpreted logically [21]. Based on his/her background knowledge and the "relevant set" as a part of it, one can infer the presence and place of some unnamed participants from the operator and the named participants of the discourse.

	P(x)	¬P(x)
$\exists x \in R_n$	Quantifier 'also' ($Q_\exists$) a. *Ili is el-jött.* IJK~~LM~~ Ili also away-came 'Ili has come too.'	Contrastive topic (K) b. *Ili (bezzeg) el-jött.* IJK~~LM~~ Ili (despite) away-came 'As for Ili, she's come.'
$\forall x \in R_n$	Quantifier 'every' ($Q_\forall$) c. *Mindenki el-jött.* IJKLM Everybody away-came 'Everybody has come'	Focus (F) d. *Ili jött el.* ~~IJKLM~~ Ili came away 'It is Ili that has come.'

Fig. 4 The system of Hungarian operator meanings:
$R_n = R \setminus R_m$, where R_m: mentioned participants, R: every participant which could have played the role played by the mentioned participants

The table above summarizes the logical implicature of the Hungarian operators (apart from *topic* (3f), whose interpretation is not of logical nature). Behind the illustrating sentences, let us suppose the following basic situation: the speaker had invited his five friends Ili, Juci, Kari, Lea and Mari; they make up the *relevant set*. Then he evaluated different outcomes of the party depending on the participants by (4a–d).

Let 'every' (4c) be our starting-point: this operator practically retrieves the relevant set; it is claimed that what is predicated is predicated of *each* member of the relevant set without naming them (IJKLM in the example). Operator 'also' (4a) refers to the *existence* of (at least) one unnamed participant satisfying what is predicated (JK). The contrastive topic (4b) refers to the *existence* of (at least) one unnamed participant *not* satisfying what is predicated (LM), whilst the focus (4d) refers to the fact that *each* unnamed participant (JKLM) is such that he/she/it does *not* satisfy what is predicated. All this information may be relevant in the course of picking the semantic content of a text.

The three-clause discourse in Figure 5 below illustrates our background knowledge concerning the central topic of this paper, *influence-chain families*, which are parts of some kind of extended lexical knowledge. The Hungarian verb form *elásat* (away-dig-Cause) in sentence (5a) exhibits explicitly only the relation between a Causer (LJS.) and a Patient-like participant (the treasures). Getting to the straightforward interpretation of the second sentence (5b–c), however, requires the retrieval of the following chain of influence: the Causer exerts some direct influence (via an order, say) upon an Agent, which is a group of men, who dig frozen pieces of soil (another Patient?) out of their original place (Source) with the aid of their mean spades (Instrument), forming holes (another kind of Patient); the treasures get into these holes due to the Causer's will. The role types, mentioned above, come from the toolbox of the theory of thematic roles [22–23], whose improved model relying on abstract role hierarchies [12–13] is applied in our project.

a. Long John Silver *el-ás-at-ta* a rabolt kincs-ek-et.
LJS. away-dig-Cause-Past the stolen treasure-Pl-Acc
'LJS. had the stolen treasures buried.'
b. *Az emberei* nap-ok-ig küszköd-tek *a fagyos föld-del*,...
the his_men day-Pl-Term struggle-Pl3 the frozen soil-with
'His men struggled with the frozen soil for days,...'
c. és alig sikerül-t *kellően mély gödr-ök-et* csinálniuk *a hitvány ásóik-kal.*
but not succeed-Past enough deep hole-Pl-Acc to_form the mean their_spades-with
'...and they could hardly produce sufficiently deep holes with

Fig. 5 Argument roles and chains of influence

We represent the standpoint, thus, that the most intelligent tasks of language technology require such modeling of communicating human minds that relies on a vast (and dynamically improvable) cultural/encyclopedic, ontological and immensely extended lexical information basis. The mathematically formalized ℜeALIS model (***Re**ciprocal **a**nd **L**ifelong **I**nterpretation **S**ystem*) offers just this, due to its "reciprocal" and "lifelong" character: its reciprocity guarantees capturing communicational interaction, and its lifelongness means huge, permanently expandable databases. In other words, we argue that the future of computational content extraction and translation lies in a functioning of the machine as if it simulated the observing and translating (interpreting) human being.

The simultaneously recursive definition of the interpreter model of ℜeALIS [6&7&20] can be sketched briefly and informally as the description of a lifelong process in the course of which in a set of referents (unstructured at the beginning, in the idealized moment of birth) four relations are expanding, mapping impacts of the surrounding world. There is an "anchoring" function α, to associate the referents pointing to the same participants (found in the course of interpretation chiefly on the basis of successfully associating arguments with their predicators and antecedents with the coreferring definite nominal expressions). In an interpreter's mind, for instance, there are many referents lying in relation α involved in formulas in pieces of information concerning W. A. Mozart, the famous musician. There is a "level function," denoted by λ, which can be

understood as a generalization of constant–variable distinction, known from classical logic; what substitutes for this two-valued system is a classification of referents — pointing to entities of the external world or representing the fictive internal objects of our beliefs, desires, intensions, suppositions, plans etc. — into classes of a rich partially ordered structure we call worldlets. The third relation, cursor κ, is to capture the interpreter's concentrated attention, fixing on temporal, spatial, topical and eventual referents in a momentarily changing way. Finally, we call the fourth relation the eventual function, denoted by σ; its task is to form formulas from referents in order to capture events and states (i.e. eventualities).

The eventual formula which serves as an example in Figure 6 below, for instance, describes essentially the event in (5a) above "in the language" of DRT [15–16], linking it to a particular point of time.

SCOPE ORDER	1 > 2 > 3 > 4 > 5 >...
ROLE LABEL IN INFLUENCE-CHAIN	e.g. Agent, Patient, Pat.', Pat.", Causer, Causer', Instr., Time, Space
CASE PROMINENCE	Central (→) > Non-central (subcat'd) argument (●) > Free adjunct (○)
ROLE IN INFORMATION STRUCTURE	{T (Topic),K (Contrastive topic}*^ {F (Focus, Q (Quantifier}*^ (M (Modifier of V))^C (Complement)
DEGREE OF REFERENTIALITY	+def > [-def,+spec] > [-spec,+ref] > [-ref,+exp] > ∅
SPEECH ACT PARTICIPATION	e.g. Sg1 > Sg2 > ... > Pl3

Fig. 6 Parameters of σ deciding argument roles, and their values, for an eventual formula, e.g. (el-ás-at: away-dig-Cause)

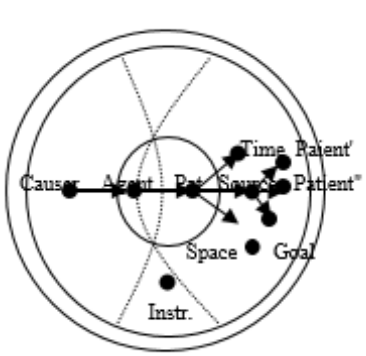
$e_{elásat} : p_{elásat}\ t_{21\text{-}08\text{-}2012}\ r_{LJS}\ r_{treasure}$

ℜeALIS captures eventual formulas by means of relation σ as follows: σ is specifically a two-argument function which maps a referent (e.g. $e_{elásat}$, which qualifies to be an *eventual referent* just by this first-argument status of σ) and a *role parameter* to a predicator (e.g. $e_{elásat}$), or a temporal referent (e.g. $t_{21\text{-}08\text{-}2012}$), or an argument-referent (r_{LJS} and $r_{treasure}$ are such referents here; whose number, otherwise, ranges from 0 to 4 or 5, or maybe more). All this, up to this point, is scarcely more than a technical innovation; the essential innovation comes now: it is the vector of values of the role parameter that selects some participants from a bipolar influence-chain family, setting all values, according to the right side of the table above in Figure 6, of the parameters listed on the left-hand side. We show in what follows how many grammatical features, restrictions, additional phonological and semantic properties of the clause or constituent forming around a given verb (or other predicator) can be captured due to this approach based on the two-

argument eventual function σ First of all, however, the concept of bipolar influence-chain families will be discussed.

2. Bipolar influence-chain families

Let us consider the visual representation of the bipolar influence-chain family belonging to a Hungarian verb stem (and its affixed variants):



a. ⟨Causer, Patient"⟩
LJS. *el-ás-at-ta* a kincs-et.
LJS. away-dig-Cause-Past the treasure-Acc
'LJS. had the treasure buried.'
b. ⟨Agent, Patient⟩
As-t-ák a fagyos föld-et.
dig-Past-Pl3mthe frozen soil-Acc
'They dug the fr. soil.'
c. ⟨Agent, Source⟩
Mély gödr-ök-et ás-t-ak.
deep hole-Pl-Acc dig-Past-Pl3

Fig. 7 A bipolar influence-chain family (ás 'dig')

The definitions are available in [12–13], which offer also arguments for considering the mathematical content of this representation to be the mental registration of the semantic character of verb meaning, with special emphasis on the meaning factors that thematic roles are intended to capture [22–23]. It is also important to note that a representation like this can be precisely calculated on the basis of a learning set consisting of transitive argument-structure versions, and then it is possible to predict in a much bigger set of transitive argument-structure versions which variants exist and which ones are excluded. In linguistics subject and object play a distinguished role (in contrast to oblique cases): they are called *central* arguments in Figure 6 above.

The principal law is that the Subject→Object direction in a transitive argument-structure version should follow the direction of the influence chain (i.e. the direction of the arrows in the representation). (7a–c) above demonstrate three Hungarian argument-structure versions of the same verb stem *ás* 'dig' keeping this law. The ⟨Agent,Space⟩ pair, for instance, also keeps the law, and does give a felicitous transitive argument-structure version with the meaning "dig the whole area over" (*felássa az egész szigetet*). A temporal participant can also occupy an object position in the family if it serves as an Incremental Theme [24]: "dig through the whole day" (*végigássa a napot;* ⟨Agent,Time⟩). The figure shows that Time, Space and Source ("dig out a hole") are incommensurate objects, in the sense that the transitive argument-structure versions formed with them offer alternative broadening of the soil-digging core meaning.

Similarly, the hole-digging meaning has also three facilities of broadening in the given Hungarian family: the ⟨Agent,Goal⟩ transitive pair refers to digging of a grave (*ás egy sírt*), the ⟨Agent,Patient"⟩ transitive pair refers to digging of something out of the soil

(*kiássa*), whilst the ⟨Agent,Patient'⟩ transitive pair refers to burying of something (*elássa*); the last meaning thus requires another verb stem (*bury*) in English, but not in Hungarian.

The two arcs in the representation decide two plum-like territories. The one on the left-hand side contains the potential *subjects* of transitive variants. In the given family shown in Figure 7 Causer and Agent are the two potential transitive subjects. The right plum consists of the potential objects. As can be seen, Causer and Instrument are excluded from this facility (NB Instrument is permitted to occupy neither central role in the family of *ás*, in contrast to other Hungarian families of verb stems). Agent is on the borderline, referring to the fact that a ⟨Causer,Agent⟩ transitive pairing is neither perfect, nor ill-formed in the family (*?ásatja az embereit* 'make their men dig'); but a ⟨Patient,Goal⟩ or a ⟨Source,Patient'⟩ transitive pairing, for instance, is totally impossible in the family. The bipolar character thus is shown by the two plum-like territories.
As for traditional thematic-roles [22], they serve only as labels (and reminiscence) in the representations of verb-stem families in the ℜeALIS model; their semantic character is defined within the rich structure of bipolar influence-chain families in de Saussure's structuralist spirit.

We hypothesize that each verb stem in each language can be characterized by a bipolar influence-chain family, which may be much smaller or perhaps even bigger but follows the same partially ordered structure. Let us compare, for instance, two typical English verb stems:

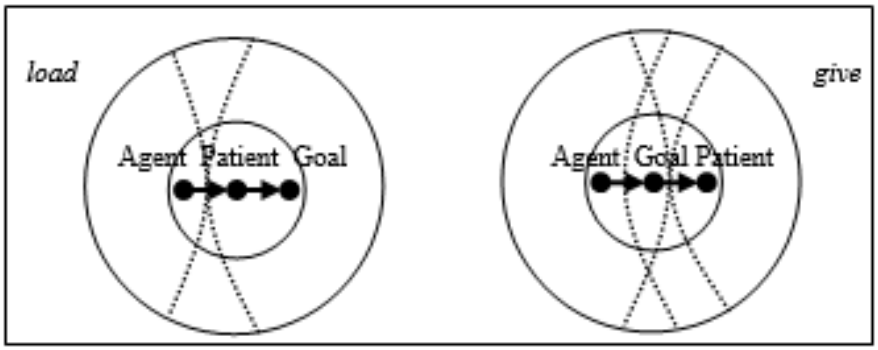


Fig. 8 Bipolar influence-chain families (partial) of two English verbs:
*⟨**Agent**,Patient⟩: **He** loaded hay onto the wagon. / **He** gave a book to Mary.*
*⟨**Agent**,Goal⟩: **He** loaded the wagon with hay. / –*
*⟨**Agent**,**Goal**,Patient⟩: – / **He** gave **Mary** a book.*
⟨Patient⟩: The hay was loaded onto wagons. / The book was given to Mary.
*⟨**Goal**,Patient⟩: – / **Mary** was given a book.*
⟨Goal⟩: The wagon was loaded with hay. / –

It can be observed that *load* and *give* provide distinct transitive argument-structure versions (see above); this fact can be attributed to their distinct bipolar influence-chain families. The double plum-like territory on the right-hand side appears due to the special character of English (at least compared to Hungarian) that it contains ditransitive argument-structure versions. Goal can be expressed as an object while Patient is also an

object, but their order is fixed: ⟨**Agent**,***Goal***,*Patient*⟩. We also show intransitive variants; the representations prove that in the family of *load* Patient or Goal is permitted to be associated with no object to form a transitive variant, whilst in the family of *give* the ⟨Goal,Patient⟩ pair provides a well-formed transitive variant, certainly in this order, but Patient can never appear as a transitive subject.

3. The vector of grammatical parameters of function σ

As was said at the end of Section I, in the model of ℜeALIS relying on four relations (anchoring α, level function λ, eventual function σ, cursor κ) it is function σ that is responsible for the organization of representation of states and events, assigning an eventual referent (and a parameter vector) the referent serving as the representative of the predicator and the referents belonging to the arguments (and a temporal referent).

As for selecting the predicator, we would like to raise the reader's attention to the advantageous approach of ℜeALIS that if one may come across any variant of, say, verb stem *ás* 'dig' in a text, (s)he need not decide promptly which argument-structure version (s)he has happened to come across; what (s)he should only establish is that the phonetic form *ás* 'dig' has retrieved the whole bipolar influence-chain family shown in Figure 7; so a basic, highly underspecified, meaning is set. It is then left to a later specifying procedure to decide the particular argument-structure version observed in the given text. In this way we adequately capture the extremely rich polysemy typical of verb stems in all languages.
This specification essentially lies in the setting of values of the vector of grammatical parameters in the second argument position of the eventual function of ℜeALIS, responsible for argument positions of the given verb; see Figure 6 above. A decisive part of the task is to select a set of *role labels* out of the bipolar influence-chain family and to select a distinguished subset (concerning *case prominence*), deciding in this way, respectively, the (explicit) arguments to be associated with the predicator and those obtaining *central* function, that is, appearance as a subject or an object. Note that only the central status should be indicated, the subject/object choice is defined by the influence direction in bipolar influence-chain family.

Figures 9 and 10 below illustrate two deliberately specified selection of argument frame (on the basis of Figure 6) out of thousands of facilities in Hungarian; they share the choice that Causer (the role of the leader who instructs others) and Patient' (the role of the treasure to be buried) have been given a central function out of the roles whose semantic system, i.e. their bipolar influence-chain family, is shown in Figure 7 above. The law of influence direction makes it unambiguous that a ⟨Causer,Patient'⟩ transitive argument-structure version has been formed.

ás	SCOPE	ROLE LAB.	CASE PR.	INFOR.	REF.DEGR.	SAP
+ref	1	Causer	→	T	+def	Sg3
$(+ref)_F$	2	Space	○	F	-def, +ref	Sg3
+ref	3	Patient'	→	C	+def	Pl3
+ref	4	Agent	●	C	+def	Sg1

Fig. 9 A distribution of argument positions belonging to ás:
a. 'LJS. egy ''lakatlan szigeten ásatta el velem a kincseket.
LJS a uninhabited island-on dig-Casue-Past away with_me the treasure-Pl-Acc
'It is on an uninhabited island that LJS. had the treasures buried by me.'
b. ?''LJS. egy ''lakatlan szigeten ásatta velem el a kincseket.
c. ?''LJS. egy ''lakatlan szigeten ásatta el a kincseket velem.
d. ??'LJS. egy ''lakatlan szigeten ásatta velem a kincseket el.
e. ??'LJS. egy ''lakatlan szigeten ásatta a kincseket velem el.
f. ???'LJS. egy ''lakatlan szigeten ásatta a kincseket el velem.

The table in Figure 6 has indicated that, in addition to its thematic label and grammatical function, formal properties of an argument getting in a sentence are influenced by some further parameter values. They all are worth uniting in a single vector of parameters, because in this way all kinds of grammatical restrictions can be formulated (and can be built in our computational implementations) as a (partly language-specific, partly universal) requirement on the well-formedness of this vector. In this approach the problem of machine translation practically boils down to the question whether a given parameter vector of the source language is well-formed in the target language, and if the answer is not, the parameter vector "nearest" to the one in the source language should be found in the target language. The interested reader is referred to Section 3 of [5], where we compare Hungarian sentences with different topic–focus constructions to their active and passive English equivalents.

ás	SCOPE	ROLE LAB.	CASE PR.	INFOR.	REF.DEGR.	SAP
$(+ref)_F$	1	Space	○	F	+def	Sg3
$(+ref)_F$	2	Agent	●		-def, +ref	Sg3
+ref	3	Causer	→	C	+def	Sg3
+ref	4	Patient'	→	C	+def	Sg3

Fig. 10 Another distribution of argument positions belonging to *ás*:
a. ?''Madeirán ásatta el LJS. a kincseket egy ''tússzal.
Madeira-on dig-Cause-Past away LJS the treasure-Pl-Acc a hostage-with
'It is on Madeira and by a hostage that LJS. had the treasures buried.'
b. ? ''Madeirán ásatta el egy ''tússzal LJS. a kincseket.
c. ?? ''Madeirán ásatta egy ''tússzal el LJS. a kincseket.
d. ??? ''Madeirán ásatta el LJS. egy ''tússzal a kincseket.

In Hungarian, in contrast to English, the scopal order of arguments is independent of their thematic roles and grammatical functions (Figure 6). Figure 9 illustrates this scopal order: Causer > Space > Patient' > Agent, while Figure 10 shows a significantly differing

one: Space > Agent > Causer > Patient'. Remember the two variants share the distribution of grammatical functions.

The roles in information structure, however, highly depend on the scopal order, though the standard hypothesis [26] concerning a strict order of [Topics > Quantifiers > Focus > Verbal modifier > Verb > Complement] has been liberalized [27–28], essentially in the way shown in Figure 6 above (a focus can be followed by other foci or quantifiers). Figure 9 illustrates the assignment of a topic role to Causer and a focus role to Space. Patient' and Agent remain simple complements, without any distinguished role in information structure. Figure 10 shows a *mirror-focus* construction [26–27] assigned to the ⟨Space,Agent⟩ pair. Remember Figure 4 in Section 1 provides a systematic summary of the additional meaning factors due to the different kinds of topics and quantifiers and the focus, also depending on such pragmatic factors as the relevant sets belonging to argument places in the interlocutors' background knowledge.

The degree of *referentiality* of arguments can be limited in Hungarian depending on thematic role [29–30]. A certain kind of Patient, for instance, is not permitted to be definite, due to its existential meaning factor: e.g. **Alakult a kórus a klubban* (formed the choir the club-in). This limitation, however, is deleted if a coargument receives the focus role, because in this case the content of the Patient will belong to the presupposed part of the sentence [14]: e.g. *A klubban*$_{Focus}$ *alakult a kórus* 'It is in the club that the choir has formed.' All the universal and language-specific knowledge concerning this phenomenon or similar phenomena can be built in the well-formedness conditions of the parameter vector of function σ. The person and number features are also to be considered, because Speech Act Prominence is known to show some correlation with topic selection, for instance, in several languages [25].

As can be seen under the tables in Figures 9 and 10, after the complete specification of the parameter vector we receive not a single word order intoned in an unambiguous way, but a quite long list of word-order variants furnished with intonational alternatives, arranged in a preference order. In Hungarian, for instance, a quantifier can "take its scope" advancing in the preverbal zone of the sentence, but it need not do that; that is, a quantifier is permitted to remain in the postverbal complement zone [30]. Sentence (4a) *Ili is eljött* (Ili also away-came) has a semantically equivalent version: *Eljött Ili is*.
Sentence (9a), compared to the less felicitous variant (9c), shows another interesting phenomenon concerning the preference list of competing word orders: in the Hungarian postverbal zone, word order and scopal order are practically independent. Phonetic factors will determine the word order: the less heavy (shorter) expression precedes the heavier one (mentioned as Behaghel's Law in [32]).

4. Conclusion

A dynamic discourse-semantic interpretation system [15–17], called ℜeALIS [6–9&20], is demonstrated in Section 1, which also provides many linguistic phenomena whose intelligent processing (translation by machine, summarization) requires such a total and sophisticated linguistic model as ℜeALIS.

Section 2 is devoted to the demonstration of a special area of lexical semantics: families of argument-structure versions "around" the same verb stem in natural languages. We register related argument-structure versions not as a list of argument frames, but in a more sophisticated form from a theoretical linguistic viewpoint [12–13, 21–32], called *bipolar influence-chain families*, which provide much more information concerning the occupation of argument position in sentences.

Section 3 discusses the integration of the lexical-semantic model in the dynamic discourse-semantic model (ℜeALIS) by means of the linguistic parameter vector of function σ of the latter.

We sketch the cornerstones of an implementation in Prolog in the counterpart of this paper in this volume. The advantage of this programming language lies in the facilities of the simultaneous application of forward- and backward-chaining inference frameworks and the reification of certain formulas.

Acknowledgment

The present scientific contribution is dedicated to the 650th anniversary of the foundation of the University of Pécs, Hungary. We are grateful to SROP-4.2.2.C-11/1/KONV-2012-0005 (Well-Being in the Information Society) for their supporting our research team ℜeALIS in 2013–2014.

References

[1] G. Alberti and J. Kleiber, "The *GeLexi* MT Project," *Proc. of EAMT 2004 Workshop (Malta)*, pp. 1–10, 2004.

[2] G. Alberti, K. Balogh, J. Kleiber and A. Viszket, "Total Lexicalism and GASGrammars: A Direct Way to Semantics," *CICLing 2003, LNCS 2588*, Springer, Heidelberg, 37–48, 2003.

[3] G. Alberti, J. Kleiber and A. Viszket, "GeLexi project: Sentence Parsing Based on a GEnerative LEXIcon," *Acta Cybernetica* 16 (Hungary), 587–600.

[4] G. Alberti and J. Kleiber, "The Grammar of ℜeALIS and the Implementation of its Dynamic Interpretation," *Informatica* 34/2 (Slovenia), pp. 103–110, 2010.

[5] G. Alberti, M. Károly and J. Kleiber, "From Sentences to Scope Relations and Backward," *Natural Language Processing and Cognitive Science* (*Proc. of NLPCS 2010 / ICEIS 2010 (Madeira, Funchal)*, SciTePress Portugal, pp. 100–111, 2010.

[6] G. Alberti, M. Károly and J. Kleiber, "The ℜeALIS Model of Human Interpreters and Its Application in Computational Linguistics," *Proc. of ICSOFT 2010, 5th International Conference on Software and Data Technologies, Athens, Greece,* Vol. 2, SciTePress Portugal, pp. 468–474, 2010.

[7] G. Alberti and M. Károly, "The Implemented Human Interpreter as a Database," *Proc. of IC3K, the 5th Int. Conf. on Software and Data Technologies*, Vol. 2, SciTePress, Portugal, pp. 468–474, 2011.

[8] G. Alberti and M. Károly, "Multiple Level of Referents in Information State," *CICLing 2012, Part I, LNCS* 7181, Springer, Heidelberg, pp. 349–362, 2012.

[9] G. Alberti and J. Kleiber, "Where are Possible Worlds? (Arguments for ℜeALIS)," *Acta Linguistica Hungarica* 59 (1-2), pp. 3–26, 2012.

[10] I. Kilián and G. Alberti, "A Metamodel-Driven Architecture for Generating, Populating and Manipulating "Possible Worlds" to Answer Questions," in ICSOFT 2013 Proceedings of the 8th Int. Joint Conf. on Software Technologies (Reykjavík, 29-31 July 2013), pp. 74–78, 2013.

[11] G. Alberti and L. Nőthig, "ReALIS1.1: The Toolbox of Generalized Intensional Truth Evalation," in INTELLI 2014, The Third International Conference on Intelligent Systems and Applications, pp. 60–66, 2014.

[12] G. Alberti, "Changes in Argument Structure in the course of Derivation in Hungarian," *Acta Linguistica Hungarica* 53 (1), pp. 1–39, 2006.

[13] G. Alberti, "Az argumentumok precedenciaelvű realizálódása," in Analógia és modern nyelvleírás, Erdélyi Múzeum-Egyesület Kiadó, Kolozsvár, Románia, pp. 9-41, 2013.

[14] G. Alberti, "Restrictions on the Degree of Referentiality of Arguments in Hungarian Sentences," *Acta Linguistica Hungarica* 44 (3–4), pp. 341–362, 1997.

[15] J. van Eijck and H. Kamp, "Representing discourse in context," in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam) pp. 179–237, 1997.

[16] H. Kamp, J. van Genabith and U. Reyle, "Discourse Representation Theory," in D. Gabbay and F. Guenthner (eds.), *Handbook of Phil. Logic*, vol. 15, pp. 125–394. Springer, Berlin, 2011.

[17] N. Asher and A. Lascarides, *Logics of Conversation*, Cambridge Univ. Press, 2003.

[18] D. R. Dowty, R. E. Wall and S. Peters, *Introduction to Montague Semantics*, Reidel, Dordrecht, 1981.

[19] P. Dekker, "Coreference and Representationalism," in K. von Heusinger and U. Egli (eds.), *Reference and Anaphoric Relations,* Kluwer, Dordrecht, pp. 287–310, 2000.

[20] http://lingua.btk.pte.hu/realispapers

[21] K. É. Kiss, *Hungarian Syntax*. Cambridge Univ. Press, Cambridge, 2001.

[22] A. Komlósy, "Complements and Adjuncts," F. Kiefer and K. É. Kiss (eds.), *The Syntactic Structure of Hungarian*, *Syntax and Semantics* 27, San Diego: Academic Press, 91–178, 1994.

[23] Ch. J. Fillmore, "The Case for Case," in E. Bach, R. T. Harms (eds.), *Universals in Linguistic Theory*, NY, 1–88, 1968.

[24] D. R. Dowty, "Thematic Proto-Roles and Argument Selection," *Language* 67/3, 547–619, 1991.

[25] W. Croft, *Radical Construction Grammar. Syntactic Theory in Typological Perspective*, Oxford University Press, 2001.
[26] F. Kiefer and K. É. Kiss (eds.), *The Syntactic Structure of Hungarian*, *Syntax and Semantics* 27, San Diego: Academic Press, 91–178, 1994.
[27] K. É. Kiss, "Multiple Focus in Hungarian Sentence Structure" (in Hungarian), in *A mai magyar nyelv leírásának újabb módszerei II.*, 47–66, 1995.
[28] G. Alberti and A. Medve, "Focus Constructions and the »Scope–Inversion Puzzle« in Hungarian," *Approaches to Hungarian* **7**, Szeged: U. Szeged, Dept. of General Linguistics & Dept. of Hungarian Language, 93–118, 2000.
[29] A. Szabolcsi," From the Definitness Effect to Lexical Integrity," in W. Abraham, and S. de Meij (eds.), *Topic, Focus, and Configurationality,* , Amsterdam: John Benjamins, pp. 321–348, 1986.
[30] K. É. Kiss, "Definiteness Effect Revisited." *Approaches to Hungarian*, Vol. 5, pp. 63–88, Szeged, 1995.
[31] B. Surányi, *Freedom of Word Order and Domains for Movement: A Flexible Syntax of Hungarian.* Acad. dr. diss., Research Inst. for Linguistics, Hungarian Academy of Sciences, 2011.
[32] K. É. Kiss, "Is free postverbal order in Hungarian a syntactic or a PF phenomenon?" in N. Erteschik-Shir and L. Rochman (eds.), *The Sound Pattern of Syntax,* Oxford Univ. Press, 2009.

Bipolar Influence-Chain Families Instead of Lists of Argument Frames. Cornerstones of Implementation in Prolog

Kilián, I. - Alberti, G.
University of Pécs (Hungary)

1. Input and output

The basic principle underlying ℜeALIS is to decompose syntactic descriptions, to store the decomposed syntactic elements in the lexicon, and to bind immediate semantic relations to lexical elements. Semantic structures are built up therefore not in a separate pass, but together, in parallel with syntax analysis. Such an approach needs careful preparations and design from the software designer and also from the implementor. Such a complexity can only be reached by a symbolic programming language, which is also called „artificial intelligence language". Because of previous experience we have chosen Prolog for this.

When designing an algorithm, the first step is to specify its input and output. In the black-box approach an algorithm is regarded as a mathematical function whose domain is the input, and the range is the output. This specification can also be the basis for a partial implementation: either if only a selected dimension of a multi-dimensional result set, or if only certain steps of a function-chain are implemented.

In the case of ℜeALIS, the relationship between the text and the discourse representation is calculated [6]. If we regard the text as an input, then the expression is calculated which describes the interpreter's internal mental state, and which can be evaluated, proven, or added to the actual knowledge base (ontology). If the representation is the input, for instance, when the inference engine answers a question, then, as an output, we expect to produce the text itself. The former is called text recognition, the latter is text generation.

2. Non-determinism

In certain cases it is not decidable in which direction the linguistic analysis should be continued, or there are more than one feasible threads for the analysis. There are many examples to demonstrate this: e.g. homonyms, like *vár* ('wait ' / 'castle' in Hungarian), or the lazy-eager accommodation/evaluation in case of intensional expressions (*In this year the president is democrat*). In both cases, when reading the sentence, it is still impossible to determine which is the proper version. Reading the sentence further on may usually make the interpretation unambiguous.

To manage such situations, nondeterministic solutions are proposed. Nondeterministic procedures produce each possible return value for a given input as a set, and callers select their preferred value (or values) from the set. Finally they may also produce their own nondeterministic return value as a set.

3. Target model

In our case the term "target model" does not cover an instruction and/or register set of a processor, but it is rather about using the entire or some subset of first order logic so that linguistic concepts are mapped to first order logic. This could be also the subset of Horn-clauses, but we might eventually extend it (e.g. with disjunction of the consequence side, or with reified modal operators, etc.). It would be a special pleasure if we still could remain by Horn-clauses so that we also can interpret the target format with pure Prolog. It is also possible that even for Horn-clauses we must choose another resolution strategy [3].

4. Describing data structures in first order logic

First order logic is basically typeless. Yet it is possible to simulate a type mechanism so that we restrict certain arguments of functional expressions. This solution is somehow similar to Prolog's mode declarations that are used primarily to specify the bound or free state of Prolog variables [5]. Though there are strongly typed Prolog implementations, a full type description language has never been specified, and such an extension to pure/standard Prolog has never been constructed or implemented.

Another special difficulty is to map second- or higher-order logical constructs to first order logic. We may get two possibilities for this:

- *Reification*, that is, predicate symbols can be also taken as data (Prolog functional expressions), so that even a logical variable can bind them as values. A precondition of this is that predicates should not be evaluated as predicates, or they are evaluated in a separate phase of the entire theorem-proving process. In such a phase certain built-in predicates (over the pure logic) may transform the functional expressions to real predicate expressions, so that later they also may take part in theorem proving.
- *Immediate utilization of tools beyond logic*. Prolog systems usually offer certain built-in predicates to create functional or predicate expressions dynamically, in run-time (e.g. =../2). A drawback of the solution is that it is not declarative enough: the lexical knowledge base applies software means that are specific to the actual Prolog implementation.

5. Identification, referents

We denote referents by variables that are undefined (free) in the beginning but later they may be bound by a concrete value, whose actual form depends on the type of the referent.

Object referents denote one individual in the world (e.g. 'William Shakespeare'). These are instances of some class in the knowledge base, which are not identified unambiguously by their natural attributes. Another William Shakespeare, for instance, in a telephone directory is not the same individual as the famous dramatist, not even if their other properties are the same. As opposed to this, data structures with

unambiguously identifying attributes are called data types. Such data types are, for example, time referents, addresses, etc.

Time referents may be logical variables that are bound by time values. Time referents describe events in past, present or future, time points and intervals as well, with arbitrary precision. The referent past(date(1-0-0)) for example denotes the " this day a year ago" time specification.

Predicate referents are used only through reification in first order logic. Predicate referents in ℜeALIS specify a predicate pattern that is bound to the free variables of the pattern by the ^/2 switch. For example the AG^PAT^TIME^dig(AG,PAT,TIME) predicate referent denotes the concept of *digging* with three arguments (agent, patient and time). Remark that predicate referents play also the role of binding the logical form of a sentence to the concepts of an underlying ontology.
Eventual referents are logical expressions that are built up based on predicate patterns, by well known logical switches. Literal expressions over predicates and expressions connecting literals by logical switches behave like data in this situation. Their actual role in the inference (extending the ontology, proving something, etc.) is determined by the program environment.

6. Implicational and relational target model

The *relational target model* utilizes Prolog's ability to evaluate relations in both directions. According to these, if the relationship between the internal and the ASCII character-based representation is denoted by a Prolog predicate, than, provided that we apply only pure Prolog, the same program may be able to recognize and to generate at the same time. We also remark that Prolog's built-in DCG parser works also according to a similar target model. This model may seem obvious for Prolog programmers, but it has also a significant drawback. The nondeterminism, the backtracking nature of Prolog programs backtracks also the input channel, which involves unnecessarily winding the input head back and forth, which causes a loss of efficiency, and in general it means a strange behavior.

The so-called *implicational target model* emphasizes the recognition, as a main task, and implements it by inference on Horn-clauses. That is, the source nodes of the inference graph are the fact statements, which represent, in this case, the elements of the discourse (words, morphs, letters) being just read. The parsing operation itself is represented by Prolog implications: these implications may lead to end-consequences, which are represented by Prolog goals. As a matter of convenience these goals might be statements to refer to the result of some mapping from the input text to the discourse structure. Intuitively, we can also say that "floating" grammatical elements and referents and their interconnections are realized much more naturally in the inferential target model.

Note, however, that the implicational target model applies a deductive way of inferencing, which starts at fact statements, and ends up by goals as hypotheses. Deductive inferencing in our case implements recognition whereas we may also want it

to generate text in our program. For generation we need the inverse approach: an abductive strategy must be applied for reasoning, which, by knowing the hypothesis, infers facts that may deduce the hypothesis. As the experiment shows, a slight rearranging of dynamic predicate handling, the deductive Prolog program may be able to run also abductive inference, and thereby generate text from its logical form.

It would be useful to define certain layers, or cuts, in the inference graph (this definition is the same for deductive as for abductive case):

- The level of *morphological analysis*. This works on the series of characters read, and transforms it to a list of words (or morphemes), just like their indices inside their word (or sentence, or discourse). Now we assume that this operation already works seamlessly, and we can start from the result of this phase. We can also assume that the first phase has already identified the word stems, the compositions, together with their prefix and suffix system, cases and agreements. Let's take the next Hungarian sentence as an example: *Petra vágyik arra a magas német úszóbajnokra* "Petra longs for that tall German swimming champion." The result of this phase can be represented by the following set of facts.

```
word(1,noun('Petra',proper,nom,sing-3)).
word(2,verb('vágyik',[], decl,
     pres, sing-3)).
word(3,noun('az',pro(point),sub,sing-3)).
word(4,det(def,cons)).
word(5,adj('magas')).
word(6,adj('német')).
word(7,adj('úszó')).
word(7,noun('bajnok',common,sub,sing-3)).
```

- The level of *word-structure transition*. This practically discovers the predicate–complement, the base–adjunct, and the anaphora–antecedent grammatical correlations, based on the demand–offer lexical relationship. The example below demonstrates how the Hungarian adjective–noun construction *úszóbajnok* 'swimming champion' is built up. (Warning! The example is not runnable by Prolog).

```
adj('úszó'):- word(XA,adj('úszó')),
  noun(XN,_STEM,common,_CASE,_NR-3),
  order(XA,XN,0,nei).
noun('bajnok',common,CASE,NR-3):-
  word(X,noun('bajnok',common,
  CASE,NR-3)), adj(ID,S,XA,_),
  order(X,XA,-5,nei).
```

Facts containing the elements of the sentence, in a forward-chaining inference framework, start inference waves, which are combined by the result of other inference waves. This also happens in case of *úszó* 'swimming' and *bajnok* 'champion ', which, through their lexical rules, mutually demand each other's proximity. The last piece of

the rule is the order/4 condition that stores information regarding the strength of mutual binding: nei means that the words should, if possible, directly join. Parameter '-5' denotes an adjective loosely preceding the noun, but parameter 0 needs the noun attached directly, by word-composition.
Besides these, other Horn rules also describe other connections of sentence elements; like:

- predicate–complement relationship
- base–adjunct relationship: predicate or base demands a complement or adjunct in a given grammatical state and rank: there must not stand a parameter with higher rank between them.
- anaphora–antecedent relationship: pronouns and definite articles demand their antecedents, which have been mentioned before.
- the dependency relationship that generalizes these relationships above

- *The level of discourse relationships.* This is also denoted by the ℜeALIS function λ, which determines the relationship between the distinct discourse representation contexts, which are like embedded box-like structures, whereas the relationship between certain neighboring layers are determined by the rhetoric intention (Explanation, Elaboration, Hint, etc. [5]).
- *The level of eventual function.* This demands the evaluation of the formerly described function σ of the interpreter's. When doing this, all interpretation patterns and models should be taken into account (e.g. the bipolar influence-chain). The example below illustrates the Horn-inference that evaluates the desire/3 eventuality relationship.

```
sigma3(ID,S,XV,verb('vágy',[],MODE,
       VTIME,AGR),TIME,SUBJ,OBJ,
       desire(TIME,SUBJ,OBJ)):-
  verb(ID,S,XV,'vágy',[nom,sub],[],MODE,
       VTIME,AGR), TIME =.. [VTIME,_],
  noun(ID,S,XN1,_SUBJN,gqd(proper),
       nom,AGR), order(XV,XN1,-7,nei),
  noun(ID,S,XN2,_OBJN,gqd(proper),sub,_),
  order(XV,XN2,7,nei).
```

The following elements from the example are worthwhile to mention:

- MODE: the mode of the verb in the current description is indifferent.
- gqd: Generalized Quantifier-Determinant.
- AGR: number and person agreement between subject and predicate. The agreement between object and predicate is indifferent.
- =../2 call: that performs transformation between Hungarian verb tenses (past, present, future) and the time-referent of the given sentence

The eventual function is evaluated according to the model of bipolar influence-chain. Practically this can be implemented according to the relational model, where the occurrence of certain word-classes is in relationship with their semantic content predicate.

```
sigma(AG^PAT^TIME^INST^INIT^
  stabWithInit(AG,PAT,TIME,INST,INIT)),
  noun(INIT,_,nom,AGR),noun(PAT,_,acc,_-3),
  verb(caus('szúr'),_PREFIX,decl,
       VTIME,AGR),
  [noun(INST,common,inst,_-3)]):-
  time2verb(TIME,VTIME).
```

In the example above we can observe a Prolog program fragment that describes the semantics of the Hungarian word: *szúr* 'stab', as occurring in the next sentence: *Mari kiszúratta az abroncsot egy szöggel* 'Mary has had the tire stabbed by a nail.' The causative verb is denoted by the predicate stabWithInit/5, which has five arguments: AG (agent), PAT (patient), and TIME, as general arguments, but also INST (instrument) and INIT (initiator). The condition of time2verb is a simple conversion utility between the general time referent of stabWithInit/5 and the tense parameter of the verb. In the implicational model, the predicate stabWithInit is reified in the consequence side of the statement. Other arguments are referred to in the precondition side of the statement.

The program is built up of statements similar to the one depicted above, and it is astonishingly simple: they are more or less mechanical transformations of the principles mentioned in the first half of the article. Testing the program, on the one hand, has brought the expected results. On the other hand, it has demonstrated that without any semantic or other environmental information the situation is not unambiguous at all. It sometimes causes an overwhelming number of formally correct, but intuitively senseless nondeterministic results. In Hungarian, for example, the pure syntactic analysis cannot decide if *szög* 'nail' is an instrument or the agent of the action. To reject the false result, a further semantic check would be necessary [4] (for example, we could check if the agent of a verb is derived from the ontology concept of Agent).

7. Conclusion

A dynamic discourse-semantic interpretation system, called ℜeALIS, is demonstrated in Alberti and Kilián's paper (this volume), which also provides many linguistic phenomena whose intelligent processing (translation by machine, summarization) requires such a total and sophisticated linguistic model as ℜeALIS.

In this paper we sketch the cornerstones of an implementation in Prolog [1–2]. The advantage of this programming language lies in the facilities of the simultaneous application of forward- and backward-chaining inference frameworks and the reification of certain formulas.

Now we focus on the "eventual" function σ of ℜeALIS, responsible for producing formulas expressing the semantic content of events and states coming from texts in natural languages. This project is a part of a major project aiming at the complete

implementation of the ℜeALIS model of the interpreter translating or performing some other intelligent language-processing tasks.

Acknowledgment

The present scientific contribution is dedicated to the 650th anniversary of the foundation of the University of Pécs, Hungary. We are grateful to SROP-4.2.2.C-11/1/KONV-2012-0005 (Well-Being in the Information Society) for their supporting our research team ℜeALIS in 2013–2014.

References

[1] W. F. Clockshin, C. S. Mellish: Programming in Prolog Springer Verlag, New York 1987.
[2] R. O'Keefe: The Craft of Prolog, The MIT Press, Cambridge, Massachusets, London England 1990.
[3] Kilián: Contralog: a Forward-Chaining, Prolog-Conform Inference Engine, and its Application to Linguistic Parsing in the Model of ℜeALIS (in Hungarian), Proceedings of SzámOkt 2011, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, pp. 199–205.
[4] Kilián, G. Alberti: A Metamodel Driven Architecture for Generating, Populating and Manipulating "Possible Worlds" in order to Answer Questions, Proceedings of ICSOFT, 2013, pp. 74-78.
[5] Nakashima, H.: Term Description: "A Simple Powerful Extension to Prolog Data Structures," Electrotechnical Laboratory, Umezono, 1-1-4, Ibaraki, Japan, 1985.
[6] Alberti, G.: ReALIS: An Interpretation System which is Reciprocal and Lifelong. Workshop 'Focus on Discourse and Context-Dependence' (16.09.2009, 13.30-14.30 UvA, Amsterdam Center for Language and Comm.), http://www.hum.uva.nl/aclc/ events.cfm/C2B8E596-1321-B0BE-6825998CFA642DB2, http://lingua.btk.pte.hu/realispapers (2009)

Capitation based resource allocation and managed care in the Hungarian health care system

Boncz, I. - Vajda, R. - Jankó-Király, A. - Csákvári, T. - Lampek, K. - Endrei, D. – Ágoston, I.
University of Pécs (Hungary)

Introduction

In each country of the developed world there are strong financial constraints because of the rising cost of health and social services. In the former socialist – or as they are referred today the transitional – countries after the political changes around 1990 the reform of health care became an important social and political issue. As part of healthcare reforms one of the most important questions is the allocation of financial resources.

There are many papers and studies on the structure and main policy issues of the Hungarian health care system [1,2,3,4,5,6,7,8]. In this paper we would like to give a practical and detailed overview on the changes in the financing system of health care in Hungary. A pilot care managing programme (CMP) was introduced in Hungary in 1999. The conceptual foundations of the Hungarian implementation of managed care is closer to what was called the GP fundholding in the UK than HMOs in the USA. The cornerstone of the Hungarian CMP is the capitation based resource allocation, which defines the per capita health expenditures according to risk factors (age, gender) and type of health care (outpatient, inpatient, CT/MRI, home care, drugs, medical devices, etc.) [9]. The purpose of the study is to analyse the changes in the number of enrolees and capitation in the care managing programme.

---

[1] Boncz I, Nagy J, Sebestyen A, Korosi L.: Financing of health care services in Hungary. Eur J Health Econ. 2004, 5(3): 252-258.
[2] Gulácsi L, Dávid T, Dózsa C. Pricing and reimbursement of drugs and medical devices in Hungary. Eur J Health Econ. 2002;3(4):271-8.
[3] Boncz I, Dozsa C, Kalo Z, Nagy L, Borcsek B, Brandtmuller A, Betlehem J, Sebestyen A, Gulacsi L. Development of health economics in Hungary between 1990-2006. Eur J Health Econ. 2006, 7(S1):4-6.
[4] Kroneman M, Nagy J. Introducing DRG-based financing in Hungary: a study into the relationship between supply of hospital beds and use of these beds under changing institutional circumstances. Health Policy. 2001; 55:19–36.
[5] Gaal P, Szigeti S, Csere M, Gaskins M, Panteli D. Hungary health system review.Health Syst Transit. 2011;13(5):1-266.
[6] Gaál P, Szigeti S, Panteli D, Gaskins M, van Ginneken E. Major challenges ahead for Hungarian healthcare. BMJ. 2011 Dec 7;343:d7657. doi: 10.1136/bmj.d7657
[7] Boncz I, Sebestyen A. Financial deficits in the health services of the UK and Hungary. Lancet, 2006; 368(9539):917-918.
[8] Betlehem J, Kukla A, Deutsch K, Marton-Simora J, Nagy G. The changing face of European healthcare education: The Hungarian experience. Nurse Educ Today. 2009;29(2):240-245.
[9] Ágoston I, Lampek K, Vajda R, Szőke G, Szarka E, Jankó-Király A, Csákvári T, Boncz I. Fejkvóta alapú forrásallokáció tapasztalatai az egészségügyi finansz\rozásban. Egészség-Akadémia, 2013; 4(4): 240-251.

Data and methods

The data derive from the financial database of the Hungarian National Health Insurance Fund Administration (NHIFA) covering the period 1999-2007. The Hungarian health insurance database contains the utilization data of the Hungarian publicly financed health care system with special details in international comparison. This database can be analysed by data mining techniques. We identified the average annual number of persons enrolled to Care Managing Organizations (CMO). The Hungarian CMOs was financed through a risk adjusted capitation fee and the health services covered by CMOs were defined in legal regulations.

Methods of financing healthcare services

There are three main types of general practice in Hungary: adult, child and mixed practices. Most GPs are responsible for a certain catchment area (territorial supply obligation) and only 3% of GPs have no catchment area responsibility. Hungarian GPs have a gatekeeper role towards the secondary care provided by specialists. The main source of general practitioners' revenues comes from a capitation fee based on patients' age, with a special point system for different age groups. To avoid GPs getting too many cards (patients), there is a degression limit for patient numbers (or points). The second part of the revenue of a general practice is the fixed fee, which covers the GP's costs, such as heating, cleaning, etc. The fixed fee varies according to the type of practice (adult, child or mixed) and the number of patients involved in the practice. The third part of financing is a supplementary fee, which equals the differences of practice locations (capital or big city, village, practice in more than one village or in outer areas). GPs receive reimbursement for being on duty and for cases not registered on their list (case fee).

In Hungarian ***outpatient care***, a German fee-for-service point system is used for financing. Medical procedures are listed according to the ICPM (International Classification of Procedures in Medicine) code system of the WHO (World Health Organization). Each procedure has a point value. Outpatient care departments make a report on each patient, listing the procedures carried out on the patient, and submit this report to the NHIFA. From the implementation of this adjusted fee-for-service point system in July 1993 until March 1999, there was a mixed financing system of input- (base) and output-related (activity-related or fee for service) financing. After gradual changes in the ratio of input (base or fix) and output (activity related) elements of financing, from April 1999, financing is based entirely on activity (adjusted fee for service).

To keep within the budgetary limit, until June 2000, there was a floating rate system in the fee-for-service financing of outpatient care with monthly changes in the forint value of one performance point. It meant that a prospective monthly budget was set up for the fiscal year and this monthly budget was divided by the total number of points reported by the outpatient institutions. In July 2000 this floating system was abolished and the forint value of one performance point is now announced in advance.

Acute inpatient care is financed through the implementation of a system similar to the American Diagnosis Related Groups (DRG) further referred to as HBCS. The Hungarian HBCS system has its own characteristics. The cost calculation of acute hospital care started in the late 1980s and it is based on representative studies carried out in the hospital sector. The regular application of the HBCS system started in July 1993. The disease groups reflect the epidemiological characteristics of the Hungarian population. The Hungarian HBCS system covers all the costs occurring during hospital care, including diagnosis and therapy, drugs, wages (doctors, nurses, etc.), accommodation costs (heating, cleaning, meals, etc.) and costs of hospital management except for investment costs.

Data structure of the Hungarian health insurance database

As described earlier, the Hungarian health insurance database contains nationwide health care providers claim data covering the whole publicly financed health care system. This database includes data on the following medical provisions: general practitioner, home care (nursing), out-patient care, acute inpatient care, chronic inpatient care, CT-MRI, renal dialysis, transport of patients, drug, medical devices, sick-pay and disability.

TYPE OF SERVICES	utilization (visits / admissions)	diagnosis (ICD-10)	interventions (procedures: WHO ICPM)	Outcome
general practitioner	YES	YES	NO	NO
home care (nursing)	YES	YES	YES	NO
out-patient care	YES	YES	YES	NO
acute inpatient care	YES	YES	YES	NO
chronic inpatient care	YES	YES	YES	NO
CT-MRI	YES	YES	YES	NO
renal dialysis	YES	YES	YES	YES
transport of patients	YES	YES	N.A.	NO
drug	YES	YES	N.A.	NO
medical devices	YES	YES	N.A.	NO
sick-pay	YES	YES	N.A.	NO
disability	YES	YES	N.A.	partly

Table 1.: *Data structure of the Hungarian health insurance database*

Results

In the operation framework (**Figure 1**), provider organizations applied to the OEP and had a virtual budget, an adjusted capitation account, which was determined by the size and characteristics of the population they cover. Enrolment is by the GPs and not the

individual patients, therefore there is no room for cream skimming (risk selection) at patient level. The GP enrols the population in his or her list into the Care Managing pilot programme. Patients who do not want to be enrolled into the CMO have the right to change their GP to another GP who is not involved in the managed care programme. The CMOs are self-selected through an application process, and then systematically selected by the OEP. Generally half of the applicants became Care Managing Organization.

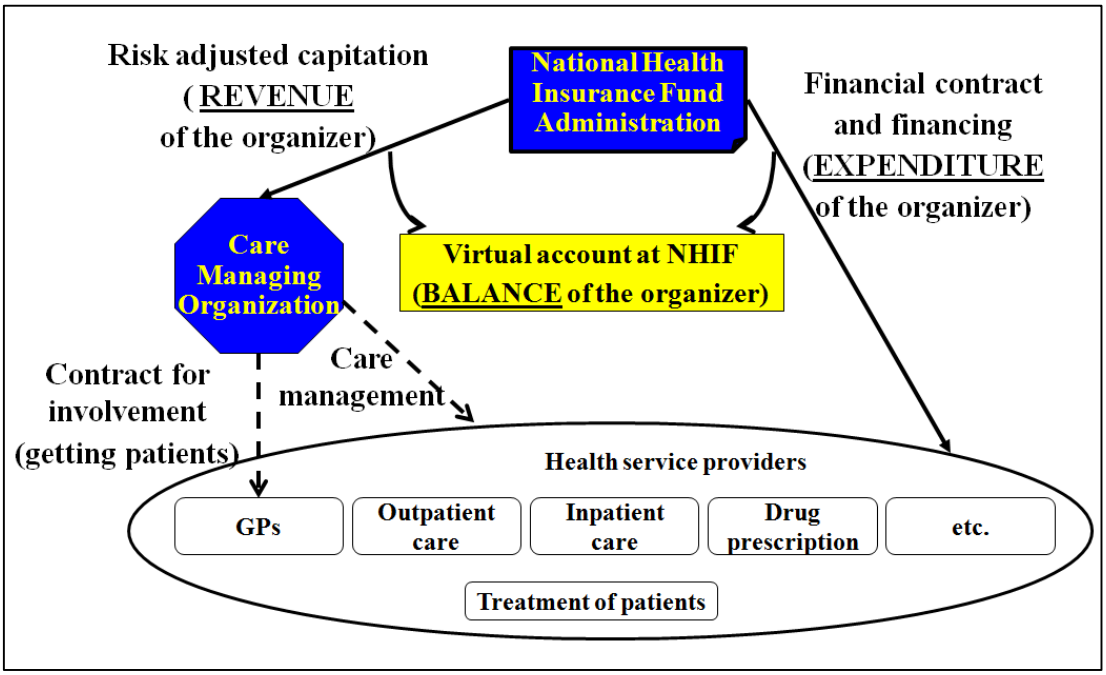


Figure 1.: *The structure of the Hungarian Care Managing Organization Pilot Programme*

In the first years of the implementation of the Hungarian care managing pilot programme, Care Managing Organizations application derived from many parts of Hungary. These CMOs usually were located in a certain geographical area, however, they did not covered neither all the general practitioners nor their patients. Later (around 2003) CMOs could have applied only from 'closed areas' which meant that all the GPs should join to the programme. (**Figure 2**)

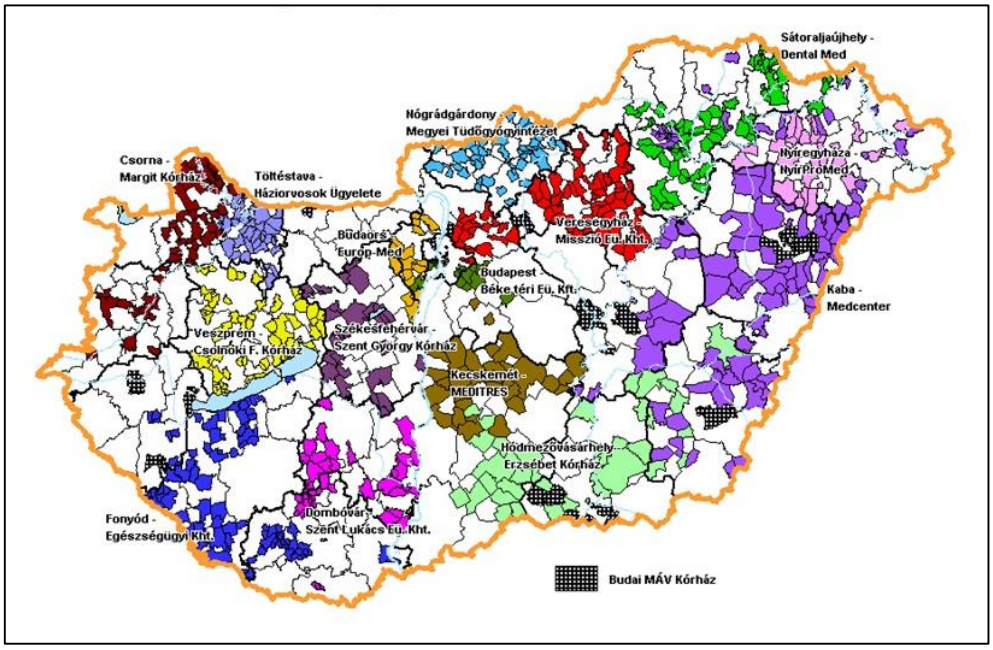


Figure 2.: *Geographical location of the Hungarian Care Managing Organizations (July 2006)*

Determination of par capita health insurance expenditures is the crucial point of defining the revenues of the care managing organizations. Among children (0-4, 5-14), the average per capita health insurance expenditure was higher for boys. Between the age 15-60, utilization of females were higher than males. Over the age 60, the average per capita health insurance expenditure was higher for women compared to men (**Figure 3**).

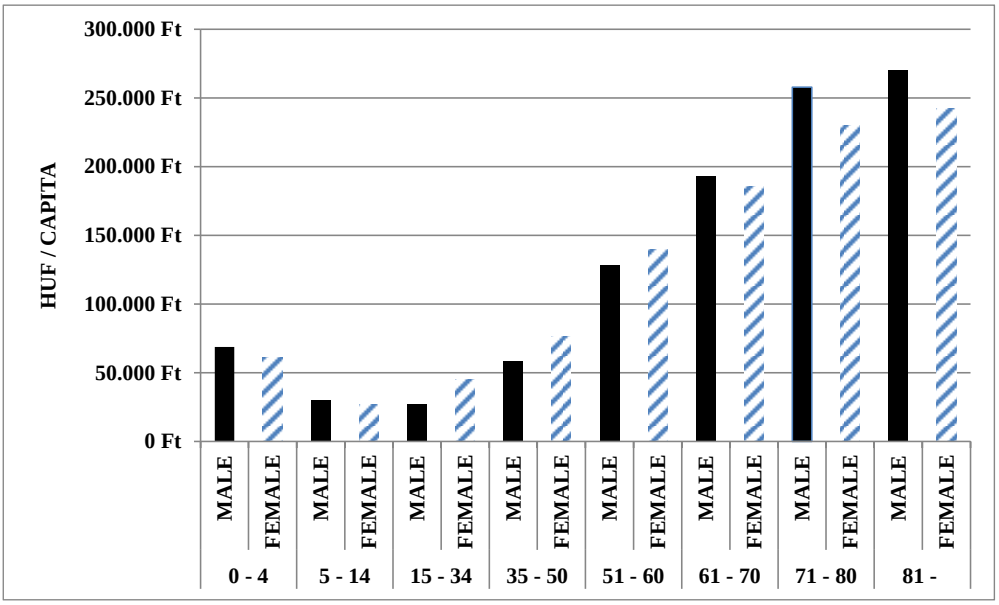


Figure 3.: *Per capita health insurance expenditures according to age and gender (2006)*

Discussion

One of the greatest challenges for the OEP was to create and operate a nationwide monitoring system to control utilization of health services. As a response to the failure to meet this challenge, the CMOs were to improve this function. The CMOs were themselves medical providers and could act locally to control utilization. Their goal was to ensure that care is delivered cost-effectively, at the right level, and doesn't use unnecessary resources. In doing so, they may intervene at many levels (out- and inpatient care, drugs and medical devices, CT/MRI examinations, etc.) and make appropriate actions to maximize efficiency. [10]

The implementation of the Hungarian Care Managing Organizations Programme could have been considered as an important step in the continuous process of health care reform. Previous elements of Hungarian health care reforms (e.g. implementation of a DRG like financing system in the inpatient care) provided a solid foundation and experience with detailed data collection, cost-sensitivity of providers and improvement of technical efficiency. The fact that the new pilot CMO programme was based on voluntary applications of organizers allows this reform to evolve from the bottom-up, and to be driven by local management incentives. [11]
There is an on-going study on the assessment of current health insurance per capita expenditures which allows us to update the experiences of health care resource allocation with capitation scheme.

References

Boncz I, Nagy J, Sebestyen A, Korosi L.: Financing of health care services in Hungary. Eur J Health Econ. 2004, 5(3): 252-258.
Gulácsi L, Dávid T, Dózsa C. Pricing and reimbursement of drugs and medical devices in Hungary. Eur J Health Econ. 2002;3(4):271-8.
Boncz I, Dozsa C, Kalo Z, Nagy L, Borcsek B, Brandtmuller A, Betlehem J, Sebestyen A, Gulacsi L. Development of health economics in Hungary between 1990-2006. Eur J Health Econ. 2006, 7(S1):4-6.
Kroneman M, Nagy J. Introducing DRG-based financing in Hungary: a study into the relationship between supply of hospital beds and use of these beds under changing institutional circumstances. Health Policy. 2001; 55:19–36.
Gaal P, Szigeti S, Csere M, Gaskins M, Panteli D. Hungary health system review.Health Syst Transit. 2011;13(5):1-266.
Gaál P, Szigeti S, Panteli D, Gaskins M, van Ginneken E. Major challenges ahead for Hungarian healthcare. BMJ. 2011 Dec 7;343:d7657. doi: 10.1136/bmj.d7657
Boncz I, Sebestyen A. Financial deficits in the health services of the UK and Hungary. Lancet, 2006; 368(9539):917-918.

[10] Ágoston I, Vajda R, Jankó-Király A, Lampek K, Boncz I. Cost savings in the Hungarian care managing programme. Value Health. 2013;16(7):A486.
[11] Ágoston I, Vajda R, Jankó-Király A, Lampek K, Boncz I. Changes in the number of enrolees in the Hungarian managed care programme. Value Health. 2013;16(3):A268.

Betlehem J, Kukla A, Deutsch K, Marton-Simora J, Nagy G. The changing face of European healthcare education: The Hungarian experience. Nurse Educ Today. 2009;29(2):240-245.
Ágoston I, Lampek K, Vajda R, Szőke G, Szarka E, Jankó-Király A, Csákvári T, Boncz I. Fejkvóta alapú forrásallokáció tapasztalatai az egészségügyi finansz\rozásban. Egészség-Akadémia, 2013; 4(4): 240-251.
Ágoston I, Vajda R, Jankó-Király A, Lampek K, Boncz I. Cost savings in the Hungarian care managing programme. Value Health. 2013;16(7):A486.
Ágoston I, Vajda R, Jankó-Király A, Lampek K, Boncz I. Changes in the number of enrolees in the Hungarian managed care programme. Value Health. 2013;16(3):A268.

Misspelled Domain Name Based Piracy in Hungary

Kruzslicz, F.
University of Pécs (Hungary)

Introduction

The definition of a domain name is cited from the regulations of the Hungarian Board of Internet Suppliers (Internet Szolgáltatók Tanácsa, ISzT 2013): „A *domain* is an administrative autonomous subset of the global computer network, managed by a databases assigning names to Internet addresses. A *domain name* is a technical identification string consisting of alphanumeric characters. Domain names helps the users easier to remember addresses of the Internet, hence this resolution is essential part of the Internet communication, replacing numeric codes with words."

Internet addresses and domain name structures are supervised by the Internet Assigned Numbers Authority (ICANN, www.icann.org) international non-profit organization from 1998. Its European regional coordinating branch is called Réseaux IP Européens (RIPE, www.ripe.net). The country code top level domain names (TLD) are managed by local registration organizations, specifically all Hungarian domain names, which end in *.hu*, and all its derivative subdomains are supervised by the Board of Internet Suppliers (ISzT, www.nic.hu) since 1997. Every subordinate institution follows its higher level recommendations, but their regulations can differ according to the local requirements and conditions. The usability question of Hungarian accented letters in domain names introduced in 2014 which is a good example of a localized divergence. In Hungary all the *.hu* top level domains and its all public subdomain (like *cegnev.hu, cegnev.co.hu*) is the property of ISZT. Any claimant receives his domain name only for usage for a fixed time period, called delegation.

At the beginning of the Internet in Hungary one could register and use just those kind of domain names which are strictly belonged and directly connected to the company or person (ie. names, trademarks etc.) The domain name liberalization happened on 1th March 2000 and the limitation mentioned above were canceled. From this date anyone can claim almost any kind of domain name under the Hungarian TLD. As a result of this newly born freedom the local domain name speculation market has evolved. Reselling of the already registered *portal.hu* domain for 5 million HUF was the first case which got a lot of publicity in the country. After the liberalization a large scale of domain name buying-up started (see fig. 1) narrowing down the number of meaningful and available names. As a consequence of this process the local domain name market run out of the usable Hungarian words by the end of 2010. The verification of proper and lawful usage of the registered domain names seemed to be hardly controllable manually.

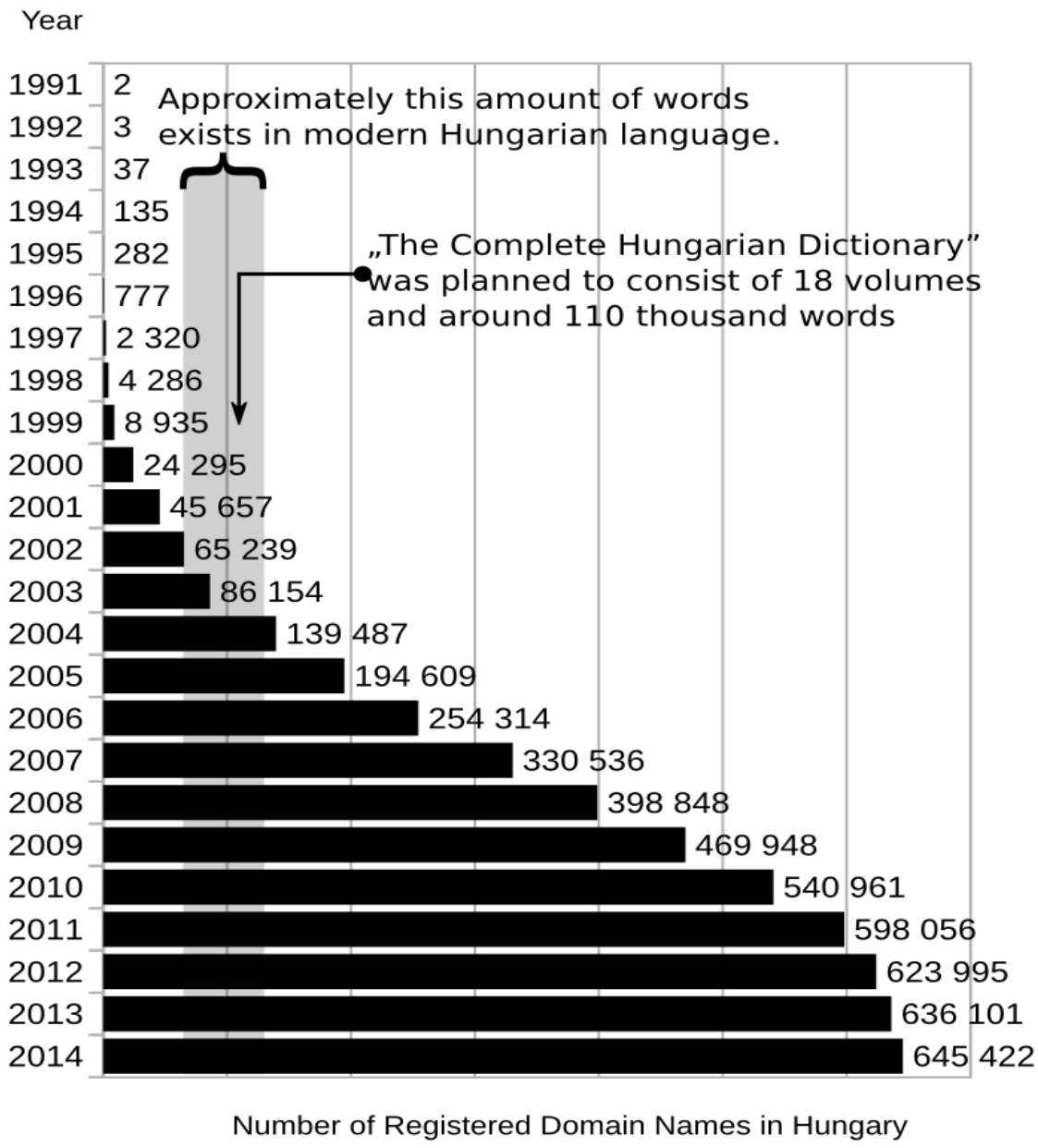


Figure 1.: Yearly number of delegated .hu public domain names
Source: http://www.nic.hu/statisztika/ (based on data available until 01. June 2014.)

Domain name abuses

The remaining free space of domain names were narrowed further by different laws and industrial regulations. In this more dense name space the judgment of improper or correct usage of a registered domain name became rather complex and difficult to decide whether a registration claim of a name similar to another one is malicious or not. For

example in the case when Google Inc. took objection to a non-priority registration of the *googol.hu* domain name by a person, the position of the Advisory Board (Tanácsadó Testület) on this question was to keep the domain name delegation for the individual in its resolution announcement of 10/2011. (V. 30.)

Following the international literature we use the traditional classification of domain name abuses, which may belong to one of the three major type isolated by FairWinds Partners, LLC (2008) and Coull et al. (2012). Those cases enrolled to the first group where itself the legality of the domain name usage is questionable, like domain name preallocation or fallacious name registration. Domain names where not the usage is problematic but the content provided under the name is not absolutely legal form the second group of abuses, like malware distribution, spam transmission, physing or displaying content with indignation or fear. Attacks against the infrastructure of the domain name resolution system go to the third group, such as domain name hitchhiking and blocking or diverting Internet addresses. Only the first type of misuses were examined in this research, which can be divided further subtypes suggested by Banerjee et al. (2010):

- *Cybersquatting* denotes all kinds of domain delegation where the allocated name is someone else's registered name or trademark and the content of the portal maliciously generates profit for the claimant (ie. *madonna.hu*).
- *Typosquatting* is that form of domain piracy where the registered name is very similar to the name of another web site, and the profit is generated from the fact that certain amount of user will mistype a word in the name (ie. *madona.com*).
- *Combosquatting* is a new phenomenon meaning combination of protected names with frequent terms, events or trends guessed from actual news, search engine keywords with high scores and other heuristics (ie. *madonnakoncert.hu*).

Considering other actors of the domain name market, not only the claimants, one can identify additional unfair behaviors. Usually domain registration offices help their clients with an online lookup interface where they can check the status of their selected domain name. Dishonest registrars collecting these domain guess searches can predict what kind of domain names will be requested in the future. By capitalizing this knowledge they are able to buy up these names in advance or sell their tips. The practice using insider domain name information is called *domain name front running*. In certain countries one can require domain names even only for a short period to try out its visibility, traffic and value. The technique when lots of potential names are tested by a claimant is called *domain name tasting*. It is worth to note, that this method is not applicable in Hungary, since there is a minimum length of delegations and every domain name request must be published for eight days to make possible put a veto on the initiated delegation.

If a delegation period is over and the current claimant do not want to renew it, the actual domain name is deallocated and becomes free again. Since these sites have previous reputation and statistics about popularity, a new claimant immediately can make profit from it. The list of abandoned domain names are published before their registration is deleted and the name is freed again finally. The delegation of these names is granted on a first come first served base, therefore claimants can lie in wait for grabbing them. When a domain name is immediately grabbed after being freed again is called *domain drop catching*, and this method is available both for registrars and claimants. Hundreds of registered domain names by a single claimant can not be easily filled with suitable contents, but it is not their primary intention either. While waiting for reselling the allocated names can be used to generate profit as well. To do this, only the name should be resolved to an existing (virtual) site which are available through *domain name parking* services, where the content of the parked domain is usually filled with automatic advertisements. Finding and parking good typo names makes money from misspellings through pay per click links embedded into the content. The detailed business model of parking services is described in a McAfee Inc. report (Keats, 2007).

Domain name dispute resolution

Complains emerged by a third party against a domain holder are examined by the Hungarian Alternative Dispute Resolution Forum (Alternatív Vitarendező Fórum, AVF), whose adjudication is accepted both the complainant and the registrant. AVF is an independent professional forum, which follows the international Uniform Domain Name Dispute Resolution Policy (UDRP). Within the framework of AVF a separate Advisory Board (Tanácsadó Testület, TT) is functioning, whose main duty is to examine the legal usage of conditionally delegated domain names, and publish policy statements and modifications on delegations. A complain submitted after the delegation of the domain is investigated by an umpire elected by the AVF committee. In order to successfully transfer a domain name to the new holder the following conditions must meet at the same time:

1. *Confusion*: that is the domain name is identical or confusingly similar to an existing trademark or service mark which the complainant has rights.
2. *Unrelatedness*: that is the holder has no rights or legitimate interest in respect of the domain name.
3. *Bad faith*: that is the domain name is registered and used in an unusual, malicious way.

Documents of TT and AVF guidances, disputes and resolutions were analyzed using text mining tools, to get a detailed overview about the working process of these organizations and to understand their decision style and efficiency. Figure 2 shows the summary of our main results proving that domain name piracy was born along with the liberalization. At the beginning most of the cases belonged to the cybersquatting group.

In year 2000 TT had to decide about 410 suspicious domain name requests, most of which was found to infringe others' rights. Typical examples of names of jurisdictions were like *hitachi.hu bbc.hu cnn.hu cartoon.hu britishtelecom.hu*, which where complained by their foreign trade holders. Next year administrative authorities were attacked by mass buying up, for example *onkormanyzatok.hu, miniszterium.hu, orszaghaz.hu, cegbirosag.hu, miniszterelnok.hu* and *miniszter.hu* names were withdrawn and similar names were nationalized and retained by the government. After a short range of decay in the number of cases the number of disputes started to increase again from 2008. The renewal of complained cases was partly due to the efficiency of the established AVF forum and its processes, but one can find new type of domain name attacks, namely typosquattings. When characters with accents had been allowed to use in domain names in 2012 the number of accent based typosquatting cases increased immediately, and AVF became the main forum of resolutions. It should be noted that last year data is not complete and contains only documents from the first term of 2014.

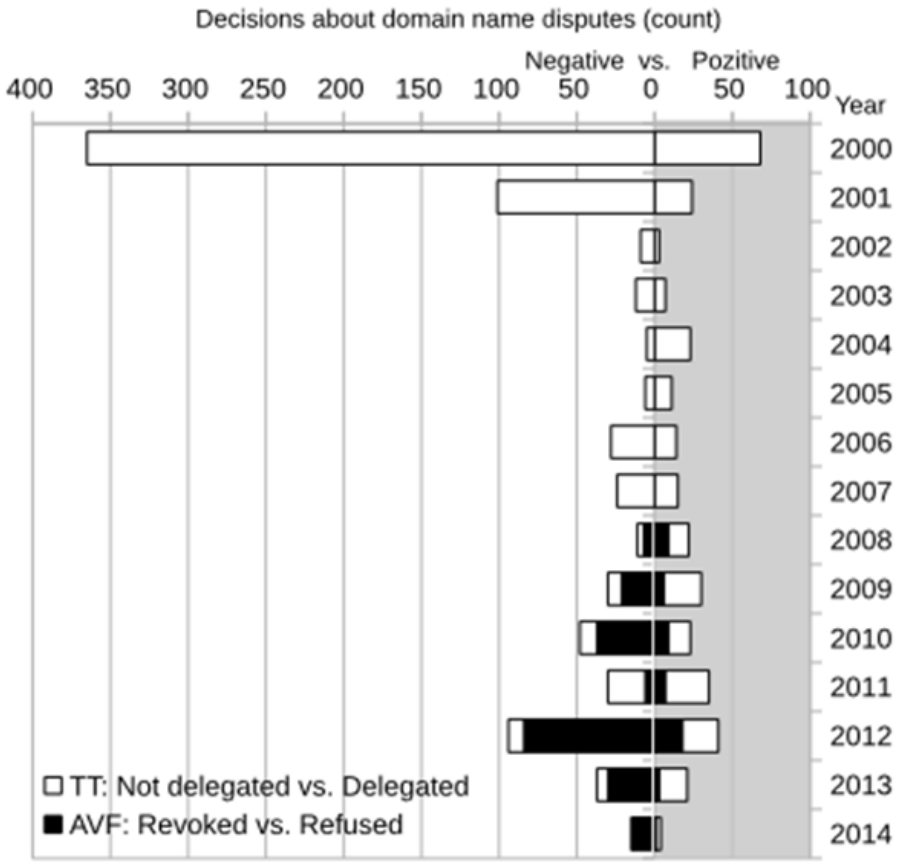


Figure 2.: Results of reported disputes by time from the viewpoint of current domain name users

Source: Public announcements of TT and AVF resolutions and decisions (until 01. June 2014)

Figure 2 exhibits not only the yearly number of domain names disputed, but their resolution as well. Methods commonly used in sentiment mining were applicable to analyze judgments of the authorities. In case of TT decision only two types of judgments were identified: either the name was delegated for the claimant, or the process was interrupted and the delegation was canceled. There were also two major types of judgments made by AVF, either the domain name was revoked from the registrar or the transfer request to the new holder was refused. Resolutions were classified into negative or positive categories for every domain name, according to the viewpoint of the current holder. Namely a decision resulting in delegation or refusing a transfer admission was taken as a positive judgment. Most of the AVF resolutions in latest years were negative and the role of TT from 2008 focused on prevention and guidance.

On figure 2. one can clearly see that the number of domain names checked by the official authorities are infinitesimally small compared to total numbers of delegated names of the same period on figure 1. Subsequently if we want to find out the current state of Hungarian domain name piracy new methods must be used to find out whether this phenomenon is an existing and important one or not.

Cases of domain name misuse

In our study the similarity between domain names were measured not only with the Damearu-Levenshtein edit distance (Navarro, 2001), but other text similarity metrics was added as well. The „fat finger" effect is one of the best known extension where the penalty score of character replacement is weighted by the physical distance of keys on keyboard or other input device. Additionally the character shape similarity (ie. letter O and number zero) and the pronounce based similar sounding were included in our method. All of these calculations were adapted for the Hungarian locale, such as the keyboard arrangement, comparable shapes with or without accents, and sound shapes close to each other. The Internet slang, replacing a sequence of characters with numbers or symbols by pronunciation (ie. *1szer1.hu*, *t-gla.hu*), was handled by a language dependent replacement table. To find a typosquatting neighborhood of a domain name, all the names in its unit distance environment was generated. Applying all the possible single letter insertions, deletions and replacements, followed by other (shape, sounding, slang etc.) types of replacements all the candidates were generated and tested against existence and content validity.

From the work of Linari et al. (2009) is known that there is a lower limit in popularity (and length) of domains which are worth to surround with squatting pages profitably. Table 1 lists the top ten most visited web sites in the country among the *.hu* TLD. All these sites has at least 300 hundred unique visitors per day. Below this popularity one can still find sites with fake pages around, but these are seasonal ones or dedicated to

special target. The general rule of thumb, that a site having at least 300 hundred daily unique visitors must have near distance squatting pages proved to be true for the Hungarian web as well according to the popularity list in January 2014. A content displayed under a domain name is taken as genuine if its content is the same as the examined site or very similar to it. A web page filled only with links pointing or redirecting to the primary web site is classified as not a fake, but a self advertisement page. There may be other pages with real content (or at least they look like) registered in the domain neighborhood by chance as well. The rest of the cases were counted as typosquatting attacks, but according to their content they can be divided into further subcategories by the level of deception. A web page with simple content filled with click per pay advertisement links was classified as a parking one. These pages usually had information about the parking services too. Less in number, but there were other type of redirections to pages of special marketing tools or forums, and after the automatic transfer we landed in real content. We found pages even with no or minimal (even erroneous) content, just to make the impression of a valid domain name content rather than a reservation. Finally there were pages exactly stating that they are for sale anyone who missed his really wanted domain.

Examined .hu domain		Type of web page content						
name	**unique daily visitor**	**own**	**self ads**	**other real**	**parking**	**other ads**	**space reservation**	**for sale**
aprod	680 000	1		3	6	1	4	2
freemail	589 357	1		4	11	3	5	
origo	579 213	1		1	3	1	6	2
startlap	497 492	3	7	7	3	1	1	
index	492 072	2		6	5		11	
blog	452 146	1		7	5		7	1
hir24	362 518	1		3	2		3	
nlcafe	336 828	2	1	4	1			
t-online	315 000	3		2			3	
idokep	300 169	3		1	2			

Table 1.: Number of fake (squatting) web pages around the top 10 Hungarian portals
Source: own collection and editing based on data at 01. June 2013.

After examining the numbers in table 1 we can state the number of its typosquatting attacks grows by the popularity of the site. In a small country we have only few sites

with high popularity. The linear connection between the popularity and the number typosquatting attacks could not be confirmed properly because of the limited number of sample, but the trend is visible. Moore és Edelman (2010) in their pioneer research found that there are 281 typosquatting pages around a very popular site in the *.com* domain on average. In Hungary this number is only approximately 9, which can be explained with smaller population and less amount of potential visitors.

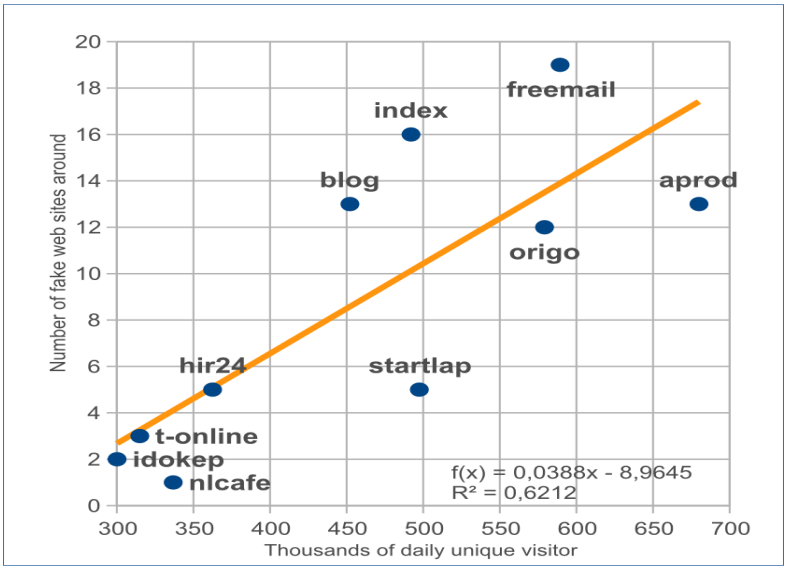


Figure 3: Typosquatting coverage of the top 10 Hungarian web sites
Source: own collection and editing based on data available until 01. June 2014.

Deceptive usage of domain names was proved to be an existing phenomenon in the Hungarian TLD as well, and it show similar occurrence conditions and rates what was found in earlier international researches on the commercial (*.com*) TLD, but at a lower scale. Among the largest Hungarian web sites a free email service was found to be surrounded with the most number of typo domains. As a result of analysis and content classification of the potential typo candidates generated by our specialized algorithm we could identify 19 malicious pages parked at different providers attacking the *freemail.hu* free email service owned by Origo Zrt.

Conclusion

In our paper we analyzed all the documents of the Hungarian Board of Internet Suppliers and its Alternative Dispute Resolution Forum authority. Using text mining approach and tools the dispute resolutions were identified and classified as the complained domain name usage was to be found of bad faith or not. Parsing the texts we could find examples of all type of domain name piracy which were published in earlier studies in Hungary as well. Algorithms to measure domain name similarity in English (specifically for the *.com* domain) were extended to the Hungarian language and

environment. Assuming that main results found on international domains should be hold for domains in Hungary as well, the most visited portals were checked against typosquatting attacks. In spite of the less population, slightly different regulations and fewer number of sites with efficiently high popularity in the country, the business model of typosquatting and its profitability has the same requirements as in the commercial TLD. As precaution some companies buy up the most probably typo versions of their domain name, but still there is a valuable niche to take of its advantage. When an attack is revealed a dispute on this case is typically resolved by revoking the domain name from the current holder, though neither the detection of typo sites or the decision support of the judgment is not automated. Finally it is worth to note, that the domain name buying-up seems to be a good defensive method, but in this way a company extends the number of possible new typos as well.

References

Banerjee, A., Rahman M.S., Faloutsos, M. (2011): SUT: Quantifying and mitigating URL typosquatting. *Computer Networks* 55. 3001-3014 pp.
Coull, S.E., White, A.M., Yen, T.F., Monrose, F.R., Michael K. (2012): Understanding domain registration abuses. *Computers and Security* 31:7. 806-815 pp.
http://www.scottcoull.com/COSE12.pdf (2014.06.01.)
FairWinds Partners, LLC (2008): The Foundation for the Online Business of Brands and of Cybersquatters. *Perspectives* 3:7. 1-27 pp.
http://www.brandchannel.com/images/papers/445_direct_navigation_final.pdf (2014.06.01.)
FairWinds Partners, LLC (2008): The Business of Cybersquatting – Typosquatting. *Perspectives* 3:8. 1-14 pp.
http://www.brandchannel.com/images/papers/452_typosquatting_final.pdf (2014.06.01.)
FairWinds Partners, LLC (2008): The Business of Cybersquatting – Combosquatting. *Perspectives* 3:9. 1-11 pp.
http://www.brandchannel.com/images/papers/461_Combosquatting_final.pdf (2014.06.01.)
Internet Szolgáltatók Tanácsa (2013): Domain Registration Regulation – operative version from 1th January 2013.
http://www.domain.hu/domain/szabalyzat/szabalyzat.pdf (2014.06.01.)
Keats, S. (2007): What's In A Name: The State of Typo-Squatting 2007. McAfee Inc. Report
http://www.siteadvisor.com/studies/typo_squatters_nov2007.html (2014.06.01.)
Linari, A., Mitchell, F.R., Duce, D.A., Morris, S. (2009): The „Curse" of Popularity. *Proceedings of the WebSci'09: Society On-Line, Athens, Greece*, March 2009. 1-5 pp.

Moore, T., Edelman, B. (2010): Measuring the Perpetrators and Funders of Typosquatting. *Proceeding FC'10 Proceedings of the 14th international conference on Financial Cryptography and Data Security* 175-191 pp.
Navarro, G. (2001): A guided tour to approximate string matching. *ACM Computing Surveys* 33:1.

Why do investors confuse companies?

Kovács, B.
University of Pécs (Hungary)

Introduction

Rashes (2001) analyzed the co-movement of stock prices of MCI Communications and Massmutual Corporate Investros during 1996-1997. The former was one of the largest telecommunications companies of the United States. Common stocks of MCI Communications were listed on the NASDAQ under the symbol MCIC. In 1996 the company started merger negotiations with other telecommunications companies. In 1997 Worldcom acquired MCI Communications for $20 billion. (Brack, Drezen, Noguchi & Schneider, 2005; Rashes, 2001) Massmutual Corporate Investors[12] filed for initial public offering on NYSE in 1971 under the ticker symbol MCI. In 1997 the net asset value of the closed-end fund was approximately $193 million. Massmutual invested in long term liabilities of privately held, small companies and was not interested in major telecommunications firms. (Babson Capital Management, 2011; MassMutual Corporate Investors, 1997; Rashes, 2001)

Rashes (2001) observed that the daily volume of MCIC and MCI showed 65.86% correlation with each other during the Worldcom-MCI Communications merger in 1996-1997. This value is higher than the correlation coefficient on the whole sample from 1994 to 1997 (55.91%). He showed that the MCI stock prices were influenced by the MCIC returns and during the merger period only the MCIC returns explained significantly the MCI returns. Rashes concluded that if two stocks have similar ticker symbols and the market of the mistaken share is illiquid then arbitrage is costly and ineffective.

Davies et al. (2007) analyzed the phenomenon on a larger sample size. They defined ticker symbol and company name similarity by several aspects. So ticker symbols are close when they differ only in the last letter. Company names are similar if they begin with the same word. They considered also similar spelling and pronunciation of ticker symbols and company names. All the similar equities in their sample belonged to different industries. They did not find any long-term systematic co-movement between similarly identified securities on their sample of 41 companies. This means that the market is efficient and it is not possible to gain excess profit from easily interchangeable equities in the long run. Davies et al. left the question open why do investors confuse some similar stock symbols while do not others. (Davies, Hillier & Thamm, 2007)

[12] The fund changed its name to Babson Capital Corporate Investors in 2011.

Research methodology

To answer the above question, it is necessary to analyze the particular cases when some confusion occurred. The next section describes the main reasons which led to investor confusion in selected cases. Cases were selected based on the credibility of the hypothesis that investors bought (or sold) a firm's shares not being conscious that they invested in the wrong company. There are many news articles presenting various cases when investors mistakenly put their money in an equity which has nothing to do with the triggering event. In some of the reported cases, unusual price movements can also be attributed to events concerning the company itself. On March 25, 2014, for example, Facebook, Inc. announced that it will acquire Oculus VR, LLC. for $2 billion. The next day, the stock prices of Oculus Innovative Sciences, Inc. peaked 15% higher than the previous closing price. The spokesman of Oculus Innovative Sciences said that their stock prices experienced enhanced volatility in the past couple of days which can be attributed to the IPO of their subsidiary, Ruthigen, Inc. (Aglionby, 2014) It is questionable whether the entire price movement is due to confused investors.

Sometimes the report on the case takes into account only the facts that support the opinion of the writer – namely, that investors misidentified a firm and therefore they bought the shares of a different company. In 1997, for example, Amazon.com, Inc. filed for IPO on the NASDAQ under the ticker symbol AMZN and after that a similarly named company Amazon Natural Treasures, Inc. experienced increase in its stock price. First, CBS.MarketWatch.com (1999) presented the case of Amazon as illustrative example to the recent VerticalNet case. Based on this source, Head, Smith and Wilson (2009) described the events as follows: "[…] stock in a company called Amazon Natural Treasures which had been using the ticker symbol AMZN before Amazon went public (and switched to AZNT afterward) rose from $1 a share to $3 and then fell back to $1 after investors realized their mistake". (Head, Smith & Wilson, 2009, p. 553)
The authors did not validate the information properly. Indeed, stocks of Amazon Natural Treasures used the symbol AMZN and were traded on the OTC market from April 1996 to March 24, 1997 (the Amazon.com IPO was accomplished later in that year). However, in its news release from March 24, 1997, Amazon.com drew the attention of investors to the potential confusion: "Please do not confuse Amazon.com, Inc., with Amazon Natural Treasures, Inc., an unaffiliated company that formerly traded on the Over-The-Counter Bulletin Board under the symbol 'AMZN.' ". (Amazon.com Inc., 1997) According to Amazon Natural Treasures' Form 10-KSB for the fiscal year ended December 31, 1998, the stock quotes AZNT did not rise much after the Amazon.com IPO on May 15, 1997. (Amazon Natural Treasures Inc., 1999) The stock quote tripled between March and May and remained at around $2.5-$3 until February, 1998. (Cavalry, 1998) The prolonged price change between the news release about the IPO and the IPO itself (and after the IPO) may or may not be attributed to investor confusion, but rather to speculation about potential confusion.

Results

Davies et al. (2007) showed that similarity of company names or ticker symbols is not a sufficient condition to confuse stocks with each other. The similarity merely of names or tickers is not either a necessary condition of investor confusion. Confusion may also occur among investors when two companies have similar identity. After General Motors, Corp. filed for bankruptcy in 2009, its name was changed to Motors Liquidation Company. The stock of the bankrupt company was traded on the OTC market under the ticker symbol GMGMQ. On July 10, 2009, as part of the GM reorganization, General Motors Company (new GM) purchased the assets of old GM. GM CEO Fritz Henderson held a press conference about the details of the reorganization on the same day. Investors not knowing the difference between Motors Liquidation and new GM (which was fully owned by the government) bought the GMGMQ shares until 2 p.m. that day, when U.S. Financial Industry Regulatory Authority (FINRA) halted trading in the equity. (ABC News, 2009; FINRA, 2009) Motors Liquidation Company and General Motors Company had nor similar names neither similar ticker symbols. Yet they had similar identities in the eyes of the investors. In other cases, similar identity means similar business activities or unclear ownership relationships between the companies. It seems that the necessary condition to confuse equities should be formulated as follows: To confuse two or more equities, they must have similar names or ticker symbols or identities.

In most cases, incorrectly identified companies were similar to each other based on more than one similarity criteria. On Friday, February 24, 2006, for example, U.S. Food and Drug Administration (FDA) seized $3 million worth of products (dietary supplements) from Hi-Tech Pharmaceuticals, a privately owned company. (FDA, 2006) The report about the seizure appeared on Monday, February 27 at Bloomberg, but the news agency assigned the ticker symbol HITK to the article. The symbol HITK belonged to Hi-Tech Pharmacal, a drug manufacturer. Both companies operated in the pharmaceutical industry and they had very similar names. HITK stock prices fell 8.25% in ten minutes on February 27 with a yearly record volume of 805,000 shares. (Morris, 2006) Another example is the case of MCI Communications, where the companies had similar ticker symbols and the name of one company contained the ticker of the other company.

In some cases, all the three similarity criteria are present simultaneously. On February 11, 1999, VerticalNet, a software company went public on the NASDAQ. Investors were expecting the IPO since the latest version of its Form S-1 was published on February 8. According to the registration statement, VerticalNet was going to trade under the ticker symbol VERT. Another software company, Vertica Software changed its ticker symbol to VERT on January 26, 1999. Vertica Software stock was trading over the counter. On February 8 and 9, the OTC shares experienced unusually high trading volumes of 238,000 and 704,800 respectively. Stock price peaked at $4,875 and $6,625 respectively in intraday trading. On February 10, Vertica Software received a new ticker symbol, VERI. In this case, both company had similar company names, same ticker symbols and similar identities as both company operated in the software industry. (CBS.MarketWatch.com, 1999; Dueck, 1999)

Davies et al. (2007) assumed that only the investors are responsible for misunderstanding, and if they evaluate the available information incorrectly, they may buy the shares of the wrong company. The authors also assumed that this phenomenon can occur over a prolonged period – as in case of MCI Communications. They did not find any long-term systematic co-movement between similarly identified securities, and their results suggests that either only particular types of events can trigger the misunderstanding or the available information may be incorrect or incomplete. Davies et al. (2007) did not consider the latter possibility, which is observable in several cases. In the case of Hi-Tech Pharmacal, for example, Bloomberg assigned the wrong ticker symbol to the news article. The error originated from the information source itself and not solely from the investors. Another example is the Oculus VR case. As shown in Table 1, Thomson Reuters keeps assigning wrong tickers to news items about Oculus VR. (Thomson Reuters, 2014) The ticker symbol OCLS belongs to Oculus Innovative Sciences and the symbol OVT belongs to Oculus VisionTech, a software company traded on the OTC market in the United States. Facebook announced the acquisition of Oculus VR on March 25, 2014, one day later, Oculus VisionTech's stock price peaked 152% higher relative to the previous day's closing price in the morning. Trading volume soared even higher, increasing by 11,677% on March 26. Trading was halted until Oculus VisionTech issued a press release wherein the company clarified that it was not associated with Facebook's acquisition of Oculus VR. (Aglionby, 2014; Alden, 2014)

Date/Time	Source	Tickers	Article title
25-June-2014			
19:01:02	RTRS	OVT.V FB.O	Oculus denies claims by ZeniMax in virtual reality tech lawsuit FB.O
22-May-2014			
21:32:04	NYTIME	OVT.V	Virtual Rift Leads to Legal Rift for Oculus and ZeniMax
26-March-2014			
10:44:01	ASNEWS	YHOO.O OCLS.O	Zuckerberg defends $2bn Oculus acquisition just month after mega WhatsApp purchase
9:16:08	LATIME	OCLS.O HAS.O	A major bet on virtual reality
25-March-2014			
22:30:06	PRN	OCLS.O FB.O	Facebook to Acquire Oculus FB.O

Table 1: Selected news items in Thomson Reuters Eikon

Besides news agencies, investment advisors, brokers, stock-picking services, etc., may also disclose incorrect or incomplete information about confusable companies.

Conclusion

This paper investigated the causes of investor confusion when the mistaken companies share some similarities with the actual companies. Rashes (2001) observed an unusual co-movement between two stocks with similar ticker symbols. Davies et al. (2007) extended Rashes' analysis and studied 41 UK equities with similar ticker symbols. They concluded that there is no significant co-movement of shares with similar names in the long run. They found valuable an investigation of the reasons why investors confuse certain similarly identified equities but do not others. To confuse two or more equities, it is necessary that they have similar names or ticker symbols or identities. I conclude that companies must meet at least two of three similarity criteria to be confused and the information sources (news agencies, investment advisors, brokers, stock-picking services, etc.) play a significant role in coordinating investor errors. If the available information about confusable companies is incorrect or incomplete, investors will misidentify the equities. It seems that only particular types of events can trigger the misunderstanding.

References

ABC News. (2009): AP: GM emerges from bankruptcy. 2009-07-10 08:38:08; http://abcnews.go.com/Business/story?id=8049767 (2014.07.31.)
Aglionby, J. (2014): What's in the name Oculus? *Financial Times*, March 26, 2014 4:41 pm; http://www.ft.com/intl/cms/s/0/3bf7b056-b4f9-11e3-9166-00144feabdc0.html#axzz37WeeBHOZ (2014.07.14.)
Alden, W. (2014): After Facebook's $2 Billion Deal, Some Virtual Unreality in the Stock Market. *DealBook*, MARCH 26, 2014 11:53 AM; http://dealbook.nytimes.com/2014/03/26/after-facebooks-2-billion-deal-some-virtual-unreality-in-the-stock-market/?_php=true&_type=blogs&_r=0 (2014.07.14.)
Amazon.com Inc. (1997): Amazon.com Files for Initial Public Offering; http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle_Print&ID=502856&highlight= (2013.08.03.)
Amazon Natural Treasures Inc. (1999): Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 (Form 10-KSB): For the fiscal year ended December 31, 1998.Securities and Exchange Commission; http://www.sec.gov/Archives/edgar/data/844055/000084405599000028/0000844055-99-000028.txt (2014.08.03.)
Babson Capital Management. (2011): MCI Quarterly Dividend Announcement and Fund Name Change; http://www.babsoncapital.com/Strategies/Trusts/MCI/News/article.aspx?ContentID=f27d9325dd654310VgnVCM200000d37106aaRCRD (2014.07.18.)

Brack, D., Drezen, R., Noguchi, Y. & Schneider, H. (2005): MCI: From Scrappy Upstart to Fallen Giant. *The Washington Post*; http://www.washingtonpost.com/wp-srv/business/daily/graphics/mci_022105.html (2014.07.18.)
Cavalry. (1998): 629. *Silicon Investor*; http://www.siliconinvestor.com/readmsgs.aspx?subjectid=15664&msgnum=639&batchsize=10&batchtype=Previous (2014.08.03.)
CBS.MarketWatch.com. (1999): Investors suffer a bit of VERT-igo 2-9-99. *MarketWatch*; http://www.marketwatch.com/story/investors-suffer-a-bit-of-vert-igo-2-9-99 (2014.08.03.)
Davies, J.R., Hillier, D. & Thamm, J. (2007): Investor confusion and similarly identified securities. *Accounting & Finance,* 47(4):693-711.
Dueck, C. (1999): Vertica Software shrs fall amid confusion over ticker symbol. *Bridge News*, Feb-10-1999; http://www.siliconinvestor.com/subject.aspx?subjectid=25752 (2014.08.02.)
FDA. (2006): FDA Requests Seizure of More Dietary Supplements Containing Ephedrine Alkaloids. February 24, 2006; http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108606.htm (2014.07.30.)
FINRA. (2009): Regulatory Notice 09-37: Trading in Motors Liquidation Company (Formerly Known as General Motors Corporation). July 10, 2009; http://www.finra.org/Industry/Regulation/Notices/2009/P119357 (2014.07.31.)
Head, A., Smith, G. & Wilson, J. (2009): Would a stock by any other ticker smell as sweet? *The Quarterly Review of Economics and Finance,* 49(2):551-561.
MassMutual Corporate Investors. (1997): *1997 Annual Report.* Springfield: Massmutual Corporate Investors; http://www.babsoncapital.com/BabsonCapital/http/bcstaticfiles/MCI/AnnualReports/files/7.pdf (2014.07.18.)
Morris, K. (2006): Amityville, N.J., drug maker suffers from confusion with similarly named firm. *Newsday*, 03/01/2006; http://search.ebscohost.com/login.aspx?direct=true&db=n5h&AN=2W61289388659&lang=hu&site=ehost-live (2014.07.30.)
Rashes, M.S. (2001): Massively Confused Investors Making Conspicuously Ignorant Choices (MCI–MCIC). *The Journal of Finance,* 56(5):1911-1927.
Thomson Reuters. (2014): *Eikon.* Desktop. [Computer program]. (2014.08.16.)

Traditional and Bayesian methods of model choice

Kehl, D. – Rappai, G.
University of Pécs (Hungary)

The analysis of household savings is a prominent area within the research of a nation's economy. In this study, we are modeling the per capita real savings of Hungarian households by investigating eight competing single-equation linear model specifications (some of which are nested). We provide a comparison between the results of the traditional (frequentist) method of model selection and the results obtained from the method of Bayesian model averaging. In order to choose between nested models, we have to understand the principles of constructing the likelihood ratio, the Lagrange multiplier and the Wald test; for non-nested models, we need to be able to follow how the information criterion is obtained. In the second part of this paper, we give an overview on the Bayesian methodology with a focus on model uncertainty, and discuss the results of model averaging. Based on our empirical analysis, the simultaneous application of the two approaches ("modeling philosophies") provides researchers with a more detailed insight regarding the problem at hand.

Savings, especially household savings, are devoted special attention within the intertwined relationships of macroeconomics. One of the most common economic objectives is the maximization of savings, be it solely driven by less theoretical political strategies or by almost entirely theory-based objectives, such as in the fundamental equation of macroeconomics where macro-level savings equal to the amount of a nation's total investments. In this paper, we discuss different approaches of modeling household savings. We provide an overview on model selection within the traditional framework of econometrics and we also discuss how the total information summarized in every possible model can be "extracted" by applying the method of Bayesian model averaging.
Providing a literature review on savings models in this study is beyond our objectives. We do not intend to assess or to provide an overview on modeling attempts regarding Hungarian savings from recent decades. For a better understanding on this subject, please refer to the manuscript by Mellár (2013) or the study by Attanasio-Weber (2010). Our model selections were based on Körösi et al. (1990) and Király (1988).
This study is structured as follows:

- First, we give an overview on the specifications of nested savings models;
- We provide a brief summary of the traditional econometric tests and the information criteria that are used for comparing competing models;
- We describe the methods of Bayesian econometrics for handling model uncertainty;

- We perform parameter estimations using the traditional and the Bayesian approach, then we select the optimal specification.

1. Savings Functions Used in our Analysis

The basic idea behind savings models discussed in this paper is that the level of household savings only depends on the household income and the previous, accumulated savings. This, from a modeling perspective, provides us with sufficient simplicity, since estimates can be made based on only three variables: total household income, cumulative previous household savings, and the current level of total savings. The time series in this analysis encompasses half a century, ranging between 1960 and 2013 during which time frame conspicuous changes have taken place regarding the size of the population and the level of the consumer price index. In order to adjust for these factors, we have been using real categories throughout our analyses, i.e. each data had been converted to per capita values calculated on the price levels of 2012. Each of these categories are accessible via official sources (KSH, MNB) and a detailed description and their interpretation can be found in a study by Rappai (2014).

Since time series data is used in our analysis, it is important to test for stationarity in order to avoid spurious regression, for which we applied the augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Scmidt-Shin (KPSS) tests. (For a detailed description of these methods, see Rappai, 2013). These tests showed ambiguous results, indicating that spurious regression is likely to be present. Therefore, instead of using the original variables, our calculations are based on their respective differences.

Model specifications applied herein are summarized in Table 1. Models are named after their originators.

Model	Specification
Zellner-model	$s_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 y_t + \beta_3 \Delta y_t + \beta_4 m_t + u_t$
Friedman-Brown-model	$s_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 y_t + \beta_3 \Delta y_t + u_t$
Klein-model	$s_t = \beta_0 + \beta_2 y_t + \beta_3 \Delta y_t + \beta_4 m_t + u_t$
Houthakker-Taylor-model	$s_t = \beta_0 + \beta_1 s_{t-1} + \beta_3 \Delta y_t + u_t$
Friend-Mack-model	$s_t = \beta_0 + \beta_2 y_t + \beta_3 \Delta y_t + u_t$
Tobin-model	$s_t = \beta_0 + \beta_2 y_t + \beta_4 m_t + u_t$
Keynes-model	$s_t = \beta_0 + \beta_2 y_t + u_t$

Table 1: Models considered

Notations used in Table 1 follow the standard customs, where s_t is the change in real per capita savings, y_t is the change in real per capita income and m_t is the standard notation for the change in cumulative real per capita savings at time period t. Random variables are denoted by u_t and Δ represents difference. According to custom notations, $\beta_0, \beta_1, \ldots, \beta_4$ are the model parameters to be estimated (obviously, each parameter with the same notation may take up different values depending on the applied model). In our notations, we assigned the same index to the parameter of each explanatory variable in order to better display the structure how models and nested. For a diagram of this structure, see Figure 1.

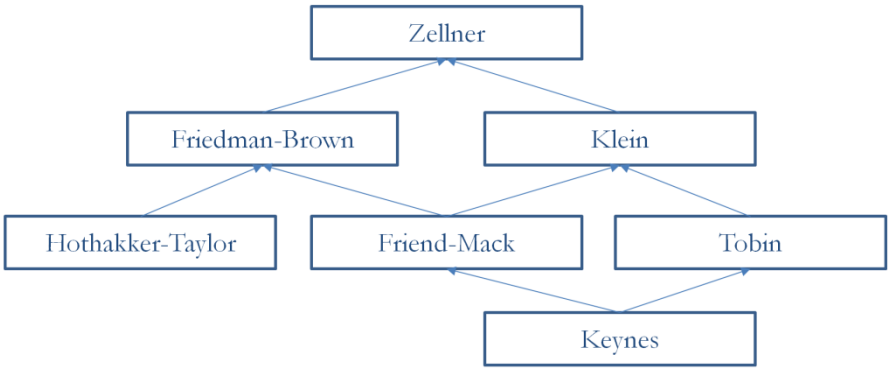


Figure 1: The nested models

In the following section of this study, we discuss the method of choosing between competing models using the traditional (frequentist) econometric approach; then, from the Bayesian viewpoint of modeling, we provide a brief analysis on how to grasp the uncertainty within model specifications.

2. Traditional (Frequentist) Econometric Methods of Model Selection

When choosing between models, an essential idea is to find one that is realistic enough yet not overly complicated. Aiming for a best possible fit oftentimes means the inclusion of many variables, whereas the aim for simplicity suggest the inclusion of few variables. During the process of model selection, the first step is to decide whether the competing logical constructions are

- nested, or
- non-nested.

In case of nested models, a larger model can be reduced to a smaller one by equating one or more of its coefficient to zero, thus reducing the number of variables. Testing becomes somewhat more complicated when non-nested models are compared, since the examination of parameter restrictions' relevance is insufficient, and additional considerations need to be made. Since Figure 1 includes both nested and non-nested models (e.g. the Keynesian model is nested within the Tobin-model, best not nested in the HT specification), we are giving an overview of the traditional principles for choosing between both nested and non-nested models in the following sections.

2.1 Econometric Tests: Aiding Choice between Nested Models

Test functions aiding our choice between nested models can be constructed following three principles:

- The principle of Likelihood Ratio (LR);
- The principle of the Lagrange-multiplier (LM);
- The Wald (W) principle.

From the well-known definition in inferential statistics, let

$$L_0 = L(s_t, y_t, m_t, \boldsymbol{\theta}_0) = f(s_t | s_{t-1}, y_t, \Delta y_t, m_t, \boldsymbol{\theta}_0)$$
$$L_1 = L(s_t, y_t, m_t, \boldsymbol{\theta}_1) = f(s_t | s_{t-1}, y_t, \Delta y_t, m_t, \boldsymbol{\theta}_1) \quad (1)$$

be the likelihood function for when the null hypothesis or the alternative hypothesis is fulfilled, where

- s_t, y_t, m_t are the empirical values of the variables included in the analysis, and
- $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ are the parameter vectors estimated with the maximum likelihood method for the null and the alternative hypothesis.

The idea behind the *LR tests* is that a likelihood ratio can be calculated from the likelihood functions which can be written as

$$\Lambda = \frac{L_{H_0}}{L_{H_1}} \quad (2)$$

when two models are compared. The principle is apparent: if the null hypothesis is correct, then the likelihood of the constrained model is only slightly less than the likelihood of the unconstrained model, i.e. the ratio is near 1. In this case, using the bigger model is unnecessary. Unfortunately, the distribution of Λ cannot be determined in many cases, and an asymptotic LR test needs to be performed. It can be shown that the test function

$$LR = -2 \log \Lambda \quad (3)$$

follows a chi-square distribution (see Hunyadi, 2001), where the degrees of freedom equals to the number of conditions.

The core idea of the second approach, the *Lagrange-multiplier (LM) priciple*, is to maximize the log-likelihood function subject to the null hypothesis (constrained model), and to investigate whether removing the condition would cause a significant increase in the value of the log-likelihood. If there is a significant increase, then the null hypothesis should be rejected, and if not, then the smaller (constrained) model is considered satisfactory. In other words, the LM principle translates into solving the following optimization problem:

$$\log L_0 + \lambda(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1) \rightarrow \max \quad (4)$$

where λ is the Lagrange multiplier. Since the first derivative has to be zero, i.e.

$$\frac{\partial \log L_0}{\partial \boldsymbol{\theta}_0} + \lambda = 0 \quad (5)$$

The same problem can be answered by deciding whether the value of the Lagrange multiplier is significantly greater than zero. In order to obtain the test statistic, the Fisher information metric needs to be determined.

$$I\left(\boldsymbol{\theta}_0\right) = Var\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}_0}\right) = -E\left(\frac{\partial^2 \log L}{\partial \boldsymbol{\theta}_0^2}\right) \tag{6}$$

Using this and the derivative of the log-likelihood, the test statistic can be determined:

$$LM = \frac{\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}_0}\right)^2}{I\left(\boldsymbol{\theta}_0\right)} \tag{7}$$

which also follows a chi-square distribution with its degrees of freedom being equal to the number of conditions.

The third principle, known as the *Wald (W) principle,* follows a slightly different logic and is directly based upon testing whether or not the constraints exist. The core idea is that in certain cases, restrictions can be applied so that the parameter vector of the extended model ($\boldsymbol{\theta}_1$) becomes identical with the parameter vector of the smaller model ($\boldsymbol{\theta}_0$). (In many practical cases, some of the parameters within the vector equal to zero. These parameter restrictions can be used to populate an appropriate $\mathbf{R}$ matrix).

Therefore, the following hypotheses need to be tested for the extended (larger) model:

$$\begin{aligned} H_0 &: \mathbf{R}\boldsymbol{\theta}_1 = \mathbf{r} \\ H_1 &: \mathbf{R}\boldsymbol{\theta}_1 \neq \mathbf{r} \end{aligned} \tag{8}$$

Rewriting the same pair of hypotheses, we can examine whether

$$\left(\mathbf{R}\hat{\boldsymbol{\theta}}_1 - \mathbf{r}\right) - \left(\mathbf{R}\boldsymbol{\theta}_1 - \mathbf{r}\right) = \mathbf{R}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\right) = 0 \tag{9}$$

holds true. If it does, it means that for the estimated parameter the inclusion of the related constraint is not significant, therefore, the constrained (smaller) model is to be chosen. If the relationship does not hold true, the unconstrained (larger) model is held relevant. The generic test

$$W = \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\right)^{\mathbf{T}} \mathbf{R}^{\mathbf{T}} \left[\mathbf{R}\mathbf{I}^{-1}\left(\boldsymbol{\theta}_1\right)\mathbf{R}^{\mathbf{T}}\right]^{-1} \mathbf{R}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\right) \tag{10}$$

derived from the relationships explained above follows a chi-square distribution where the degrees of freedom equals to the number of constraints. Having only one constraint allows major simplifications: in this case, we only have to examine whether the additional parameter of the estimated larger model is equal to zero.

2.2 Model Selection Based on Information Criteria

The model selection methods described in the previous section are based on the assumption that one model can be converted into another with appropriate parameter restrictions. However, it is possible that one is forced to choose between models where such conversion is not feasible (i.e. non-nested models).

Generally, given two linear models, it is usually the one with more variables that has a higher coefficient of determination. Using its adjusted alternative is recommended in

many cases (see Mundruczó, 1981). For linear regression models including a constant parameter fit to time series data, its value can be calculated using:

$$\tilde{R}^2 = 1 - \left(1 - R^2\right)\frac{T-1}{T-k-1} \tag{11}$$

where $\tilde{R}^2$ is the adjusted coefficient of determination, R^2 is the "unadjusted," deterministic coefficient calculated from the sum of squares, T is the length of the time series and k is the number of variables included on the right hand side of the regression model. Model selection, from this point, is straightforward: the model specification with the higher adjusted coefficient of determination is more favorable.

Despite its popularity, the latter method is not unconditionally accepted. One objection is that in certain extreme cases it is possible for this measurement to take up negative values, which, considering its meaning, appears contradictory. Others criticize the ad hoc nature of the applied correction and miss the relationship with probability theory. As a result of these criticisms, the use of information criteria has become a widely accepted means to select between non-nested models. Model selection based on the information criteria, as we will see later on, does not exploit the "internestedness" or "non-nestedness" of the models, which means that this procedure can be applied as a model selection tool between any two kinds of models.

In the past, several methods were proposed to help researchers select the most parsimonious model. The three most commonly used information criteria are discussed below.

- Information criterion proposed by Akaike (1974)

$$AIC = \frac{2\log L}{T} + \frac{2k}{T} \tag{12}$$

- The Bayesian information criterion, as formulated by Schwarz (1978)

$$SBIC = \frac{2\log L}{T} + \frac{k}{T}\log T \tag{13}$$

- The *HQIC* indicator, as proposed by Hannan-Quinn (1979)

$$HQIC = \frac{2\log L}{T} + \frac{2k}{T}\log\left(\log T\right) \tag{14}$$

Each of the three indicators suggests that the model with the smallest value of information criterion is the most parsimonious. When choosing between model specifications, note that the applied penalty for each criterion is different therefore, only information criteria of the same kind should be compared.

The strategy of traditional (frequentist) econometrics from this point on, again, is simple and straightforward: we select the model that best passes the aforementioned tests and has minimal information criteria; then, we base our analyses and predictions on the selected model.

3. Model Uncertainty and the Bayesian Approach

A fundamental idea of the Bayesian approach is that every information related to a parameter is included in the posterior distribution. A formal representation of this relationship, using the notations and framework of Koop (2003), is:

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} \tag{15}$$

where $p(\theta|y)$ is the posterior, $p(y|\theta)$ is the likelihood, $p(\theta)$ is the prior, and $p(y)$ is the marginal likelihood (these notations and the terminology are valid in case model uncertainty is not considered). The prior incorporates parameter-related information not related to the sample, and the likelihood carries information from within the sample. Therefore, the posterior as a sum of these, incorporates the total information. The role of marginal likelihood, which in case of a given sample is a scalar, is to ensure that the posterior is a proper density function, meaning that the area below the curve equals to one. Since this scalar has no influence on the information that can be extracted from the posterior, it is often considered irrelevant; however, because of the subject of this study, its value plays a considerable role in our analysis.
Equation (15) indirectly implies that the model is given, therefore, we assume that model uncertainty is not present. If the number of possible models (R) is greater than one, (15) has to be rewritten in a way that this fact remains explicit, i.e. for the rth model (M_r, $r = 1,\ldots,R$):

$$p(\theta_r|y, M_r) = \frac{p(y|\theta_r, M_r)\,p(\theta_r|M_r)}{p(y|M_r)} \tag{16}$$

from which it is clearly noticeable that each model can have a different posterior, likelihood and prior; also, the parameter vector can be different.

Handling model uncertainty is a heavily researched field within Bayesian literature, and is not free from scientific debates. Two fundamental approaches exist, hence we use the phrase "handling model uncertainty" instead of the less complicated but somewhat less descriptive expression of "model selection." These two fundamental approaches can be summarized as follows:

- Bayesian model selection is a traditional interpretation of selection, i.e. given certain criteria the best model is chosen and used for analytical purposes. Note that this approach is strongly related to the traditional (frequentist) mindset.
- According to Bayesian logic, if (parameter) uncertainty is present, this uncertainty can be described with a random variable, and with the inclusion of our data the posterior will contain every available information related to the uncertain values. Once prior probabilities are assigned to each competing model, the Bayes' theorem is to be applied. The most common terminology used to refer to this approach is Bayesian Model Averaging (BMA), since it takes all

information from each competing model into consideration, and not only the information included in the most likely specification (Hoeting et al., 1999).

3.1 Bayesian Model Selection

According to the Bayesian logic, a probability statement is to be made by using Bayes' theorem referring to an unknown phenomenon (which model is valid), given the information that we already know (empirical data). Formally,

$$p\left(M_r \middle| y\right) = \frac{p\left(y \middle| M_r\right) p\left(M_r\right)}{p\left(y\right)} \tag{17}$$

where $p\left(M_r \middle| y\right)$ is the posterior model probability, $p\left(y \middle| M_r\right)$ is the marginal likelihood described earlier, and $p\left(M_r\right)$ is the prior model probability. The numerator of the expression on the right hand side is often hard to determine, but it is also irrelevant from the perspective of model selection, since in the numerator neither models nor parameters are included.

The comparison of models *i* and *j* can be performed by calculating the posterior odds ratio, a fraction of (17) for both models, i.e.

$$PO_{ij} = \frac{p\left(y \middle| M_i\right) p\left(M_i\right)}{p\left(y \middle| M_j\right) p\left(M_j\right)} \tag{18}$$

where the second factor is the prior odds ratio. From this, each posterior model probability can be calculated assuming that all models are analyzed, i.e. the sum of posterior model probabilities equals to one.

Often, we are interested in model selection when prior model probabilities are equal; in other words, when the prior odds ratio equals to one. In this case, the expression can be simplified such that only the respective marginal likelihoods remain on the right hand side. This fraction is most commonly referred to as the Bayes factor.

$$BF_{ij} = \frac{p\left(y \middle| M_i\right)}{p\left(y \middle| M_j\right)} \tag{19}$$

Multiple values of the Bayes factor are referred to as rule of thumb thresholds. One of the most commonly cited taxonomies is shown in Table 2.

Evidence against $\mathbf{M}_j$	$2\log BF_{ij}$
negative	<0
weak positive	0-6
strong	6-10
very strong	>10

Table 2: Interpretation of Bayes factor values
Source: Kass-Raftery (1995).

Besides the Bayes factor, other ways exist to support and aid model selection. Gelman et al. (2004) criticize this methodology for applying probabilities that are based on the assumption that a given models is valid. Therefore, the authors suggest using methods that rely on the distances between the data and the model. The logic behind their measurement is similar to the information criterion, and is called Deviance Information Criterion (DIC). Another path is described by Kruschke (2014), suggesting the model selection process to be considered as a hierarchical model.

3.2 Bayesian Model Averaging

The method of Bayesian Model Averaging (BMA) applies the same core idea as the technique of model selection aided by the Bayes factor, i.e. the posterior model probabilities defined in (17) will be used. Let's assume that we are interested in determining ϕ vector with each of its elements having the same interpretation for all models. In a regression model, they can be parameters of focal interest, and every known information about ϕ can be summarized in the related posterior, according to BMA and the Bayesian logic. Based on the total probability theorem, we can state that

$$p(\phi|y) = \sum_{i=1}^{R} p(\phi|y, M_i)\, p(M_i|y) \tag{20}$$

If g is any arbitrary (mostly identity) function and the expected value of $g(\phi)$ is of interest, the following expression can be formulated:

$$E[g(\phi)|y] = \sum_{i=1}^{R} E[g(\phi)|y, M_i]\, p(M_i|y) \tag{21}$$

From these, it follows that the results we intend to analyze (posterior distribution, expected value) need to be determined and their averages have to be calculated in a way that the weights used for averaging equal to the respective posterior model probabilities.

As we have seen, the theoretical framework of Bayesian analyses is simple and appealing; however, its practical applicability may be problematic. Once all posterior model probabilities are determined, the process is straightforward, but the number of possible models (R) can be very large, which causes problems from a practical point of view. When the expressions within (20) and (21) cannot be calculated using analytical methods and simulations are needed, the number of models we are able to handle becomes very limited. Fortunately, many types of models, including linear regression models allow the use of analytical methods given that proper priors are chosen. However, the large number of possible models in many cases makes the evaluation of all models nearly impossible. If k is the number of all potential explanatory variables, then each model can either include or exclude each variable, i.e. the number of possible linear models is

$$R = 2^k \tag{22}$$

If $k = 30$ and we have the ability and the resources to evaluate 100 models per second, the total time required to assess every model would amount to 4 months. In many

practical cases the number of explanatory variables is even greater, therefore a number of algorithms have been developed in order to cope with this problem. One of the most commonly applied algorithms is the Markov Chain Monte Carlo Model Composition (often referred to as MC^3, Madigan-York, 1995).
Before we apply the BMA method and discuss the results we obtained in the sections below, we provide an overview of the linear regression model and formulas that are needed to determine important relationships such as (20).

3.3 Applying the BMA approach for Linear Regression Models

For the most commonly used technique applied in linear regression BMA, the normal linear regression framework is assumed to hold true (Fernandez et al., 2001). We have N observations, and R is the number of linear regression models ($M_r, r = 1,\ldots,R$) that only differ in their explanatory variables. Generally, we assume that each model includes a constant, i.e. each model can be written as:

$$y = \alpha\iota_N + X_r\beta_r + \varepsilon \tag{23}$$

where ι_N is an $N \times 1$ vector consisting of 1s, X_r is a matrix consisting of a subset of the explanatory variables with the dimensions of $N \times k_r$ (k_r is the number of explanatory variables in the r[th] model), and vector ε is assumed to consist of elements with normal distribution. Koop (2003) avoids the direct use of variance, and applies the term precision (by definition, $h = \frac{1}{\sigma^2}$) for this expression makes derivations more manageable, since $\varepsilon \sim \mathrm{N}\left(0_N, h^{-1}I_N\right)$.

Choosing the prior is a vital step in all Bayesian analyses. This statement is especially true for model averaging. The conjugate normal-gamma prior is a standard choice since the calculations require large computational power. By doing so, we can use analytical tools to derive probabilities related to posterior parameter moments and to posterior models. When posterior odds ratios are applied for model comparison, it is important to remember that non-informative priors can only be used for parameters present in all models. For all other parameters, informative, proper priors are to be used. In order for the intercept to carry the same interpretation in every model, Fernandez et al. (2001) recommend that the average of each explanatory variable be subtracted from the respective variable itself. This does not affect the β values but the interpretation of each alpha intercept becomes identical and equal to the average of the outcome variable. From these, it follows that the intercept and the precision are identical in all models, therefore α and ε in (23) do not carry the lower index r. Because of the identical interpretation the use of standard non-informative priors is commonly accepted:

$$p(\alpha) \propto 1 \quad p(h) \propto \frac{1}{h} \tag{24}$$

where $\propto$ denotes proportionality; e.g. the prior for α is a horizontal line. In the next step, the exact form of the normal-gamma prior for the β parameters needs to be

determined. Besides having the expected value equal to zero, the g-prior (Zellner, 1986) is a widely applied choice:

$$\beta_r | h \sim \mathrm{N}\left(0_{k_r}, h^{-1}\left[g_r X_r^T X_r\right]^{-1}\right) \tag{25}$$

where the only step left is to choose the appropriate g_r for which many recommendations have been made. The logarithm of the Bayes factor resulting from different choices approximates various information criteria. For more on this subject, see Fernandez et al. (2001).

The posterior, as explained earlier, can be determined using analytical methods, and it can be shown that the posterior for β_r follows multivariate t-distribution. Applying the g-prior, the marginal likelihood for the r[th] model can be written as:

$$p\left(y|M_r\right) \propto \left(\frac{g_r}{g_r+1}\right)^{\frac{k_r}{2}} \left[\frac{1}{g_r+1} y^T P_{X_r} y + \frac{g_r}{g_r+1}\left(y-\bar{y}\iota_N\right)^T\left(y-\bar{y}\iota_N\right)\right]^{-\frac{\nu}{2}} \tag{26}$$

from which posterior model probabilities can be obtained by using the equality in (17). (For details, see Kehl 2014.)

4. Empirical Results and the Comparison of Methodologies

Following the theoretical overview, we applied the OLS method to estimate the parameters of the models in Table 1. The most important results are listed in Table 3.

Parameter	*Zellner*	*F-B*	*Klein*	*H-T*	*F-M*	*Tobin*	*Keynes*
$\hat{\beta}_0$ $\left(s_{\beta_0}\right)$	-5.191 (5.24)	-5.569 (5.35)	-6.115 (5.27)	-1.124 (4.94)	-6.360 (5.36)	-6.354 (5.01)	-7.001 (5.09)
$\hat{\beta}_1$ $\left(s_{\beta_1}\right)$	0.220 (0.15)	0.193 (0.15)	-	0.260 *(0.15)	-	-	-
$\hat{\beta}_2$ $\left(s_{\beta_2}\right)$	0.279 **(0.11)	0.199 *(0.11)	0.309 ***(0.11)	-	0.232 **(0.10)	0.329 ***(0.09)	0.267 ***(0.09)
$\hat{\beta}_3$ $\left(s_{\beta_3}\right)$	0.105 (0.11)	0.131 (0.11)	0.033 (0.10)	0.248 **(0.10)	0.065 (0.10)	-	-
$\hat{\beta}_4$ $\left(s_{\beta_4}\right)$	-0.062 *(0.03)	-	-0.057 (0.04)	-	-	-0.059 **(0.03)	-
$\log L$	-254.420	-256.125	-255.659	-259.996	-257.035	-260.150	-261.720
$\tilde{R}^2$	*0.179*	0.142	0.157	0.097	0.129	0.172	0.139
AIC	9.978	10.005	9.987	10.038	10.001	*9.930*	9.952
SBIC	10.165	10.155	10.137	10.151	10.114	10.042	*10.026*

HQIC	10.050	10.062	10.044	10.081	10.045	*9.973*	9.980
DW	1.903	1.802	1.692	1.828	1.628	1.691	1.637

Table 3: Parameter estimates of the models

Besides providing the standard errors in the table, we are also referring to the results of the partial t-tests: * indicates that the value of a parameter is different from zero at 10% level of significance, ** indicates a 5%, and *** indicates a 1% level of significance. We have also provided the test statistics for the Durbin-Watson test. Maximum values of adjusted coefficients of determination and minimum values of information criteria have been italicized.

The results of our estimations give heterogeneous impressions. Although the coefficients of determination are moderate, note that the calculations are based on high-volatility difference variables and not on the original data, which negatively affects the goodness of model fit. We performed model selection among the nested models using the *LR* and *W* tests. For the results, see Table 4.

Larger model (est. parameters)	*Smaller model*	*Restriction*	*LR*	*p-value*	*W*	*p-value*
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*F-B*	$\beta_4 = 0$	1.481	0.224	3.185	0.081
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*Klein*	$\beta_1 = 0$	1.076	0.300	2.293	0.130
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*H-T*	$\beta_2 = \beta_4 = 0$	4.843	0.089	6.929	0.031
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*F-M*	$\beta_1 = \beta_4 = 0$	2.271	0.321	4.972	0.083
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*Tobin*	$\beta_1 = \beta_3 = 0$	4.977	0.083	2.406	0.300
Zellner $\beta_1,\beta_2,\beta_3,\beta_4$	*Keynes*	$\beta_1 = \beta_3 = \beta_4 = 0$	6.341	0.096	5.425	0.143
Friedman-Brown β_1,β_2,β_3	*H-T*	$\beta_2 = 0$	3.362	0.067	3.581	0.058
Friedman-Brown β_1,β_2,β_3	*F-M*	$\beta_1 = 0$	0.079	0.374	1.710	0.191
Friedman-Brown β_1,β_2,β_3	*Keynes*	$\beta_1 = \beta_3 = 0$	4.860	0.088	2.143	0.342
Klein β_2,β_3,β_4	*F-M*	$\beta_4 = 0$	1.195	0.274	2.609	0.106
Klein β_2,β_3,β_4	*Tobin*	$\beta_3 = 0$	3.901	0.048	0.110	0.741
Klein β_2,β_3,β_4	*Keynes*	$\beta_3 = \beta_4 = 0$	5.265	0.072	3.050	0.210
Friend-Mack β_2,β_3	*Keynes*	$\beta_3 = 0$	4.069	0.043	0.426	0.513

Tobin β_2, β_4	*Keynes*	$\beta_4 = 0$	1.364	0.243	3.052	0.080

Table 4: Model choice with LR and Wald tests

The test results, in our case, are less than unequivocal. As a general statement, the smaller models are not significantly worse than the larger ones, with the Houthakker-Taylor model being one of the only exceptions, which, compared to both the Zellner and the Friedman-Brown model, resulted in a relatively low p value. Altogether, the Tobin model appears to be the most acceptable one, supported by two information criteria (by the adjusted R^2; SBIC ranks this model to the second place). Additionally, hypothesis testing for nested models prefers smaller models to larger ones. Another evidence supporting the choice of the Tobin model is that both regression coefficients significantly differ from zero and have appropriate signs.

Now, let us examine the results obtained by using the Bayesian model averaging method. The BMA methodology, as we explained earlier, demands high computational capacity. From the various packages available, we chose the R framework (R Core Team, 2014) and the BMS software package (Feldkircher-Zeugner, 2009) that was specifically designed for linear model averaging, and supports a wide range of parameter and model priors, as well as several MCMC methods. The package, besides performing the necessary iterations, also helps users generate figures and interpret the results.

First, let us give an overview of the results obtained by applying the uniform model prior and the benchmark g-prior (suggested by Fernandez et al., 2001). The posterior probabilities for each model are depicted in Figure 2, where the horizontal axis represents the cumulative model probabilities. Alongside the vertical axis, explanatory variables are sorted by their respective posterior inclusion probability (PIP) with their size on the axis being proportionate to this probability. For each variable, PIP is the sum of all posterior model probabilities in which models the given variable is included. Parameters with a positive sign are in grey, the ones with a negative sign are depicted in black, and white represents that the given variable is excluded from the model at hand. In every model, the parameters of income, lagged savings and income change have a positive sign, and the parameter of capital has a negative sign (given that the specific parameters are included in the model).

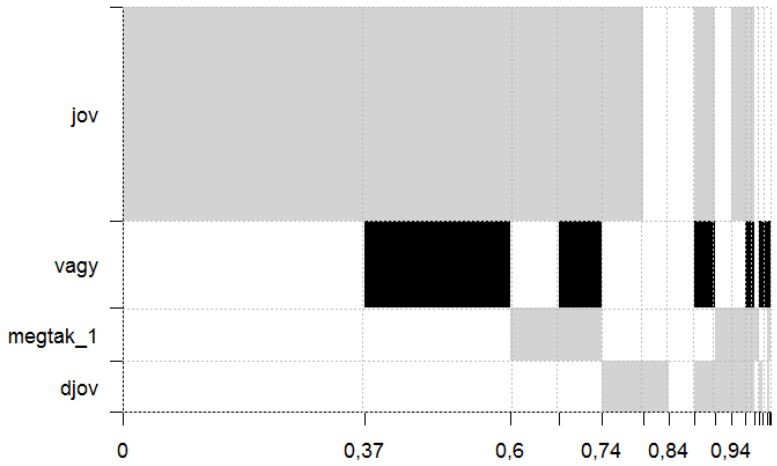


Figure 2: Cumulated posterior model probabilities

In the model with the greatest posterior probability (0.3732) only the income variable is included, whereas the model with the second-largest posterior probability (0.2247) includes both income and the value of assets. No explanatory variables are included in the model with the seventh-largest posterior probability, and the total probability of all remaining models to its right amounts to only 0.12. In essence, the variable of income is included in all relevant models.

Besides the PIP values, Table 5 includes posterior averages and standard deviations, in descending order with respect to PIP.

Variable	*PIP*	*unconditional*		*conditional*		*cond. pos.*
		exp. value	st. dev.	várh.ért	szórás	
y_t	0.8712	0.2447	0.1316	0.2808	0.0987	1.00
m_t	0.3562	-0.0203	0.0345	-0.0571	0.0352	0.00
s_{t-1}	0.2155	0.0348	0.0944	0.1614	0.1446	1.00
Δy_t	0.2076	0.0249	0.0738	0.1199	0.1218	1.00

Table 5: PIP and posterior moments of parameters

Using the BMS package, posterior expected values and standard deviations can be calculated in two different ways:

- As unconditional posterior expected values (this being the usual approach in BMA), and
- As posterior expected values given that the variable at hand is included in the model.

The difference between the two approaches is that the first option assigns zero parameters to a given variable for the models it is excluded from. The second method (conditional posterior expected value) only averages coefficients where the models include the given explanatory variable. The relationship between the parameters calculated by using the two different approaches can be written as:

$$\beta_j^{uncond} = PIP_j \beta_j^{cond} \quad (27)$$

meaning that the expected value of the unconditional parameter equals to the product of the conditional expected value and its respective PIP.

The last column in Table 6 (cond. pos.) shows the probability that conditional expected values have positive signs. If a variable is included in a model, then its sign remains identical regardless of what other variables are included, therefore, all probabilities are either 0 or 1.

Figure 3 depicts the posterior marginal distributions for all four parameters. Besides plotting the posterior distributions, we displayed the conditional expected values and used dashed lines to indicate two standard deviations' distance. Each of the plots include the expected values for each individual model, which represents how model uncertainty affects the parameters in question.

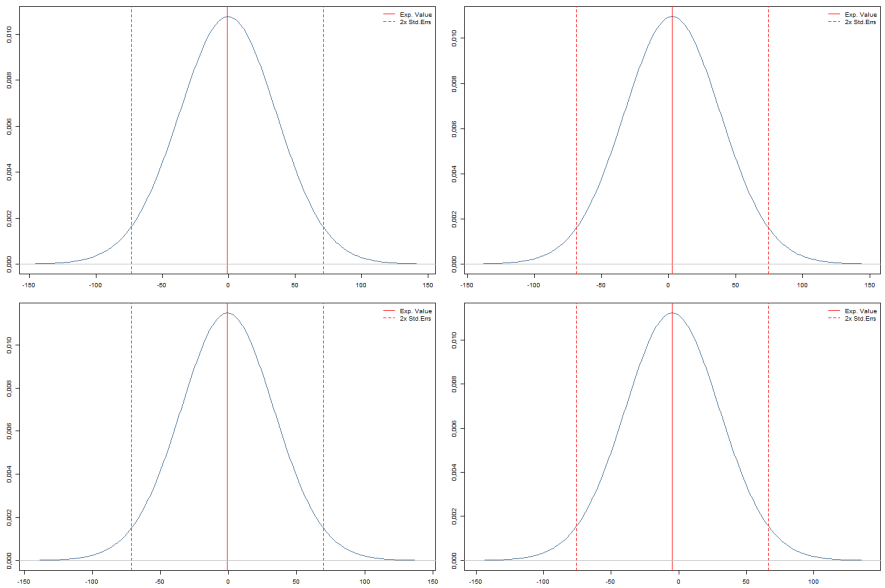


Figure 3: Posterior marginal parameter distributions

The BMS function provides us with the opportunity to choose an appropriate g_r. The results we have shown before had been calculated using the benchmark prior where $g_r = \max\left(N, K^2\right)$, which, in our case gives the same results as the default $g_r = N$ setting. By selecting $g_r = K^2$ the most likely model remains the one where income is considered as the only explanatory variable; however, the model where both income and the value of assets are included is assigned an almost identical probability. The posterior expected values of the parameters would remain similar regardless of our choice of g_r.

5. Conclusions

In this study, we have been focusing on the comparison of the traditional econometric methods (LR, LM, and Wald tests for selecting between nested models) and the method of Bayesian model averaging aiding appropriate choices between competing model

specifications. Despite our initial hopes that relatively simple model specifications would result in a straightforward and a unique and optimal choice, the different criteria and tests left us somewhat uncertain whether the selected model is "the" best choice to be used for forecasting purposes. In order to overcome this specification related uncertainty, the Bayesian model averaging is an effective method with at least two benefits: it provides us with the posterior inclusion probability of each potential explanatory variable, and it generates predictive distributions for the estimated values of each regression coefficient by taking every possible model specification into consideration.

A key conclusion of this study is that both the frequentist and the Bayesian approaches resulted in indicating that the per capita real income and the cumulative per capita real savings are the most likely explanatory variables. Additionally, the related regression coefficients (or as per the Bayesian approach, the conditional expected values of the parameters) had similar values and identical signs. This supports the observation that the Bayesian approach that puts more emphasis on model uncertainty results in little change regarding the parameter values, which is an important argument for its application.

Another important conclusion is that once every underlying aspect is considered and a decision is made on which model to be selected and used for prediction, the traditional methods of model selection provides a theoretically sound logical construction, which is not solely suited to forecast per capita real savings, but it can also be applied in simulations aiming to determine the value(s) of the explanatory variable(s) necessary to reach a certain level of savings, which may also be a part of an objective function.

One of the main arguments that this study aimed to underline is that the traditional and the Bayesian econometric methodologies are not necessarily competing concepts, but they are in a symbiotic relationship complementing one another, providing researchers with an enriched toolkit.

References

Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* Vol. 19. No. 6., 716–723 pp.
Attanasio, O. P., Weber, G. (2010) Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy. *Journal of Economic Literature*, Vol. 48 No. 3, 693-751.pp.
Feldkircher, M., Zeugner S. (2009): Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging. IMF Working Paper, 09-202
Fernandez, C., Ley, E., Steel, M. (2001): Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, Vol. 100. No. 2., 381-427 pp.
Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004): Bayesian Data Analysis. Chapman & Hall/CRC

Hannan, E. J., Quinn, B. G. (1979): The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society B*, Vol. 41. No. 1., 190–195 pp.
Hoeting, J. A., Madigan, D., Raftery, A. E., Volinsky, C. T. (1999): Bayesian Model Averaging: A Tutorial. Statistical Science. 14, 4, 382-417.
Hunyadi László (2001): *Statisztikai következtetéselmélet közgazdászoknak*. KSH, Budapest, 484 old.
Kass, R. E., Raftery, A. E. (1995): Bayes Factors. *Journal of the American Statistical Association*, 90, 430, 773-795.
Kehl Dániel (2014): *Modell-bizonytalanság a bayes-i ökonometriában*. (In Rappai-Schepp (szerk): *A válságtól a jóllétig – A múlt tanulságai, a jelen kihívásai és a jövő útjai*) PTE KTK, Pécs, 131-146. old.
Király Júlia (1988): *Aggregate Household Saving: Empirical Evidence for Hungary*. OT-IMI 2181/1/XX/1988.
Koop, G. (2003): Bayesian Econometrics. John Wiley & Sons
Kőrösi Gábor, Mátyás László, Székely István (1990): *Gyakorlati ökonometria*. Közgazdasági és Jogi Könyvkiadó, Budapest, 481 old.
Kruschke, J. K. (2014): Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan. 2nd Ed. Academic Press / Elsevier
Madigan, D., York, J. (1995): Bayesian Graphical Models for Discrete Data. International Statistical Review. 63, 2, 215-232.
Mellár Tamás (2013): *Szemben az árral*. PTE KTK, kézirat, 262 old.
Mundruczó György (1981): *Alkalmazott regressziószámítás*. Akadémiai Kiadó, Budapest, 259 old.
R Core Team (2014): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rappai Gábor (2013): *Bevezető pénzügyi ökonometria*. Pearson Ed., Harlow 149 old.
Rappai Gábor (2014): *Modellválasztás az ökonometria hagyományos eszközeivel a magyar megtakarítási függvények példáján*. (In Rappai-Schepp (szerk): *A válságtól a jóllétig – A múlt tanulságai, a jelen kihívásai és a jövő útjai*) PTE KTK, Pécs, 113-130. old.
Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, Vol. 6. No. 2., 461–464. pp.
Zellner, A. (1986): On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. Goel, P. K., Zellner, A. (eds.) Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti. Amsterdam. North-Holland

SROP-422c "Well-being in information society", no. III/2: "Non-classical logics and geolocation" results

Jenei, S. – Brachmann, F. – Rébay, V.
University of Pécs (Hungary)

The topic of location-based services has been a hot topic in the world's leading ICT publications for about 10 years. The vast majority of articles focus on practical matters with some questions in theory being discussed. The research area has somewhat been revitalized by the new progress shown in network theory: the number of articles dealing with independent networks and location-based social networks have been increasing in the last couple of years. These include positioning technologies, implementing issues, conceptual, architectural and performance aspects and IT security, privacy issues. This boom created a situation that enabled academia to discuss the technological aspects of location-based social networking. From several communication forms widely used today telephony still has an enormous share of all communication activities worldwide. With landline coverage close to the maximum and mobile coverage reaching ever-higher penetration and ever-lower cost, the importance of this form of communication is constantly increasing. These advances will enable users to connect to the network with a high-speed connection in any services area in an affordable way and to constantly use geolocation services. Sensor-based geolocation is a challenge even in the times of GPS, as most of the time is spent indoors where no GPS coverage is available. The location of the devices and users that constantly communicate is essential. As GPS is highly energy inefficient and requires complex hardware with indoor coverage non-existent, indoor positioning with inexpensive hardware capable of operating in indoor environments is a highly interesting topic of research. As sensor networks become ever more widespread, indoor geolocation systems become feasible.

Our goal was to develop results in the area of the theoretical background indoor positioning with algorithm development and investigations in the research domain. The communication devices can transmit information that can be used for the task of positioning. The using the signal's strength, intensity and latency can be beneficial to indoor localization. Other inexpensive sensor readings can also be available (like car parking radar). But possible reflections and shadowing can hamper localization.

Our goal was dual: first, a sensor measurement app was developed to enable the team to gather sensor readings from all popular mobile operating systems to allow the researchers to gather data for further investigation. Secondly, a model and process was developed and implemented to enable sensor readings to be aggregated, to increase precision. In our demo, we present the algorithm's simple to use, efficient form.

The theoretical base for the results were made possible by advances in non-classical logics. This is one of the most dynamically developing field of logic with several unsolved problems. Our theoretical results can be applied on a wide spectrum of fields from control-problems to game theory from code theory to data mining or web search. Because of these traits, our equal goal was to develop results in this field as well.

We have obtained strong theoretical results on the structure of certain residuated lattices, thus gaining new knowledge on some related non-classical logics. Using the theory of non-classical logics, we have developed a successful theoretical model for the indoor localization problem. We also developed software, which is capable to collect and organize all sensor data, too. This has supported our further research on making indoor localization more precise.

Low impact development practices in urban stormwater management

Czigány, Sz.- Flury, M. – Hegedüs, P. - Hinman, C. - Bembenek, R.
University of Pécs (Hungary) Washington State University (United States)

Introduction

Due to the intensification of the hydrologic cycle, triggered by the global climate change and expansion of impermeable surfaces due to urbanization, many developed areas suffer by the catastrophic consequences of extreme weather phenomena (supercells and convective storms), while simultaneously intense stormwater runoff and flash floods cause substantial economic losses, pollution of water bodies and in cases loss of human lives. Several studies have recently investigated the potential changes in the intensity of the hydrologic cycle and rainfall regime due to climate change, specifically in terms of the occurrence of extreme storm events (e.g.: Mueller and Pfister, 2011; McMillan et al., 2011). The changing frequency of torrential rainfalls not only alters stream flow characteristics, but may also carries nutrients, silts and hydrocarbons, chlorinated organics and heavy metals from surfaces of buildings directly into watercourses and other water bodies (Hinman, 2005). The second reason for increased runoff and higher frequency of urban floods today is that the Earth's natural land cover and land use have been dramatically changed (Warburton et al., 2012). Up to 95% of the ground surface in cities is now sealed due to urban development and this is ground space through which rainwater cannot be lost or stored by infiltration. This leads to increased runoff, up to 75% of rainwater may become run-off in urban areas. To prevent localized flooding, built-up areas need to be drained of excess rain water or flood retention basins need be built. In most cases, conventional drainage systems are not built to cope with such instant changes in flow rate and volume, leading to the flooding of the drainage system or areas further downstream. Lately, sustainable stormwater management (low impact development, LID) techniques have been gaining popularity, however their practical application and implementation is still limited. Nonetheless, LID facilities commonly function as tools for both runoff and contaminant collection. LID facilities include raingardens, bioretention cells and permeable pavements with primary focus on stormwater management and contaminant adsorption (Hinman, 2005). The ratio of runoff to infiltration is profoundly influenced by antecedent soil moisture contents, soil texture and hydraulic conductivity, thus awareness of soil hydraulic and water retention properties is essential when sustainable stormwater mitigating facilities are designed. Urban runoff may contain large amounts of contaminants (Pb, Zn, Cu, pesticides, PAHs and alkylphenols) that is washed off by stormwater runoff and usually originates from human activities (Bougon et al., 2011). As bioretention media need to fulfill the two aforementioned requirements, namely (a) having high infiltration capacity, and (b) high pollutant adsorption capacity simultaneously, its composition and the proportion of the individual components need to be designed carefully. To achieve high infiltration

capacity, soils of sandy texture are commonly used. However, sandy media has a low specific surface area and adsorption capacity to filter pollutants. To decrease pollutant transport to water bodies, appropriate amendments (clay minerals, compost and biochar), with relatively high adsorption and water retention capacities, may serve as ideal candidates to retain and accumulate pollutants and water in the vadose zone and to prevent the excessive contamination of groundwater and other water bodies. However, the physical, hydraulic and chemical properties of composts may vary largely according to their origin and the proportion and ratio of the individual constituents.

The objective of the current study was to test the water holding and water retention capacities of five soil-compost/biochar mixes developed for stormwater mitigation and surface runoff reduction. To achieve this goal silica sand was mixed with cedar grove compost, water treat residuals, bark, biosolids, sawdust and biochar at various ratios.

Materials and Methods

Twenty mesocosm units, with 5 different treatments (sand and organic matter composition) were set up at the Washington State University Research and Extension Center, in Puyallup, Washington state (United States) using five different sand-compost composition types (Figure 1). Each individual mesocosm has a diameter of 1.4 m and filled with the various media at a depth of about 69 to 83 cm (Figure 2). Underneath the soil mixtures a gravel layer is found at a thickness of 20 cm (Figure 2). The bottom of each mesocosm is drained by having a drainage pipe inserted at the bottom of the gravel bed providing free drainage at the bottom. The mineral constituent-compost composition of the 5 different treatments are shown in Table 1.

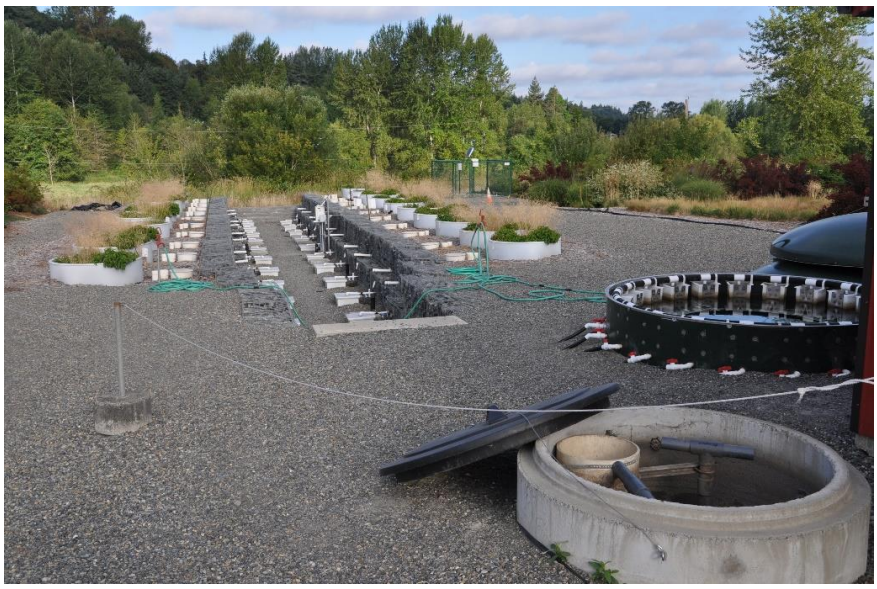


Figure 1. The mesocosm site at the Washington Stormwater Research Facility in Puyallup, WA

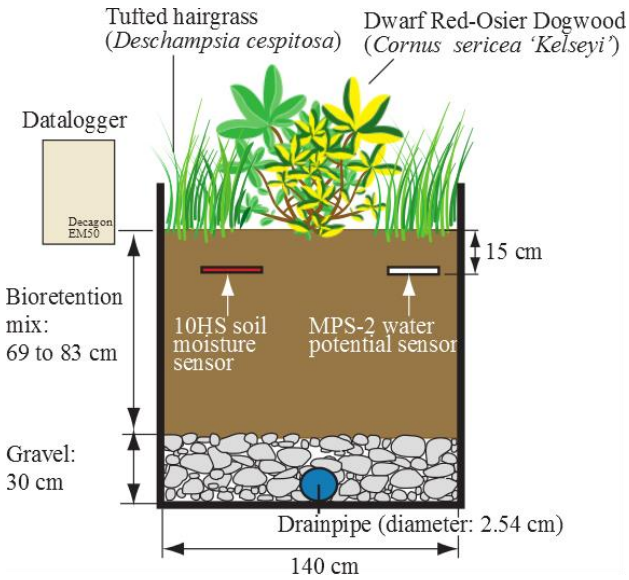


Figure 2. Schematic of the bioretention cell design

	Walrath Sand	*Cedar Grove*	*Water Treatment residuals*	*Bark*	*Sawdust*	*Biosolids*
Treatment name	*%*	*%*	*%*	*%*	*%*	*%*
C20	80	20				
C40	60	40				
C30W	60	30	10			
C40	60	15	10	15		
BS	60		10	15	10	5

Table 1. Constituents and the volumetric proportion of the five soil-compost treatment types at the beginning of the experiment in 2011

On March 21 2014, two Decagon MPS-2 water potential (measurement range: -9 to -100,000 kPa, pF 1.96 to pF 6.01) and two Decagon 10HS TDR-type large volume soil moisture sensors (both sensor types are manufactured by Decagon Devices Inc., Pullman, WA, United States) were inserted into Mesocosms 11 and 12. Similarly, two MPS-2 and two 10HS sensors were inserted into mesocosms 13 and 14 three days later, i.e. on March 24, 2014.

Mesocosms 11, 12, 13 and 14 correspond with the C20, C15WB, C40 and C30W treatment types, respectively. In all cases sensors were inserted horizontally at a depth of 15 cm. After leaving the sensors equilibriating for a period of seven (Mesocosms 13 and 14) and ten (mesocosms 11 and 12) days, data collection started on April 1st, 2014 with Decagon EM50 dataloggers in 10 minute collection time intervals. Data collection ceased on July 29, 2014. The 10HS soil moisture sensors were laboratory calibrated using about 5 kg soil

samples from each of the Mesocosm 11 to 14, and water was added to the samples in 0.05 kg kg^{-1} increments. On June 16th, 2014 two 10HS soil moisture and two MPS-6 water potential sensors were inserted into each of the Mesocosms 21, 24, 32, 35 and 44 at depths of 15 and 30 cm (data are not shown in the current study).
Water retention measurements were carried out on all 20 Mesocosm soil samples, thus 4 repetitions were taken for each treatment type. Water content values were then averaged and standard deviations were calculated for each samples. Measured data and the inverse of the standard deviation values (weights) were then used as input data in RETC modeling environment (University of California, Riverside, CA, United States) to determine the Van Genuchten parameters of a and n by fitting equation 1 (van Genuchten equation):

$$\frac{\theta_i-\theta_r}{\theta_s-\theta_r} = \frac{1}{(1-(\alpha h)^n)^m} \quad [1]$$

, where θ_i is the actual volumetric soil water content, θ_s and θ_r are the saturated and residual water contents, respectively; h is the corresponding tension, while a, n and m are fitting parameters. Water holding capacity between -33 and -1,500 kPa (i.e. plant available water) was then calculated with the following, rearranged form of the van Genuchten equation:

$$\theta_i = \left(\frac{\theta_s-\theta_r}{(1-(\alpha h)^n)^m}\right) + \theta_r \quad [2]$$

For the model simulation experiments the van Genuchten retention model was combined with the Mualem conductivity model. For the conductivity model the Mini Disk infiltrometer results and the K_s values were used as input data.

Unsaturated hydraulic conductivity *K(h)* was calculated from the cumulative infiltration (I) data measured in each mesocosm with a portable tension mini disk infiltrometer (Decagon Inc., Pullman, WA, United States) of 2.25 cm diameter. Measurements were taken in the field at a tension of h = -4 cm. Unsaturated hydraulic conductivity was calculated using equation 3, proposed by Zhang (1997):

$$I = C_1(\mathrm{h})\mathrm{t}^{1/2} + C_2(\mathrm{h})\mathrm{t} \quad [3]$$

, where C_1 (m s^{-1}) and C_1 (m $s^{-1/2}$) are fitting parameters. C_1 is related to hydraulic conductivity and C_2 is the soil sorptivity. Unsaturated hydraulic conductivity ($K(h)$) was then calculated with the following equation:

$$K(h) = C_1 \cdot A \quad [4]$$

, where C_1 is the slope of the curve of the cumulative infiltration vs. the square root of time and A is a value relating the van Genuchten parameters for a given soil type to the suction rate and the radius of the infiltrometer disk. From Table 2 of the Decagon mini disk infiltrometer's manual A = 0.8926 and a = 0.145, while n/h_0 = 2.68 for sandy soils at a tension of h = -4 cm. Calculated *K(h)* values were then averaged for all five treatment types (Decagon Mini Disk Infiltrometer manual).

Results and Conclusions

Our results indicated that bioretention mixes are typical for sandy media, but the different amendments, although added in different proportions, had little effect on physical and hydraulic properties. The largest change occurred in the second year (2012) after installation of the lysimeters (lyimeters were installed in 2011). As the mineral fractions are likely not changing their particle size, the overall change of the saturated hydraulic conductivity must be due to decomposition of organic amendments in the bioretention mixes.

Also, the time evolution of the saturated hydraulic conductivity values indicate an increase over the period of 2011 to 2014 for the C20, C40, C30W and C15WB mixes.

Acknowledgement

This research was funded by the “TÁMOP 4.2.1.B-10/2/KONV-2010-0002” Scholarship (Developing Competitiveness of Universities in the South Transdanubian Region) and the KEOP-2.2.2/B/2F/09-2009-0001 project

References

Bougon, N., L. Auterives C., Aquilina, P. Marmonier, J. Derrider, and P. Vandenkoornhuyse. 2011. Nitrate and sulphate dynamics in peat subjected to different hydrological conditions: batch experiments and field comparison. J. Hydrol. 411, 12-24

Dane, J. H., and G. C. Topp. 2002. Methods of Soil Analysis, Part 4, Physical Methods. Soil Science Society of America, Madison, WI.

Hinman, C. 2005. Low impact development. Technical guidance manual for Puget Sound. Washington State University, Puyallup, Washington.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and H. Woods. 2011. Increasing occurrence of high intensity rainstorm events relevant for the generation of soil erosion in a temperate lowland region in Central Europe. J. Hydrol. 411, 266-278

Muelller, E. N. and Pfister, A. 2011. Increasing occurrence of high-intensity rainstorm events relevant for the generation of soil erosion in a temperate lowland region in Central Europe, J. Hydrol. 411, 266-278

Warburton, M., R. Schulze, and G. Jewitt. 2012. Hydrological impacts of land use change in three diverse South African catchment. J. Hydrol. 414, 118-135

Zhang, R. 1997. Determination of soil sorptivity and hydraulic conductivity from the disk infiltrometer Soil Sci Soc. Am. J 61, 1024-1030

Web resources:

Decagon Mini Disk Infiltrometer manual
http://www.decagon.com/education/minidisk-infiltrometer-manual/ last accessed: November 30, 2014

The South Transdanubian Hydrologic Information System (STHIS)

Horváth, G.- Pál, I. – Czigány, Sz.
South Transdanubian Hydrologic Directorate, University of Pécs, (Hungary),

Introduction

Multiple water management techniques modifies natural runoff at the management area of the South Transdanubian Water Directorate (in Hungarian: Dél-Dunántúli Vízügyi Igazgatóság, hereafter abbreviated in the text as DDVIZIG). Real-time information on the water budget of the management area of 10,000 km² is available for the water management professionals. A large number of water courses and water facilities are present at the operational area of the DDVIZIG. Furthermore, facilities and gages at the operational area of the water directorate require a constant, year-round maintenance operations. However, data management and storage hitherto was hardcopy based and lacked structured storage and spatiotemporal tagging and information and processing of any maintenance operations has not yet been documented in a digital database. Similarly, archive, paper-based management documentations have not been digitized yet. To overcome these management shortcomings, flaws and challenges, a systematized and clearly-structured Hydrologic Information System (Hereafter STHIS, stands for South Transdanubian Hydrologic Information System) has been developed at the DDVIZIG to provide real time information on any water and hydrology related issue within the operational area (South Transdanubian region of Hungary that covers the SW part of Hungary). The STHIS is able to automatically provide relevant and quasi-real time hydrologic data (including water level, flow or hydraulic head) for water professionals, decision makers, stakeholders, homeland security officials and local administrative leaders. The program also archives all data in a searchable format for future reference (Figure 1).

Figure 1. The opening page of the website of the STHIS (source: http://ddvir.ddvizig.hu:8000/ddvir/flex/ddvir.html#)

Major modules and elements of the STHIS:

1. *Maintenance module:* Future planned and completed data and spatial information on any maintenance and construction operation carried out on the water courses of the operational area.
2. *Live Cam module:* In select cross sections of the rivers Dráva and Danube, including the public bridges in Barcs, Drávaszabolcs and Mohács publicly accessible live images are obtained with web cameras. This way local stakeholders can access direct information on water conditions, including ice cover, suspended sediment load, potential contamination and percentage of ice cover and contamination compared to the total surface area.
3. *Water management module:* Preset pages continuously shown in real time for the professionals of DDVIZIG. Data are directly available at the headquarters via LCD screens. Such data include water level and flow values, web cam images, flood data and relevant maps.
4. *Archive data module:* This module contains images, photos, video films and other audio-visual data, maps and miscellaneous hydrology-related documents.

In the following chapters the water management module will be discussed in detail.

Water management module

The current paper provides a detailed overview on the water management module of the DDVIZIG. Thee water management module provides a complex dataset on water balance and budget over the entire operational area of the water directorate. The following units and submodules are included in the water management module:

1.1 Water course delineation

The raster based 1: 10,000 basemap of the operational area has been updated and actualized, and then digitized providing the base map for the stream network of the operational area (Figure 2). Missing and erroneous information were also corrected in these maps.

Figure 2. Stream network of the operational area of the DDVIZIG

1.2. Delineation of watersheds, drainage areas, and subcatchments

Delineation was carried based on the 10-meter resolution digital elevation model (DEM) of the operational area. By using topographic data, this module is capable of the delineation of any watershed area upstream form any outlet point in every 100 m cross section or the drainage network. Watershed delineation, like in other similar softwares, has a higher accuracy in areas of high reliefs compared to the floodplain areas. Water balance data is demonstrated on the example of the watershed of the Fekete-víz, SW Hungary (Figure 3).

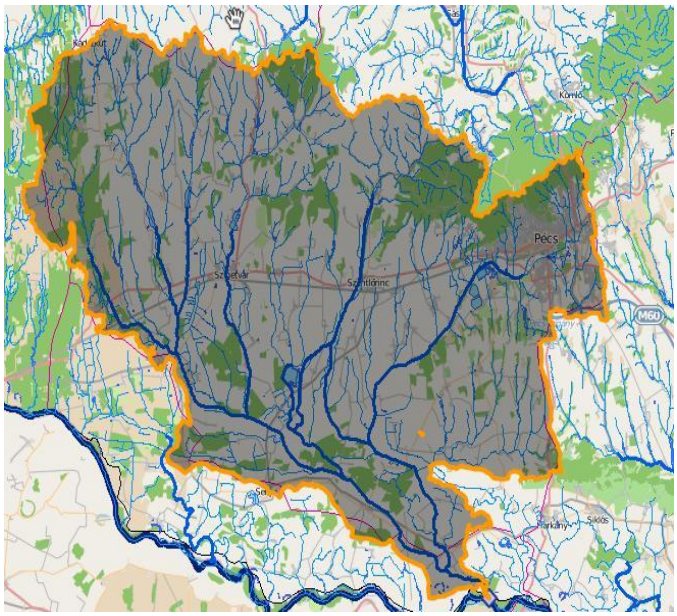


Figure 3. Drainage area of the Fekete-víz

1.3. Estimation of water balance and water demand

Runoff and flow from small catchments can be characterized with the following three parameters

- Mean discharge for the period between 1991 to 2000
- Mean daily flow during August with an 80% probability
- Ecological low water level.

Determination and calculation of the mean flow

Mean flow values for small streams are calculated from the 30-year specific runoff data, which is further corrected by the climate factor value (z).

$$KÖq_{i\,[1991\text{-}2000]} = z \cdot KÖq_{i\,[1971\text{-}2000]}$$

where:

$KÖq_{i\,[1971\text{-}2000]}$ where *KÖqi* is the specific runoff for cell i for the period of 1971 to 2000 [l/s/km²];

Where *z* is calculated as follows:

$$z = 1 - \frac{C_i^{[1991-2000]} - C_i^{[1971-2000]}}{C_i^{[1991-2000]}} \cdot 3.5$$

where:

$C_{i\,[1971\text{-}2000]}$ is the mean annual precipitation for the period of 1971 to 2000 in cell *i* (mm y^{-1})

$C_{i\,[1991\text{-}2000]}$ is the mean annual precipitation for the period of 1991 to 2000 in cell *i* (mm y^{-1});

Calculation of the annual low flow value

The $Q_{aug80\%}$ as a characteristic water balance quantifier parameter has been used in the Hungarian water balance calculations for more than 35 years. For watersheds of high relief the following formula was used:

$$Q_{aug80\%} = KÖQ * f * m$$

where:

f proportionality factor based on the types of the water bodies: $f = (A*a+b)$;

A is the land area of the watershed or subwatershed, [km²]

a, b linear fitting parameters between the area of the watershed and the proportionality factor *f*

m coefficient for stream order (based on the number of forks and confluences)

Ecological low water flow

Stream gage name	Gage ID	Monitoring period	Area (km²)	Stream km	EOV x	EOV y
Büdösgáti vf. Szólád	150154	1967-2009	48,32	5,54	159143	557989
Koppány, Törökkoppány	856	1956-2009	260	36,2	139986	574037
Kis-Koppány, Ádánd	150153	1956-2009	262,49	3,472	168200	582800
Surján, Szentbalázs	857	1956-2009	109,5	4,6	109738	560724
Baranya, Csikóstöttös	841	1935-2009	461	3,2	110701	580919
Vízfő Spring, Orfű	948	1970-2009	15,2	18,3	88654	581582
Vögységi p., Magyaregregy	150148	1969-2009	28	45,1	100100	592900
Hodácsi p., Magyaregregy	150014	1969-2009	2,4	0,13	99201	592929
Csele p., Mohács	150149	1952-2009	94,65	0,1	76307	621693
Karasica, Szederkény	838	1949-2009	204	56,416	72452	603902
Karasica, Villány	837	1949-2009	647	42,157	58914	604980
Dráva, Barcs	835	1960-2009	33977	154,1	68688	525528
Dráva, Drávaszabolcs	836	1960-2009	35764	77,7	49181	584215
Babócsai R., Babócsa	851	1952-2009	813	5,312	78617	518814
Fekete víz, Kémes	845	1960-2009	1185	7,2	53025	574923
Egyesült Gyöngy., Kétújfalu	850	1951-2009	430	10,6	71687	548137
Bükkösdi-víz, Szentlőrinc	844	1954-2009	137	6,521	78154	567201
Pécsi-víz, Kémes	150031	1951-2009	576	2,5	53922	57918
Ecseny-diósi, Somodor	150033	1989-2009	22,4	0,9	124900	557800
Dombó cs., Somogyudvarhely	150030	1990-2009	292,16	5,6	92593	505711
Köröshegyi Séd Köröshegy	150022	1978-2009	19,42	2,9	165890	562092

Table 1. Characteristic hydrologic parameters of the individual stream gages

The individual values of the ecological low water values were calculated with the following formula:

$$Q_{ökol} = Q_{aug80\%} \cdot e$$

where:

e is the proportionality factor for the ecological low water flow adapted for the operational area (SW Hungary) of the DDVIZIG. Runoff calculation and analysis for the operational area is based on the data of 21 stream gages monitored over a 30 year period. The characteristic hydrologic parameters of the individual monitoring stations are shown in Table 1.

To obtain specific runoff for a unit area, first we generated an isoline vectorized map from mean discharge values for all watersheds in the operational area of the DDVIZIG. Flow values were determined at the stream gages and were estimated for the watershed upstream. Afterwards, flow data were interpolated with the inverse distance weighted function in ArcGIS environment. Subsequently data was rasterized generating grid cells of 1 by 1 km (Figure 4). All grid cells then were assigned a unique ID number.

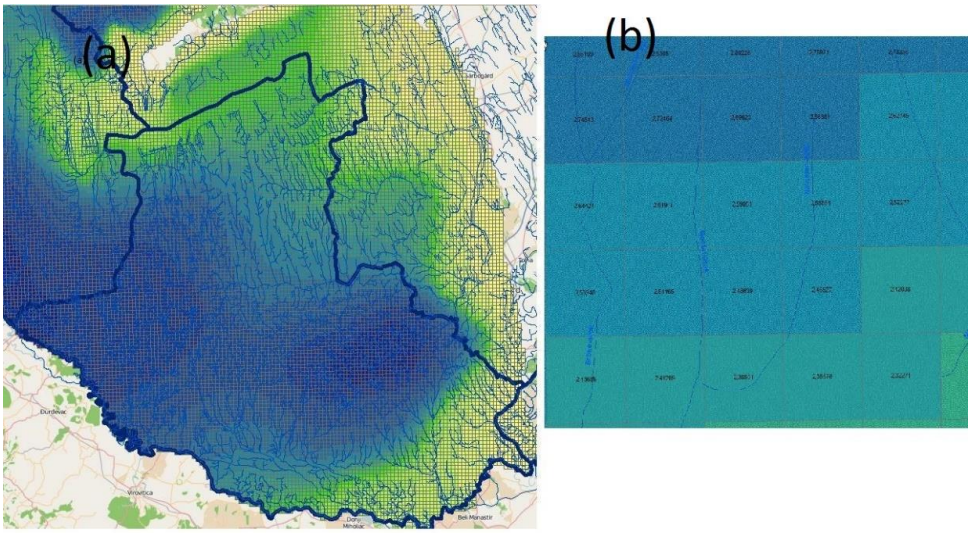


Figure 4 Specific mean flow values in the operational area (a) and the zoomed in raster grid (b) [l/s/km²]

The total contributing area (watershed area) upstream from any given outlet point was derived from the digital elevation model of the area. The obtained flow accumulation function then calculated the mean runoff and flow of the given watershed. Following the above-described calculation method, both the 80% August flow and ecological low flow are calculated (Figure 5).

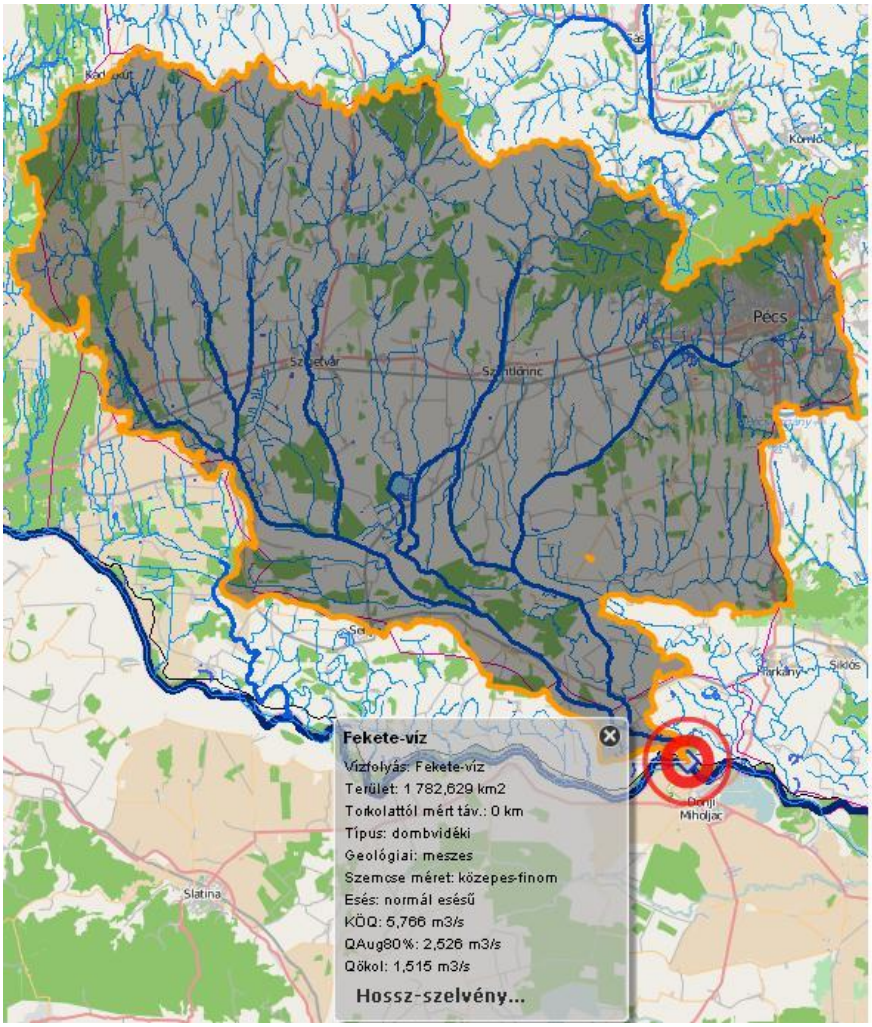


Figure 5 Calculated natural water budget values (from top to bottom: stream name, watershed area, distance from outlet, catchment relief type, geology, soil texture/mean particle size, slope, mean flow, mean August 80% flow and ecological low flow) for the outflow point of the Fekete-víz

1.5. Other miscellaneous map layers

In order to obtain spatially relevant data, various other map layers can be accessed and viewed in the STHIS. Such layers include for instance the location of the hydrometeorological stations and gages. Figure 6 indicates the location of the hydrogeographic stations (stream gages for water level and flow measurements), gages used for the determination of the shallow groundwater table and the location of the weather stations.

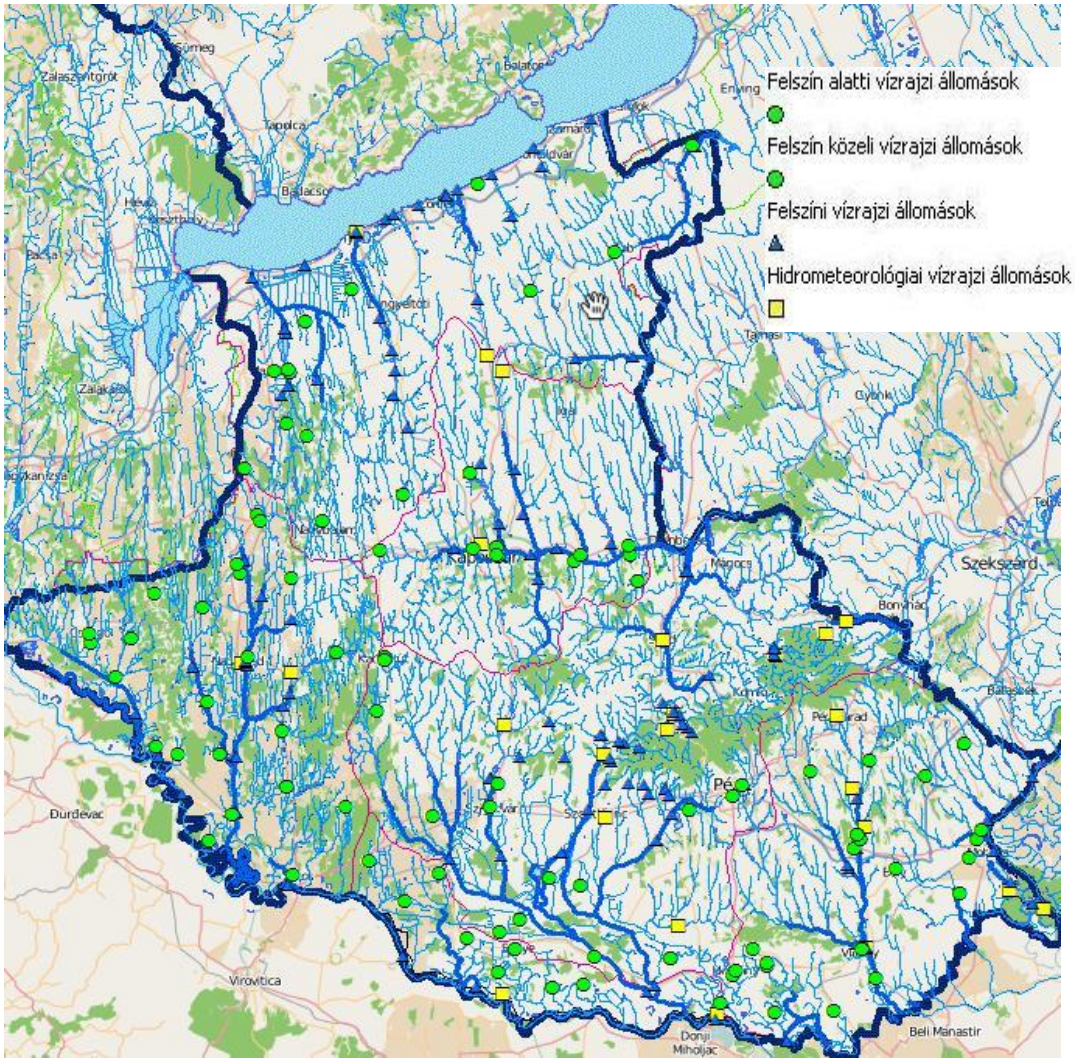


Figure 6. Location of the hydrometeorological monitoring stations within the operational area of the DDVIZIG

2. Calibration and validation of the water budget module

The output data of the water budget module was calibrated using the 30-year field data obtained by the DDVIZIG monitoring network. Data derived from the digital elevation model were also compared with field surveyed data. However, due to the terrain preprocessing method, and insufficient resolution of the digital elevation model, the digital elevation model only provides reliably data in area of relatively high relief (hilly and low-mountain areas). Watershed delineation in the floodplains and area so low slope value are burdened with a relatively large error.

3. Future directions

The current version of the STHIS explores the water budget of the area in a complex and sound way. Nonetheless, it solely focuses on water balance and budget quantification and does not include data and information on water quality. This shortcoming primarily roots in the low density of the water quality monitoring stations in the area and the uncertainty in calculation of contaminant transport as contaminant transport is heavily influenced by diffusion, advection and dispersion in natural water bodies.

The ultimately objective of the STHIS is providing state-of-the-art and readily accessible data on both water balance and water quality, as well as water usability for the entire operational area of the DDVIZIG.

A future field of improvement should include the monitoring of the groundwater supply, where additional data on hydrogeology, extent of subsurface water bodies would also be included in the database. Also, communal water contribution, as well as water conveyance and addition into the sewerage should also be mapped for the optimization of the sewage management.

Results

The currently described STHIS informational system is based on numeric approach and GIS techniques which provides hydrologic data through a user friendly interface. Nonetheless, the software itself is not yet devoid of errors and shortcomings: still, with future developments, it could provide a practical, cutting edge platform for a nationwide water balance model that provides a quantified spatial and temporal information on the individual elements of the hydrologic cycle.

Benefits of the STHIS model:

- Databases are systemized and structured and can be overlain, providing a complex information on a single surface layer for water professionals;
- Water balance and budget data in the Water Management Module supplies data for decision makers and relevant officials in making operative decisions; Measured data is readily available for water budget and flood forecasting.

Acknowledgement

This research was funded by the "TÁMOP 4.2.1.B-10/2/KONV-2010-0002" Scholarship (Developing Competitiveness of Universities in the South Transdanubian Region) and the KEOP-2.2.2/B/2F/09-2009-0001 project.

Reference

http://ddvir.ddvizig.hu:8000/ddvir/flex/ddvir.html# (last accessed: November 30, 2014)

Geomorphologic, hydrologic and tectonic modelling on a novel, computer-controlled sand table at University of Pécs

Pirkhoffer, E. – Halmai, Á.
University of Pécs (Hungary)

Technical specifications

Our sand table at University of Pécs is a steel framed table with a four-footed base frame and a moveable steel table with enclosing walls on the top of it (Figure 1.). This upper table and its walls provide the surface and volume for the experiments. The experimental area has an elongated, hexagonal layout. The longest inner length of the surface is 4.00 m and the longest inner width is 2.30 meters. The table can be arbitrary tilted at any angle between ±7.5 degrees along the longitudinal axis and ±10.0 degrees along the transversal axis. The required angle—measured along booth axes—is set by eight fully computer-controlled electric engines. The engines have variable speed from 0.2 degree per day to 1 degree per minute. Between these extremities the engines provide continuous movement. The maximal load of the experimental area is 2.5 metric tons.

Figure 1.: The sand table with no experimental material inside. The vertical tectonically plates are clearly visible (in the middle of the table with a handler on the top of them). There are horizontal tectonic plates on the left and right side of the table. There are two electric engines on the left side (black boxes with silver rings).

It is well-known in hydrology that the interval forces of the earth—especially the tectonism—has a serous influence on any morphologic processes. So there are four separately moveable plates on the bottom of the experimental area to mimic vertical displacement of the surface. But the tectonic faults has usually a second, horizontal component, so four vertically moveable plates are also included. The maximal vertical displacement of the plates is ±1.20 meters and the horizontal one is +1.0 meter respectively. These moveable parts are also driven by computer-controlled, programmable electric engines. These engines provide quasi-continuous movement between 0.01 m/day and 0.001 m/second.

The third part of the modelling device is the photographic array is mounted on the ceiling above the experimental area (Figure 2.).

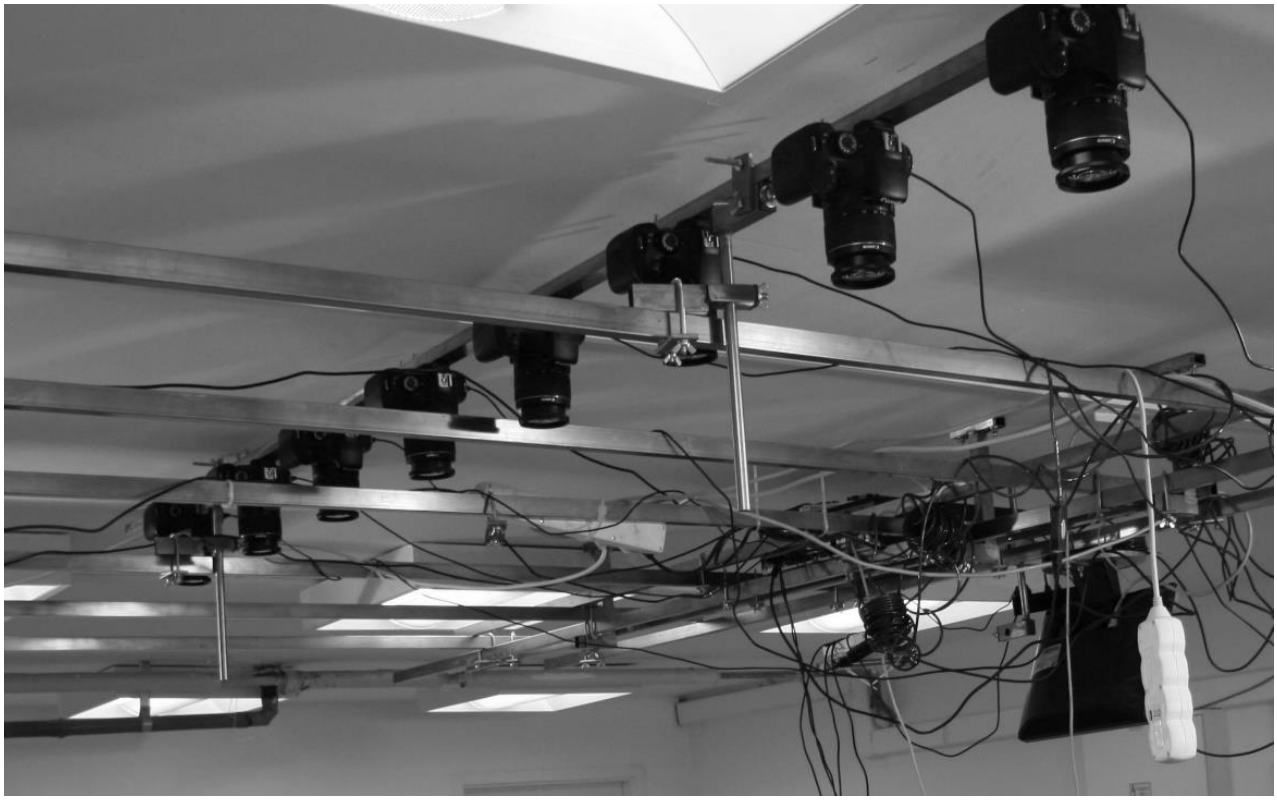


Figure 2.: The photographic array above the sand table. Eight Canon 1100D (12 MPix) with standard, EF-S 18–55 mm f/3.5–5.6 III kit lenses.

It consists of eight consumer-grade, entry-level DSLR cameras (Canon EOS 1100D), equipped with standard EF-S 18–55 mm f/3.5–5.6 III kit lenses. By 20 mm focal length the 75% of the experimental area is covered by the photographic array. The individual images are highly (≥65%) overlapping. By these settings the field resolution is around 0.6 mm. The cameras are connected to the central computer. Each parameter of the exposure on each camera can be altered programmatically and synchronously with the engines. The default exposure cycle takes a picture—with each camera—of the experimental area in every 10th second. After the experiment was ended, the images are

taken and downloaded they are transferred to a photogrammetric software. The photogrammetry provides orthophotos and digital elevation models—as the 3D reconstruction of the surface (Figure 3.).

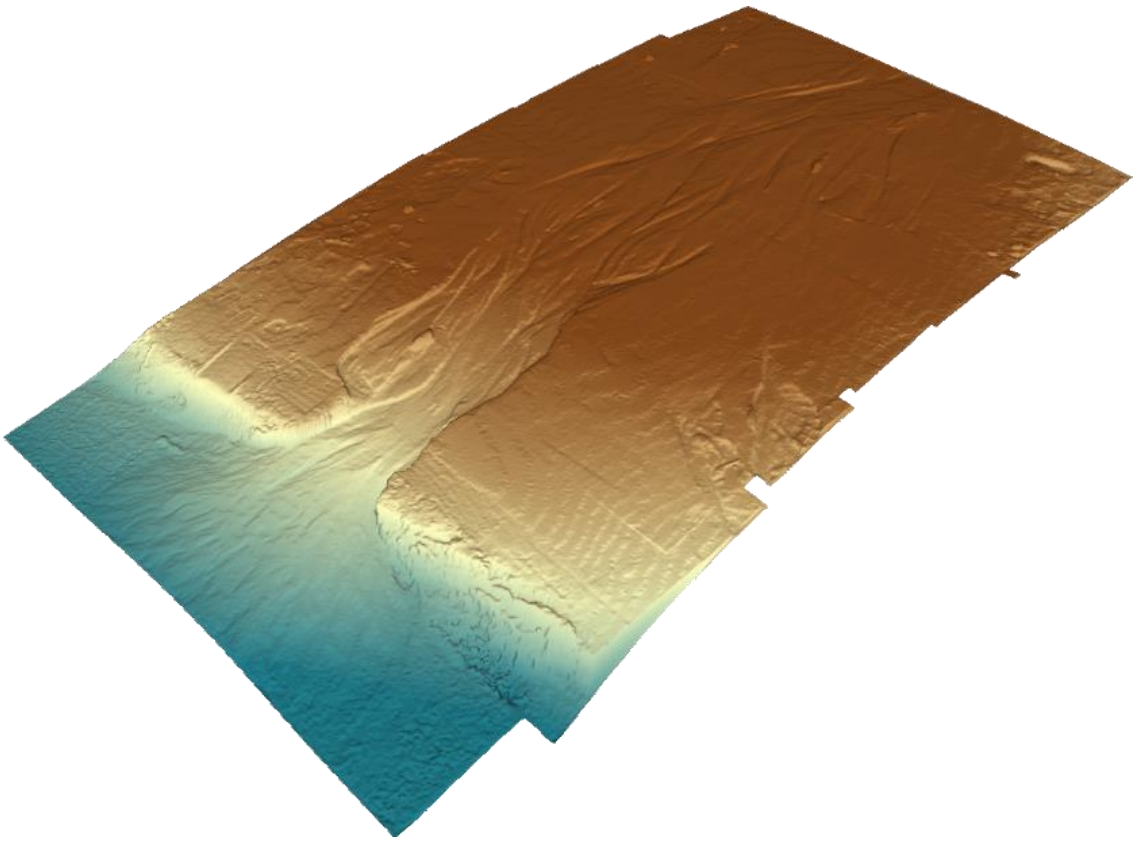


Figure 3.: 3D reconstruction of the surface. This digital elevation model was made from eight individual images with the help of close-range photogrammetry. On the left side of the image (blue colored) is a classical river delta. The upper, brown part is a typical floodplain with braided channels and meander cut-offs.

These reconstructions are not only spatial but also temporal ones, so we can record the digital footprint of an experiment. These photogrammetric derivatives of the experimental surface are totally identical to other photogrammetric derivatives are collected in the real-world. So the digital footprint of an artificial surface made on the sand table could be the input of any hydrological, or GIS modelling software.

There is also a pouring and sink point on the table. The water—or any density modified liquid—steps in at the pouring point and leaves the table on the sink point. It is the only way to leave the table, so it is an isolated system. A small filter prevents the sand or other solid material from leaking.

The experimental area could be filled up with any material. It could be sand, or different mixtures of grinded stones types, silts, soils, or grinded glass debris.

Purpose of the sand table

If someone is speaking about a physical modelling device this question always arises: "Why do we need physical devices at all? The world is full with computer aided numerical models and prediction systems." The answer is quite simple: a numerical model involves only well-known phenomena even they are deterministic or non-deterministic. But the nature has a haphazard manner. So the numerical models—even the Monte Carlo simulations—only pursuing the reality. A physical modelling technique or device includes or involves the reality, the real haphazard manner of the nature and—occasionally some unknown phenomena.
This sand table was built to build a bridge between these two modeling approaches. We can stop the experiment anywhere in the reality (on the sand table) and then continue it in a numerical, computerized hydrologic model or we can stop a numerical model and continue the experiment in the reality. This makes this device extremely flexible and capable of modeling.

If the table is tilted and we add some water than we can simulate natural hydrological processes such as river- or delta formation, meandering, sediment accumulation, linear and lateral erosion, stability of riverbanks and dykes, evolution of the cross and longitudinal section of a water flow or river. We can easily change and control the environmental settings, so we can experimentally predict the alteration of a river caused by climate change. We can also estimate the effects of the tectonic or lithological changes on a river—and with the help of photogrammetry—both in the real and in the digital world.

The importance of floodplain landforms along the Drava River in Hungary

Lóczy, D. - Czigány, Sz. – Pirkhoffer, E. – Gyenizse, P. - Dezső, J. – Halász, A.
University of Pécs (Hungary)

1 Introduction

As a consequence of river regulation and land drainage measures on the one hand and the impact of global climate change on the other, the alternation of floods and drought periods became typical of the environment of the Hungarian Drava floodplain. In 2012 a research project was launched at the University of Pécs with the principal purpose to assess the potentials of the physical environment after the completion of floodplain rehabilitation. *Rehabilitation potential* is defined as the opportunity for the simultaneous performance of as many as possible of floodplain ecosystem services (MEA 2003), including flood protection, groundwater replenishment, sediment and nutrient retention, water purification, resilience and recovery of river ecosystems after accidents, biodiversity/habitat, river-floodplain products (wood, fish, game, reed), cultural values, recreation and tourism, and climate change buffering capacity (WWF International 2010).

2 The hungarian drava river and its floodplain

The Drava is an important tributary of the Danube. Its length is 720 km and total catchment area is estimated between 40,150 and 41,810 km^2, of which 6,160 km^2 (ca 15%) lies in Hungary, where it is a border river with Croatia (except for a 29-km section) (VKKI 2010). (Although for historical reasons the international border does not follow the thalweg of the present river channel.) The lower catchment in the Carpathian Basin is mostly of hilly and lowland character with a wide floodplain (of which ca 486 km^2 belongs to Hungary), bordered by high bluffs in shorter sections, and elongated shape, open to the east. The latter circumstance has important implications for precipitation conditions (see e.g. Kiss, T. et al. 2013). For instance, in 2010, a year with extremely humid weather, no significant flood was measured on the Lower Drava River. The climate of the catchment is characterized by winter drought (January to March), wet summers (Atlantic influence in June-July) and autumns (Mediterranean influence in October-November) (Lovász, Gy. 1972). Higher discharges in the period April to June are due to snowmelt in the southern ranges of the Eastern Alps. Average precipitation on the Hungarian catchment is around 720 mm and runoff is 435 mm. Due to global warming, the duration of river ice has been considerably reduced: in the 1930s the river was commonly frozen for 28-30 days, while in the 1990s only for 2-3 days. It was almost free of ice cover in the 21st century. But in the winter 2013/14 record amounts of snow (5.9

km³) accumulated on the Alpine catchment by 27 February (National Water Management Service). A flood wave, however, was caused by high autumn precipitation.

2.1. Hydromorphology

In Hungary the mean *discharge* of the river is 496 m^3s^{-1} at the Barcs gauge, with long-term (1901-2000) low flow 190 m^3s^{-1} and high flow 1,433 m^3s^{-1} (for the Drávaszabolcs gauge, where hydrological observations began in 1936, the same parameters are 525, 220 and 1,365 m^3s^{-1}, resp.) (OFTE 2007). (The considerable difference between discharges measured at the upper and lower river gauge is explained by floodplain storage during flood waves.) The highest flood discharge was 3,070 m^3s^{-1} at Barcs (recorded on 19 July 1972), while lowest ever discharge in December 2001-January 2002. High daily fluctuation in water level (110-130 cm at Őrtilos) results from the peak-time operation of the Dubrava hydroelectric plant in Croatia.

2.2. Active floodplain

The Drava is a regulated river with a narrow (80 to 1,800 m wide) active floodplain. On the average floodplain width is 650 m, which compares with the 3,500 m average width and 14,500 m maximum width of the morphological floodplain, which includes the active floodplain and the former, now flood-protected, floodplain outside the levees. The sheer width of the floodplain could be reckoned as a major factor in the assessment of rehabilitation potential (Lóczy, D. 2013). The vegetation of the active floodplain of the Hungarian bank is softwood forest plantations and subordinately wet meadows, mostly in the vicinity of oxbows (Ortmann-Ajkai, A. et al. 2003).

2.3 Oxbows

Due to the common anastomosing channel pattern, the active floodplain of the Drava River abounds in *side arms*, some in a progressed state of siltation (OFTE 2007). Three of them have already been revitalized, i.e. they were dredged and flushed through after their closures being (partially) removed. A relief of 2-3 m is typical of the morphological floodplain. The total number of channel reaches of the Drava (beyond the flood-control levees) which were *cut off* through natural processes in historical times or artificially in the course of river regulation originally amounted to 18, but most of them has been completely infilled and shows no water surface today. In 2014 eight of them are registered as oxbow lakes (Table 1). Eutrophication rapidly reduces the depth and open water surface of oxbows. The utilizaton of oxbow lakes is mostly for angling and they are also important plant species and bird refuges (Pálfai, I. 1998).

Oxbow	Lake	Length (km)	Width (m)	Depth (m)	Open water surface (hectares)	Settlement
Cún-Szaporca	Kisinc	1.3	100	1.0	20	Szaporca
	Kishobogy	0.6	50	0.5	6	Szaporca
	Szilhát	0.6	80	0.5	6	Cún
Majláth-puszta	Majláth-puszta	1.8	40	0.5	4	Kisszentmárton
Bresztik	Bresztik	1.5	70	0.8	1.9	Drávasztára
Verság	Verság	1.3	60	1.0	7	Piskó
Matty	Matty	1.2	100	1.2	10	Matty
Matty Old Drava	Hótedra	1.1	40	1.5	4	Matty

Table 1 Major oxbow lakes in the Drava floodplain (updated after Pálfai, I. 1998)

3 The ancient drava programme

The ongoing research also serves as a baseline survey for a large-scale rehabilitation scheme, the so-called "Ancient Drava" Programme, which affects 43 settements on 45,000 hectares of land and will cost close to EUR100 million. By 2020 it is envisoned that the rehabilitated wetlands will attract ecotourism (for instance, anglers), improved water supply will allow irrigated agriculture and local inhabitants to earn their living from traditional economic activities.

The project can only be successful if the plans of water recharge are prepared in view of the rehabilitation potential of the floodplain, which serves as the basis to judge the success of rehabilitation. Authors are surveying the present-day (baseline) hydrogeological, hydromorphological and landscape ecological conditions of the floodplain with special regard to oxbow lakes and attempt to estimate the rehabilitation potential of the floodplain.

Water replenishment in the upper Ormánság is based on water transfer along the 44-km long canal system named after the *Korcsina* Stream. At higher water levels gravitational flow is possible through a sluice built into the flood-control levee and water recharge will affect 81 km^2 area. The canal is also suitable to collect excess water from the inundated floodplain and conduct it to pumping stations, where it is discharged to the channel. In gravitational operation the conduction capacity of the Korcsina canal is 5.5 m^{3}s^{-1} and in pumping 1 m^{3}s^{-1}. The bottom width of the canal is 1 m and its average slope ranges from 0.054% on the upper section to 0.014% on the lower. For six kilometres the canal follows the course of a triple oxbow system to promote groundwater recharge for agricultural, forestry and nature conservation purposes.

The flood retention function of the studied oxbows is hindered by the low-capacity of the outdated sluices and the siltation of the oxbow lakes. Because of the complex alluvial sequence, the oxbow lakes will not be able to raise groundwater levels regionally. In the environs of the oxbow groundwater flow heavily depends on the water stage of the Drava (Fig. 2), but at low water groundwater flow is invariably directed towards the river channel. Low waters are increasingly common as a consequence of the subsidence of the Drava basin and the incision of the river channel. The clogging (biofilm development) of the oxbow bed may adversely influence groundwater flow. Well testing indicates that with a 3-m deep fluvial sequence of highly variable grain size under the oxbow bed, 15 days are needed for 0.5-m drop in water level.

4 Conclusions

The water recharge of the studied oxbows has to be strictly regulated as the inflow of too great amounts of water may be disastrous for the riparian ecosystem of the oxbows. The precise estimation of environmental/ecological flow values is critical for success and constitutes a separate part of the project. Although exceptionally high-water stages on the Drava River may last as long as 200 days (as in 1999) for most of the year (particularly in July and August), water recharge will only be possible for limited periods, but lake levels could be impounded through engineering solutions and thus ensure favourable groundwater levels at least in the immediate environs of the oxbows.

Acknowledgements

This research was partly funded by the "SROP-4.2.2.C-11/1/KONV-2012-0005" (Well-being in the Information Society) grant with financial support from the Hungarian Scientific Research Fund (OTKA, contract K 104552) and the Visegrad Fund (contract no 31210058).

References

AQUAPROFIT 2010, Ős-Dráva Program – Összefogással az Ormánság fellendítéséért (Ancient Drava Programme: Joint Action for Development in the Ormánság). Executive summary. AQUAPROFIT Zrt., Budapest, 29p. [in Hungarian]

Dénes, A. & Ortmann-né Ajkai, A. 2006, A Dráva baranyai holtágai 2. Vegetáció-térképek (Oxbows of the Drava River in Baranya county 2: Vegetation maps). Natura Somogyiensis, 9, 27-38. [in Hungarian]

Lóczy, D., Dezső, J., Czigány, Sz., Gyenizse, P., Pirkhoffer, E., Halász, A. 2014, Rehabilitation potential of the Drava in Hungary. In: Gâştescu, P. & Breţcan, P. (eds): Water resources and wetlands. Conference proceedings, 11-14 September 2012, Tulcea, Romania. Editura Transversal, Târgovişte. 71-80.

Lóczy, D. 2013, Hydrogeomorphological-geoecological foundations of floodplain rehabilitation: Case study from Hungary. Lambrecht Academic Publishing, Saarbrücken. 382p.
Lovász, Gy. 1972, A Dráva-Mura vízrendszer vízjárási és lefolyási viszonyai (Hydrography and runoff in the Drava-Mura water system). Akadémiai Kiadó, Budapest, 158p.
MEA 2003, Millennium Ecosystem Assessment. Ecosystems and Human Well-being: A Framework for Assessment. Island Press, Washington, D.C.
http://www.millenniumassessment.org/en/products.aspx
OFTE 2007, Ős-Dráva Program. Vízügyi műszaki terv (Ancient Drava Programme: Water Management Technical Plan). Ormánságfejlesztő Társulás Egyesület, Sellye. 172p. [in Hungarian]
Pálfai, I. (ed.) 1998, Magyarország holtágai (Oxbows in Hungary). Közlekedési, Hírközlési és Vízügyi Minisztérium, Budapest, online version: http://www.holtagak.hu [in Hungarian and English]
VKKI 2010. Vízgyűjtő gazdálkodási terv. Dráva részvízgyűjtő (Drainage basin management plan: the Drava partial catchment). Vízügyi és Környezetvédelmi Központi Igazgatóság, Budapest. 162p. [in Hungarian]
http://www.vizeink.hu/files/Reszvizgyujto_VGT_Drava.pdf
WWF International 2010, Assessment of the restoration potential along the Danube and main tributaries. Working paper for the Danube River Basin. Final Draft. World-Wide Fund for Nature,Vienna, 60p.
http://assets.panda.org/downloads/wwf_restoration_potential_danube.pdf

Figures

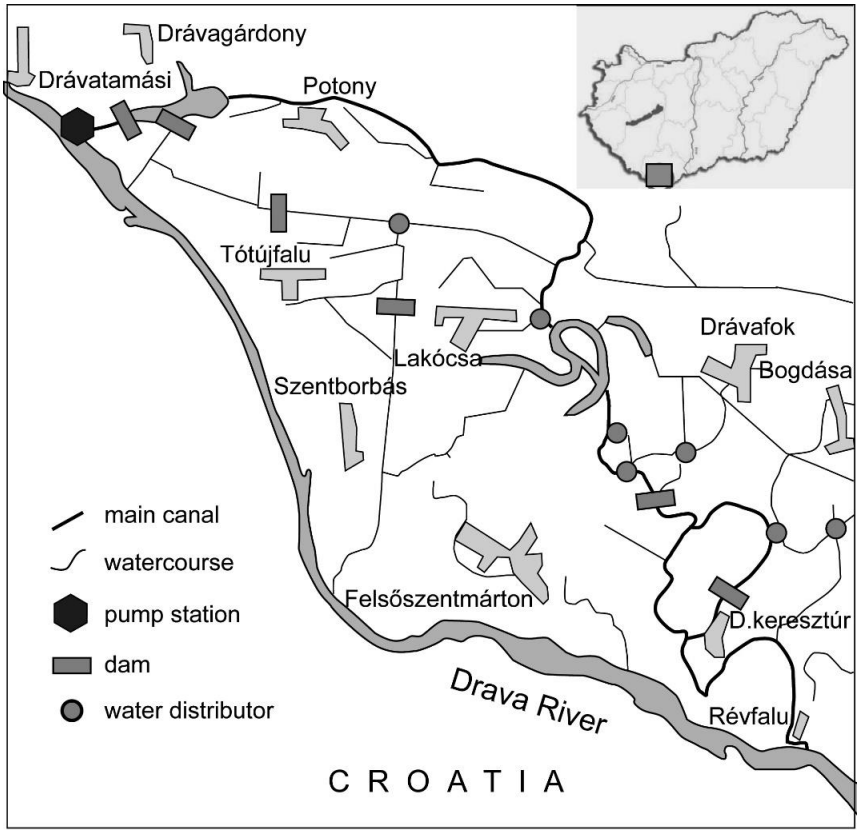


Fig. 1 Detail of the water recharge scheme in the upper Ormánság section of the Hungarian Drava floodplain (after OFTE 2007)

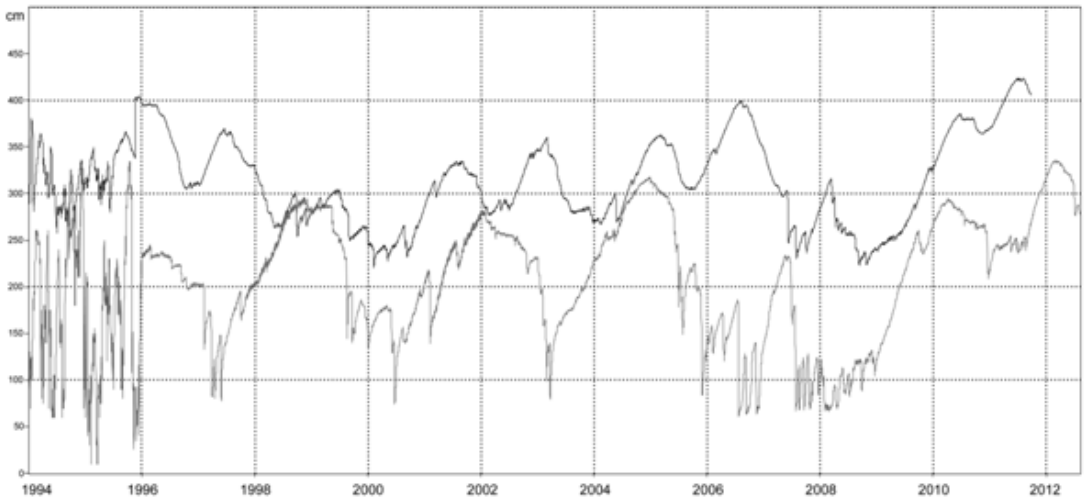


Fig. 2 Groundwater table levels at Drávaiványi, in the morphological floodplain of the Drava River (lower curve), and at Drávafok, on a sand mound, at 8.5 km distance from the channel (upper curve)

Cost reduction and density optimization of hydrometeorological monitoring systems in small catchments

Balatonyi, L., Czigány, S. and Pirkhoffer, E.
University of Pécs (Hungary)

1.Introduction

Flash floods occur when specific meteorological and hydrological conditions exist together. Although heavy rainfall is usually the main factor, a given amount and duration of rainfall may or may not result in a flash flood, depending on the variable hydrologic characteristics of the watershed where the rain is occurring. Given the importance of meteorological data and forecasts to the generation of flash flood forecasts, it is very important that there is a close collaboration between National Meteorological and Hydrological Services. Flash floods are localized phenomena that occur in streams and small river basins with a drainage area of few hundred square kilometers or less. It is difficult to realize high levels of forecast reliability for localized thunderstorms occurring on small and medium size watersheds. Nevertheless, flash flood forecasts do not necessary need the same accuracy as for large system to be effective. Insurance claims associated with floods and extreme rainfalls in Hungary totaled 835 cases over the past 25 years, out of which 39 cases were considered typical flash floods. On June 27, 1987, for example, several houses and part of the railroad was swept away in the Bükkösd Valley (Mecsek Mountains, SW Hungary) when 71 to 88 mm rain fell over a 6-hour period (Eszéky 1987). Flash floods caused traffic jams and overflown sewage systems when a rainstorm swept through the city of Győr (NW Hungary) on March 27, 2005. Perhaps the largest economic loss associated with flash floods was recorded in Mátrakeresztes, when a flash flood inundated the valley of the Csörgő and Kövicses Streams on April 18, 2005 (Horváth 2005). International research on flash floods have been relatively extensive in the US, UK and Italy (Gruntfest 1987; Gruntfest & Ripps 2000). However, today, when the consequences of climate change, such as the higher frequency of extreme rainfall events, are clearly notable, the need for sound knowledge of the hydrological responses given for such extreme meteorological events has emerged. The objectives of the current paper is to analyze the impact of (i) rainfall totals, (ii) rainfall intensities and (iii) percentage of canopy cover on interception and soil moisture values in small forested catchment in SW Hungary.

2.Materials and methods

2.1.Hydrometeorological monitoring and measurements

The Bükkösd Watershed has been relatively extensively monitored, primarily hydrologically, and to a lesser degree meteorologically, since 2000. Detailed meteorological data is provided by an automated meteorological station located at the outflow point of the Sás Stream.

One tributary catchment of the Sás Stream (Pósa Valley), located in the SE section of the Bükkösd Drainage area, has been studied in details since September, 2010. The Pósa Valley currently (as of September 30, 2011) is monitored by 11 hydrometeorological monitoring stations (Figure 1). This extensively monitored subcatchment covers a land area of 1.7 km^2.

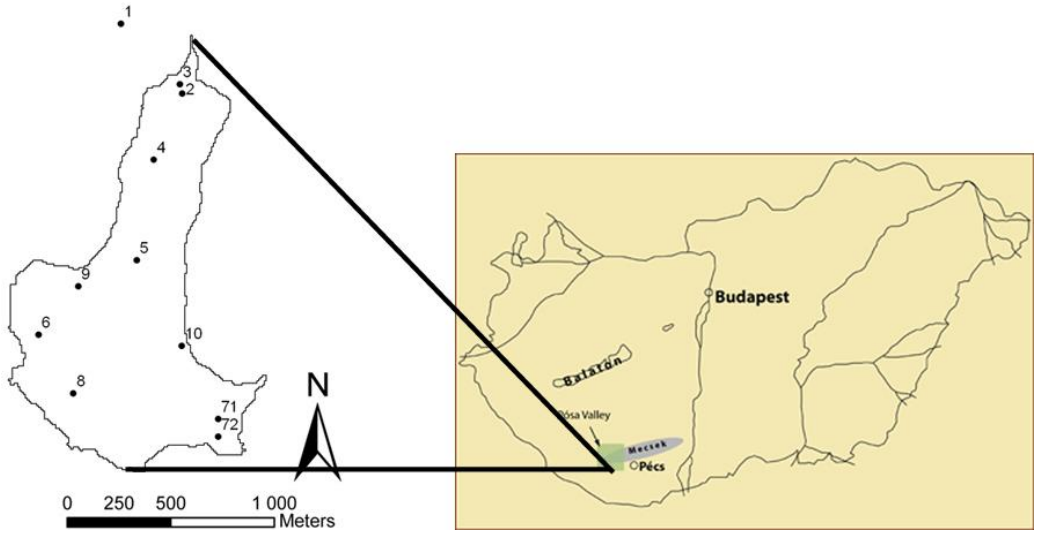


Figure 1. The hydrometeorological monitoring network and locations of the study area in the Pósa Valley. The shaded circle indicated the location of the two monitoring stations from which data was analyzed

2.2.Precipitation data

In September 2010, ten hydrometeorological monitoring stations were deployed in the Pósa Valley at an about even spatial distribution (Figure 1). Each individual hydrometeorological monitoring station includes a soil moisture sensor, a soil temperature sensor, a rain gage, and a datalogger. Three monitoring station includes shielded air temperature sensors and psychrometers to measure relative humidity. All sensors and loggers are manufactured by Decagon Devices Inc. (Pullman, Washington Unites States).

Data is usually collected in 10-minute intervals. The Pósa Valley is also equipped with a stream gage measuring stage at the outflow point of the subbasin. The TDR-type soil moisture sensors were inserted into undisturbed soil horizontally at a depth of 20 cm. Until May 30, 2011 the data collection interval was 10 minutes.

The onset of an individual rainfall event was any 10-minute data that was higher than 0 (0.2 mm in our case, as the resolution of the ECRN-100 precipitation gage is 0.2 mm). The end time of the given precipitation event was more complicated. The end time of the given precipitation event was identified of the last detected precipitation was followed by at least 4 precipitation-free collection intervals.

To account for the rate of interception, we further analyzed the data collected at the 7.1 (N574786, E84386) and 7.2 monitoring stations (N574786, E84286, Figure 2). The two-

selected monitoring station is situated in a relative proximity. Monitoring station 7.2 is found under a 40-year old dense hardwood (hornbeam and beech) canopy, while 7.1 is located in a 6-year-old clearcut, that is dominated by low, bushy saplings. Both monitoring stations are located in undulating surfaces at an elevation of 350 m in the close vicinity of the watershed divide. In the current paper we studied the temporal changes of canopy cover, the correlation between the canopy cover and the rate of interception, and the impact of rainfall totals, mean rainfall intensity and maximum rainfall intensity on the rate of interception.

2.3.Determination of canopy cover and interception

The proportion of canopy (leaves and branches) covered sky-view of the camera was detected in 2-week intervals. Black and white photos were taken from identical heights and locations each time, from the top of the rain gages.
One crucial contributing factor to the accelerated generation of flash floods is land use. Land use determines the rate of interception, i.e. the ratio of supra-canopy rainfall to ground precipitation, as well as influences the ratio of infiltration to surface runoff. Studies on interception rates have been carried out in both deciduous and coniferous forests. Literature-based interception rates in most cases range between 0,75 and 4,30 mm (Table 1). In NW Hungary rainfall lost to interception was found to be 34.6% in an oak forest, of which 27.0% was attributed to canopy and 7.6% to interception by litter (Führer 1994).

Name of the species	Canopy storage (mm)	References
Pseudotsuga menziesii	2,7-4,3	Link et al. (2004)
Pseudotsuga menziesii	1,2	Rutter et al. (1975)
Pseudotsuga menziesii	2,4	Klaassen et al. (1998)
Pinus nigra	1,0	Rutter et al. (1975)
Pinus sylvestris	1,02	Gash et al. (1980)
Pinus sylvestris	0,8	Gash és Morton (1978)
Picea stichensis	Approx. 2,0	Hutchings et al. (1988)
Picea stichensis	*0,75*	Gash et al. (1980)
Picea stichensis	1,2	Gash et al. (1980)

Table 1. Rate of interception is different species

3.Data analysis

3.1.Rainfall and soil moisture data

During the study period (February 28, 2011 to May 15, 2011) 13 and 19 rainfall events were identified at CS and UCS, respectively. In each month, spatial distribution of rainfall showed a heterogeneous pattern. Rainfall totaled 66.6 mm (Figure 3/a.) at the CS (7.1).
Both the number of defined rainfall events and the cumulative rainfall totals were higher at the UCS. Rainfall totals of 91.4 (Figure 4/a) mm was observed for the 7.2 monitoring station, which exceeds that of the covered site 2 times. The difference of ground precipitation accounts for an interception rate of 27%. In spite of the expected higher mean higher values, soil moisture was higher at the 7.2 station than at the covered station (Figure 4/b).

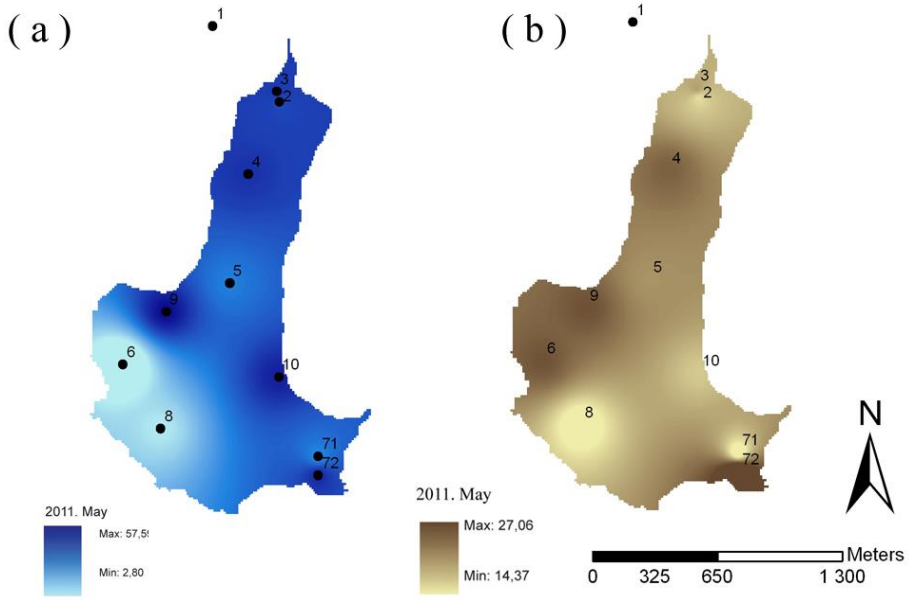


Figure 3. Spatial pattern of (a) monthly total ground precipitation and (b) an monthly soil moisture values in the Pósa Valley during May, 2011

The mean seasonal soil moisture value for the 7.1 station was 0,147 m3 m-3, while for the 7.2 station this value was almost exactly two times higher (0,294 m3 m-3, Figure 4/b)

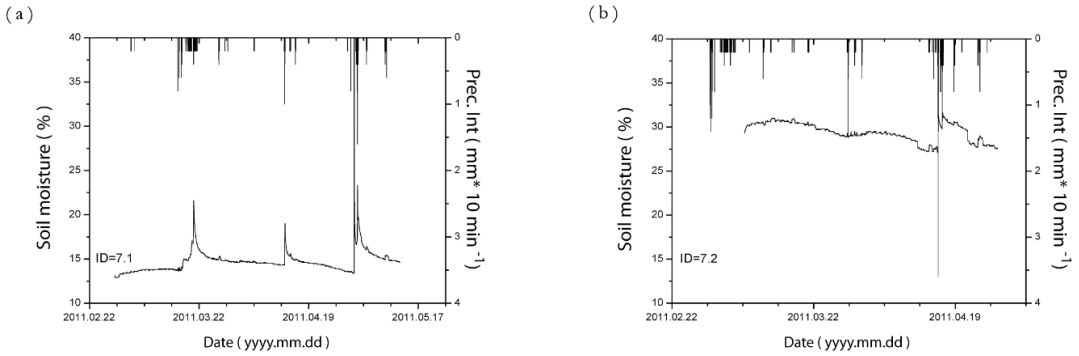


Figure 4. 10-minute precipitation and soil moisture values at the 7.1 monitoring station (a) and at the 7.2 monitoring station (b) observed between February 28 and November 25, 2011 in the Pósa Valley, SW Hungary

The shortest rainfall event at the 7.1 station lasted for 30 minutes, during which 2 mm rain fell. The longest rainfall event lasted for 1270 minutes, during which rainfall totaled 6.2 mm. The maximum rainfall total per rainfall event was 13.4 mm that fell over a period of 170 minutes.

Figures 6 indicate the time lag between the onset of the rainfall event and the abrupt increase of soil moisture values, which accounts for canopy interception. However, the time lag is also influenced by soil textural properties, mainly by infiltration rate. On average in the Pósa Valley pilot catchment, the abrupt increase of soil moisture began 90 minutes following the onset of the rainfall event.

Nevertheless, soil moisture changes are profoundly influenced by several other environmental factors, such as rainfall intensity.

The shortest rainfall event lasted for 10 minutes during which 1.2 mm ground precipitation was observed. The longest event lasted for 700 minutes with 11.4 mm precipitation. The highest rainfall totals per rainfall event were observed on 30 th of April, when 15.2 mm rain fell in 180 minutes. Interestingly, the data observed at the 7.2 site show moderate extremities compared to the canopy covered 7.1 sites (Figure 4).

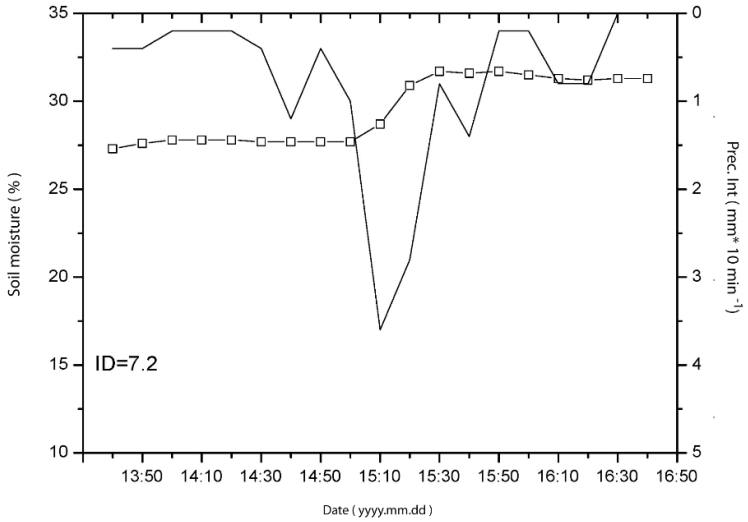


Figure 6. Time lag between the beginning of a rainfall event and the soil moisture maximum at the 7.2 monitoring station (UCS) during the April 30, 2011 rainfall event in the Pósa Valley

4.Conclusions

Numeric model based runoff models are ideal tools for comprehensive environmental analysis of watersheds. However, in order to obtain sufficiently accurate output data, detailed and thorough environmental monitoring is required that provides high-quality input data for hydrological models. Although some data is directly available and obtainable, others can only be obtained through indirect measurement methods or correlation functions.

During the studied period, in an experimental watershed in SW Hungary we found a strong correlation between canopy cover ratio and rate of interception in a low mountain deciduous forest. Mature canopy delayed runoff in all cases compared to premature and initial canopy.

Relatively close inverse relationship was found between interception rate and cumulative rainfall totals over the studied period. However, above about 2 to 3 mm per rainfall event, the amount of canopy-retained rainfall did not change. Thus, by measuring rainfall total, we can estimate ground precipitation at discrete points of the watershed where the actual land use type is known by antecedent survey. Interception rate may increase lead-time for the purpose of flood mitigation or prevention. Consequently, by knowing exact interception-generated runoff delay and time of concentration, flood generated losses and damages could be mitigated.

The results of the current study may significantly contribute to the increased accuracy of runoff models. This contribution is especially pronounced in the case of unexplored or partially unexplored watersheds, and may also provide sufficient insight for the seasonal differences of runoff. Nevertheless spatial differences in canopy cover not only impact the magnitude of runoff, but also profoundly affect soil moisture contents and storage capacity of the near-surface soil horizons. As storage capacity, soil moisture contents and initial soil moisture deficit are crucial input parameters for runoff models, awareness of exact initial soil moisture values will once again increase the output accuracy of the hydrological models. The influence of antecedent soil moisture values on surface runoff was simulated in the Pósa Valley experimental watershed in details by Pirkhoffer et al. (2009).

Acknowledgement

This research was funded by the “TÁMOP 4.2.1.B-10/2/KONV-2010-0002” Scholarship (Developing Competitiveness of Universities in the South Transdanubian Region) and the KEOP-2.2.2/B/2F/09-2009-0001 project.

References

Eszéky, O. 1987: A Bükkösdi-víz felső vízgyűjtőjén tervezhető árvízcsúcscsökkentő tározók hidrológiai tanulmányterve Pécs, kézirat

Gruntfest, E. 1987. What We Have Learned since the Big Thompson Flood, Proceedings of the Tenth Anniversary Conference, 17-19 July 1986, Special Publication NO. 16. Natural Hazards Research–applications Information Center, Boulder, CO.

Gruntfest, E. és Ripps, A. 2000. Flash floods: Warning–mitigation efforts–prospects, in: Parker, D.J.: Floods. Vol. 1. Routledge, London, pp. 377-390.

Pirkhoffer E., Czigány S. és Geresdi I. 2009. Modeling of flash flood event in a small low mountain watershed in SW Hungary.
Proceedings of joint MAP D-PHASE Scientific Meeting COST 731 mid term seminar: Challenges in hydrometeorological forcasting in complex terrain. Bologna:2009. Pp. 139-146

GIS based modelling of foreign waters in the sewerage of Pécs, SW Hungary

Ronczyk, L. – Czigány, Sz. – Nagy, G.
University of Pécs (Hungary)

1. Introduction

Urban hydrological issues are extremely complex. There are multiple types of human, infrastructural and environmental factors that may contribute to the hydrologic consequences of torrential rainfalls and other extreme weather event. These infrastructural damages may impact the flow properties of the surface water sewer system, while during heavy rainfall events, the sewage system may also be affected by flash floods. The proper maintenance of these human-made hydrologic structures is essential in order to manage stormwater and in general, urban environment form a hydrologic viewpoint. The plethora of contributing factors is further exacerbated by the heterogeneous spatial and temporal pattern of urbanization-generated processes.

An urban watershed can be analyzed in a complex way according to its risk potential if the contributing environmental and human factors are parameterized. By mapping these environmental and anthropogenic parameters with GIS tools, a watershed-based hazard map can be generated, that will provide a rapid analysis for areas where drainage and other problems may occur (Czigány et al. 2010). One special issue in case of Pécs, SW Hungary is the illegal stormwater canalization into the separated sewage system. The excess runoffs, as high-flows and floods, although invisibly, but also appear in the underground sewage system (Ronczyk & Lóczy 2006). The sewage system, due to its territorial development by EU funding, in the majority of time is capable to manage illegally canalized surface water. However, these excess waters originating from surface runoff may pose excessive pumping, transportation and management costs within the sewerage. This foreign water (at least from the viewpoint of sewage system) will increase the pumping costs of the water supplying company (Tettye Forrásház Ltd.), while ultimately the extra costs will burden the local taxpayers (Ronczyk & Czigány 2011).

The main goal of the present study is twofold; (a) to characterize the types of foreign water sources that enter to the sewerage and (b) to localize the major regions where foreign water enters into the sewerage and exerts additional load on the entire drainage infrastructure. With the establishment of a cutting-edge surface and subsurface environmental monitoring system, the overloaded sections of the sewerage were determined. To identify overloaded sections of the sewerage system of Pécs, University of Pécs, in cooperation with the Tettye Forrásház Ltd water supplying company, established an underground flow gage network at the pumping stations and a rain gage monitoring network. By employing integrated all-source catchment modeling schemes, overloaded periods and locations could be analyzed and forecast when torrential rainfall events are expected.

2. Materials and Methods

In the current work, from a hydrological viewpoint we have analyzed and studied the urbanized area of Pécs within the administrative borders of the city, however, the watershed delineation was expanded beyond the city borders. The total administrative area of the city covers 16,261 ha, of which the interior land zone covers an area of 6,381 ha. The total number of residents according to the 2012 census is 146,581 (KSH 2013).

First we determined the types of foreign waters that enter into the sewerage. Generally, foreign water types were classified into four major categories. Intentional water addition to the sewerage is considered as direct input, while accidental (unintentional) and natural input is defined as indirect input to the sewerage. Within both the direct and indirect categories, two subcategories were created, one being surface waters the second one being subsurface water *(Table 1)*.

	Direct	*Indirect*
Surface	Direct entry of stormwater runoff from impervious surfaces	Entry of stormwater runoff into the sewerage trough manhole drains
Subsurface	Direct flow conveyance from wells and springs into the sewerage	Decay of pipeline walls due to corrosion may trigger entry of shallow groundwater, especially at the boundary of sediment and soil layers of different hydraulic conductivity

Table 1. *Types of foreign waters that enters into the sewerage of Pécs*

3. Results

During the exploration of the problems of the individual sewage catchments, the above mentioned categories were used. Direct surface water input into the sewerage is likely at the highly urbanized districts of Pécs where high relief may contribute to decreased infiltration and enhances runoff. Here, the complete lack of drainage infrastructure and the high proportion of impermeable surfaces results in significant water addition to the sewerage. Indirect input is observed in the lowland part of the city, where, during heavy rainfalls, water enters into the sewerage through the manhole lids. Direct surface input occurs around springs and wells, in an estimated radius of 50 meters. Additional indirect input may origin from high level of groundwater table as well. Each category was mapped individually during the current project.

Our complex hydrological analysis primary focused on the 1-0 main sewage catchment, as this drainage area is the most complex in hydrological aspects. This catchment comprises the core of the old city center, where the degree of urbanization is the highest, and the land use type is extremely complex. Here the sewerage has been continuously

developed over the past decades and it connected areas of various level of development. Along the main axis of the sewerage (the trunk sewage drainage pipe), districts of high population density are found. This district also receives water inputs from the former suburban regions, where former vineyards were converted to a family house district. Here, and basically at the entire land area of the city, foreign water may enter to the sewerage in a multiple ways.

The 1-0 sewage catchment was further divided into 21 subcatchments for more detailed analysis. Here in this sewage catchment, surface water potentially enters into the sewerage through 6242 manhole drains. The communal infrastructure here was elaborated in the early 1930s. Here, steep slopes toward the higher ground with limestone rocks at or near the surface may also contribute to intense runoff. The highest manhole lid is located at an elevation of 526 meters, thus a hydraulic head of 414 m is built up while water reaches the lowest elevation of the sewerage at 112 meters (Fig. 1). The mean slope (drop) of the sewerage here is 12 m/ 100 meter horizontal distance. This value is coupled with an average slope of 6°.

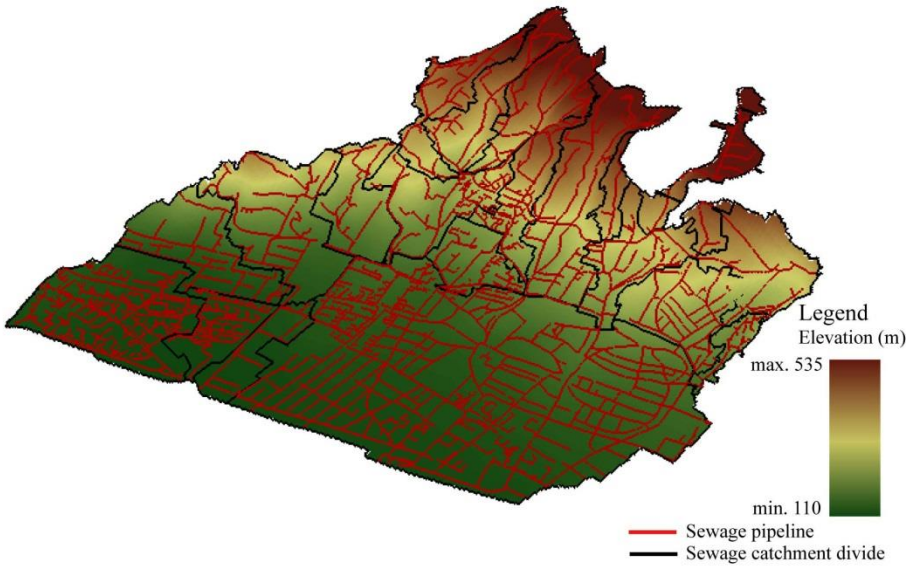


Fig. 1. Topography of the central part of the sewerage of Pécs

In the subsurface, all major groundwater types are present, including shallow groundwater, deep groundwater and karst water. The general flow direction, due to the tip of the limestone layers on the southern slopes of the Mecsek Hills is from the north to the south. At the joint of the individual limestone layers, there are numerous surface springs that may significantly contribute to excess stormwater runoff during heavy rainfall events in town *(Fig. 2)*. A major group of springs is found in the Bálics Valley, in the NW part of the city. Spring discharge values for the most voluminous Pacsirta Spring

vary greatly with a 2010 and 2011 mean value of 80 and 42 L min^{-1}, respectively. Toward the Pécs Basin, where the lowest point of Pécs is found, shallow groundwater often causes inundation and sewerage overflow problems, where paved surfaces and clayey fine quaternary sediments may restrict infiltration into the subsurface *(Fig. 3)*. In 36% of the main sewage catchment stormwater management and infrastructure are poor, as drainage holes are located at within a distance of more than 20 meters. Higher concentrations of properties that are devoid of appropriate stormwater management are found at the higher, northern part of the city, where steep slopes may significantly contribute to excess runoff. With the exception of the Uránváros Sewage catchment at least one of the above mentioned foreign water input categories are present *(Fig. 4)*. One way for stormwater retention here would be rainwater harvesting from roofs. According to a local legislative bill, since 2010 all newly built houses need to be equipped with cisterns used for rainwater harvesting.

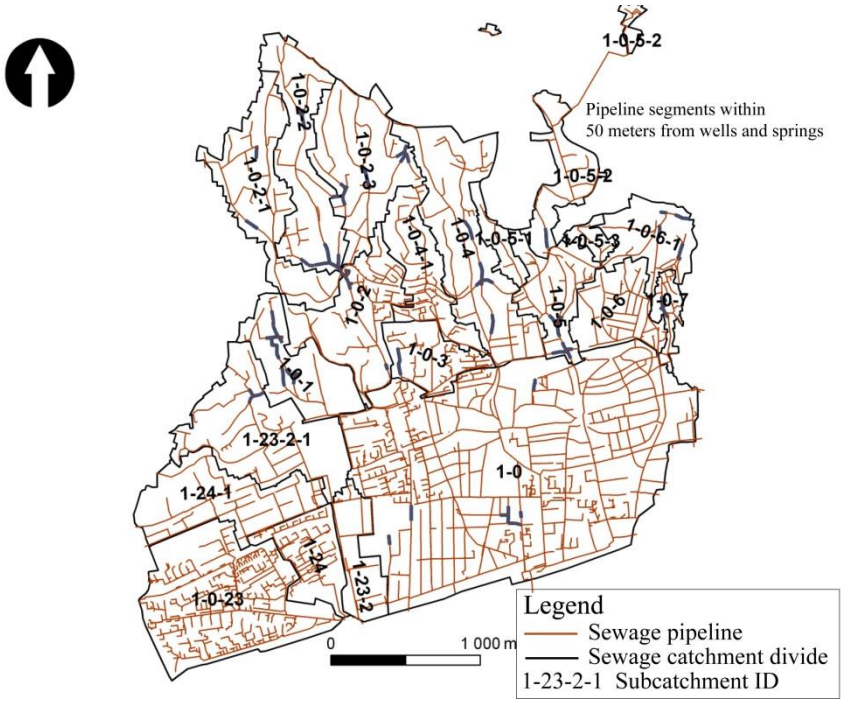


Fig. 2. Sewerage pipeline segments in Pécs with distances less than 50 meters from wells and springs

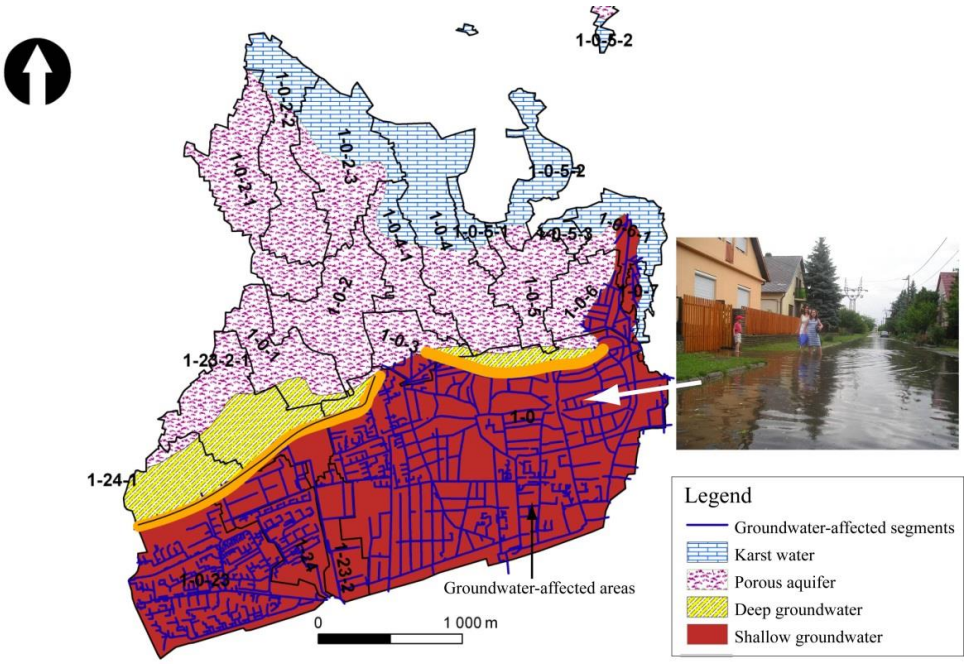


Fig. 3. Potential groundwater entry locations into the sewerage pipeline system in Pécs

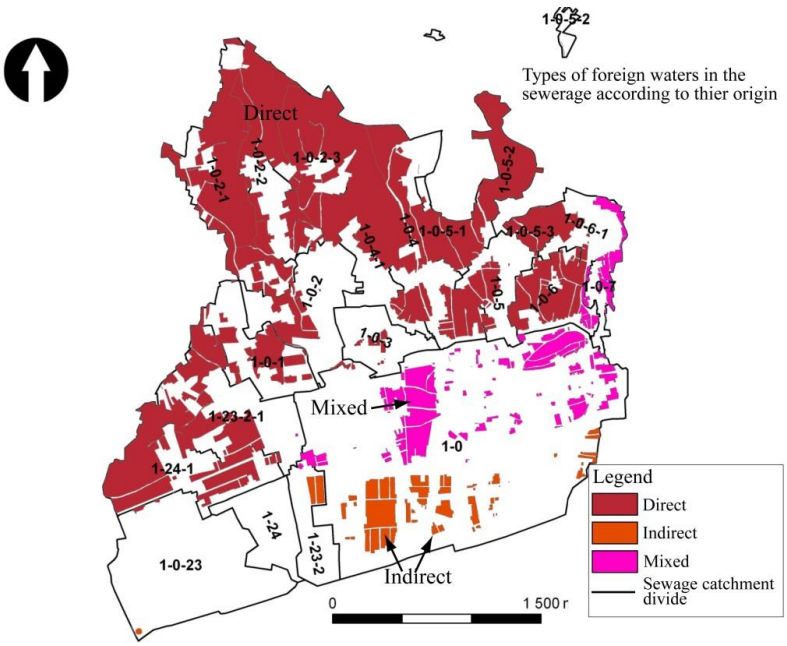


Fig. 4. Major types of foreign waters in the sewerage of Pécs

4. Conclusions

Defense procedures, in order to prevent the entry of foreign water are complex tasks for urban development planners. Undesired excess load of waters (e.g.: stormwater or sewage) may enter to the pipeline system in multiple ways (DRV 2012). As they are detected and mapped with various methods, their detection is challenging. In the current study we pointed out the various ways, through which foreign water contributes to the increased flow in the sewage system. We identified the types of foreign water entry into the sewerage and associated them with areas and sewage subcatchments in the city. The analysis presented in the current paper, highlighted that 32% of all subcatchment area is affected by foreign water entry into the sewage system. To reduce the extra costs associated with excess foreign water entry into and transfer in the sewage system, establishment of a hydrometeorological and hydrodynamic network may provide appropriate solution. A monitoring network of this type for the quantification of selected elements of the hydrologic cycle includes rain and flow gages. Based on the current collaboration between University of Pécs and the Tettye Forrásház Ltd. a numeric model supported urban master plan will be elaborated to achieve cost-efficient waste water management in Pécs. The master plan would assist on predicting the load on individual pumps in the case of substantial stormwater entry into the sewerage and distribute extra load among the individual pumps. On the long run, a stormwater management approach of this kind will be beneficial for Pécs and its principle, in general, for local municipalities as localized flooding would become avoidable, or at least mitigated this way. Additionally, with the integration and incorporation of infiltration models into the urban watershed models, increased accuracy of rainfall-runoff models could be achieved and could form the establishment of a high output predictions for urban flood guidance system to mitigate or even prevent stormwater runoff triggered catastrophic socio-economic losses.

Acknowledgements

This research was funded by the "**SROP 4.2.2.C-11/1/KONV-2012-0005**" Scholarship (Well-being in the Information Society). The authors are indebted to the Tettye Forrásház Ltd for providing financial support and pumping station flow data to the current study.

References

Czigány, S., E. Pirkhoffer, L. Nagyváradi, A. Szebényi, P. Vass & I. Geresdi 2010a. Potential areas of flood impoundment in Human environments, Bükkösd Valley, SW. Hungary. Géomorphologie: relief, processus, environnement. 3:301-310.

Központi Statisztikai Hivatal (KSH) 2013: Magyar Köztársaság helységkönyve 2013. január 1. KSH, Budapest, http://www.ksh.hu/docs/hun/hnk/hnk_2013.pdf - last accessed: November 25, 2014. (in Hungarian)

Ronczyk L. & Czigány Sz. 2011: Geoinformatikai adatok hasznosítása a pécsi szennyvízcsatorna öblözeteinek komplex monitoring rendszerének kidolgozásához,

(Use of GIS data on the elaboration of complex monitoring network of the sewage catchments in Pécs) Scientific report 161 pages
Ronczyk, L. & Lóczy, D. 2006: Alternative stormwater management in Pécs, Publicationes Instituti Geographici Universitaties Tartuensis 101. pp. 113-12

The Discourse-Semantic and Syntactic Background Behind ReALIS

Nőthig L. - Alberti Gábor
University of Pécs (Hungary)

1. Introduction

This paper claims in harmony with the spirit of our earlier papers that, on the basis of discourse-semantic relations "in the background", the corresponding word orders—or more precisely, *sets of synonymous word-order variants*—can be accurately calculated.[13] These representations and calculations form the basis of our current computational implementations (Alberti & Nőthig 2014, Nőthig & Alberti & Dóla 2014).

Section 2 demonstrates the relevant properties of these "discourse-semantic relations in the background". Initially they are considered underspecified *conditional chains of operators* (2.1), to be specified later due to some principle of "Maximize Discourse Coherence" (2.2).

Then, in Section 3, it is shown how to calculate word-order variants (in Hungarian) in this approach where syntax is based upon semantics, and not *vice versa*, as in the mainstream generative approaches (Section 4). Hungarian "surface" word-order variants will be calculated on the basis of five demands concerning order and adjacency in word order. The extreme freedom will be attributed to the fact that four of these five demands happen to be of the *same rank* (Alberti *et al.* 2002), and only the demand concerning the first focus is ranked higher (3.1). Subsection 3.2 offers a systematic review of straight and inverse scopes in Hungarian in the interaction of two operators (Topic/Quantifier/Focus). In 3.3 we sketch the basic differences between Hungarian and English in order to point to the universal character of our semantics-based approach.

In Section 4 we compare this approach to current minimalist (4.1-2) and OT (4.3-4) approaches in order to show that they (are to) converge toward a similar model of grammar.

[13] The present scientific contribution is dedicated to the 650th anniversary of the foundation of the University of Pécs, Hungary. We are grateful to SROP-4.2.2.C-11/1/KONV-2012-0005 (Well-Being in the Information Society) for their supporting our research team ℜeALIS in 2013–2014..

2. The Discourse-Semantic Background

2.1 Chain of Generalized Conditionals

In a representational dynamic discourse-semantic framework (e.g. ℜeALIS[14]), an operator zone belonging to a verb or other head[15] can be captured, as a first step, by a chain of *generalized conditionals* with the first focus (F_1) as the last element (if K>0):

(1) GENERAL SCHEME OF THE OPERATOR ZONE **[O*^F_1^O*...^F_K^O*^X]** OF A HEAD X:
- a. [... [$O_1 \rightarrow_{\omega 1}$ ($O_2 \rightarrow_{\omega 2}$ (... ($O_{J-1} \rightarrow_{\omega J-1}$ ($O_J \rightarrow_{\omega J}$ (X$\rightarrow_\forall F_K$))) ...))] $\rightarrow_\forall F_{K-1}$ ...]$\rightarrow_\forall F_1$
- b. Where X = V, A, Adv, Inf, N; J = 0, 1, 2,...; K = 0, 1, 2,...
- c. F_i: foci
- d. O_i: other operators, i.e. diff. sorts of topics, quantifiers and adjuncts
- e. Arguments are regarded as elements in the chain of O's (with "vacuous" logical content)[16]
- f. Specifications of '$\rightarrow_\omega$': $\rightarrow_\exists$, $\rightarrow_{\text{A FEW}}$, $\rightarrow_{\pi\%}$, $\rightarrow_{\text{MOST}}$, $\rightarrow_{\text{default}}$, $\rightarrow_\forall$, $\rightarrow_{\text{EXACTLY n}}$, ...

The analysis of sentences in (2-4) below will elucidate the details of (1) and its application. Let us start with sentence (2), whose (simplified) semantic background is the specified form shown in (2a) of the general scheme in (1); where e_i stands for *eventuality* in the formulae, and t_j for time; other referents are marked with r_k.

(2) GENERIC–HABITUAL AND ACTUAL READINGS:
A cseh férfi a barátnő·jé·t húsvét·kor el·visz·i a kedvenc söröző·jé·be.
the Czech man the girlfriend-Poss3Sg-Acc Easter-At away-take-3Sg the favorite pub-Poss3Sg-Into
e1[t1,r1] e2[t2,r21,r22] e3[t3,r3] e4[t4,r41,r42,r43] e5[t5,r51,r52]
'At Easter the Czech man takes his girlfriend to his favorite pub.'
- a. e1 $\rightarrow_{\omega 1}$ (e2 $\rightarrow_{\omega 2}$ (e3 $\rightarrow_{\omega 3}$ (e5 $\rightarrow_{\omega 5}$ e4)))
  e1: "r1 is a Czech man at some point of time t1" O_1 (topic)
  e2: "r21 is a girlfriend of r22 at t2" O_2 (topic)
  e3: "r3 is the Easter at t3" O_3 (adjunct)
  e5: "r51 is one of r52's favorite pubs at t5" O_4 (argument)
  e4: "r41 takes r42 to r43 at t" X (verb head)
- b. *generic–habitual*: e1′ $\rightarrow_{\text{default}/\exists}$ (e2′ $\rightarrow_{\text{default}}$ (e3′ $\rightarrow_{\text{default}}$ (e5′ $\rightarrow_{\text{default}/\exists}$ e4′)))
  e1′: "r_m is a Czech man at some point of time t"
  e2′: "r_g is a girlfriend of r_m at t"
  e3′: "r_E is the Easter at t"

[14] Alberti (2011a-b) and Alberti *et al.* (2010a-b, 2012) offer a thorough demonstration of ℜeALIS.
[15] See É. Kiss and Kiefer (1994), Szabolcsi (1997), Alberti (1997), Alberti and Medve (2000), É. Kiss (1995, 2001), Surányi (2011).
[16] Subsection 3.2 will provide an interesting approach to definite (and specific) arguments remaining in the complement zone of their heads: they are topics exhibiting their argumenthood instead of their scope. Indefinite arguments belong to some "predicative core" with the head in its center (details are postponed to future research).

e5′: "r_p is one of r_m's favorite pubs at t"
e4′: "r_m takes r_g to r_p at t"

c. If a man has a dime in his pocket, he throws it in the parking meter.
d. *actual* (eventualities e1, e2, e3, e5 have been anchored):
e4*: $p_{take}(t_{8April2012.10.30}, r_{Bohumil}, r_{Eliška}, r_{U\ Fleků})$

What comes *directly* from applying scheme (1) to a sentence like (2) is (2a): a *generic/habitual* reading, whose meaning is expressed by the simplified formula in (2b). Its interpretation requires considering event(ualitie)s in which somebody proves to be a Czech man and then their (more-step) extension to complex eventualities with a Czech man (r_m), his girlfriend (r_g), an Easter (r_E) and his favorite pub (r_p); and the given reading is to be evaluated as true if in (almost) each relevant extension of such complex eventualities[17], *r_m takes r_g to r_p at t* – provided that there is no explicit information to the contrary. This additional clause and the "almost" in brackets are intended to capture the *default* character of the conditionals ('$\rightarrow_{default}$').
The existential alternative ('$\rightarrow_{\exists}$') is interpreted so that, relative to the premise eventuality, a single extension is required to be true. This alternative is called the *weak reading* by Dekker (1996: 4.2), often illustrated by (2c) above: the man with a dime is not required, according to the preferred reading, to throw more dimes in the parking meter. Alike, sentence (2) has reasonable readings according to which "only one of the favorite pubs of the man in question is visited", and "only one girlfriend (out of the possibly more) is taken to the pub".

We return to the question of the *priority* of generic/habitual readings in 2.2. Nevertheless, sentences to be interpreted as generic are rare, due to the principle of Maximize Discourse Coherence (Asher & Lascarides 2003). It is typical that *in the relevant context* the eventuality expressed by the premise of a conditional should be matched to a single situation, *anchoring* certain referents to given entities. Hence, (2) may mean that *my friend Bohumil takes Eliška to U Fleků on 8th April 2012* (2d). Anchoring results in the disintegration of certain parts of the initial conditional scheme in (1).

2.2 *On the Priority of Generic Readings and the High Frequency of Anchoring*

Our starting-point is Vogel & McGillion's (2002: 163-164) observation: "There is a profound lack of syntactic cues to indicate when a sentence is to be interpreted as a generic or what sort of generic is intended. The paucity of syntactic encoding is cross-linguistic." They are led to this conclusion: "That a sentence may be understood as compatible with some concept that it does not express is not an issue of semantics, but of background conceptual structures or pragmatics."

[17] During the procedure of *extension*, the elementary eventualities should be *unified* on the basis of grammatical cues; which practically means an appropriate identification of their referents. See Kamp *et al.* (2011), for instance; and our method is illustrated in Alberti &Kleiber (2010). Let's consider the class of r_m, the Czech man: r_m is identical to r1, r22, r52 and r41, that is, the Czech man coincides with, respectively, the man whose girlfriend and whose favorite pub are mentioned, and who takes the girl to the pub. (Notice that referent r_{52} can also be identified with r_g, the referent belonging to the girl; the grammatical cues allow this ambiguity. Here we ignore this reading.)

We base the following hypothesis on this: the conditional formula in (1) will *directly* (or *primarily*) provide *generic* readings.

What is to be explained then is that sentences to be interpreted as generic are very rare.

The principle of "Maximize Discourse Coherence" (Asher &Lascarides 2003), which serves as the cornerstone of formal attempts to account for pragmatic phenomena, can explain this phenomenon as well. *In the relevant context* the eventuality expressed by the premise of a conditional (in scheme (1)) *can* typically be matched to a particular situation, *anchoring* certain referents to given entities; and according to this principle this possibility ("can be matched…") will "sharpen" to obligation ("must be matched…").

Anchoring results in the *disintegration*, or "crash", of the concerned parts of conditional scheme (1). (2d) above, for instance, shows the formula expressing the actual reading of sentence (2): "My friend Bohumil takes Eliška to U Fleků at the Easter of 2012". This proposition pertains to a single quadruple of a man, a girl, a pub and an Easter, while any of the four variants of formula (2b) pertains to millions of such quadruples. In other words, the truth value of (2d) depends on a single quadruple of entities, while that of (2b) depends on millions of quadruples of entities. We have meant this reduction of the space of search by the "crash" of conditional substructures owing to "Maximize Discourse Coherence".

It is worth noting that the very sentence *My friend Bohumil takes Eliška to U Fleků at the Easter of 2012* should primarily be considered as expressing a multiple conditional structure whose interpretation depends on a set of quadruples (with every friend of mine named Bohumil as its first argument, for instance); but proper nouns are inherently such that the space of search should be a singleton. Anchoring, thus, is practically obligatory.[18]

2.3 *Blocking Anchoring*

The role of explicit quantifiers lies in blocking the disintegration of conditional relations, that is, they retain a major space of search in the sense used in 2.2. They *set* a certain variant of generalized conditional $\rightarrow_{\omega}$ in scheme (1). From the viewpoint of interpretation, it is the degree / way of *extendibility* (of input eventualities to output ones) that is set in dynamic semantics, which defines the space of search relevant in the course of truth evaluation.

Most ($\rightarrow_{\text{MOST}}$), for instance, means that the premise eventuality can be *extended* to the conclusion eventuality *in most (relevant) cases* (>50% or 60%). *Every* ($\rightarrow_{\forall}$) permits no exception from the *extendibility* of the input eventuality; while *a(n)* ($\rightarrow_{\exists}$) means that a single *extension* of the input eventuality to the output one is enough. Intricate quantifiers can also be described by means of extendibility: *more than three but less than eleven and not*

[18] The interested reader is referred to Chapter 6 of Elbourne's (2005) book for a more detailed approach.

exactly nine, for instance, requires that the input eventuality can be extended to an output one 4, 5, 6, 7, 8 or 10 times.

The three sentences in (3) below, for instance, rely on the same space of search in truth evaluation, which contains certain triplets of friends, girls and Easters of a definite decade.[19] (3a) seems to require a much greater space of search than (3b) but that is not the case, because if (3b) is false, this fact can be proved only by reviewing the same vast set of triplets as the one to be reviewed if (3a) is true and we intend to prove this fact. In a felicitous case we soon find a counterexample in the case of (3a) and we soon find an appropriate visit in the case of (3b). As for (3c), the space of search is divided into two disjoint subsets. A triplet ⟨x,y,z⟩ belongs to the one or the other subset depending on the truth value of the proposition "x visited y at z". The truth value of sentence (3c) depends on the way of the above mentioned division. It is true, for instance, if at each Easter in question the speaker had six friends (not necessarily the same ones), out of which exactly 5 ones (5/6: *most*) were such that they visited exactly 5 girls (5 ≥3). We mention a case in which (3c) is false: there are seven Easters when every friend of mine stayed at home.

(3) NON-DISINTEGRATING CONDITIONALS

e1 $\rightarrow_{\omega 1}$ (e2 $\rightarrow_{\omega 2}$ (e3 $\rightarrow_{\omega 3}$ e4))

e1: "r1 is an Easter in the eighties"	O_1
e2: "r2 is a friend of mine at t2"	O_2
e3: "r31 is a female friend of r32 at t3"	O_3
e4: "r41 visits r42 at t4"	X (verb head)

a. Minden húsvét·kor a 80-as év·ek·ben minden barát·om minden lány·ismerős·é·t fel·keres·te.
every Easter-at the eighty year-Pl-in every friend-Poss1Sg every girl-acquaintance-Poss3Sg-Acc up-seek-Past
'At every Easter in the 80's every friend of mine visited every female friend of his.'
e1 $\rightarrow_{\forall}$ (e2$_{[t2=r1]}$ $\rightarrow_{\forall}$ (e3$_{[r32=r2,\ t3=r1]}$ $\rightarrow_{\forall}$ e4$_{[r41=r2,\ r42=r3,\ t3=r1]}$))

b. Valamelyik húsvét·kor ... egy barát·om fel·keres·te egy lány·ismerős·é·t.
some Easter-at ... a friend-Poss1Sg up-seek-Past3Sg a girl-acquaintance-Poss3Sg-Acc
'At some Easter in the 80's a friend of mine visited a female friend of his.'
e1 $\rightarrow_{\exists}$ (e2$_{[t2=r1]}$ $\rightarrow_{\exists}$ (e3$_{[r32=r2,\ t3=r1]}$ $\rightarrow_{\exists}$ e4$_{[r41=r2,\ r42=r3,\ t3=r1]}$))

c. Minden húsvét·kor ... a legtöbb barát·om legalább három lány·ismerős·é·t fel·keres·te.
every Easter-at ... the most friend-Poss1Sg at_least three girl-acquaintance-Poss3Sg-Acc up-seek-Past
'At every Easter ... most friends of mine visited at least three female friends of his.'
e1 $\rightarrow_{\forall}$ (e2$_{[t2=r1]}$ $\rightarrow_{>50\%}$ (e3$_{[r32=r2,\ t3=r1]}$ $\rightarrow_{\geq 3}$ e4$_{[r41=r2,\ r42=r3,\ t3=r1]}$))

[19] It cannot be said that the space of search is simply the Cartesian product of certain friends, certain girls and certain Easters, because time-dependent changes should also be considered. It is not sure that the set of my friends in 1983 is the same as in 1988; nor is it sure that Bohumil's female friends in 1983 coincides with his female friends in 1988.

2.4 *Foci: Last Elements in the Conditional Chain*

If there is only one focus in an operator zone belonging to a head, its piece of information counts as the novel assertion according to the information coming from other parts of the operator zone in question, which serves as the presupposition that the novel information relies on. (1a) in the general formula in 2.1 captures this by setting the focused piece of information in the last position of the conditional chain followed by the symbol of the universal extension ($\rightarrow_\forall$).

Let us check this method. In (4a) below, after anchoring referents r1, r3 and r4 to particular persons, the unification of eventualities e3, e4 and e2 will serve as the presupposition, whose content can be formulated as a complex eventuality e* expressing that "r21 introduced a certain John to a certain Mary". The universal extension then captures the meaning typical of focus as follows: every (relevant) extension of e* is such that r21 coincides with the particular Peter's referent r1. That is, nobody else introduced John to Mary.

(4) SINGLE AND MULTIPLE FOCUS

a. Csak Péter mutat·ta be Jóská·t Mari·nak.
only Peter show-Past-3Sg in Joe-Acc Mary-Dat
'It is only Peter that introduced Joe to Mary.'
$[e3 \rightarrow_{\omega 1} (e4 \rightarrow_{\omega 2} e2)] \rightarrow_\forall e1$
e1: "r1 is Peter" F_1
e2: "r21 introduced r22 to r23" X (verb head)
e3: "r3 is Joe" O_1
e4: "r4 is Mary" O_2
After anchoring e3, e4: $e2_{[r22 = rJoe,\ r23=rMary]} \rightarrow_\forall e1_{[r1=r21]}$

b. Csak Péter mutat·ott be öt vendég·et csak Mari·nak.
only Peter show-Past-3Sg in five guest-Acc only Mary-Dat
'It is only Peter that introduced five guests only to Mary.'
$[e3 \rightarrow_{\omega 1} [e2 \rightarrow_\forall e4]] \rightarrow_\forall e1$
e1: "r1 is Peter" F_1
e2: "r21 introduced r22 to r23" X (verb head)
e3: "r3 is a guest" O_1
e4: "r4 is Mary" F_2
$[e3 \rightarrow_{\geq 5} (e2_{[r22=r3]} \rightarrow_\forall e4_{[r4=r23]})] \rightarrow_\forall e1_{[r1=r21]}$

It is also permitted that a focused construction gets in the scope of another focus (e.g. É. Kiss 1995, Alberti and Medve 2000; see also 3.2). The general conditional formula in (1a) in 2.1 captures even the entire generalization containing any number of foci.[20] The focus providing the new assertion, denoted by F_1 in (1a), is the one immediately preceding the stem of the verb (if X is a finite verb). The remainder of the operator zone in question will serve as a presupposition.

[20] Two foci in an operator zone, nevertheless, practically count as the upper bound of perception.

In (4b) above, for instance, the content of the presupposition is the following conditional relation e*: "person r21 is such that there are five guests introduced by him to no one else but Mary".[21] The last element of the conditional chain adds the following contribution to this: every (relevant) extension of e* is such that r21 coincides with the particular Peter's referent r1. That is, nobody else introduced five guests exclusively to Mary.

3. Syntax – Based on Semantics

3.1 Calculating Word Order (in Hungarian)

The principal theorem of this paper is that on the basis of the conditional chain in (1a), which expresses a fixed operator hierarchy in the given scope order, the list of the corresponding word-order variants can be calculated.

The radical consequence of this approach is no less than that we need no recourse to operator heads and +/– interpretable features (e.g. É.Kiss 2001, Szabolcsi 1997), or Raising, Scrambling and A-reconstruction (Surányi 2010), or the even complex apparatus proposed by Brody & Szabolcsi (2003) in order to account for all cases of inverse scope. Hence, this semantics-based solution makes the minimalist machinery (Chomsky 1995) unnecessary, providing an "even more minimalist" solution to word-order problems (see also Alberti 2011a-b). In 4.1 we view this radical turn in the historical perspective of the half century of generative linguistics, arguing that it should rather be regarded as a "point of convergence" within generative linguistics. A similar turn is argued for in Bobaljik &Wurmbrand's (2012) paper on LF-to-PF derivation, discussed in 4.2, and optimality theoretic syntax (see Heck *et al.* (2002) has also proved to require an input containing highly sophisticated semantic information (4.3-4).

Our semantics-based approach relies on the following universal hypothesis: There are some semantic relations (5) that, in every language, are *potentially* expressed by *adjacency in word order* and/or *direction in word order*.

Language-specific differences lie in this "potentially": the demand that w′ be immediately followed by w" has a language-specific *rank*. A demand of rank α can be satisfied either *directly* due to realizing the adjacency required, or *indirectly*, by satisfying a competing demand of a rank at least as high as α (Alberti *et al.* 2002, 2003, 2010a-b).

Demands, like those listed in (5a-e), may get in conflict in the sense that two demands may require distinct word orders. A conflict like this will not necessarily result in

[21] I am analyzing here the reading in which e3 is assumed to serve as a topic, and not as a focus (cf. *Öt vendég-et **be**-mutat-t-am Mari-nak* 'five guest-Acc **in**-show-Past-1Sg Mari-Dat' vs. *Öt vendég-et-mutat-t-am **be** Mari-nak*). It is also to be noted that the expression 'five guests', viewed as a topic, means that "at least five guests". In the course of truth evaluation it is enough to find five appropriate guests, and finding a sixth one (ore even more ones) will not destroy the positive truth value.

ungrammaticality, but the corresponding *ranks* will decide word order. ℜeALIS, hence, can be regarded as a special Optimality Theory (see 4.4).
Turning to Hungarian, syntacticians' hottest topic for decades has been the question of the source of the extremely flexible Hungarian word order, which, however, is quite rigid in the preverbal zone, at least in a certain sense (Surányi 2010). We claim that in our semantics-based approach the source of this double nature of word order can be regarded as the following adjustment of (numerical) parameters of Universal Grammar in the case of Hungarian: four semantic relations, namely (5a-d) below, come with the *same ranks* in the above mentioned sense, so the corresponding demands can be satisfied *directly* or *indirectly* (also in the above mentioned exactly defined sense), providing alternatives.

The extreme (but restricted) flexibility of Hungarian word order, thus, can be attributed to a quadruple coincidence of the strength of (genuinely discourse-semantic) demands concerning word order. Each language can be characterized, viewing from the perspective of UG, by different values of strength of essentially the same demands. Less coincidence yields stricter word order.

(5) FIVE DEMANDS CONCERNING ADJACENCY AND DIRECTION IN HUNGARIAN:[22]

a.	predicate ↔ argument	rank 7
b.	host ← adjunct	rank 7
c.	predicate ↖ "stolen complement"	rank 7
d.	[operator > scope]	rank 7
e.	F1 ⇒ verb stem	rank 3

It is the primary focus, denoted by F1 in the general scheme in (1) in 2.1, that is distinguished in Hungarian (5e): its high rank ensures the corresponding focused expression a fixed position in word order, the one immediately preceding the verb stem, and this requirement cannot be overridden (cannot be "satisfied indirectly").[23]

[22] It is assumed here that the head X in question whose arguments are studied is a verb. Other cases are not discussed here. The precise values of ranks will not be discussed here either (but see Alberti *et al.* (2002, 3.1); only their relative strength counts: the equally high first four ranks in (5a–d) are less high than the quite high rank 3 in (5e). Rank 1 is the highest rank.

[23] Preverbs and other *aspectualizers* (immediately preceding the verb stem) are not discussed thoroughly in this paper. If you consult, say, the book edited by É. Kiss &Riemsdijk (2004), it will be clear that this topic would require at least another paper like this. I claim (without proving it now) that the analyses in my paper in the above mentioned book (Alberti 2004) can be reformulated in the semantics-based approach of this paper (included all kinds of "roll-up" an "climbing preverb" constructions).

The solution relies on the observation that F1 and expressions held responsible for aspect occupy the same place in word order, the position immediately preceding the verb stem (e.g. É. Kiss and Kiefer 1994). The following sentence in (i) can serve as an answer to any of these questions *What did you do? / Where did you travel? / Did you travel home?*

(i) [Haza ⇒ utaz·t·am]. (home travel-Past-1Sg) 'I traveled home.'
(ii) Haza.
(iii) [Utaz·t·am⇒] ↔ haza. 'I was traveling home.

The sentence in (ii) above can also serve as an answer to the last two questions. Demand (5e) thus can be generalized by replacing F1 with 'primary (element of) assertion' (*NB* in a sentence with more foci, F2, F3, ...,F_K belong to the presupposed part (see (1a), so the distinguished role of F1 in (5e) is not surprising).

Let us start with studying the competition between demands (5a), (5b) and (5d). The arrow types in (6-8) show the reader which demands in (5) are satisfied directly in producing which word order variant. The annotated first sentence is taken to serve as a fixed one with a fixed meaning, relative to which alternative word-order variants are being searched, firmly insisting on the fixed meaning in the discourse-semantic background.

(6) WORD ORDER VARIANTS IN HUNGARIAN DUE TO QUANTIFIERS AND ADJUNCTS
Tegnap mind·három lány·nyal két·szer is táncol·t·am.
yesterday all-three girl-with two·times too dance·Past·Sg1
'Yesterday I danced with all three girls twice.'

a. [Tegnap > mindhárom lánnyal > kétszer is > táncoltam].
b. *[Tegnap > kétszer is > mindhárom lánnyal > táncoltam]. *diff. scope order!*
c. [Táncoltam] ↔ mindhárom lánnyal ← tegnap ← kétszer is
d. [Táncoltam] ← tegnap ← kétszer is ↔ mindhárom lánnyal
e. [Tegnap > mindhárom lánnyal > táncoltam] ← kétszer is
f. [Tegnap > kétszer is > táncoltam] ↔ mindhárom lánnyal
g. [Tegnap > táncoltam] ↔ mindhárom lánnyal ← kétszer is
h. [Tegnap > táncoltam] ← kétszer is ↔ mindhárom lánnyal
i. [Mindhárom lánnyal > kétszer is > táncoltam] ← tegnap
j. ??[Mindhárom lánnyal > tegnap > kétszer is > táncoltam] *wrong scope order (?)*
k. ...

(6a) is the variant where *scope relations* are exhibited by directly satisfying (5d) in each relevant case, to the detriment of demands (5a-b), which are regarded in this approach as "satisfied indirectly by having satisfied a stronger (or at least not less strong) demand directly".

(6b) is an ill-formed alternative to (6a) because it expresses another scope order, hence, another meaning. This variant cannot be attributed to the direct satisfaction of demands (5a-b) because these demands require the complements (arguments or adjuncts) to follow their predicate or host. It is in this way that our semantics-based approach accounts for the rigidity of the preverbal zone of Hungarian sentences: demands (5d-e) are responsible for the word order in this zone, and not (5a-c), so the word order in this zone truly shows the scope order, which determines the meaning.

The precise semantic content of this concept in cases like the one when sentence (i) is an answer to the first question is postponed to future research.

In (iii) above there is a sentence in the progressive aspect. In this case it is not the preverb that is responsible for aspect because it could be responsible for the perfect aspect. Therefore, it is the verb stem itself that can be regarded as responsible for the aspect of the sentence. The arrow in (iii) attached to the verb stem is the denotation of the fact that in this word-order variant the verb stem is "self-aspectualizing", and the generalization of demand (5e) is to be understood as follows: "the aspectualizer should immediately precede the predicate if they are distinct; otherwise, their coincidence is to be considered the satisfaction of the demand".

Variant (6c), however, is a well-formed alternative to (6a).[24] In this variant it is the relation of the argument to its predicate that is shown primarily (5a); and what is satisfied then, secondarily, is the relation of the two adjuncts to their host (5b). The demands concerning scope hierarchy (5d) are "satisfied indirectly", that is, practically ignored.

Variant (6d) is another well-formed alternative to (6a). The difference between this (6d) and (6c) is that in the case of (6d) it is the demand concerning the adjacency of adjuncts to their host (5b) that is satisfied primarily, and demand (5a) is satisfied only secondarily, hence, partly. (5d) is ignored ("satisfied indirectly") also here.

Variants (6e-i) are all well-formed, which can be attributed to mixed strategies of directly satisfying (5d), on the one hand, and (5a-b), on the other. Anyway, these variants are all ambiguous, because if the place of an argument / adjunct exhibits its argumenthood / adjuncthood (5a-b), then its position in the scope hierarchy (5d) is not exhibited, so more possibilities concerning scopes are compatible with the given word order.

This semantics-based approach thus clearly accounts for both the phenomenon that more meanings may correspond to the same word-order variant and the phenomenon that numerous word-order variants correspond to the same meaning. Another promising possibility of the approach is that if competing word-order variants are qualified by native speakers as preferred or less preferred, there will be a tool at our disposal enabling us to analyze the reasons behind distinctions on the basis of directly or indirectly satisfied demands.

The word order in (6j), where *tegnap* 'yesterday' occupies an inappropriate position (relative to the intended scope order) without resulting in an undoubtedly different meaning, is intended to illustrate two questions. Are there potentially interchangeable meaning components in the scope hierarchy (cf. (6a))? And is there a fixed order for such interchangeable elements? The answers to the two questions seem to be positive, at least to a certain degree. Adverbials in a clause seem to make up an invariable hierarchy, which appears to be more or less stable across languages (esp. Cinque 1999). In our approach this skeleton is held to be rooted in the discourse-semantic background. This topic is postponed to future research.

The annotated sentence variant in (7) below serves as an illustration of demand (5c), which is to account for word-order variants in which an argument or adjunct semantically belonging to a predicate appears in the complement of a higher predicate (e.g. É. Kiss 2002:205-210). We call an argument or adjunct like this a *stolen complement*. In our semantics-based approach these stolen complements should be captured "semantically". An appropriate definition is obviously to rely on the following recursive extension of complementhood: the extended complement C* of a predicate p consists of the set C_1 of arguments and adjuncts belonging to p, and the set C_2 of arguments and adjuncts

[24] This does not exclude, nevertheless, that it is a well-formed alternative to (6b) at the same time. That is, (6c) is ambiguous because it does not exhibit its scope hierarchy (5d).

belonging to elements of this former set C_1 of arguments and adjuncts (which serve as predicates), and so on.

(7) SOME WORD ORDER VARIANTS WITH STOLEN COMPLEMENTS

A dékániá·nak tegnap nem sikerül·t sajnos a nyitva tartás·ról újra megszerez·n·em a tájékoztató·já·t.

the dean's office-Poss yesterday not succeed unfortunately the open hours-about again obtain-inf-1Sg the brochure-Poss3sg-Acc

a. ?[Nem > sikerült] ← sajnos ← tegnap ↔ megszereznem ← újra ↔ a táj. ↔ a dékániának ↔ nyt.

a.1 sikerült ← sajnos ← tegnap ↔ megszereznem

a.2 megszereznem ← újra ↔ a táj.

a.3 a táj. ↔ a dékániának ↔ a nyt.

b. [Sajnos > tegnap > nem > sikerült] ↔ [újra > megszereznem] ↔ [a dékániának > a táj.] ↔ nyt.

c. ??[Sajnos > nem > sikerült] ↖ újra ↖ a dékániának ← tegnap ↔ [megszereznem ↖ nyt.] ↔ a táj.

d. ?[A dékániának$_T$ > tegnap > nem > sikerült] ← sajnos ↔ [nyt'ról$_T$ > újra > megszereznem] ↔ a táj.

'Unfortunately yesterday I could not obtain again the brochure of the dean's office about open hours.'

(7a.1-3) show the semantic complements belonging to *sikerül* 'succeed', *megszerezni* 'to obtain' and *tájékoztató* 'brochure', respectively. The word order shown in (7a), thus, is such that all demands concerning argumenthood or adjuncthood (5a-b) are directly satisfied, at the expense of blurring scope relations (5d), as each complement set immediately follows its predicate/host.

In (7b), however, in numerous cases it is the demand concerning scope relations (5d) that is satisfied; here the cost is the potential blurring of complementhood. *Újra* 'again', for instance, at the same word order, could be regarded as belonging to the predicate *sikerül* 'succeed' instead of the predicate *megszerezni* 'to obtain', as is intended (7a.2). According to this alternative meaning the speaker has attempted many times to obtain the brochure, without any success; while the primarily intended meaning is that once the speaker could obtain the brochure, but he may have lost it, and yesterday he attempted to obtain it again.

What is demonstrated in (7c) above is not a preferred word-order variant of the meaning fixed in (7a) but is not entirely unacceptable either. *Újra* 'again' is the adjunct of *megszerezni* 'to obtain' (7a.2), but as it belongs to the extended complement of *sikerül* 'succeed', it is permitted to be adjacent to *sikerül* 'succeed', due to demand (5c). It is also due to demand (5c) that *a dékánia* 'the dean's office' is also quite near to *sikerül* 'succeed', having left the zone of *tájékoztató* 'the brochure'. The expression *a nyitva tartásról* 'about open hours' has also left the zone of *tájékoztató* 'the brochure', but it rests satisfied with a position in word order belonging to *megszerezni* 'to obtain' (which contains it in its extended complement).

The weak acceptance of (7c) can be explained in our approach as follows. The discourse-semantic significance of demands (5a-b) and (5d-e) are obvious. In the case of an argument it is its predicate that determines, partly or entirely, its semantic contribution in a certain sentence (5a). As for (5b), it was discussed in connection with (7b) that it counts in the case of an adjunct which predicate serves as its host. And we have seen more examples proving that scope relations between operators count very much (5d-e). It is demand (5c) whose significance; i.e. discourse-semantic contribution, is still not clear at all. What is clear is that a predicate (in Hungarian) dominates its entire "extended complement" in the above defined semantic sense.

This domination of the extended complement becomes relevant semantically in cases where special scope relations are attained between expressions that belong to distinct complements due to their belonging to the extended complement of a certain higher predicate. By (7d), for instance, the speaker is likely to intend to speak something about the dean's office. This expression can perform as a topic operator belonging to the highest predicate *sikerül* 'succeed' in spite of its origin as an argument of *tájékoztató* 'brochure'. The expression *a nyitva tartásról* 'about the open hours' also appears as a topic, but on a lower level. (7d) is fairly acceptable due to its useful discourse-semantic contribution.[25]

We argue, thus, that the hundreds of more or less acceptable word permutations belonging to a semantic content like the one fixed in (7a) can be accounted for, syntactically as well as discourse-semantically, by means of what is formulated in (5) concerning the relation between form and meaning. Moreover, this approach is promising in capturing even the preference order in sets of sentence variants sharing the same meaning, while minimalist approaches say almost nothing about the factual cost coming from Merge and Move belonging to the derivation of elements of such sets of more or less acceptable sentences sharing the same meaning (e.g. Surányi 2010). *Costs* of Merge and Move are often mentioned in minimalist explanations of the grammaticality status of sentences but no complete algorithms are provided for the exact calculation of the cost of competing word-order variants. Section 4 will discuss this topic.

The last example in this subsection is devoted to the illustration of the distinguished demand in (5), which has a stronger rank: it pertains to the focus, or more precisely, the focus having the widest scope out of the potentially, but very rarely, non-singleton set of foci belonging to the same predicate. This F1 should immediately precede the verb stem in order to satisfy (5e), as is shown in (8a) as well as in (8b-c). Further foci and expressions with other operator functions, however, can form a wide variety of word-order variants again, due the coincidence of four ranks in (5).

(8) TWO WORD ORDER VARIANTS OF SENTENCE (4B) CONTANING MULTIPLE FOCUS
a. [Csak Péter⇒mutatott] ↔ be ↔ öt vendéget ↔ csak Marinak.

[25] It is to be noted that in our model the word order in (7d) is not due to some kind of transformation through "extrapositions" but demands (5a-b) are to be regarded as holding true for all arguments and adjuncts in the semantically defined extended complement of a predicate/host.

b. [Csak Péter⇒mutatott] > öt vendéget > csak Marinak > be
c. [Csak Péter⇒mutatott] > csak Marinak ↔ be ↔ öt vendéget.

It is to be noted that the secondarily considered demand (5b) is not ignored, but is satisfied "partly": the adjuncts should be adjacent to their host, but as this host coincides with the predicate that the preferred argument is to immediately follow, demand (5b) can be regarded as having satisfied if the argument in question, but nothing else, can be found between the adjuncts and their host.

3.2 Straight and Inverse Scopes in Hungarian

Brody & Szabolcsi (2003, Section 1) argue that the overt syntax of scope in Hungarian, without considering preverb and contrastive topic, is (to be) based on four distinct devices: movement to operator-feature checking positions, feature inheritance via specifier–head agreement, i.e. pied piping, reconstruction to successive cyclic A-bar positions, and a way to alternate the order of specifier and complement (this last element is their supreme innovation).

We claim that ignoring preverb and contrastive topic does not necessarily make things simpler. The system is incomplete and less transparent without them. The table in (9) below shows the system based on the equally ranked demands (5a) and (5d), and the stronger demand (5e).

If only the two equally ranked demands cooperate (see the grey zone (Q»Q), and the T»T and T»Q quartets of rubrics), five word orders are predicted to belong to the same scope hierarchy, which is a correct prediction (if both arguments are chosen to exhibit their argumenthood, two word orders are possible in the postverbal zone, whilst if both arguments are chosen to exhibit their scopes in the preverbal zone, a rigid order is given). If demand (5e) is involved, the position of the focused argument is rigid immediately before the verb stem, and the other argument can be chosen to exhibit its scope (5d) or remain in its postverbal argument position (5a). The T»F and Q»F quartets show that this prediction is also correct.[26]

[26] The table below serves as the background of Table 9. Each operator pair is illustrated by a Hungarian sentence in which both arguments are chosen to exhibit their scopes. Alternative word orders should be computed by the reader on the basis of the permutations of letters in Table 9, where stresses are also provided (X: (normally) stressed expression, X: focus-stressed, x: non-stressed, *X*: stressed in the special way typical of the contrastive topic). The stresses in the postverbal zone are somewhat idealized according to the author's intuition/dialect, but they are in harmony with Brody & Szabolcsi's (2003) comment on their example (5); there is still no consensus on the slight and unreliable differences in stress in the right periphery of the Hungarian sentence.

	T	Q	F
T	[Valamelyik munkatársam]^{M} [valamelyik könyvet]^{K} el^{E}olvasta^{o}. [one_of colleague-1Sg] [one_of book-Acc] away-read-Past 'One of my c's read one of the books.'	$^{*?}$[Mindhárom munkatársam]^{M} [valamelyik könyvet]^{K} el^{E}olvasta^{o}. [all_three colleague-1Sg] [one_of book-Acc] away-read-Past 'All my three c's read one of the books.'	$^{??}$[Csak két munkatársam]^{M} olvasta^{o} [valamelyik könyvet]^{K} el^{e}. [only two colleague-1Sg] read-Past [one_of book-Acc] away 'Only two of my c's read one of the books.'

(9) WORD ORDERS IN HUNGARIAN ASSOCIATED WITH PAIRS OF T/Q/F OPERATORS (OP1»OP2):

op1 (m)→ ↓op2 (k)		T op	T arg	Q op	Q arg	F op	F arg
T	op	MKEo	KEoM	*?MKEo	KEoM	??MoKe	–
	arg	MEok	EokM/EoMK	MEok/*K*MEo	EokM/EoMK	Moek	
Q	op	MKEo	KEoM	MKEo	KEoM	?MoKe	
	arg	MEok	EokM/EoMK	MEok	EokM/EoMK	Moek/*K*Moe	
F	op	MKoe	Koem	MKoe	KoeM	?MoKe	
	arg	–				?Moek/*K*Moe	

There are four remaining pairs of operators: Q»T, F»T, F»Q, F»F.

The Q»T hierarchy is the case thoroughly discussed by Brody & Szabolcsi (2003) in connection with their examples (50–51) because scope transparency is not permitted to be realized here (*?MKEo). In our system this fact can be attributed to a basically discourse-semantic constraint on operator types: the well-known (e.g. É. Kiss 1994) T»T»...»Q»Q»...»F hierarchy: this order serves (in Hungarian) as a default which must not be violated, *at least to the left of (the surface position of) the verb stem*. The permission is to be construed as follows: the cases violating the discourse-semantic hierarchy are logically possible, but "forced" / artificial; and hence they can be expressed not via the most transparent way (where both T and Q exhibit their scopes) but via the alternative ways.

This explanation can also be applied to the other three pairings of operators F»T, F»Q, F»F. There is a new factor to be considered, however: as F1 immediately precedes the verb stem, there arises a potentially available zone between the verb stem and the preverb for the other argument to exhibit its scope (above the preverb); see the word order denoted by *MoKe* in Table 9. The question mark(s) in the appropriate rubrics, thus, can be attributed either to the dispreferred transparent exhibition of the artificial scopes, or to the violation of a phonetic constraint often referred to by É. Kiss (e.g. 2009), *Behaghel's Law*. According to this law, the heavy expression (compared to the small preverb) should occupy the right periphery of the sentence, so the argument with the narrow scope should exhibit its argumenthood instead of its scope (*Moek*). All in all, We consider it an interesting result that our approach (based on (5a) and (5d) in interaction with (5e)) associated with the manner of *not* ignoring the

Q	[Valamelyik munkatársam]^{M} [a könyvek többségét]^{K} el^{E}olvasta^{o}. [one_of colleague-1Sg] [most books] away-read-Past 'One of my colleagues read most books.'	[Mindhárom munkatársam]^{M} [a könyvek többségét]^{K} el^{E}olvasta^{o}. [all_three colleague-1Sg] [most books] away-read-Past 'All my three c's read most books.'	?[Csak két munkatársam]^{M} olvasta^{o} [a könyvek többségét]^{K} el^{e}. [only two colleague-1Sg] read-Past [most books] away 'Only two of my c's read most books.'
F	[Valamelyik munkatársam]^{M} [egyetlen könyvet]^{K} olvasott^{o} el^{e}. [one_of colleague-1Sg] [a single book] read-Past away 'One of my colleagues read a single book.'	[Mindhárom munkatársam]^{M} [egyetlen könyvet]^{K} olvasott^{o} el^{e}. [all_three colleague-1Sg] [a single book] read-Past away 'All my three c's read a single book.'	?[Csak két munkatársam]^{M} olvasott^{o} [egyetlen könyvet]^{K} el^{e}. [only two colleague-1Sg] read-Past [a single book] away 'Only two of my c's read a single book.'

preverb can enable us to account for and distinguish between two (not entirely ill-formed) word-order variants (*MoKe/Moek*).
Let us turn to the three word orders in Table 9 containing a contrastive topic ('*K*MEo'/ '*K*Moe'). In the light of what has been said above concerning the dispreferred transparent exhibition of artificial scopes, the well-known *scope-inverting* character of contrastive topic (e.g. É. Kiss & Kiefer 1994) is another straightforward way out of the tension: the operators left to the verb exhibit their default order T»Q»F, while the special intonation (denoted by '*K*' in italics) calls our attention to the „forced" opposite scope hierarchy.[27]

It is not a mistake in Table 9 that the discussed cases containing a contrastive topic are set in rubrics where the concerned argument is claimed to exhibit its argumenthood instead of its scope. What is relevant now is that the contrastively topicalized argument does not exhibit its scope. Presumably it exhibits a third factor: its special topic character, independent of the intended scope hierarchy. Topichood also explains the occupied sentence-initial position.

3.3 Towards the Calculation of Word Order in English

The semantics-based demands concerning word order in (5) in 3.1 are proposed as universal rule types with language-specific parametric differences in rank values. Each human language, thus, can be described uniformly in this approach, and their differences can be pointed out as well.

Let us have a look at English, for instance. There are a few straightforward significant differences relative to Hungarian, which severely limits the flexibility in word order, best illustrated in (6) in 3.1. In English there is no multiple coincidence of rank values in (5), but the demand concerning arguments (5a) is stronger than the demand concerning adjuncts (5b), and the demands concerning scope relations (5d-e) are even weaker.

Because of the former distinction, a word order like this one is ill-formed in English:[28] **I introduced yesterday Hubert in the library to my sister*; as *yesterday* and *in the library* are adjuncts between arguments. Inclination to scrambling, thus, is typical of Hungarian but not of English.

The English preference of demands concerning complementhood over those concerning scopes manifests itself in two ways: operator roles are exhibited in English either by intonation, while there is a fixed word order expressing argumenthood, or in a "tricky" way, illustrated in the table in (10) below. What else should be considered is the sentence-

[27] Cree inverse derivation (Croft 2001:319) can serve as an analogous case in morphology. In this language the expression of speech act prominence is claimed to be prior to that of grammatical functions. In examples (i–ii) feature 1Sg has a fixed word-initial position, and it is an inverting morpheme in (ii) that is responsible for the opposite distribution of grammatical functions relative to (i).

(i) ni-wāpam-ā-wak 'I see them'
1Sg- see -dir -3Pl

(ii) ni-wāpam-ikw-ak 'They see me.'
1Sg- see - **inv** -3Pl

[28] I thank Mark Newson for the precise form of this example (p.c., 12.12.2012).

initial place of topic and the sentence-final place of focus; this latter property also radically differs from the property of Hungarian in the given respect.

Suppose then that it is decided in the discourse-semantic background which argument should serve as a topic, and which one should play the role of the focus. In a felicitous case, the word order required by (5a), responsible for distinguishing between subject, object and further arguments, accidently coincides with the word order determined by the intended information structure. (10a') below shows a case like this. This "felicitous" state of affairs, however, can be generated by force – having recourse to the competing argument structure versions due to derivation. The verb *give*, for instance, has at least four ASVs: a ditransitive one (10b'), an active (10a') and a passive (10c'-d') transitive one, and an intransitive one (10e'-f').

(10) PREVERBAL TOPIC AND SENTENCE-FINAL FOCUS IN ENGLISH

neutral version: Péter oda·ad·ta a könyv·et Mari·nak. 'Peter gave Mary the book.'
Peter to-give-Past the book-Acc Mari-Dat

a. Péter$_T$	Marinak$_F$	adta oda a könyvet.
a'. Peter$_T$	gave the book	to Mary$_F$
b. Péter$_T$	a könyvet$_F$	adta oda Marinak.
b'. Peter$_T$	gave Mary	the book$_F$
c. Marinak$_T$	Péter$_F$	adta oda a könyvet.
c'. Mary$_T$	was given the book	by Peter$_F$
d. Marinak$_T$	a könyvet$_F$	adta oda Péter.
d'. Mary$_T$	was given (???by Peter)	the book$_F$
e. A könyvet$_T$	Péter$_F$	adta oda Marinak.
e'. The book$_T$	was given to Mary	by Peter$_F$
f. A könyvet$_T$	Marinak$_F$	adta oda Péter.
f'. The book$_T$	was given by Peter	to Mary$_F$

The reader can check above in the table that, with these argument structure variants at our disposal, several scope hierarchies can be exhibited by satisfying (5d-e) but at the same time satisfying (5a).

4. Optimality

4.1 The Autonomy of Syntax, which should be Abandoned

We consider it an important argument in favor of our semantics-based approach that it can serve as an excellent viewpoint for interpreting, from stage to stage, the history of

generative linguistics, with special respect to its central axiom, the principle of the Autonomy of Syntax (ASP). In what follows we sketch a review like this.

Obviously, the semantics-based approach of ℜeALIS is explicitly against ASP, according to which syntactic structures of natural languages can be described in a formal system without any reference in this description in any way to the meaning that syntactic structures bear.

In the first period of transformational generative linguistics, in the 50's and 60's of the last century, ASP served as a brilliantly useful working hypothesis – in the lack of any kind of formal semantics (Partee 1996). ASP ensured the legitimacy of Chomsky's (1957) syntax-based Language Faculty; and this approach made it possible to reveal an enormous amount of exact, falsifiable results in syntax.

And it is just on the basis of these exact results in syntax that a discipline of formal semantics could be developed in the 70's (e.g. Montague 1970, Dowty *et al.* 1981).
In the 80's the freely—hence, blindly—functioning Move-α was in total harmony with ASP (Chomsky 1981, 1986). What this apparatus suggests in respect of explanatory adequacy, however, is nothing else but a weak cooperation between syntax and semantics, which is simply counter to intuition. Why would the formal side of language be such that it works against the exhibition of the semantic side, instead of exhibiting as much as possible. Numerous problems arose in the area of descriptive adequacy as well (Surányi 2010:11-12).[29]

The current minimalist research program, initiated in the 90's (e.g. Chomsky 1995), already captures the tension between what we consider to be demands concerning the exhibition of the two basic aspects of the discourse-semantic background, i.e.

[29] "… in pursuit of the twin objectives of descriptive and explanatory adequacy some of the basic notions and principles became increasingly non-natural and complex (like government and the ECP, or the notion of Local Domain in Binding Theory). This gave cause for growing concern in the field, in no small part because the question of why UG is the way it is became disappointingly elusive. The ultimate source of the emergent complexities, beyond the strive for ever-improving empirical coverage, was the fact that GB lacked an actual theory of possible principles or, for that matter, of possible parameters. As for the latter, continued in depth research on cross-linguistic variation has shown many of the macro-parameters, among them the null subject parameter, to be unsustainable in the strong form they were originally proposed: several of the linguistic properties correlated by macro-parameters turned out to be cross-linguistically dissociable. Even though the idea of parametric linguistic variation was upheld, parameters themselves needed to be scaled down. In addition, as GB relied on massive overgeneration resulting from the fundamental freedom of basic phrase structure and transformations, downsized by declarative constraints imposed (mainly) on syntactic representations, the computational viability of the model was often called into question."

complementhood (5a-c), on the one hand, and scope relations (5d-e), on the other, when it permits only "triggered moves" and prefers the least possible amount of moves, in total opposition with the great freedom of Move-α in the 80's. Merge is practically responsible for complementhood, and Move is primarily responsible for scope relations. This model thus practically satisfies the intuition behind our semantics-based model. What remains counterintuitive is that triggers of Move are features to be checked in appropriate syntactic positions; hence, they are "syntactic", or rather, they are pretended to be "syntactic" in the first, "cartographic" period of MP. We claim that it is just at the cost of this pretension that ASP can be retained.

The 2000's is the age of the "anticartographic" Minimalist Program (see e.g. van Craenenbroeck 2009), whose decisive property is that many features/positions are held to be "methodologically unsound" (Surányi 2010:17) and hence to be avoided in the name of some Semantic Economy,[30] which can be captured in the form of a generalization of *Last Resort*: reference to some semantic function is sufficient for move. The earlier syntactic disguise is not required any more, because it counts as a superfluous cost, which is against the minimalist spirit. What is to be regarded as the most economical grammar in the minimalist spirit is practically the approach of ℜeALIS, in which both Merge and Move are dispensed with.

The intuition behind Semantic Economy, thus, already almost coincides with that behind our semantics-based approach. We consider it obvious that Generalized Last Resort is in explicit conflict with the Autonomous Syntax Principle, or at least makes it vacuous. We think that the only reason why ASP has not been abandoned is the anxiety about loss of prestige, the loss of what seems to be the cornerstone of the enormous building of generative linguistics.

What we have intended to prove is that there is no danger of this loss. We can declare that semantics is prior to syntax in the sense that it is syntax that should be based on semantics, and not *vice versa*. All descriptively or explanatory adequate results due the generative paradigm, however, can be quite easily reformulated in our semantics-based approach with the universal demands in (5) in its center.

[30] "Accordingly, pairs of agreeing uninterpretable and interpretable abstract morphosyntactic features like [top(ic)] and [foc(us)] have been posited by analysts of pertinent constructions in different languages, conforming in this manner to the working hypothesis of the MP that all (displacement) operations are triggered. But such an implementation of the notion of trigger is methodologically unsound, since, while it applies the same mechanism of trigger throughout, it substantially weakens the predictive power of the hypothesis itself (the general prediction being that all movements are triggered), to the degree that makes the argumentation almost circular."

4.2 *LF-First*

In the "Economy" frameworks of the last two decades of different branches of generative linguistics the initial step of derivation can be characterized as follows (Bobaljik & Wurmbrand 2012, Section 2): "economy conditions only evaluate competing derivations from the same numeration (input)".

The authors review several problems concerning the rules of this competition. Do active and passive sentences compete with one another? Does a topicalized construction compete with its with its topicless counterpart? What about variants modulo scrambling?

It would go far beyond the size of this paper to enter into these problems. What is relevant now is the authors' conclusion (Section 3), which is in total harmony with the spirit of our framework based on a semantics-based syntax: "... the examples considered will provide key support for the claim that the LF:PF isomorphism condition (Scope Transparency[31]) is asymmetric and gives LF a privileged status. PF representations compete to find the best expression of a given LF, not the other way around." In other words (Abstract/3.1), "LF (broadly construed) is calculated first, and determines PF (surface word order)"; "the LF representation constitutes the input, and the best PF representation is chosen to realize it". In Section 4 they speak about their "LF-First" framework as follows: "the choice of PF is relative to a particular LF"; "PF "spells-out" LF"; "our interface economy conditions (ScoT) regulate competitions with a common LF and choose among competing PFs to express that LF". They prove (4.2–3) that "the machinery Reinhart [(2005)] posits by starting from the PF representation in fact serves to mask a simpler LF-to-PF account" because the LF-to-PF account "is arguably simpler than Reinhart's, involving a subset of the mechanisms that she needs to invoke, and with fully deterministic derivations that at no point involve backtracking..."

They write in their Summary: "... different PF representations (word orders, in the general case) compete for the realization of a fixed LF, and not the other way around"; "this model stands in conflict with common proposals that inherit (sometimes tacitly) the GB ordering of covert operations after Spell-Out"; "by giving up that commitment, we believe we offer a fresh, and possibly simpler, perspective on these interactions". "In sum, we suggest that we have offered some incremental progress towards a theory of the interaction of word order and scope, but it will take a good deal of further work to see just how far one can make good on the promise of this particular approach."

The model we suggest is obviously one of the possible realizations of the authors' model: the simplest one, or the most radical one:

(11) COMPARISON:
B&W's model: ... → LF → ... → PF

[31] ScoT: If the order of two elements at LF is A»B, the order at PF is A»B. The symbol » represents the canonical manifestation of hierarchical order at the relevant level: roughly scope at LF and linear precedence at PF. That is, scope rigidity is the "most economical" state of affairs.

The model offered in this paper: LF → PF

What we argue for is that the underspecified mechanisms in the course of derivation in B&W's model ("...") simply do not exist in the human language faculty. LF is first not only relative to PF, but LF is the input itself, construed as follows: lexical items built together according to a fixed discourse-semantic content. And there is no non-trivial chain of derivational steps between LF and PF. No derivations compete, but *demands* compete (see (5)), and the "winners" will directly determine the possible intoned sequences of words. Or rather, all demands can "win" simultaneously, at least to a certain extent, because each demand is associated with the rules of its direct and indirect ways of satisfaction in the grammar proposed in this paper.

No word has been said about Bobaljik & Wurmbrand's (2012) method of proving the privileged status of LF-to-PF yet, based on "the ¾ signature". Now we would like to show the application of their method to competing Hungarian word orders / scope hierarchies. Let us consider the grey zone of Table 9 in 3.2: the interaction of two quantifiers is demonstrated there. There are six relevant permutations of word order, as is shown below in (12), and a fixed scope order is considered:

(12) THE $^{5}/_{6}$ SIGNATURE BELONGING TO A PAIR OF QUANTIFIERS (Q1»Q2):

		V ignored I ScoT *MOVE	**V ignored II ScoT *MOVE**	**V considered I ScoT *MOVE**	**V cons'ed II ScoT *MOVE**
V Q1 Q2	✓	0 0	✓ ✓	2 0	* ✓
Q1 V Q2	✓	0 1	✓ *	1 1	* *
Q1 Q2 V	✓	0 2	✓ *	0 2	✓ *
V Q2 Q1	✓	1 0	* ✓	3 0	* ✓
Q2 V Q1	✓	1 1	* *	2 1	* *
Q2 Q1 V	*	1 2	* *	1 2	* *

Out of the six potential PF representations, five are suitable for expressing the scope order fixed. This 5/6 signature has been easily accounted for in an LF-to-PF framework, by assuming that in Hungarian a quantifier can freely choose between expressing its scope to the left of the verb or expressing its argumenthood to the right of the verb. Here two times two makes five as there is no canonical word order to the right of the verb (e.g. É. Kiss & Kiefer 1994) when argumenthood is exhibited by both quantifiers. What is exclusively excluded, correctly, is a word order with both quantifiers exhibiting their operator status in the left periphery but in the incorrect order.

Table (12) above shows potential versions of the PF-to-LF approach. Neither version provides a homogeneous cluster for the five well-formed word orders as the rich variety of numbers, ticks and stars puts on view. We have considered the two conditions considered by Bobaljik & Wurmbrand (2012): Scope Transparency and "Avoid Move". The word order V»Q1»Q2 (in the first row), for instance, satisfies both conditions if V is not considered, but violates ScoT twice if V is also considered. In the fourth column and

the last column the number of violations is not considered. As another example, let us consider the order Q2»Q1»V: it violates both ScoT (once) and *MOVE (twice).

All in all, Q2»Q1»V at PF is correctly predicted to belong to the "worst" expressions of the Q1»Q2 scope order at LF, but it cannot be predicted in this approach that the other five permutations are well-formed, and if there is a preference among them at all, this preference has nothing to do with that suggested by the table in (12).

4.3 Optimality and Optionality

This subsection and the following one offer a comparison of our endeavors to the endeavors of those working in the framework(s) of Optimality Theory (e.g. Prince & Smolensky 1993/2002, Archangeli & Langendoen 1997, Newson 1994, Gáspár 2005[32]).

First, let us consider some ideas of Mark Newson, who is known as the most devoted representative of OT syntax in Hungary. His starting point in his 1994 paper (Newson 1994, 1.2) is that "although there is nothing in OT itself that would suggest that it cannot 'do syntax', the framework is too general to suggest specific ways in which to implement this". The input, for instance, is highly problematic (Section 4): "We also have to assume a notion of a legitimate input, where legitimacy is determined in terms of possible satisfaction of the lexical requirements of the elements contained in the input set". This requirement is trivially satisfied by using our discourse-semantic basis as the input.

A specific problem concerns optionality, which has already been discussed in connection with Table (11) (Newson 1994, 1.2.3): "... it represents the situation where the evaluation allows more than one optimal candidate. This, in itself, is the cause of a potential problem. ... The solution to this problem lies in the possibility of not ranking the relevant constraints with respect to each other. ... [It] is a way of neutralizing the effects of both constraints and thus allowing both sets of phenomena to be optimal." Instead of a simple ordering C1»C2»C3»... of conditions, thus, a significantly more complex ordering is needed, like this one: C1»{C2,C3}»{C4,C5,C6}»C7»...

We would like to take in here an advantage of our approach: optionality (occurring very frequently in languages) is simply the consequence of coincidence of rank numbers—which is a natural state of affairs in our system (see (5)), whose realization requires no complex tool.

What is more, Newson points to a very infelicitous general consequence of permitting non-ordered conditions in OT (p95): "... as long as there remains more than one optimal candidate [in the course of Evaluation], the evaluation of these candidates must continue until ... the bank of constraints is exhausted. ... [thus] each grammatical candidate must satisfy all constraints. What this means is that no constraint can be "hidden" by placing it law in the ranking, which is one possible way of getting rid of inapplicable constraints in

[32] Gáspár (2005) devotes a subsection (1.3.3) to a comparison between my semantics-based totally lexicalist approach and his optimality theoretic aims.

the OT framework. Thus, either all constraints are applicable to all languages, or we assume that languages can differ in terms of the constraints they select. ... there are arguments that the first should be rejected."
In the OT framework, thus, capturing the so frequent optionality in languages requires the double cost of introducing a meta-level of ordering of conditions and weakening an axiom of OT concerning UG, according to which all constraints are of relevance to all languages, and it is their ranking that is the only source of linguistic variation. In our approach, however, optionality is due to a straightforward fact: if ranks belonging to the demands are characterized by a very small set of numbers, these numbers will often coincide. Lack of optionality, thus, would require a thoroughgoing explanation.

4.4 The Input in OT

Heck *et al.* (2002:345/346) also confirm that "the nature of the input in syntax is notoriously unclear", and "it is still an open question what the input consists of in optimality theoretic syntax". We will argue that their radical conclusion—"the concept of input can be dispensed with in optimality theoretic syntax"— practically coincides with the identification of the input with an appropriate discourse-semantic representation.

The authors (3.2) point out that (if there is an input at all) the input has to be highly structured to avoid unwanted competition between candidates having nothing to do with each other semantically. It should consist of "a predicate-argument structure, together with some specifications about tense and aspect", for instance. Moreover, the input should contain information about the target scope positions. Hence, the appropriate input practically contains the "target LF". That is, "two competing candidates must have the same target LF", which is such that "the target LF in the input is encoded by abstract scope markers ... in the output candidates; one then distinguishes between the target LF and the actual LF (the one which is finally interpreted) of each output" (p368).

They also prove (p371) that "syntactic faithfulness constraints can systematically be reanalyzed as constraints on outputs that do not refer to the notion of input. Even though the individual analyses that we considered differ both with respect to their theoretical orientation and the empirical domain that they cover, there is a single unifying reason why an input-free reformulation of the pertinent constraints is possible throughout: Syntactic output representations are richly structured objects that can provide all the information that faithfulness constraints locate in the input".

They argue (p372) that " there is a crucial difference between phonology and syntax [in OT] that forces the existence of an input in the former type of system, but not in the latter: Syntax is information preserving, phonology is not. To make this claim clearer: Syntactic systems can be easily construed in such a way that no information is lost during the derivation. All properties of the input are recoverable from the output. However, in phonology this is not always the case. Consider the prototypical syntactic transformation: movement. If a head's argument is moved away from the head in the syntax, the output-tree still contains the information that the argument was there before movement applied:

It is encoded in the selectional properties of the head (usually, and in many cases redundantly, also in traces)."

Heck *et al.* (2002), thus, point out in the first half of their paper that the input ought to be extremely highly structured to avoid unwanted competition, and raise the question (p362) "Where does the highly structured input come from?" Then they cut the Gordian knot by saying (p394): "All this amounts to the same conclusion: Inputs can be dispensed with in syntax but not in phonology because syntax is information preserving and phonology is not".

We agree with the authors that there is an inevitable inherent redundancy between the input and the LF of the output in standard OT syntax, which can be ceased, among others, by deleting the input (13a):

(13) COMPARISON: a. Heck *et al*'s model / b. the model offered in this paper
a. ~~INPUT=?~~ → CANDIDATE SET → OPTIMAL CANDIDATE → LF
b. INPUT=LF → CANDIDATE SET → OPTIMAL CANDIDATE

It can be said again (see 3.3), however, that this solution (13a) "masks a simpler LF-to-PF account", because this mechanism involves backtracking: the set of competing candidates are determined on the basis of their "would-be" interpretations, encoded in some (quite obscure) way in the competing candidates themselves, which are "information preserving" syntactic trees.

(13b) shows the much simpler alternative this paper offers. The input is nothing else but an entire LF representation. The competing candidates then are not highly complex syntactic trees as information preserving is simply needless. Nor are they intoned sequence of words, however. In our approach the competing candidates are demands associated with the lexical items participating in the LF representation. These demands "compete" in the sense that certain ones will be satisfied by their "direct way of satisfaction" while others by their "indirect way of satisfaction". Hence, the optimal candidate is a system of "satisfactions", which determines a set of intoned sequences of words. They are expected to make the complete set of synonymous sentences belonging to a fixed meaning.

5. Concluding Remarks

This paper argued that the principle of Semantic Economy in the Anticartographic Minimalist Program (4.1) is very near to a semantics-based approach of syntax, which can be regarded as the "most minimalist" model of Language Faculty. Current considerations concerning optimality in minimalist (4.2) and OT (4.3-4) approaches also converge toward a similar model of grammar ("LF-to-PF" / "LF target" in input). We have based the syntax of a model like this on language-specifically ranked universal demands (3.1) capturing the relation between some operator-based discourse-semantic background structure (2.1) and sets of synonymous word-order variants. Hungarian (2.1-3.2) and English (3.3) sentences have illustrated the novel approach to such hot topics of current

generative linguistics as actual and generic readings (2.1-4), operators with straight and inverse scope (3.2), scope transparency (4.2), optimality and optionality (4.3) and the general problem of the input of grammar (4.4). Beyond these theoretical results, the given formal approach to syntax and semantics forms the basis of our current computational implementations.

References

Alberti, Gábor 1997. "Restrictions on the Degree of Referentiality of Arguments in Hungarian Sentences". *Acta Linguistica Hungarica* 44 (3-4). 341–362.
Alberti, Gábor 2001. "Morphological Relations and Idioms in a Totally Lexicalist Grammar". *Argument Structure in Hungarian* ed. by István Kenesei. Akadémiai Kiadó, Budapest, Hungary, 201-235.
Alberti, Gábor 2004. Climbing for Aspect – with no Rucksack. In É. Kiss & Riemsdijk (2004), 253–289.
Alberti, Gábor 2011a. *ℜeALIS: interpretálók a világban, világok az interpretálóban* [ℜeALIS: Interpreters in the World, Worlds in the Interpreter]. Budapest: Akadémiai Kiadó.
Alberti, Gábor 2011b. "ℜeALIS, avagy a szintaxis dekompozíciója [ℜeALIS, or the Decomposition of Syntax]". *Általános Nyelvészeti Tanulmányok* XXIII., 51–98.
Alberti, Gábor, Kata Balogh, Judit Kleiber & Anita Viszket 2002. "Towards a totally lexicalist morphology". In *Approaches to Hungarian* 9. 9–33.
Alberti, Gábor, Kata Balogh, Judit Kleiber & Anita Viszket 2003. "Total Lexicalism and GASGrammars: A Direct Way to Semantics". In *LNCS2588* ed. A. Gelbukh. Berlin: Springer-Verlag. 37-48.
Alberti, Gábor, Márton Károly & Judit Kleiber 2010a. "From Sentences to Scope Relations and Backward". *Natural Language Processing and Cognitive Science* (*Proceedings of* NLPCS 2010 / ICEIS 2010, Madeira, Funchal), ed. by B. Sharp & M. Zock. SciTePress Portugal. 100–111.
Alberti, Gábor & Judit Kleiber 2010b. "The Grammar of ℜeALIS and the Implementation of its Dynamic Interpretation". *Informatica* 34/1 ed. by Y. Ledeneva and G. Sidorov. 103–110. http://www. informatica.si/PDF/34-1/10_Alberti...
Alberti, Gábor & Judit Kleiber 2012. "Where are Possible Worlds? (Arguments for ℜeALIS)". *Acta Linguistica Hungarica*, Vol. 59 (1–2). 3–26.
Alberti, Gábor & Anna Medve 2000. "Focus Constructions and the "Scope–Inversion Puzzle" in Hungarian". *Approaches to Hungarian* 7. Szeged. 93–118.
Alberti, Gábor & László Nőthig 2014. "ℜeALIS 1.1: The Toolbox of Generalized Intensional Truth Evalation". In Jerzy Grzymala-Busse & Ingo Schwab (eds.): *INTELLI 2014, The Third International Conference on Intelligent Systems and Applications*, ed. by Jerzy Grzymala-Busse & Ingo Schwab. 60–66.
Archangeli, Diana & Terence D. Langendoen eds. 1997. *Optimality Theory: an overview*. Blackwell, Oxford.
Asher, Nicholas & Alex Lascarides 2003. *Logics of Conversation*. Cambridge U. Press.
Bobaljik, Jonathan David & Susi Wurmbrand 2012. "Word Order and Scope: Transparent Interfaces and the ¾ Signature". *Linguistic Inquiry* 43.3. 371–421.
Brody, Michael & Anna Szabolcsi 2003. "Overt Scope in Hungarian". *Syntax* 6.1. 19–51.

Chomsky, Noam 1957. *Syntactic Structures.* The Hague, Mouton.
Chomsky, Noam 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
Chomsky, Noam 1986. *Barriers*. MIT Press, Cambridge, Mass.
Chomsky, Noam 1995. *The Minimalist Program*. MIT Press, Cambridge, Mass.
Cinque, Guglielmo 1999. *Adverbs and functional heads: A cross-linguistics perspective*. Oxford: Oxford University Press.
Craenenbroeck, Jeroen van (ed.) 2009. *Alternatives to Cartography.* Berlin: Mouton de Gruyter.
Croft, William 2001. *Radical Construction Grammar. Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.
Dowty, David R., Robert E. Wall & Stanley Peters 1981. *Introduction to Montague Semantics.* Dordrecht: D. Reidel Publishing Company.
Elbourne, Paul D. 2005. *Situations and Individuals*. MIT Press, Cambridge, Mass.
Gáspár, Miklós 2005. *Coordination in Optimality Theory*. PhD dissertation, ELTE BTK, Budapest, http://www.pressflex.com/gasparPhD/GasparThesis.pdf.
Heck, Fabian, Gereon Müller, Ralf Vogel, Silke Fischer, Sten Vikner & Tanja Schmid 2002. "On the nature of the input in optimality theory". *The Linguistic Review* 19. 345–376.
Kamp, Hans, Josef van Genabith & Uwe Reyle 2011. "Discourse Representation Theory". In *Handbook of Philosophical Logic* 15, ed. by D. Gabbay & F. Guenthner. Springer, Berlin. 125–394.
Kiss, É. Katalin 1995. "Többszörös fókusz a magyar mondatszerkezetben [Double focus in Hungarian Sentence Structure]". *A mai magyar nyelv leírásának újabb módszerei II. Néprajz és Nyelvtudomány* **37**. 47-66.
Kiss, É. Katalin 2009. "Is free postverbal order in Hungarian a syntactic or a PF phenomenon?" In *The Sound Pattern of Syntax* ed. by N. Erteschik-Shirand & L. Rochman. Oxford Univ. Press.
Kiss, É. Katalin 2002. *The Syntax of Hungarian.* Cambridge: Cambridge Univ. Press.
Kiss, É. Katalin & Ferenc Kiefer eds. 1994. *The Syntactic Structure of Hungarian, Syntax and Semantics* 27. New York: Academic Press.
Kiss, É. Katalin & Henk van Riemsdijk eds. 2004. *Verb Clusters; A study of Hungarian, German and Dutch*. Linguistics Today 69. Amsterdam / Philadelphia, John Benjamins.
Montague, Richard 1970. "English as a formal language". In Richmond H. Thomason ed. 1974: *Formal Philosophy: Selected Papers of Richard Montague*. New Haven and London: Yale Univ. Press. 188–221.
Newson, Mark 1994. "Negation, Double Negation and Optimality Theory". In *The Even Yearbook 1994*. ELTE SEAS Working Papers in Linguistics. Budapest: ELTE. 87–136.
Nőthig László & Alberti Gábor & Dóla Mónika 2014. "ReALIS1.1". *X. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. Tanács Attila–Varga Viktor–Vincze Veronika. Szeged: SzTE Informatikai Tanszékcsoport. 364–372
Partee, Barbara H. 1996. "The Development of Formal Semantics in Linguistic Theory". In *The Handbook of Contemporary Semantic Theory* ed. by Sh. Lappin. Oxford: Blackwell. 11–38.
Prince, Alan & Paul Smolensky 1993 (ROA Version: 2002). *Optimality theory. Constraint interaction in generative grammar*. Ms., Rutgers University and University of Colorado at Boulder.

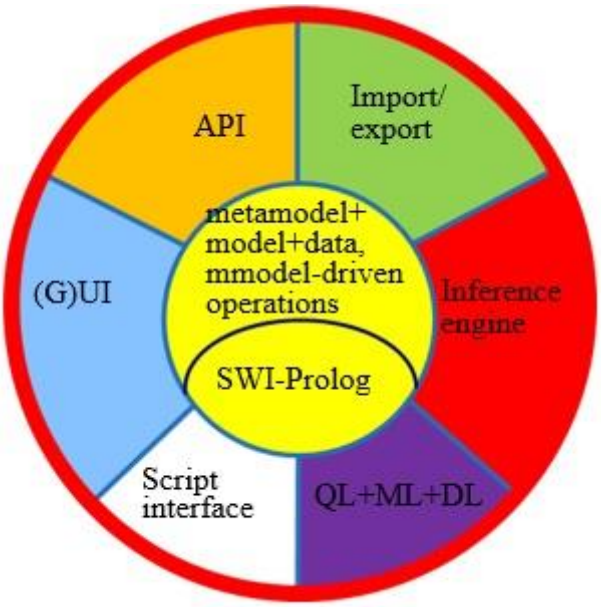
Reinhart, Tanya 2005. *Interface Strategies.* Cambridge, Mass.: MIT Press.
Surányi, Balázs 2011. *Freedom of Word Order and Domains for Movement: A Flexible Syntax of Hungarian.* Acad. dr. diss., RIL HAS.
Szabolcsi, Anna ed. 1997. *Ways of Scope Taking*. SLAP 65. Kluwer, Dordrecht.
Vogel, Carl & Michelle McGillion 2002. "Genericity is Conceptual, not Semantic". In *Proceedings of the Seventh Symposium on Logic and Language* ed. by G. Alberti, K. Balogh & P. Dekker. Univ. of Pécs. 163–172.

ℜEalkb: towards of a semantic platform

Kilián, Imre
University of Pécs, Hungary

Architecture of A SEMANTIC PLATFORM

The general architecture and the most important components of the ℜeALKB semantic platform can be seen in Fig 1.



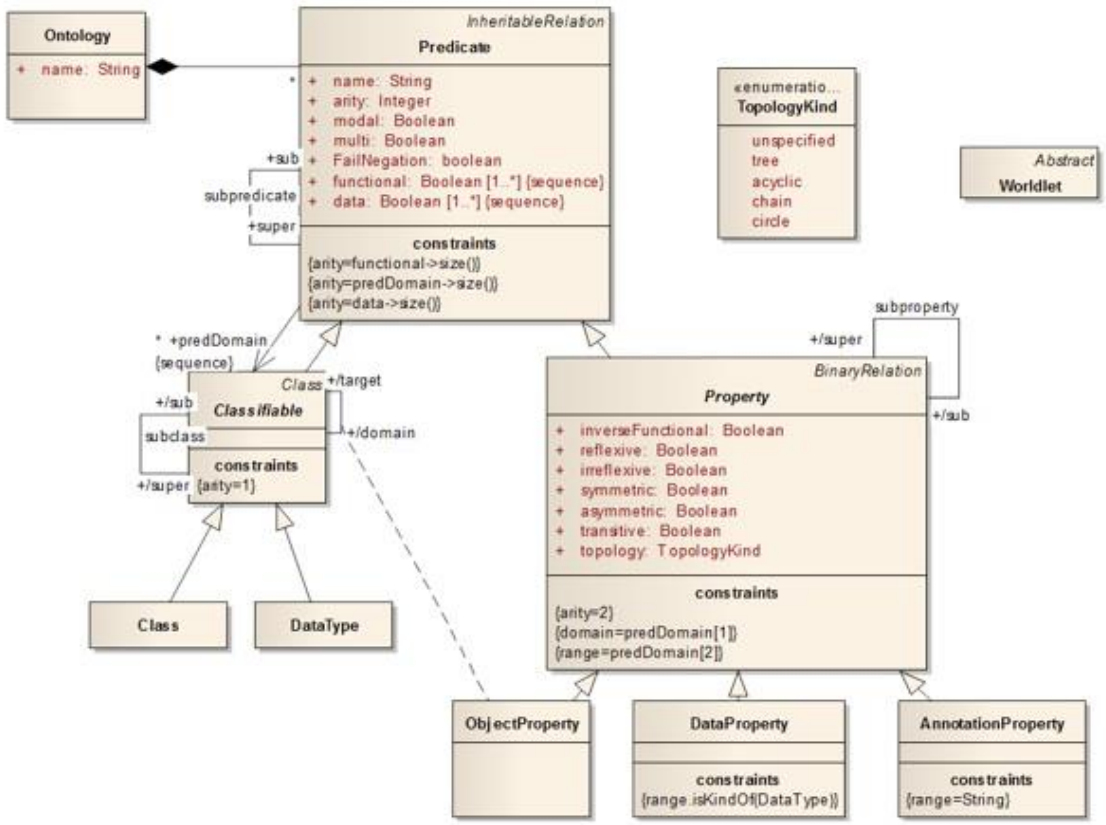
1.Fig. The outline of the architecture

The core of the software is a knowledge repository with SWI-Prolog, as an implementation language. The core of the system implements the metamodel-driven knowledge access operations. The structure of the knowledge repository itself has also two layers: we distinguish data and model layer, and we also map the metamodel itself as a module of the application model. In the design of the system we are following the existing de-facto standards: the structure of the knowledge base is upward compatible with the OWL ontology description language of the Semantic Net project.
The components of ℜeALKB that are built (or only planned) upon the above mentioned core software:

- Programmable interface (API): the core system can be programmed through Prolog procedures.
- We have created a Graphical User Interface (GUI) for the system, which also uses the above mentioned API. The GUI has been created in Java, and uses its Swing GUI package.

- We intend to provide an import/export compatibility with different ontology description languages. At the moment this is realized only for importing OWL, but even this is done in an off-line way.

- Inference engine. There are various inference packages available, which, through a well-defined programmatic interface can be built into our system, or, can be used remotely as a Web service. One basic problem with these is that they apply modeless logic, and they work in the framework of Description Logic (DL), and can therefore be used only for an OWL subset or dialect. In the current status of works we don't have other inference engine Prolog itself.

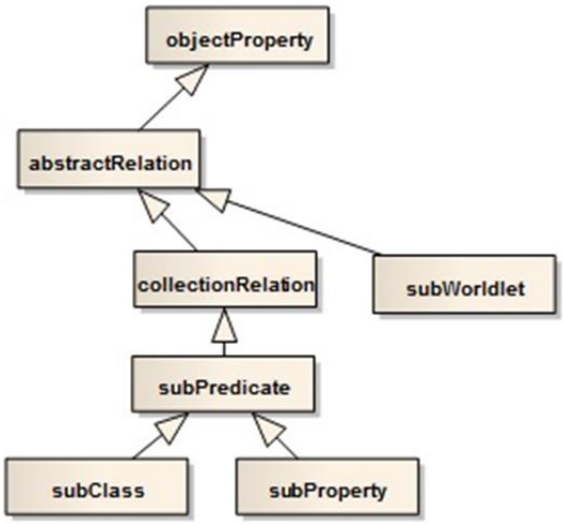


11. Fig. ℜeALM: an OWL compatible metamodel

- *Query (Data Modification/Data Definition) language* would make the access and modification of data easier. Such languages are used primarily for conjunctive queries, but without inference operations. At the moment Prolog itself can also be used as a query/ modification/ definition language.
- For easy and quick programming purposes a *script language* would also be useful. Anyhow, Prolog can be, and in fact is used for such purposes as well.

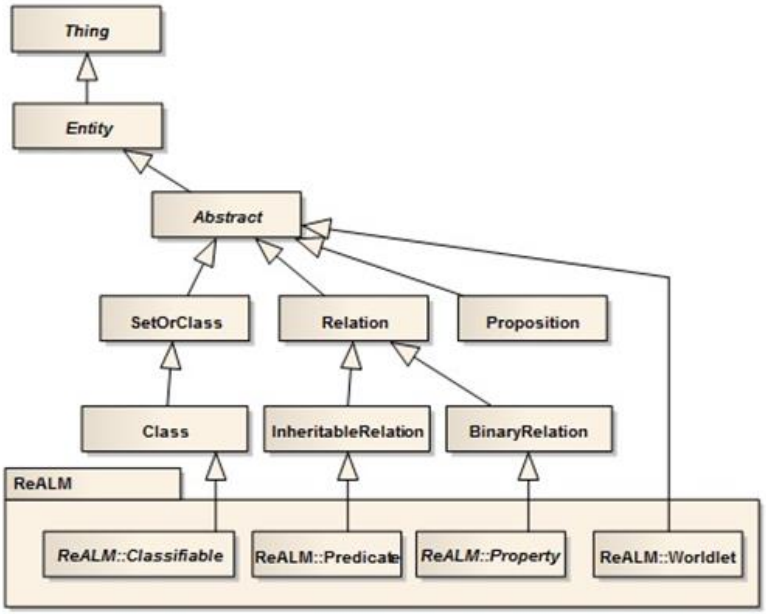
Two-level repository and the compatibility with owl

- The language of the knowledge repository is called ℜeALIS Modal Logical Language (ℜeALM), which is from upward compatible with OWL. By omitting representation and syntax details, it can be defined in the most exact way by its UML metamodel (see below).
- Compared to OWL-DL, the language applies the following extensions:
- Certain knowledge-subsets are organized here in named modules, so called ontologies.
- *Beyond OWL's knowledge elements: concept (here: Classifiable) and Property, we have introduced their superset for describing n-ary relations. This is described by metaclass* Predicate. Classes and properties are sub-metaclasses of predicate.
- We have enhanced predicates, and thus each inherited knowledge elements, by the following meta-attributes:
    - modal: the knowledge-element is understood modally, i.e. it is mapped into the modal worldlet structure. Modeless knowledge elements are not bound to any worldlet, and can be understood as a sort of global knowledge.
    - multi: the knowledge element is understood in a multi-valued logic. There is however no inference over multi-valued logics implemented.
    - failNegation: the negation-as-failure (incomplete negation) operation, as usual in Prolog, can be applied freely for the knowledge element. The complete negation operation is at the moment not implemented.
- We also have extended Property knowledge elements (binary relations) with a new metaproperty. This is an enumeration and is used for graphical presentation: in case of tree and acyclic we provide a tree browsing possibility, in case of chain and cycle this is implemented as a drop-down list.
- For the description of modalities, the metaclass Worldlet is responsible. Its elements may be immediately given identifiers, or the instances of individuals, or their subworldlets, being created for their beliefs, desires and/or intentions (BDI). We are also defining the subWorldlet binary relation over worldlets. This can also be given immediately to model *superindividual worldlets*. Knowledge is *inherited over superindividual worldlets*. On the other hand subWorldlet can be defined also by a Prolog rule over epistemic worldlets of the *supraindividual layer*. A basic difference is that *no knowledge is inherited over epistemic worldlets*.



3. Fig:. *Extending SUMO155 properties with ℜeALM metaproperties*

The target model of ℜeALM is similar to OWL's one: it has two levels that are realized in separate Prolog modules. A-boxes, i.e. the data layer is used to store fact data immediately. T-boxes are stored in a separate Prolog module. It is called the model layer, because the structure of the ontology is stored here.
The model level can also be modelled: the model of the model is the metamodel, and the model or ontology is the instance of the metamodel.



4.Fig. Extending SUMO155 classes with ℜeALM metaclasses

The primary significance and use of the layer structure described above is in connection with first order logic. Namely, in one layer we have only first order logical means, i.e. not even second order questions can be answered. To manage more than first-order logic structures the means of reification is used. If the second order questions are understood as first order questions over first order relations and predicates, then practically there is no need for more than first order (hence undecidable) logical means and algorithms. The OMG's four-level metamodel stack therefore practically separates first order layers, thus helping reification, and enabling a practical escape from the closed world of first order logic.

Crossing metalevel boundaries

Beyond the mere separation of meta-levels, we also intend to enable the cross-cutting of them. We fit therefore the metamodel itself into a region of the application model. So the elements of the application model are in also instances of certain meta-classes that are elements of the model itself. A system with this feature has a self-description ability. Our hypothesis is that in biological and technical systems the ability for evolution is only due to self-describing structures. The biological world, or even the history of modern computation is a perfect proof for that.
Beyond the theoretical point, the metalevel-cut has also practical uses. If we implement, an instance browser in our GUI, we may use the above-mentioned topology metaproperty. This enables the system to browse for example family, or geographical relationships in a tree-driven way. At the same time, classes and relationships as instances of metaclass Classifiable or Property, over the subClassOf and subPropertyOf metaproperties, can also be browsed by the same browser. The only difference lays in parameterisation, and metamodel-elements must be read-only.
Metalevel-cut can be defined only in relation to a concrete ontology. In our case we have downloaded SUMO155 from the internet (Niles and Pease, 2001), and we extended it by the ℜeALM metamodel. The result can be seen in Fig 3. and 4.

Metamodel-driven data access. The Assert-Query-Retract pattern.

In case of relational databases, and the relevant persistence technologies it is widely popular to use the Create-Read-Update-Delete (CRUD) pattern for data access. The CRUD semantics was developed for classical (sequential) programming languages. For Prolog, a non-sequential, non-deterministic programming language, the Assert-Query-Retract semantics (AQR) is used. The basic principles of the AQR semantics are the following:

- For querying already existing knowledge elements, the operation may contain free (unknown) variables. The operation is nondeterministic: it finds the possible values for the variables, binds them and delivers the possible solutions of the query in a backtracking manner. The strategy however follows from the depth-first-like general strategy of Prolog, and it is therefore fixed.

- The assertion of new knowledge elements is a deterministic procedure. The result is logically false, if, due to the violation of any consistency rules, the new instance cannot be asserted.
- The retraction operation may also contain free variables. The operation is non-deterministic: it finds all the possible values for the variables, and as a side effect it deletes the addressed knowledge element.

The AQR pattern is not a brand new strategy. Its relation is quite obvious in case of Create, Read and Delete. Updating however is not possible; asserted knowledge elements are not modifiable any more. A similar result can be reached by a consecutive Retract-Assert operation pair.
The non-deterministic nature of Retract can also be used to delete each matching knowledge element in a backtracking cycle.

Metamodel-driven software and the meta-metamodel

The ℜeALKB knowledge repository, as a software works on ℜeALM logical statements, i.e. on the previously described data and model layers. The most obvious way to implement the repository therefore is to take the model of the model level, and implement it immediately.
However, such direct solutions may have various drawbacks. The solution is not abstract enough: certain algorithms for similar data structure may still be implemented differently. On the other hand it is not flexible enough: a little modification on the metamodel may need rather significant change in the program code itself. It may also be difficult, that any such small change in the metamodel need a complete regeneration and installation of the software.
Instead of the immediate programming of the metamodel, we also apply the two-level software approach. For such a software we depict the metamodel in form of a data structure, and programming efforts are invested to program the model of the metamodel, i.e. the meta-metamodel.

Conclusions, future work

We have built up a metamodel-driven architecture, which enables to create unified and metamodel-independent data accessing algorithms. As a consequence, the metamodel is handled as a data, thus changes in the metamodel do not imply a change in the software itself.
The metamodel-cut in the application ontology enables to manage model elements as instances of certain other model elements. A concrete result of this is the instance browser in the GUI package, and the rule based implementation of static (early) inferences.
One further possibility is to complete the implementation regarding its incomplete packages (inference engine, high level query language, script language). Another very important development could be the integration of probabilistic or even fuzzy logic. One similarly significant further step could be to define and implement thematic and context

dependent inference operations and/or a sort of meta-programming feature that is able to control all these.
From the linguistic side there are a number of studies and/or theses on the modelling issues and the application for language theoretic questions. At the time of finishing the present article, another thesis in informatics tries to transform these theoretical considerations into computer programs, compiling thereby a series of test cases.

Acknowledgements

The authors are grateful for the Hungarian Research Fund TÁMOP 4.2.2.C-11/1/KONV-2012-0005 (Well-being in the informational society) for the support of their research activities.

References

Alberti, G., 2009. *ℜeALIS: An Interpretation System which is Reciprocal and Lifelong*. In Workshop 'Focus on Discourse and Context-Dependence' (16.09.2009, 13.30-14.30 UvA, Amsterdam Center for Language and Comm.), http://www.hum.uva.nl/aclc/events.cfm /C2B8E596-1321-B0BE-6825998CFA642DB2.
Alberti, G., Károly, M., 2012. *Multiple Level of Referents in Information State*. In LNCS7181: 349–362.
Alberti, G., and Kleiber, J., 2012. *Where are Possible Worlds? (Arguments for ℜeALIS)* In *Acta Linguistica Hungarica* 59 (1-2): 3–26.
Károly, M., Alberti, G., 2013. *The Implementation of a ℜeALIS-based Method of Static Intensional Interpretation*. In *KEOD 2013, 5th International Conference on Knowledge Engineering and Ontology Development, Proceedings*, Vilamoura, Portugal, 19-22 September 2013, ed. by Joaquim Filipe and Jan Dietz. pp. 393–398.
Kilián, I., Alberti, G., 2003.: *A Metamodel-Driven Architecture for Generating, Populating and Manipulating "Possible Words" in order to Answer Questions*. In *ICSOFT 2013 8th International Joint Conference on Software Technologies Proceedings*, Reykjavik, Iceland, 29-31 July, 2013. pp. 74-82.
Niles, I., Pease, A. 2001. *Origins of the Standard Upper Merged Ontology: A Proposal for the IEEE Standard Upper Ontology*. In *Proceedings of Measuring Intelligence and Performance of Intelligent Systems Conference*.
Grosof, B. N., Horrocks, I., Volz, R., Decker, S., 2003. *Description Logic Programs: Combining Logic Programs with Description Logic* In **Proceedings of the Twelfth International World Wide Web Conference**, pp. 48-57. ACM.
Ohlbach, H. J. 1988: *A Resolution Calculus for Modal Logic*, Research Report, FB Informatik, University of Kaiserslautern, Germany. (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.5003, accessed: 25-Jun-13.)

Possibilities of refining the ReALIS world and language model and extending its semantic postulate set – from the scope of cognitive and functional linguistics, illustrated by interpretations of some static Hungarian verbs

Károly, M.
University of Pécs (Hungary)

1 Introduction

Precise information extraction, automated text analysis and reliable machine translation are some of the most important and longest-sought goals of computational linguistics. The search for usable algorithms brought formal linguistics as a whole into existence, beginning with the generative syntax of Chomsky, continuing with the branches of formal semantics (Montague 1970, Groenendijk–Stokhof 1990, 1991) as well as discourse theory and the attempts to formalize pragmatics (Dekker 2000, Kamp 2011, Alberti 2011).

But, in parallel to this, along with the sociolinguistical approach, the cognitive paradigm broke through in linguistics, on which functional linguistics is based. In spite of the above-mentioned results of formalization, cognitive linguists often put its effectiveness into question ("That no attempt has yet been made to formalize Cognitive Grammar reflects the judgment that the cost of the requisite simplifications and distortions would greatly outweigh any putative benefits", Langacker 2007). This also involves questioning the possibility of any sort of computer implementation, in which, even if far from complete, cognitive language theory would be used in a consistent way, and it would even be possible that one could recognize some concepts of older paradigms in the structure of the actual program.

It is known that the any attempt to implement Cognitive Grammar would involve even more compromises. So the basic question of the implementation should be reworded: one wants to know whether there is an implementable algorithm based on both formal and cognitive linguistics which could be used for the above-mentioned NLP tasks, or, at least, for creating a computer-assisted NLP system. It has yet to be clarified, exactly where and how Chomsky's "ideal native speaker" should be "weakened" to get an implementable system which is still usable for NLP.

More to say, the development of hardware and software tools gave a boost to implementation efforts concerning some linguistic models (even if most of them were not at all or primarily not rule-based), and, of course, it also inspired the creation of newer, more precise language models, some of which are based on formal linguistics. According to our current knowledge, the well-known complexity limits of deterministic algorithms (e.g. in the case of dependency grammars) cannot be overcome easily; so, assuming a purely rule-based program as starting point, our real medium- or long-term goal is to upgrade our system with corpus-based (i.e. statistical) heuristics. The result

would be a "reversed hybrid" NLP system whose data flow model is sketched below, converting linguistic data to model data: it is not the coverage and/or accuracy of a statistics-based system which would be optimized by applying some (more or less complex) rules but the choice between applicable rules (including stopping the analysis) would be influenced by applying functions based on statistical data extracted from the corpus and/or previous analyses.

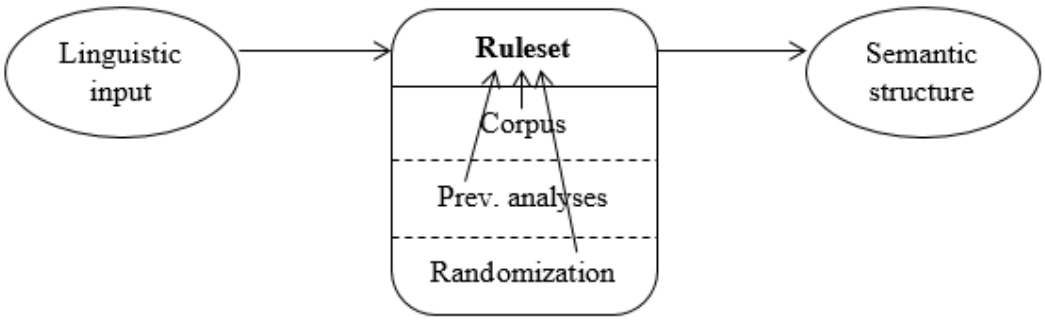


Figure 1: Simplified data-flow model of a "reversed hybrid" NLP system

It is generally thought by cognitive linguists that formal semantics is not compatible with cognitive linguistics. Indeed, even Bańczerowski (1999) states that knowledge representation (KR) still plays an important role among programmers who accept and adopt the cognitive paradigm: KR is used in semantic frames which are considered identical to lexemes and whose contents may be more or less formalized. (Apart from the frame data, contextual background data is also input and assigned to the lexemes.)

Since knowledge representation, knowledge engineering (KE) and ontology development (OD) is a large multi-disciplinary area with many practical applications, the real question seems to be how computers can be used in KR and a functional KR-based NLP system. Moreover, it also seems to be too radical to completely get rid of formal logics because the human brain itself is able to do reasoning and logical thinking. We also think that formal logic can somehow be "fuzzified" to make it able to model more aspects of non-discrete predicates like *love*, *resemble*, *seem* etc.

Originally, as it was explained in Alberti 2011, ReALIS intended to model epistemic modality, evidentiality and interpersonal knowledge. Kugler (2012) shares similar thoughts in her study, looking from the scope of functional linguistics which is largely based on the cognitive paradigm. In this paper I took some examples from her (speech) corpus, too, modeled by me using ReALIS. My goal is to demonstrate some of the most important theoretical problems (with Kugler's and our own examples) which encountered during the project and which are necessary to solve (at least partially) if one intends to develop the ReALIS software any further.

In ReALIS, semantic postulates (formalized with ReALIS' Kamp-style boxes called **worldlets**) play the role of semantic frames while ReALIS' worldlets correspond to mental spaces. Each worldlet has its own parameter list and level label (Alberti 2011:146)

which also describes its position and role in the speaker's mind and which is fairly easy to change if it appears as part of a semantic postulate. (Even the process of metaphorization could be modeled to some extent by applying systematic changes in the semantic postulates while maintaining their integrity.) An open question is the possibility of automatic creation and/or refining of semantic postulates by extracting contextual background data from a corpus.

ReALIS also overcomes the problem of representationalism by putting the worldlets back into the interpreter's mind (Alberti–Károly 2011:380) which is also part of the world by itself.

This paper describes five main problematic areas to which Kugler's examples attracted my attention while I also try to show the way to (at least partially) solve these problems. They are: the problems with ReALIS' data structure and the order its logic, the problem of perception, the problem of static/dynamic dichotomy, the problem of "fuzzy" predicates, and, last, but not least, the problem of ReALIS' contact with "classical" pragmatics illustrated with Kugler's example of subjectification (Langacker 2006) and objective construction. It is known that specific linguistic forms with the same illocutionary force can have different perlocutionary effects; in our last example, the correlation between the degree of objective construction and the actual perlocutionary force is illustrated.

2 A proposed experiment to describe Kugler's (2012) linguistic data with ReALIS

Computational implementation was not among Kugler's goals but her semi-formalized approach to epistemic modality and inferentiality as well as her extensive speech corpus data inspired the author to compare her functional approach to ReALIS, and even to try to reconcile the two approaches. This is also important because, in the last few years, ReALIS seemed to lose its main research object, that was, the language as it was being used – Saussure's *parole*. By analyzing Kugler's examples, the importance of *parole* as the primary linguistic data source is also brought back.

Because of the length of Kugler's study (almost 250 pages), it is not at all possible to model every linguistic element of her corpus. However, it was obvious to choose some linguistic elements concerning the construction of knowledge, assumption and belief because of their importance in the ReALIS model itself (which is largely based on worldlets describing beliefs, desires and intentions).

3 Data structure problems and the order of ReALIS' logic

Kugler's and our own predicates are analyzed the same way as we already did in Károly–Alberti 2013 where extensional and intensional predicates were modeled. But, in general, the results were so incoherent that applying a "program in the program" approach seems to be inevitable: semantic postulates must be built from pre-defined elementary structures by applying construction operations on them in a programming-like manner.

Construction operations may include quantification (example: *pretty*), prototype assignation (*bald*), worldlet transformation (modal and temporal operators, polarity changes/negation), direct (*to love*) and indirect fuzzification (*to resemble*), assembling verbal semantics (assigning starting and cumulative points, taking any of the 5 phases [InPre, AtSta, InCum, AtCum, InRes] into account) etc.

Quantification and indirect fuzzification quickly leads us to the question of the order of ReALIS' logic. "Direct" fuzzification is used for modeling predicates that are fuzzy by nature. Fuzzy logic is necessary for modeling **embodiment** as one of the main pillars of cognitive linguistics while "indirect" fuzzification involves defining at least one Θ: $\mathbf{\Sigma} \rightarrow$ [0,1] function to model a certain non-discrete predicate where $\mathbf{\Sigma}$ is a (in theory arbitrary) set of ReALIS' elements. And, as it is in the case of the third, facultative argument of the verb *hasonlít*(r_1/NOM,r_2/SUB, [r_3/INE])/*to resemble*(r_1/Subj,r_2/DirObj,[r_3/*in*]) which describes the nature of the resemblance and which is, in fact, a property set, the elements of $\mathbf{\Sigma}$ may be eventualities themselves which contain at least one predicate. Although, in most cases, $\mathbf{\Sigma}$ is largely constrained (let the constraint be marked with **E** as a set of possible relevant eventualities), but the order of our logic may increase to the infinity in the unrestricted case. How $\mathbf{\Sigma}$ is constrained for verbs like *hasonlít/resemble*, is crucial. (More to say, one should work with a weighted $\mathbf{\Sigma}$ to get a more accurate model.)

Let P_1 and P_2 be the set of predicates present in the **property-describing** eventualities of some α-ancestor (Alberti 2011:146) or copy of r_1 and r_2 respectively which can be evaluated as "true" at the reference time *t* (The term "property-describing" is not trivial either, but it can be defined by **E**.):

$$P_1 := p^*|\exists e(p,t,r_1)|\sigma(\langle Pred, PropDes/\mathbf{E}\rangle, e) = p, \alpha^*(r_1) = r^*, \alpha^*(t) = t^*, \alpha^*(p) = p^* \; where \; p^*(t^*, r^*) \; is \; evaluated \; as \; 'true' \; (after \; defuzzification).$$

P_2 is defined in a similar way. Moreover, let P_1' és P_2' be P_1 and P_2 constrained by **E** (which means that only predicates from **E**'s eventualities are taken). For the prototypical case of *hasonlít/resemble*, where r_3 stays unmarked, **E** can be derived from the (in most cases, visually, in the case of, for example, two musical works, audibly) perceivable properties of r_1 and r_2. An open question is whether P_1, P_2 and **E** can be constrained to first-order in this case. If yes, most linguistic phenomena can be described by using second-order ReALIS-logic and we shall not need to apply embedded reifications (Alberti–Károly 2011:383, Alberti 2012) to handle higher orders.

Now let us define $n := |P_1' \cap P_2'| - |P_1' \Delta P_2'|, \Theta(P_1, P_2) := 0.5 \; if \; n = 0, \Theta(P_1, P_2) < 0.5 \; if \; n < 0 \; and \; \Theta(P_1, P_2) > 0.5 \; if \; n > 0$. After defuzzification, *hasonlít*(r_1/NOM,r_2/SUB)/*to resemble* (r_1/Subj,r_2/DirObj) without its third argument may be evaluated as *true* if $\Theta(P_1,P_2) > 0.5$. This means that r_1 *is similar to* r_2. For r_1 *is very/a bit similar to* r_2, the threshold level of Θ may be raised to 0.75 or lowered to 0.25 respectively.

4 The problem of perception

The core relation PERCEIVE is an often neglected but still integrant and important part of any ReALIS model. Its importance lies in conveying the information from the outer world to the interpreter's brain. It is often confused with one of the five senses (↖.see etc.), but, more importantly, it is modeled in an oversimplified way which does not take any possible deception and/or distortion into account (see (1)):

(1) Mary thought that a castle is hidden in the forest.
The castle turned out to be a huge oak tree.

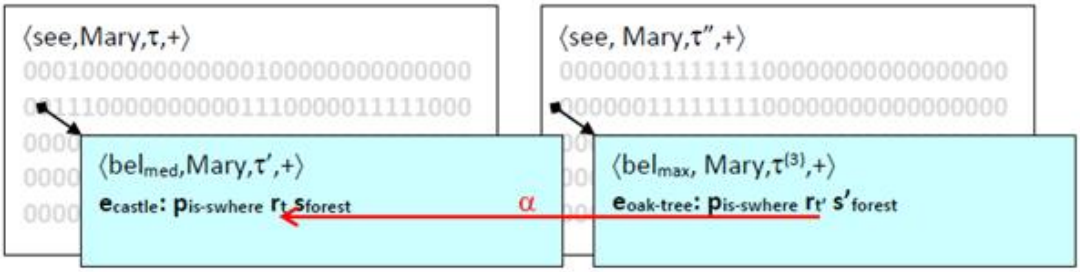


Figure 2: ReALIS model of Mary's mind in case of an illusion

In the "simple" model in which the the core relations of $\mathbf{W_0}$ (the outer world) are just projected into the interpreters' minds (e.g. into the ↖.see worldlet), the core relations must be categorized according to how they can be sensed and perceived as predicates and eventualities. Moreover, the projecting function between ↖.see and ↖.bel (belief) where any distortions could be modeled, is impossible to define, rendering the linguistic phenomenon shown in (1) indescribable.

As shown in Figure 2, a more accurate, but very complex model takes **computer vision** into account with which the cognitive concept of embodiment is also modeled. (The picture is coded with 0's and 1's.) However, if we apply this model, a radically new approach to $\mathbf{W_0}$ is also necessary: large parts of $\mathbf{W_0}$ become superfluous because the entities of the outer world have hardly anything to do with the picture seen by Mary (the interpreter).

The worldlet transformation between ↖.see and ↖.bel is made by the visual cortex of the human brain. One can never perceive directly that a given entity is a *castle* or an *oak tree* but the processor function F: ↖.see → ↖.bel may recognize individual elements of the picture "seen" by a computer camera and (algorithmically) convert them into ReALIS' particular **referents** (here, r_t, $r_{t'}$, s_{forest} and s'_{forest}), just as human interpreters do. This may be done by introducing a "prototypical" *castle, oak tree* and *forest* via RL or similar methods and comparing the picture of the camera to the "prototype".

5 Private fuzzy predicates and degree adverbs

Predicates like *szeret*(r_1/NOM,r_2/ACC)/*to love*(r_1/Subj,r_2/DirObj) are fuzzy by nature. They represent the state and/or changes in the state of the human hormonal and neural system. Unlike *pretty*, *to love* somebody depends only on two people, and as such, it can be regarded as an extensional predicate. Although *love*, too, gets dynamically into existence (*to fall in love with sy*) and, as such, it can be regarded as a result of past events; the state of *love* can be described as a static combination of hormonal levels and the state of the neural system. This, however, cannot be handled discretely (*to love sy very much* is not the same as *to love sy*): a fuzzy metric is needed. So we introduce a third, implicit θ argument (∈[0,1] fuzzy variable), changing the original predicate to *szeret*(r_1/NOM, r_2/ACC, θ)/*to love*(r_1/Subj,r_2/DirObj, θ). Using this, expressions like *to love sy very much* can be truth-evaluated after defuzzification by setting the truth limit of θ higher than 0.75.

One should notice the similarity between *to love, pretty* and *to resemble.* In the last two cases, the role of the θ fuzzy variable is taken over by a Θ: **Σ** → [0,1] fuzzy function. As already mentioned in Chapter 3, **Σ** contains (property-describing) **eventualities** in the case of *to resemble*, while in the case of *pretty*, **Σ** contains interpreters chosen from a relevant set which corresponds to **E** in Chapter 3 (the acquaintance of the subject who is said to be *pretty*), and, if **Σ** contains enough elements of **E**, Θ is higher than a threshold value (for example, 0.5 in the case of *to resemble/be similar to sg*, 0.67 in the case of *pretty* or 0.83 in the case of *very pretty*). In this case, the expression can be interpreted as true.

6 Static and/or dynamic interpretation? The (open) question of presumption and subjective evaluation

Kugler does not write too much about the static or dynamic nature of the predicates collected in her corpus. Indeed, it can be said that Kugler's approach is basically static, so it points out some problematic points in ReALIS' static interpretation which were analyzed in this paper. (Even predicates like *eszébe jut*(DAT,NOM)/*to remember*(Subj,DirObj), *kiderül* (NOM)/*to come to light*(Subj), *rájön*(NOM,SUBL)/*to find out*(Subj,DirObj) (Kugler 2012: 139) lack the cumulative phase so they can be described by using only two static predicates, one for the preparatory phase and one for the result phase.)

The verb *tűnik*(r_1/NOM,e_2/DAT)/~*seem*(r_1/NOM,e_2/*to*) seems to be static. Indeed, it can only be evaluated by bringing back earlier experiences (another set of eventualities) that puts the sense of the differentiation between static and dynamic interpretation into question.

(2) Péter fáradt-nak tűn-ik. 'Peter looks tired' (Peter shows visual signs of fatigue).
Peter tired-DAT seem-s

According to the speaker's experiences, he assumes that Peter is tired because he shows certain (mostly, but in Hungarian *tűnik*, not exclusively visual) signs of fatigue (set **E**' of

eventualities, see below). Peter may be pale, he may not listen any more, he may yawn, his movement may become uncoordinated etc. As in the case of *resemble*, the subject is the stimulus and the utterer's evaluation is reflected in the second argument.

Even though the usage of Hungarian *tűnik* is more constrained that that of English *to seem*, both verbs reflect subjective experience (of fatigue, in our case) which in fact does have some kind of a prototype (set **E** ($\supseteq$**E**′) of eventualities). But, symptoms of being tired are not unique, so ↖.bel (belief) is not necessarily the right worldlet for its eventuality. If somebody is *sick*, s/he may also show symptoms similar to those of fatigue despite the fact s/he is not actually tired.

Here, two new worldlet types are introduced which are not present in Alberti (2011): one for **assumption** (which is similar but not identical to belief) and another one for **expectation** which is "soft" conclusion. Both worldlets reflect Kugler's subjective facts which can be regarded as "true until the opposite (e.g. sickness) is shown". Even what the worldlet ↖.bel on Figure 2 represents is more like a subjective fact (which also has a fuzzy strength argument modeled by some Θ fuzzy function) than a belief because it is based on perception (in that case, sight), and it is not always the case that an assumption turns into a belief (although this happens quite often: ↖.see(**E**′) → ↖.assume(e, Θ(**E**′,**E**)) → ↖.bel$_{\Theta(\mathbf{E}',\mathbf{E})}$(e)).

In the case of (2), the interpreter's background knowledge is accessed: ⇗.back(↖.supp (e_{tired}) → ↖.expect(**E**, θ)) where **E** contains all the symptoms of fatigue as eventualities. Now let be **E**′$\subseteq$**E** (or, to be even more precise, a fuzzy subset of **E**) where **E**′ represents the actually perceived symptoms (and their strength if we work with fuzzy sets). In a more accurate model, a metric has to be defined to describe the "distance" between **E**′ and **E** on which the value of the fuzzy function Θ(**E**′,**E**) depends.

The main problem with this approach is that it is not trivial to unify information originating from multiple sources. In the case of *tűnik*, the source is mainly, but not exclusively visual (↖.see), but, if we introduce ↖.supp (supposition), the actual source cannot be recognized any more. Even the worldlets ↖.supp/cons are problematic because Kugler postulates that the supposition/result dichotomy of inferences is untenable: *I suppose he knew her. Supposedly he was in a hurry.* Even with the verb *to suppose*, both of these examples actually describe **conclusions**, just the source of information is omitted.

Although Kugler mentioned subjective facts, there are legal cases in which (legal) **presumption** is obligatory, and, as such, objective. Even when the judge does not believe what is presumed, he may have no choice but to decide according to the presumption (e.g. presumption of service/delivery in certain cases) if he is obliged to do so by the law. In such cases, the defendant may rebut the presumption, but he has to provide evidence.

The question of presumption and subjective evaluation according to past experience in verbs such as *tűnik/seem* is still open: "past experience" means a very complex and dynamically changing weighted fuzzy set of eventualities (memories) that can be modeled by using neuron networks or other complex structures.

7 Subjectification (Langacker 2006, cited by Kugler 2012:19), objective construction and "classical" pragmatics

It is important to note here that the same illocutionary act can be expressed with multiple linguistic forms. Perlocutionary acts and cultural differences, however, can still be formalized via ReALIS. In Alberti 2011:255, *Throw me that rotten/… banana!* is formalized:

(3) a. $r_{rotten} \sim_{\alpha} r_{banana}$, e_{rotten}: p_{rotten} r_{rotten}, e_{banana}: p_{banana} r_{banana}
b. e_{rotten}: p_{rotten} κ(You)

In (3b), even the principle of co-predication is violated: it is the *addressee* who is insulted by *rotten* and not the banana. A linguistic form without insult would be *Excuse me, could you throw me that banana, please?* Using this form, the addressee will feel that s/he is being respected and politely asked to do something, and, of course, the speaker knows that his/her partner is addressed in a polite manner. The mental state of the addressee may look like $bel_{med}(e_{respect}: p_{respect}\ r_{speaker}\ \kappa(Ego), int_{med}(e_{throw}: p_{throw}\ \kappa(Ego)\ r_{banana}\ r_{speaker}))$.

Objective construction may also play a role in the perlocutionary force. We hypothesize that, if the speaker chooses the more elaborate linguistic form (*I think Peter has got lost*) compared to *Peter might have got lost*, s/he believes to a lesser degree that other interpreters are possessing the same information and/or think the same about the eventuality.

Both sentences realize subjectification, only the degree of objective construction is different. But this difference can also be interpreted as the linguistic representation of the difference between the nature of the source of information (personal or collective, e.g. gossip).

The examples in this chapter could lead us very far away: both demonstrate that choosing the inappropriate linguistic form may lead us to violating the cultural norms. It is also worth mentioning that a simpler linguistic structure is often too direct and rude compared to a more elaborate one because it lacks the honorific/polite expression which emphasizes that the partner is respected and handled equally. If this is not done, our illocutionary act may "go wrong" in the long run because the perlocutionary force of causing the partner to feel remorse and worthlessness is much stronger.

8 Conclusion

Having read parts of Kugler 2012 after Alberti 2011, it can be concluded that, apart from handling second-order logic via reification (Alberti 2012), extending ReALIS' logic with

fuzzy predicates and functions seems to be inevitable. A perfect world model does not exist, only specific linguistic phenomena can be modeled accurately. Perception and subjective evaluation are among the problems which are the hardest to solve. To implement ReALIS, the programmer and/or potential user must specify what exactly s/he wants to do with the system. Although very complex structures (perhaps even neural networks) can be modeled and/or emulated with ReALIS and its worldlets, subjective evaluation and prototype assignment (such as modeling *bald* in a more precise way than in Károly–Alberti 2013) seem to raise many more problems than expected.

Acknowledgements

We are grateful to SROP-4.2.2.C-11/1/KONV-2012-0005 (Well-being in the Information Society) for their contribution to our costs.

References

Alberti, Gábor 2011: *ReALIS [in Hungarian].* Academic Press, Budapest.
Alberti Gábor 2012: Where are Possible Worlds? II. Pegs, DRSs, Worldlets and Reification,
in G. Alberti – J. Kleiber – J. Farkas (eds.): *Vonzásban és változásban [Being Attracted – Being Changed].* Doctoral School of Linguistics, Faculty of Philosophy, University of Pécs. 308–323.
Alberti, Gábor – Károly, Márton 2011: The Implemented Human Interpreter as a Database, in *Proceedings of the International Conference on Knowledge Engineering and Ontology Development.* SciTePress, Funchal, Madeira, Portugal. 379–385.
Bańczerowski, Janusz 1999: A kognitív nyelvészet alapelvei [The basics of cognitive linguistics], in *Magyar Nyelvőr.* 123. 78–87.
Dekker, Paul 2000: Coreference and Representationalism, in K. von Heusinger – U. Egli (eds.): *Reference and Anaphoric Relations.* Kluwer, Dordrecht. 287–310.
Groenendijk, Jeroen – Martin Stokhof 1990: Dynamic Montague Grammar, in László Kálmán – László Pólos (eds.): *Papers from the 2nd Symposium on Logic and Language.* Academic Press, Budapest. 3–48.
Groenendijk, Jeroen – Martin Stokhof 1991: Dynamic Predicate Logic, in *Linguistics and Philosophy 14.* 39–100.
Kamp, Hans – Josef van Genabith – Uwe Reyle 2011: Discourse Representation Theory, in D. Gabbay – F. Guenthner (eds.): *Handbook of Philosophical Logic 15.* Springer Verlag, Berlin, 125–394.
Károly, Márton – Alberti, Gábor 2013: The Implementation of a ReALIS-based Method of Static Intensional Interpretation, in *Proceedings of the International Conference on Knowledge Engineering and Ontology Development.* SciTePress, Funchal, Madeira, Portugal. 393–398.
Kövecses, Zoltán 2005: *A metafora [The metaphor].* TypoTEX, Budapest.
Kugler, Nóra 2012: *Az evidencialitás jelölői a magyarban, különös tekintettel az inferenciális evidenciatípusra [Markers of evidentiality in Hungarian, focused on the inferential evidentiality].* Eötvös Loránd University, Faculty of Philosophy, Budapest.

Langacker, Ronald W. 2006: Subjectification, grammaticization and conceptual archetypes, in Athanasidou – Canakis – Costas (eds.): *Subjectification. Various paths to subjectivity.* Mouton de Gruyter, Berlin – New York, 17–40.
Langacker, Ronald W. 2007: Cognitive Grammar, in Dirk Geeraerts – Hubert Cuyckens (eds.): *The Oxford Handbook of Cognitive Linguistics.* Oxford Univ. Press, 421–462.
Montague, Richard 1970: English as a formal language, in Richmond H. Thomason (ed.): *Formal Philosophy: Selected Papers of Richard Montague.* Yale Univ. Press, New Haven – London 188–221.

New organisation possibilities of network media collections

Mester, T.
University of Pécs (Hungary)

Collections, databases, media repositories – digital archives are part of our everyday environment. The omnipresence of archives is just another proof about the coming of age of ubiquitous computing.[33] Nowadays, our daily routine includes fiddling about our digital files, as we back up, organize, share documents, photos, mp3s, videos. Not surprisingly, digital content management is supported by a plethora of software tools, web services, and mobile applications.

We all have become archivists to certain degree, which process can be hailed democratization of a social praxis (i.e. displaying and distributing cultural content in public), but it also brings about the inflation of archiving as profession (cf. György 2002). Meanwhile, we have begun talking about the archives in a different manner, as they are no longer confined within four walls, do not operate on an institutionalized manner, and the actors connected to them do not have the traditional fixed roles (curator, donor, archivist, contributor, visitor etc.).

In this paper, I will highlight some of the new possibilities of digital archives in social research and new ways of organizing and presenting documents. I would also address some issues of the digitalization processes, and the challenges that new global archives imply. I will describe a new type of digital collection, that I call participation-based community archive. My point of view is that of a researcher, that means I will only touch upon on the theoretic debates involving disciplinary borders, or institutional question. My aim is to present examples, and raise questions about the existing practices and the new possibilities.

Digital cultural heritage

The term the cultural heritage – that became widespread in scholarly and political discourses the past decades – refers to a set of physical and abstract objects viewed as worth passing on to future generations (Lowenthal 1998). It has become a wide-ranging concept that includes artifacts of the physical environment (buildings, memorials, works of art, museum collections), intellectual achievements (texts, inventions, artistic expressions), community practices (local and national traditions, belief systems and values), and natural environment (national parks, protected species, landscapes).[34] Cultural heritage is layered in many senses – the levels of definition and/or presentation

[33] Ubiquitous computing is a term coined by Mark Weiser (1991), often used to describe the process of networked information systems becoming part of our everyday environment.
[34] http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/

(world, national, local), the elements belonging to the popular or the elite domains, the open or restricted properties of access.

In Europe, the preservation of cultural heritage has become a public issue by the millennia. Heritage was more and more often seen as an important resource for economic growth, social integration, and reasserting identities; the concluding policies have been taking the form of regulations, recommendations, directives.[35] The new language of preservation implied that there is or there should be a common practice that regulates collection and exhibition of this legacy that is based upon a social consensus, and executed through formal institutions. The making of a common policy of the preservation process is difficult not only because of the diversity and the layered nature of cultural heritage. If we take a closer look at real-world processes, we often observe professional and public debates (on issues of canon, authenticity, and priorities), legal shortcomings (due to somewhat anachronistic and contradictory regulations), and conflicting considerations about the financial commitment of the state budgets.

Some analyses that take a more critical stance identify this process as production of touristic spectacles in order to make profit by increasing attraction of a certain region or town (Herbert 1995). While heritage industry has become a new branch of economy, its financial contribution to the preservation is often dubious. In the public eye, the preservation issue is often narrowed down to the process of digitalization. Local actors debate who is responsible delivering such tasks (the old memory institutions or newly found committees), how well the money is spent on certain projects. The discussions rarely included the point of integration of digital collections and the importance of availability of the digitalized assets.

Digitalization seems to be a cost-effective way of preservation and publication heritage elements. The digital object can substitute the original, and in some contexts the digital copy (a digital text, a high resolution image, 3D model) is more accessible and manageable than the original. Digital objects, or more precisely authentication, preservation and searchability (Kallinikos – Aaltonen – Martin 2014) has changed the way we think about document management.

Is it important to note, that at the same time very little attention is paid to the born-digital elements of cultural heritage.[36] There are some noticeable exceptions, namely those parts of the artistic domain where technical apparatus is part of the artistic work (i.e. Net Art, Media Art, or electronic music). However, when we think about born-digital culture it is only the tip of the iceberg: the real issue is how the often fragile and non-reproducible items (such as electronic publications, journals, blogs, and shared audio and video data) are selected, preserved and how they are distilled and made part of our common cultural heritage. We might have an illusion that all born-digital cultural elements are automatically preserved, but the case is more complicated. Archiving often limited to

[35] http://ec.europa.eu/research/environment/index_en.cfm?pg=cultural

[36] The problems of archiving born-digital culture are discussed in detail in Dekker (2010.).

making backup copies, and it is also difficult to tell when certain documents achieve "archival state."

Issues of digitalization

Public collections (archives, libraries, museums, galleries) have long focused on providing availability to some, rather than access for everyone. While their main purposes had been defined differently, they share the main principles of collection building and the same archival model (consisting of conservation, organization, authentication and contextualization of the documents gathered). When digitalization had first made its move to the public collections, it was translated as making digital copies of the original collection items. It took long years for the specialist to realize that the whole model was challenged by the digitalization process.

Digitalization is a means and not an end. Successful initiatives do not replicate the original collection, but rearrange, rethink, refashion its content. There is a need for renewal and adjustment of the old memory institutions that deal with collection, preservation, distribution of the cultural heritage in the new media environment (Parry 2005, Cameron – Kenderdine 2007). Slowly, it becomes clear to public actors then significance of digitalization stems not from preservation, but from new ways of accessibility.

Recent critical assessment on the digitalization of public collections in Hungary has found (Berényi 2012) many problems on past and ongoing projects, some of its points can be generalized as below:

- Although the digitalization process of public collections in Europe is going on for about two decades, it is very hard to make an assessment of its state. There process is far from transparent; available numbers (of items, pages, gigabytes) reveal almost nothing about quality, usability, accessibility.
- In most of the cases, collections do not have up-to-date, attractive, and easy-to use interfaces; the actual documents are hidden behind old-fashioned catalogues and databases with obscured structure.
- Digitalization programs often lack complexity; there is no emphasis on long term usability and sustainability of the resulting informational systems. These cannot be explained solely be referring to financial constraints, although they do play a part in unfinished end-products.
- Even the collections that are well-developed and have rich content are operating in isolation, work on one platform only, are difficult to discover, and they are not integrated with other collection of the national and international level. The lack of interoperability is often connected with the lack of available standards of digitalization and content description.
- Although there are many complaints about the infrastructure, human resources are scarcer, there are still few professionals able to design information systems for these purposes.

The Europeana project (commonly called the European Library) has addressed most of these issues, and has a central role in providing guidance and working models of practice. Levels of national involvement may differ greatly, in the post-socialist states of Central Europe the number of participating institutions is lower, and their breath of contribution is slighter. In Hungary, only a handful of public archives updated their digital collections in order to offer a sample to be integrated in Europeana. The centralized state enterprise for the digitalization cultural heritage is still to take off after years on preparation. On the other hand, there are some promising initiatives, namely the international and local projects organized by the Open Society Archives, or the examples of interactive display and publication worked out by the Hungarian Jewish Museum and Archives.

Meanwhile public collections as content providers has to face a new challenger in the form of global archives of digital media (such as YouTube, Flickr, or Google Books) that work on very different principals. The offer 'documents' that are easy to find, and free to use with few constraints; users can easily share, comment, recommend content. Collection organization is based upon loosely defined content descriptions, or rather, on connections the users themselves create. The main agenda of global archives is clearly providing instant access to content.

Traditional and global collection can be easily contrasted. The traditional collections are more structured and homogenous and place-bound, while the networked global archives are more chaotic and heterogeneous. Traditional collections act as a mediator, offering a context for their collection items, global archives merely provide with content without giving much background information to their users.

The most significant difference between the 'old' and the 'new' archives is content organization. While traditional institutions work with their laid out principles of provenance and pertinence, and they use standards of content organization such as

- systems of classification (taxonomies, thesauruses, ontologies);
- name spaces, authority lists, terminologies;
- fixed time-space categories (age, century, decade / place, region);
- rich content description (document format, origin, history).

Global media repositories have their own framework of content organization, which derives from informational technologies (digital asset management). Its principles of content description is based more on common sense than anything else, the tools they use are:

- folksonomy (open set of descriptors: tags, flags, keywords);
- relations in time (timeline, order, concurrence);
- relations in space (place, path, proximity);
- rich feedback (scoring, commentary, correction);
- database structure (re-usable, transferable, remixable).

If you take a closer look at the some of the instances when the public collections go global, you will notice that the two archival paradigms can work together quite well in practice (just think of Google Art Project with numerous distinguished institutes participating, the Flickr photo stream by the Library of Congress, the YouTube channel of BBC archives, etc.). It is also obvious, the seemingly different content organization models can blended in order to prepare collections both for online and real world display.

The arrival of global media archives has also signified new types of media consumption, and changing attitudes towards cultural contents, that were described in the theory of participatory culture by Jenkins and others (Jenkins at al. 2005). This means an engagement with relatively low barriers, a vague idea of community based on informal membership, and the idea that anyone can be a contributor, and all contributions matter however small they could be.[37] One of the many new developments brought about by these changes is participatory archiving. It is used to denote the projects when archives recruit volunteers to perform different archival tasks, most often content description, tagging, georeferencing. In the next section, I would like to focus on the type of collection that is based on this kind of participation, and constitute a middle ground between traditional public and networked global archives.

Participation-based community archives

The community archives based on user cooperation offer their content freely just as the global archives, but their collection are more organized, since there are some rules and norms of their activity (Flinn 2010). It is hard to give a firm definition for them, there are many different examples that are connected by family resemblance. Some of the more common features could be summed up as follows:

- *openness*. anyone can contribute ,
- *participation*: work is motivated by social goods (crowdsourcing / crowdfunding)
- *simplicity*: easy to use, rich in content
- *democratization of access:* made for different groups of users (random guest, visitors, enthusiast, professionals)
- *publicness*: content made public is free to use („public domain" / „some rights reserved")

In order to give an idea of these collections, I will demonstrate these features on some Hungarian examples. Köztérkép [a word mashup of public and map] is one of the earliest public archive based on peer-to-peer contribution, it is a site that maps out open air sculptures. Started out as a personal blog in 2006, an individual's subjective photographic account of his impressions of street public art. When the original founder was joined by a handful of other art enthusiast, they developed their proprietary

[37] Naturally, there is a strong similarity with the dynamics of the participation based online activities in general (Haythornthwaite2011).

database format, and the site started to evolved rapidly, soon it became the largest online inventory of public sculptures in a geography information system.
The operation of the site very similar to the Wikipedia's, anyone can register and make new entries by filling out the form, uploading photos, writing a short description. Members can follow each other's entries, leave a comment suggesting corrections, providing missing data, or simply giving feedback to the original contributors.

During the years the project managed to explore streets and squares of most Hungarian cities, and the information offered is often as meticulous as in any museum catalogue. The site contains more than 22000 entries, has 2000 registered and about 200 active users with a few major contributors entering more than 2000 entries. Köztérkép has remained a civic project without any outside funding, and despite the efforts from members, it has not managed to gain mainstream attention so far. This may be attributed to its forerunner status: the site was started before the web 2.0 and social networking started to bloom, and the current system is still not integrated well with social media. While some innovations such as GIS-integrated database, or an early mobile web edition were ahead of their time, still generated a low buzz even in subcultural circles.

Fortepan[38] is an online photo collection currently providing free access to more than 43000 high resolution images through its website. The starting collection was based on the 5000 images selected from the private collection that the two founders who literary picked them up from the streets of Budapest. The thrown-away images gathered during city clearances were often missing any clues about the time and the place they were taken, the people on the image, and identity of the photographer. After scattered media coverage, the site grew a solid fan base, and the images started to appear on blogs, in seminar papers, were posted social network sites (the founders considered these as public domain, so they were offered free of charge). One of the visitors, without previous knowledge of the owners, started a topic on the largest Hungarian forum system calling for participation in working out the image-puzzles, and replies started to pour instantly. The crowdsourcing was successful, hundreds were willing to spend their time participating in the project, the resulted in thousands of identified images, the data is now included.

The collection was multiplied by donations (institutions, families, inheritors of amateur and professional photographers, anonymous contributors), which signified another motivation for participation: those who lent images were willing to share their personal possessions in order to make them visible for others. The contributions were important for the collection as a whole, as it grew it offered a ground for comparisons.

The original layout has not changed significantly since the beginning. The interface places the individual images in the foreground, while a short description is in a hidden layer. A simple search enables a free text search in the descriptions, but they can be viewed as part of streams (sets can be configured by the donators, based on the

[38] One the founders of Fortepan, a historian working for the Open Society Archives gave several lectures on the story of the collection (Tamási 2011)

geographical names, or timeline), or just random series. The popularity of the collection is closely connected to its attributes, it is easy-to use, simple, intuitive interface, quality of content, and inventive use of the built in automatic image streaming. Fortepan project is still ongoing project showing signs of further development. Nowadays it also a blog (for announcing new contributions), a Wikimedia page (containing the supposed interpretations of images) and a map tool of mapping the geographical location of the scenes portrayed.

My last example is one of the many collaborator mapping projects. Egykor ('once upon a time') a community project that gathers visual imaginary of urban places. The site grew out of a simple image collector community blog, where images of retro objects were posted with a short description. The new site had been started by the same internet community to collect and archive urban images, with the intention to spot timelessness and/or difference of the scenery. The site was re-released in 2011 with the original content, but with more detailed content description possibilities. The main reason behind the rework was the intention of joining Wikitudes, a global places-description network. The site is also more integrated well with other social media sites, shares content with personal blogs and photo collections. Altogether, it is more resourceful and flexible than the usual fixed structure and registration-bound archives. Although the site content is underdeveloped at the moment, it demonstrates the quick adaptation of new technologies into existing archives.

There are many more examples of participation-based community archives that revolve around different themes such as subcultural, genres, types of media. They are easy to be created and maintained, a simple Facebook group can serve the archival purposes. The more successful initiatives have some informal organization, more or less worked-out methods of content description, and some ideas about the possible future of the collection itself. These rules and norms should be flexible enough in order to attract new contributors, so the collections grow as their communities do. Of course, my examples have all have their strengths and weaknesses, and there are no golden standards for types of archives.

Archives of/for research

Digital archives are important for social research from three interconnected perspectives. Urban research makes use of (digital and paper-based) archives as sources for material. Examples show, that digitalization brings about a more manageable process of data collection (digital repositories can be searched, managed and evaluated more easily); and it is also makes the public collections more accessible for everyone.

Obviously, the resources coming from the ever-expanding global archives cannot be ignored by researchers, but they pose their own challenges. Finding relevant information or media content is often hindered because of missing or inaccurate metadata, documents uploaded are fractured or fragmented, parts of certain sets or series are missing. This case can be improved by using methods adapted participatory archiving, crowdsourcing some of the work to the interested public. Community archives, on the

other hand, usually offer a level of organization 'out of the box', with a group of users likely to help the research process all the way.

In most cases, the research itself generates a considerable amount of data that were previously stored in paper or digital folders (often also archived in some way). Quite often large amount of audio and visual data is pulled together, but as a collection it is usually casual, gappy and disrupted, it does not seem to be suitable raw material for a real archive. Some new methodologies offer a way to give new life to these documents gathered though the research; since they are digital, or digitalized, they are also very easy to share across platforms (research sites, blogs, and social networks). These new methodologies (such as digital ethnography) transport the whole research process into the digital domain, making it more visible and transparent for the interested parties. A research archive could serve these purposes well, while it is also a practical tool for group work.

Finally, the research processes usually end with the publication of a number of written texts (closing reports, papers, and volumes), expect there are other outcomes as well. The research could be also (re)presented in real and virtual exhibitions, interactive sites with rich media content, and can be made open for commentary for an extended period via open access digital archives. The motivation behind the creation of such archives is not preservation (usually they contain digital copies and no original items), but publication. That includes sharing with the public, making the content available, encouraging its further use for purposes different form the original intent. When the archiving is done systematically, there is chance that these archives (or parts of them) may be re-used in other archives in the future.

There are many lessons we can learn from the participation-based community archives, these understandings have already started to filter into the practice of traditional institutions (cf. Tiballi 2012). New content management tools and standards are in the works, and there a new wave of open source archiving to be expected. Legal issues may not be settle for a long time, but at least, the conversation about 'fair use' and the 'public domain' have already begun. As researchers we must also understand how communities on social networks can act as resource and an audience for research at the same time, we need to explore new ways of combining online and offline methods.

References

Berényi Marianna (2012) Múzeumaink műtárgyai - gigabájtban mérve In *Múzeumcafé* 6 (27), 48-57.
Cameron, Fiona – Kenderdine, Sarah (2007). Digital Knowledgescapes: Cultural, theoretical, practical, and usage issues facing museum collection databases in a Digital Epoch. In Cameron – Kenderine (eds.) *Theorizing Digital Cultural Heritage: A Critical Discourse (Media in Transition)*. The MIT Press. 165-191-
Dekker, Annet (ed.) (2010) Archive 2020. Sustainable archiving of born digital cultural content. Virtuel platform.

https://www.academia.edu/5994281/Archive2020._Sustainable_Archiving_of_Born-Digital_Cultural_Content (retrieved on 14/11/2014)
Flinn, Andrew (2010) Independent Community Archives and Community-Generated Content:'Writing, Saving and Sharing our Histories.' In *Convergence* 16 (1), 39-51.
György Péter (2002) *Memex. A könyvbe zárt tudás a 21. században*. Budapest: Magvető.
Haythornthwaite, Caroline (2011). Online knowledge crowds and communities. In Echeverria, Javier Alonso – Oiarzbal,Pedro J. (eds.) *Knowledge Communities*. Reno, NV: Center for Basque Studies. 193-210.
Herbert, David T. (ed.) (1995) Heritage, tourism and society. London: Mansell.
Jenkins, Henry – Puroshotma, Ravi – Clinton, Katherine – Weigel, Margaret, – Robison, Alice J. (2005) Confronting the Challenges of Participatory Culture: Media Education for the 21st Century,
http://www.newmedialiteracies.org/wp-content/uploads/pdfs/NMLWhitePaper.pdf. (retrieved on 2/10/2014)
Kallinikos, Jannis – Aaltonen, Aleksi Ville – Marton, Attila (2010) ***A theory of digital objects.*** In *First Monday*, 15 (6).
http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564 (retrieved on 2/10/2014)
Levy, David M. (1994) Fixed or Fluid? Document Stability and New Media. In *European Conference on Hypertext Technology: 1994 Proceedings*. New York: Association for Computing Machinery, 24-31.
Lowenthal, David (1998) *The Heritage Crusade and the Spoils of History*. Cambridge University Press, Cambridge, 127—147.
Parry, Ross (2005) Digital Heritage and the rise of theory in museum computing. In *Museum Management and Curatorship*, 20(4), 333-348.
Tamási Miklós. (2011) The Fortepan Open Archive. In **The archive as project**, edited by Krzysztof Pijarski. Warsaw: Archeologia Fotografii, 310-341.
Tiballi, Thomas (2012) Commercialism, Coercion and Community in Archives (elektronikus dokumentum) University of Glassgow, Humanities Advanced Technologies and Information Institue
http://www.academia.edu/5881362/Commercialism_Coercion_and_Community_in_Archives (retrieved 14/11/2014)
Weiser, Mark (1991) The Computer for the 21stCentury. In *Scientific American* 265(3), 66—75.

Encyclopedia-project – in Berlin. „The art of involvement": a community – cultural experiment

Szijártó, Zs.
University of Pécs (Hungary)

How can we present a transnational, posttraditional, real-virtual, divided and homogenous community truly from „ inside", namely with the active participation of the community members – this issue/question is in the focus of this project.

Although, nearly for a century, the sociological-antropological researches have been dealing with the ethnic communities living in big cities, it seems that this issue still has not lost its attraction and interest[39].

A wide range of new notions has appeared in the past decades (just to refer to the concept of „diaspora community" which has become an separate field of research and has had a great career, or to the theory of „communicational communities" which has also had significant influence), along which a varied research activity - extending to many fields - have been evolved[40].

This project wishes to join these initiations, and, at the same time, is trying to make an attempt to realize this kind of description with the help of the community itself. This kind of self-presentation (based on personal rights) has several reasons - one of which I would like to emphasize. In the past decade Hungary could have faced (again) the phenomenon, that within certain age groups – because of different economic, political and cultural reasons – the migrating activity to European big cities has increased significantly. The image of youngsters spending short or long periods in big Western European cities hoping to get a job, study or find a „better life" has become such a common experience, that it has become perceptible for the whole society.

These intense discourses – generated also by politics, and standing in the focus of media attention – are quite similar to one another in two aspects: the members of these communities (apart from a few exceptions) hardly get involved in them, and most of the time there are the different goals and interests of the home country behind the representational processes. At the same time, it seems that these communities are the most divided concerning generation and culture, and are more varied concerning their motivations, relations to their surroundings, and so they cannot be pushed into the

[39] Lauser, Andrea – Peter Bräunlein: Grenzüberschreitungen, Identitäten. Zu einer Ethnologie der Migration in der Spätmoderne. In: Lauser Andrea – Peter Bräunlein (Hrsg.): *Kea. Zeitschrift für Kulturwissenschaften*. Ethnologie der Migration 10 (1997), 1-17.; Hauschild, Thomas: Vorwort. In: Arbeitskreis Ethnologie und Migration (ArEtMi) e.V. (Hrsg.): *Migration – Bürokratie – Alltag*. Ethnographische Studien im Kontext von Institutionen und Einwanderung. Berlin 2011, 9–18.

[40] Römhild, Regina: Alte Träume, neue Praktiken. Migration und Kosmopolitismus an den Grenzen Europas. In: TRANSIT MIGRATION Forschungsgruppe (Hrsg.): *Turbulente Ränder*. Bielefeld 2007, 211–228.

category systems offered by the representational-systems (just to mention one of the most spectacular ones: the haunting slogan „stay or leave" [41]).

The proposed project is trying to change this situation by counting on the activity of the community members during the visualization of these community/communities (using the statements of the technical books accumulated so far[42]) and is relying on the various inner self-descriptions, self-interpretations when creating its category-system. To achieve these goals – using the opportunities offered by the public media – we would like to launch a community -cultural process based on involvement in Berlin in the next period whose goal is to set up the structure of the „Hungarian encyclopedia" and create its entries.The encyclopedia would incorporate and reveal the different areas of life of the Hungarian communities in Berlin – from the field of economy through the leisure activities to the various artistic activities. We hope that as a result of this project, a kind of (subjective) map of the Hungarian communities in Berlin would be created and the categories and experiences according to which the members of these communities interprete the every day life would be visible.

The emphasized aim of this project is to create a subjective encyclopedia which borrows this form, turns it inside out and queries its most important principles[43]. The preliminary structure of the entries – rather with a temporary feature – would be formed, but the members of the community can alter them, suggest new entries, and can also delete not relevant notions. The structure of the encyclopedia would follow the structure of the traditional antropological technical dictionaries, as its subject it actually the stock-taking and presentation of every day life (objects, customs, places) of the community/communities[44].

The visualization of the continuous negotiational process between the home and foreign culture, and capturing the features of the new social-cultural medium from an inner, (but still) outer perspective would play a significant role in creating and forming the entries.

At the same time, the entries themselves would have a specifically subjective perspective – the members of the community would present the phenomena, which they consider

[41] Schiller, Glick – Ayşe Çağlar – Thaddeus C. Guldbrandsen: Jenseits der „ethnischen Gruppe" als Objekt des Wissens. In: Helmuth Berking (Hrsg.): *Die Macht des Lokalen in einer Welt ohne Grenzen*. Frankfurt a. M. 2006, 105–144;

[42] Römhild, Regina: Welt Raum Frankfurt. In: Sven Bergmann – Regina Römhild (Hrsg.): *global heimat. Ethnographische Recherchen im transnationalen Frankfurt*. Frankfurt a.M. 2003, 7–20; Rolshoven, Johanna: Multilokalität als Lebensweise der Spätmoderne. In: *Schweizerisches Archiv für Volkskunde* 103 (2007), 157–170; Hess, Sabine: Aus der Perspektive der Migration forschen. In: Hess, Sabine - Maria Schwertl (Hrsg.): *München migrantisch – migrantisches München. Ethnographische Erkundungen in globalisierten Lebenswelten*. München 2010, 7–24.

[43] de Vet, Annelys - Bujdosó Attila (szerk.): *Magyarország szubjektív atlasza.* Budapest, Kitchen Budapest, HVG Könyvek 2011.

[44] Rogers, Alisdair – Steven Vertovec: Introduction. In: Rogers, Alisdair – Steven Vertovec (Hrsg.): *The Urban Context. Ethnicity, Social Networks and Situational Analysis*. Oxford und Herndon 1995, 1–34.

relevant, through short, thought to be typical stories, and these would be included in the data base after being edited stylistically. We would like to offer a possibility to process the entries not only in a textual way, but also in a visual way – with the help of different objects, pictures, miniature art-projects. The antropological forms would be „certified" by the different visual elements - „reference" photographs of the informant (the person writing the entry) and short descriptions related to them would be made which would be an integral part of the dictionary.

An internet (intranet) surface would act as the medium of the encyclopedia – at the same time it is important to show the entry samples at the different levels of making them, and publish the final text (considered final text which is extending continuously) in pressed version. It is still an open question whether to consider a monolingual or bilingual dictionary.
The time period is also an open question – creating the first version would be desirable by the end of the year.

The outline of an optional preliminary structure:
a) economy
- workplace (the different stages of getting a job)
- finances (salary, tax, social insurance)
- benefits, state allowances
- connection with the state burocracy

b) social organizations
- social relationships, inteactions (the stories of integration/ isolation)
- associations, institutions
- non-governmental activities

c) lifestyle
- the transformation of the schedule (new festivals, new schedules)
- free time (trips, the appearance of new free time customs
- communicational channels/media
- housing (lodgings)
- use of urban areas (pubs) - food
- objects

Issues of the Archiving of the Underground-Alternative Culture Pécs, 1980-2000

Havasréti, J.
University of Pécs (Hungary)

1.Alternative Culture

Alternative culture is a widely used concept. According to its meaning and etymology it refers to such culture that alters, differs from certain kind of normative order significantly. The normative cultural establishment from which the alternative culture differs, may include many things, but the basic cases are easy to be distinguished.
(1) In dictatorial political relations (for instance the Kádár Era in the 80s) endeavours being different than the *official* (culture accepted and approved by the single party state) shall be considered alternative in a general sense (Havasréti 2006). As well understood, until the end of the Kádár Era the valid cultural policy was built on the triad of prohibiting, tolerating, supporting, according to which the leaders of the party supported the works of art representing the ideology of the system, tolerated those with neutral attitude, who acknowledged the principals of the system, avoided criticism and also affirmation, and prohibited those who criticized the hegemony of MSZMP (HSWP) and the Soviet Union, disregarded the most fundamental political and cultural taboos. However the 'tolerating' category was heterogeneous, was not restricted only to the alternative culture. The alternative culture existed in the cross section of 'prohibited' and 'tolerated': partly based on the institutions of the official culture (clubs, cinemas, community centers, university scenes), party created own, autonomous scenes, primarily private scenes, typically at private homes, that were not affected by the official cultural policy (Havasréti 2006). On one hand as they were not provided with resources (that could have been withdrawn as a punishment), on the other hand as there was no connection between the state and the private scenes (except the state security agents). However the breakthrough and popularity of alternative culture in the eighties contributed to the abolition of 3T (prohibited, tolerated, supported) and the official publicity, also as getting closer to the change of regime, the place of alternative culture between 'prohibited' and 'tolerated' became less and less important.
(2) After the change of the regime, the meaning of alternative culture changed, the political dimension of meaning overlaid during the eighties became mostly part of the past. Mostly, because the people (involved either in creation or in operation) of the 80s, were still key figures in the 90s, and several segments of the alternative culture at that time was considered as the continuation of the eighties. The continuation was significant. However the conception that the alternative culture did not alter from the culture of dictatorial powers, but from the mainstream culture of the democratic, market leader dominance got more significant. This mainstream is represented by consumer culture or any kinds of culture supported by the hegemony of great publishers, museums, chain distributors, and media. The representatives of alternative culture think these kinds of monopolies and hegemonies obstruct cultural innovation: suffocating or distracting

(buy up, almost 'delute') it. Endeavours partly opposing this mainstream as central power, partly seeking their alternatives have significant traditions in the capitalist Western countries (Perneczky 1991). These endevaours often related to certain counter cultural, autonomous, ecological, etc. movements. These scenes are significantly interwoven with the categories and characteristics of participation, innovation, activism, devotion, and relating ideological contents.
In the 80-90s, the breakthrough of alternative culture - including the new wave in the 80s - did not occur only due to opposing the conform nature of the 'official' or 'mainstream' culture. Alternative culture became a trend, determinate part/medium of youth lifestyle(s), found not only in music, films, art, but also in street wear, clothing, and hairdo. Following fashion determined that the local audience was not satisfied with the supply of either mainstream or local cultural services, but preferred to consume the local productions. And finally: alternative culture is often some kind of *marketing category*, a unique market segment, that is created and sold to those (typically educated, qualified young adults with sophisticated preferences) who are not satisfied with commercial and homogenized popular culture.

2.Periodization

It is necessary to explain why this research focuses on these two decades that are divided with the change of regime. The first point of view was to separate the 80s from the previous decade. As the alternative registers of the culture in the Kádár Era were created more or less against or disregarding the official culture policy, the stricter political atmosphere of the 70s did not favour the creation of such activities. From the beginning of the 80s till the end of the decade, due to the constantly softening cultural control and media policy, and to the growing second publicity, the alternative scene became stronger and diverse, and considering some genres-media (film, art, pop music) it entered the mainstream culture. These processes had their positive effects also in Pécs.
Depending on the above mentioned facts, the 80s can be considered as the golden age of Hungarian alternative culture. A unique conglomerate of the cultural productions and scenes was formed that included the 'postmodern' literature published by state publishers, the film art of 'new sensitivity' of Béla Balázs Studio and great film studios, and the fine art of the 'new picturesque' supported officially. In the beginning alternative pop music first distributed on duplicate cassettes appeared besides these scenes, and so did 'underground' fine art and literature excluded from state financial support (K. Horváth – Havasréti 2003; Havasréti 2006). (Still it has to be seen that these alternative scenes and initiatives were often built on or in some way took advantage of the cultural infrastructure held by the state. This circumstance will be a significant point when interpreting the scenes of Pécs.)
It is reasonable to analyze the history of the alternative culture of Pécs going back to the beginning of the 80s because (aside from some exceptions) those connective lines go back to these times that connect the figures of the 90s to the foregoers and the history. From another perspective: those figures and initiatives that would directly-indirectly determine the alternative culture of the 90s appeared in the 80s.
The same reasoning can be followed in terms of the role of the 90s. The cultural boost, moreover blast is very important during the change of the regime. The political barriers

were eliminated, and the economical barriers only appeared later. In these times despite the critical economic situation, artists were provided with the proper resources. During this period the underground-alternative culture of the former period became accessible to the wide audience: banned films were shown, banned books and albums were published. The new cultural-alternative infrastructure is formed concurrently. In Pécs new publishers, journals, bookshops, alternative fine art galleries appear, the local television broadcasts alternative mass programs, and soon the community radio stations of the university start their work. A scene of clubs is created on local level where the significant artists, groups of the era regularly perform. University education covers fine arts, and the Department of Fine Arts is established at the University of Pécs (JPTE later PTE) the students and tutors of which present a significant base of the alternative culture in Pécs (Lengvári – Polyák 2013).

3. Archives

The fundamental problems of archiving the alternative culture of Pécs derive from the dispersion of resource material which is sometimes owned by private collectors. These private collections are also archives and their nature, orderliness, function is strongly connected to the specificity of alternative culture. The most important elements of these is that the alternative culture - by its nature refuses the traditional forms of institutionalization - prefers to create its own, so called 'alternative' archives to record some of its own operation (Perneczky 1991, Galántai – Klaniczay 2013). As according to the democratic conception of archives, the documents of the archives are equal, this form of organization suits the hierarchy-free ideology of alternative culture. However these archives are accessible mostly to a closed group, or members of some subculture, or even only to the owner, thus in these forms they cannot meet the requirements of either the cultural representation and accessibility, or function as some kind of a medium in the cultural recollection. Otherwise the breakthrough of the social media has had its impacts from this point of view, as private websites also can function as certain kinds of archives, furthermore archive owners also tend to operate websites in order to publicize their collection. It often includes some sort of ideology, because the Internet activity is closely related to that *cultural activism* that created the different products of alternative culture (in the pre-Internet era). (Coyer – Dowmunt – Fountain 2007)

What are these archives like? Some of these collections are more or less *organized*, at best including register or catalogue. There are collections *kept together* that might not be systematically organized but the owner aimed to keep the documents together. In many cases this only means that the documents are kept separately in boxes, files, photo albums. Finally, many collections cannot even considered to be archives, only *scattered documents*, hiding among the private records of the owner and the documents of the family. From the point of view of publicity and accessibility, it is a significant problem that researchers do not even know some private archives.

These archives raise several other problems. The first issue is *publicity*. On one hand it is problematic whether researchers know about the existence of certain archives, on the other hand how they can access them. The problem of publicity and accessibility is facilitated by the reluctance of collectors to make their documents accessible for research, firstly because they affectionate about this part of their past embodied by the collection,

and also because this way they should let researchers into their privacy. In addition there is a (mostly righteous) mistrust: fearing the deterioration, loss, incompetent use, or misuse of the collection. Collectors often offer their private collections to some institutes or public collections - that do not intend to accept them due to negligence, lack of space, lack of competence, or just because they do not want to take the responsibility for taking care of and recording the documents. (In terms of reluctance collectors might have high hopes that as time goes by some pieces of their collections may gain high price which can be a reasonable hope as early Hungarian punk fenzins can be purchased in book auctions, and the posters of alternative groups of the 80s are sold for 15-20-30.000 HUF by booksellers.

A further problem is the disorganization and the lack of philological, filmographic, discographic data. The procrastination of data recording makes it even more likely that some of the data will not be detectable or replaceable. Disorganization and scattering are connected to the lack of knowledge, as the owners do not know the significance of their collection or its cultural historical-social historical context. The treatment and conservation reveals several problems as the documents in private collections can be damaged in the course of moving, clearance, etc. However these unfortunate events may occur in professional archiving circumstances, and digitalization does not provide solution for every problem. Because of the fast development of digital technological culture it is sometimes inconveniently difficult to transfer data, documents, texts, images from a data medium used decades ago.

Accordingly the creation of an archive network covering the documents of alternative culture is a multi-stage process. During this process concepts and aspects should be defined - firstly what is considered to be (local) alternative culture and what is the concrete collectible, the actual corpus of the research. Regarding Pécs, a dilemma is whether the *local performance* of artists not from Pécs is part of the alternative culture of Pécs? The answer is definitely yes, as a local scene gave the frame and the audience to these performances, furthermore these performances were related to such diverse contexts (friendships, professional relations, solidarity, common values and aims, etc.) that cannot be separated from the fact of the performance. It is also in question, to what extent can a performer be regarded as part of the alternative culture of Pécs, who first had performed in Pécs but later moved to a different location? It is another issue whether performers, productions coming from the alternative scene of Pécs but later becoming part of the mainstream can be considered to be the part of the *alternative* culture of Pécs? The answer might be a cliché, but these points of view should be considered differently in every case. In my opinion the boundaries should be drawn in a wider area than the local scene, as this way relations, connections, narratives may be depicted that could be invisible in too close and narrow focal points in terms of collection and research.

Finally it is also very important either for the researchers of present project and also for future users of the archive, what interpretation context the local alternative culture and the archive data are placed to. One is a *local historical interest*: the city can get to know, preserve, interpret certain momentum-sequels of its own past. The fact that the underground-alternative groups have been included in the Pécs Lexikon is a sign of this endeavour. The next aspect is *historical-typological* (history of literature, history of art, history of theatre, history of media, etc): How can the greater, national, regional, international scales be found in the life of Pécs? Another important question if the *history*

of institutions: the activities of the alternative culture of Pécs are partly the part of the history of institutions of Pécs, partly they are alternative institutions themselves (smaller, bigger or micro-level) being part of the cultural infrastructure of the city. A significant political aspect: on one hand how the macro-political processes determine the activities of alternative culture, on the other hand how the anatomical, anti-political, casual devices and aspects (ideologies)of cultural opposition and subcultural defiance interweave the scenes and practices of the alternative culture.
Finally from the aspects of cultural theory and aesthetics, alternative culture is a collection place and site of what Paul Willis calls *symbolic creativity* (Willis 1990). From this point of view, alternative culture is the laboratory of such innovative practices that later might play a significant role in a wide variety of mainstream or market-oriented cultural activities. However it is important to note that an archive (in principle) is neutral and free of hierarchy: due to which any kinds of investigations-researches should find their empirical material in it, even if they have not appeared on the horizons of the present yet.

References

Coyer, Kate – Dowmunt, Tony – Fountain, Alan (Eds.)
2007 *The Alternative Media Handbook.* London – New York, Routledge.
Galántai, György – Klaniczay, Júlia (szerk.)
2013 *Artpool. The Experimental Art Archive of East-Central Europe.* Budapest, Artpool.
Havasréti József – K. Horváth Zsolt (szerk.)
2003 *Avantgárd: underground: alternatív.* Budapest – Pécs, PTE BTK – Kijárat - Artpool.
Havasréti József
2006 *Alternatív regiszterek.* Budapest, Typotex.
Kalmár Melinda
1998 *Ennivaló és hozomány. A kora kádárizmus ideológiája.* Budapest, Magvető.
2014 *Történelmi galaxisok vonzásában. Magyarország és a szovjet rendszer.* Budapest, Osiris.
Klaniczay Júlia – Sasvári Edit (Szerk.)
2003 *Törvénytelen avantgárd.* Budapest, Artpool – Balassi.
Lengvári István – Polyák Petra (szerk.)
2013 *A pécsi bölcsészkar bölcsője.* Pécs, PTE BTK Dékáni Hivatal.
Perneczky Géza
1991 *A háló. Alternatív művészeti áramlatok, folyóirat-kiadványaik tükrében.* Budapest, Héttorony Kiadó.
Szőnyei Tamás
2005 *Nyilvántartottak. Titkos szolgák a magyar rock körül.* Budapest, Magyar Narancs – Tihanyi Rév Kiadó.
Willis, Paul
1990 *Jugend-Stile. Zur Aesthetik der gemeinsamen Kultur.* Hamburg, Argument.

Media archives of Pécs – a case study

Doboviczki, A.
University of Pécs (Hungary)

Present case study was made in the framework of Pécs University's research programme„ „Well-being in the informational society" under its „New approaches in organising web-based media archives" subprogramme. During the research cycle lasting nearly one and a half years, I strived to map the institutions and the more relevant institutional persons which and who played a role in the production of media in the city, and, not least, to get an overview of the bodies of media that may be excavated and processed. Since, during the university research process our aim was to create a demo-archive as a pilot project *(Undeground-alternative culture of Pécs in the 1980s and 1990s)*, my own research also focused mainly on these two decades. At the same time, I couldn't disregard the organisations and institutions created in the Seventies or even earlier. The period at the forefront of this examination may be regarded as transitory from various perspectives: on one hand these decades saw the fall of the Communist regime during which fundamental societal and cultural changes took place. It suffices to mention the institutional and political background of the organisations partaking in the production of media in this case. Furthermore, media technology may also be regarded as transitory since predigital, but already transitory analogue image- and video-making processes were replaced by digital technology as the Nineties progressed. The limited technical availability that characterised documentation of the Eighties civil society was replaced by a much wider tendency to create images.

It is not too trivial a statement that in Pécs the processing/excavation of the city's media history, of various bodies of work and broadcast material is in its infancy. In 2012, systematic work begun in Pécs Central Library's Győző Csorba Local History department (Baralib/Digitár) which works on the material of Pécs Lvov-Kertváros's community television (a city part of approx. 30.000 inhabitants living mainly in blocks of flats) community television (Szekfű 1983; 2008). The library has collected all of the broadcasts for archiving purposes so it possesses a significant wealth of material. As of now, approximately 500 broadcasts have been processed which resulted in a filing system similar to that of the Opac system used in libraries and a web surface with objective, detailed descriptions of the recordings with notes on the production team, contents, and those appearing in the videos. However, the quality of this image processing raises further questions: on one hand the quality of VHS material has deteriorated significantly through the years. Furthermore, a large part of the digitised version has suffered from drop-out: the images tend to stall. In the past few months, the server has also become inaccessible.
We have to mention here private initiatives such as Károly Kismányoky's pecskep.hu portal. Kismányoky was chief editor and director of Pécs City Television between 1992 and 1998, so his private collection is substantial. He shares material concerning the cultural-artistic life of the city with regularity.

Research into the recent past has been conducted by the Department of Communication and Media Studies also. In 2005/2006 we studied the archive material of the Hungarian (National) Television's Pécs Regional Studio, mainly VHS and SVHS transcriptions. The studio, founded in 1976 was unfortunately terminated in 2010, and the material stored there were transported to Budapest, to the storage faculty of the Television (Békés 1986; 1996).
Concerning institutional infrastructure we have to mention the Hungarian (National) Radio's Pécs Regional and Minority Studio, which commands an immense wealth of material recorded between 1952 (the founding of the Studio) in various formats and quality until its 2013 closure.

A similarly exciting body of material is stored in the Film Archive at the Baranya Country Cultural Centre. This consists mainly of 16mm film, primarily informational material and educatory films from the State Educational Material Corporation, plus several videos of relevance for local history. Following the closure of this institution the material was housed in various state storage facilities, but currently there are talks with the Győző Csorba Library and the Manda Film Archive concerning the handing over and the professional processing of this body. These videos are mainly unprocessable, so in this case, one cannot talk of „archiving" them.
In the Seventies, new institutional players entered the moving image recording scene in Pécs. In 1974, as one of the workshops of Pécs Youth House, Ixilon Stúdió („Studio Ixilon") was founded under the leadership of artistic director Károly Kismányoky, where members partook in creating animation but also in producing documentaries and experimental films until the Nineties. The body of work created here was scattered, however, some of the creators took their pieces with them and so are in private hands (such as the creations of Tamás Baksa and Csaba Varga). During the time of the institutional changes the remains of the workshop's oeuvre were passed on to Control Studio Film Association. After the closure of Pécs Youth House (later Youth Centre) in 2011, this body of work migrated to Pécs Cultural Centre (the successor of the Youth Centre). Members of the workshop salvaged two boxes of recorded material from here, as they were scheduled to be disposed of.
Pannónia Studio's Pécs Atelier was created in January 1979 by, on one hand the professionals based in Studio Ixilon, and on the other by the famed artist grouping Pécsi Műhely's („Pécs Workshop") visual artists under the leadership of Csaba Varga. This body of work, mainly animation works and experimental films, remains also unprocessed. We may conclude that in all these cases, these bodies of work, these „collections" are just partially known to us, are unprocessed, and in many cases, lie scattered.
The 1980s and the 1990s (especially from 1995 on), were characterised by a heightened activity in cable television and local commercial televisions: the archives of the already mentioned Garden City Community Television, Baranya Television, Mecsek TV and Pécs City Television also remain unexplored and are in a state of disintegration.
An especially timely example is that of Pécs City Television: the studio was closed in 2013, the storing and processing of its material is, in itself problematic. Material recorded before 2000 (several thousand VHS and SVHS cassettes) was stored in an attic in a wholly unfitting way. Several U-matic, Beta, Hi8, Mini, DV, VHS-c casettes were also stored

there, dumped and mixed up in boxes wth hand-written notes regarding their contents. From 2000 on, the studio archived its material in a fairly organised manner due to the switch to digital recording processes and storage solutions. The systemization, and documentation of the recorded material was undertaken by a colleague, László Weisz, using a software- and barcode-based system.

The programmes were broadcast on the Internet also from this time on, and the majority of programmes were published on an online data base. This is not accessible any longer. We also have to mention the archive of the university television, UNIV Tv, some of its post-1993 material is accessible. We find news and reports mainly on university life here.

As we have seen, even in the case of larger, more „serious" institutions there was a tendency for various employees – editors, cameramen, reporters – to „take home" copies or musters of various materials (typically those produced with their co-operation).
This was an understandable reaction: the safekeep of own material, the creation of a portfolio are acceptable motives. Also, these employees felt that during a time of institutional changes and general insecurity, this was the only way to have these in safaty. This is how various recordings of the Hungarian Radio migrated into private hands (to Attila Koszits, former reporter of the Radio), similarly to the video material of Sándor Doktor Community Centre (Károly Wavra, technician of the Centre) the Hungarian Television's Pécs Regional Studio (György Bárány, cameraman), Pécs City Television (Károly Kismányoky and Kálmán Molvay, editors at the station). We may find certain archive material held by successors of some of these institutions; for example the documentations of festivals undertaken by Pécs Cultural Centre are currently in the hands of Zsolnay Heritage Management Public Company. These private archives obviously do not cover the whole spectrum of material of these erstwhile institutions, however, together they make up a significant and unique collection. It is to be noted that in several instances the owners of these archives themselves referred to the problematic nature of their legal standing. We have consulted with legal experts associated with the research – Zsófia Lehoczky and Gábor Polyák have been advising our work in this regard.
As porous the condition, the availability, the grades of prcessing of the materials stemming from institutions is, just as problematic is the presence of civil media – a phenomenon that came alive after democratic change. The city's first and still, to this day only pirate radio, Szubjektív Rádió („Radio Subjective") – made up primarily out of Pécs university students and based on Tilos Rádió in Budapest - was born at the beginning of the Nineties. It was in operation from 1992 to 1994 under this name. Its various mutations – ex- Szubjektív, Publikum, GFM – were the local representatives of university radio up into the Noughties. The archive materials of these – despite the fact GFM offered internet streaming – are accessible only in fragments. Mirror pages of its chat-box are, for example available on the webpage 'archive.org', but the deeper structure of its contents is, naturally not graspable. From private collections (i.e. that of András Pap) however, several valuable material surfaced from the column Papírrozmár („Paper Walrus"; 1996-97), these are reports and recorded broadcasts with demo tapes. The study of this period in radiomaking remains a task to be completed.
Perem Tévé („TV Periphery"), a television station in operation between 1993 and 1995 and operating as an independent station was a refreshing exclusion to this rule. The

crew, made up of university students created a peculiar representation of university culture, the city's cultural life, and shot experimental films and reports on social topics. Perem Tévé is trivial for the reason also that nearly the whole material has been digitised, and a significant part of this precisely notated thank to the efforts of István Szász, one of the directors of the station.
The recordings are of interest not merely due to the bands involved (Cool Dream, PMD – Pécsi Magas Fiúk Dalárdája, Wee Dot, Oktogon, Dr. Watson, TACS, Öregek Otthona…) and due to the musical style, staging or technique used, but also in regards to the venues and the audiences present. Similarly exciting is the way that the conjuncture between university life and alternative culture is represented in various artistic/ visual arts events. László Vidovszky plays the piano with the assistance of Kristóf Weber, performances by Imre Bukta, László Lugossy Lugó, Dénes Hizsynik, and experimental videos by Viktor Gálos and József Lukács. Their artists' portraits feature the youngest artist generation (Barna Leitner, Gabriella Csoszó, László Kotormán), along with interviews of their masters/teachers: Ilona Keserü, Gyula Konkoly, even a a fashion show organised by art students (Pécs Gallery,1994) with the cooperation of Eszter Szabó, Borbák Székely, Márta Csille.

From the Nineties on – due to the easier technical avaialability mentioned above – the role of private documentation is greatly increased. During the study and classification of the archive material of Approach Art Assocation it was illuminating to see how, from the turn of the Millennium the readily available stock of images and videos grew exponentially. This can be traced back to an increased, expanded and available apparatus, and to the effects of digitalisation. (Approach Art Association was founded in 1995, a university seminar undertook to prepare its archive material for further processing in 2012).
There is no well-functioning strategy as to the excavation/exploration of private material – home recordings, private photography-; it is a prevailing trend for people to appropriate social media surfaces to make these available on Youtube, Facebook, on blogs and in tweets. Pécs has several of these online surfaces of mention.
Apart from the above mention pecskep.hu, the portal regipecs deals with the visual/imagery which showcases images from primarily early photography.
User gls shares various historical material in moving and still images. On Facebook we se several groups and the memorial pages of those already passed away; these often have adequate material as regards to the alternative culture of the Eighties and Nineties. The Facebook profile of the front man of the erstwhile Demokratikus Cirkusz (Democratic Circle), Gruppensex („Group Sex"), Péter Hardy, the memorial page of aesthete and art historian Sándor Kiss (aka. Sanya Kis) are to be mentioned here. Imre Schmieder, a technician at the House of Arts and Literature, a sound technician of its events, also possesses a significant body of documentation.
These fora are to be understood as spaces of publication and representation with all the advantages and problems this entails, they are characterised by a subjective selective process and periodical activity, it is difficult to search for names, specifics in them, furthermore one is often faced with inprecise documentation and annotation, etc. Finally, there is no prevailing strategy as to the exploration of these.

A further corpus – one that also aids to contextualise the events of those times, is that of critical life – itself gathering steam from the end of the Eighties. This also represents an unexplored, unprocessed part of the past, but one that is of heightened importance from a musical or visual artistic point of view. It is made up of fanzines, alternative magazines and periodicals or unversity publications close to this subculture (university monthlies, Pécsi Campus, Pécsi Est).
These latter can be regarded as chroniclers of events, as their short reports carried a critical edge also. We need to point out here that András Pap, who was editor of Pécsi Est (a publication in the vein of „Time Out") at the beginning of the Nineties, has made available to our research all issues of the magazine, which is not otherwise accessible in any manner. The periodical Déli Felhő (Southern Cloud) became an important point of reference in the second part of the Nineties (from 1997 on), which published reviews of exhibitions by Approach Art Galelry and Alapzaj projekt (Basic Noise project) and other current visual arts events alongside being a significant forum for musical and cultural criticism. As a counterculture, the alternative/underground scene also tends to, or is, rather, forced to create its own institutions, in this case in media (typical examples of which are: fanzines, and publications/exhibitions in music and visual arts).
We may see from the above that documents of relevance from the point of view of researching the alternative/underground culture of the Eighties and Nineties are very rarely found in so-called institutional media: these events took place hidden from the media operating within the institutional framework or were featured in certain, extreme cases.
However the already existing, mainstream institution were, themselves the arenas of a certain progress: Ixilon Studio and Pannónia Filmstúdió integrated local artists where they produced experimental/one-off films in the second and third generation neo-avantgarde period of Pécs, at the beginning of the Eighties (i.e. Ferenc Ficzek, Károly Kismányoky, Ágnes Pásztor). These bodies of work are only accessible in fragments and their physical condition is constantly deteriorating.

Several years ago, in 1999, the idea of a universtiy course emerged from my friend Péter Lovas's mind, which we then proceeded to hold with PhD student Gyula Lengyel for two semesters. The course – in the framework of a kind of oral history research – undertook interviewing and talking to those who had an active part in the civil/independent cultural life. We strove to include emeblematic figures, often people with a formative effect on the given scene. Our primary aim was to establish a corpus of texts which could serve as a pivot, a basis for cultural and societal researches into the time just after the democratic changes. It soon emerged that the corpuses linked to various events and scenes/places, corpuses that could be visualised (that is: externalised), differ greatly. We also found out that private archives as such had a very porous nature: they were up of fragmented material.

Of the private archives, special examination is due to the material owned by individual bands and scenes of music (clubs, concert venues).
These are bodies of varying depth and complexity, primarily demo cassettes, newspaper clippings, photographs by and of the bands, video documentation, posters, CD and cassette covers. „Varying depth" refers to the fact that these persons had a unique

approach to archiving and documenting the activity of their respective bands or concert venues.

We do find, however, some refreshing exceptions here also, such as that of the late-Nineties band Burzsoá Nyugdíjasok: in their case we managed to process over 10 GB of images, several dozens of video- and audio cassettes (generally recorded concerts and demo tapes), their own fanzine, and posters - all in significant numbers.
In other cases, however, one may mention grave omissions with regularity so that, to form a picture of a venues or a band would entail systematic data collection and a mosaic-like reconstruction of events and trends.
During this one-year research project, the important goal is to bring to life an archiving workshop which is capable, in a technical sense, of playing back, of bringing to life the prevailing formats of media.
Our current archiving culture demands – and this is not meant as an empty phrase – new modi of visibility, and along with it, the visual reconstruction of the period in question.

Bibliography

Békés Sándor: Doing Local Television. In Soós Pál (ed.): *Mass Communication and Regionality.* – Debrecen, MTA DAB, 1985 (101-109)
Békés Sándor: *The first decade: Facts, data and events from the History of the Regional Television Studio: 1976-1986.* Pécs, Baranya M. Művelődési Központ, 1986
Békés Sándor, Bubryák István (szerk.): *Twenty Years:Jubilee Publication of the Regional Studios in Szeged and Pécs.* Budapest, MTV Rt., 1996
Pilkhoffer Mónika (szerk.): *Region or province? Pécs cultural life of the sixties, seventies and eighties: Conference publication, 6-7 June 2005, Pécs, the House of Arts.*. Pécs, Művészetek Háza - Pécs Története Alapítvány, 2006
Szekfű Balázs: Ont the local communication in the beginning of the cable television. *Jel-Kép* 1983/3. (128-133)
Szekfű András: The reception of the cable television broadcast in Pécs-Kertváros, 1983. In u.ő.: *The last Tango of the Shark. Media sociology studies of the semi-hard dictatorship.* Budapest, Gondolat, 2008 (158-162)
Bajomi-Lázár Péter: Hungarian Media History from the late Kadar-regime to the Millennium. Budapest, Akadémiai Kiadó, 2005, 19-53
Tim O'Sullivan – Brian Dutton – Philip Rayner: *Studying the Media: An Introduction.* Budapest, 2003
Briggs, Asa – Burke, Peter: *Social History of the Media: From Gutenberg to the Internet.* Budapest, Napvilág, 2004
Stokes, Jane.: *How to Media & Cultural Studies.*. Gondolat Kiadó – PTE BTK Kommunikáció- és Médiatudományi Tanszék, 2011

Well-being, crowdsourced social mapping and territorial intelligence

Miedes-Ugarte, B. - Moreno Moreno, A. - Sánchez López, C.
University of Huelva (Spain)

1. Multi-scale networks for a sociecological transition.

One of most important effects of recent economic and financial crises in southern Europe, with its impact on the collapse of the most vulnerable social groups and the greater pauperization of the middle classes is that they have expanded the social space of reflection as well as the depth of discussion on the nature of these crises and on the possible alternatives to overcome them. This does not mean that this space did not exist before, in fact, since decades ago many voices from social activism, academic criticism and green political parties had been warning about the huge costs and environmental and social risks of a model of development based on unlimited economic growth. Today this debate has become widespread. From ecological and ethical perspectives, various forums, movements and international, national and local networks have been emphasising the need to undertake a profound socio-ecological transition toward another model of social organization in harmony with the natural environment, more focused on the quality of life, more just and more genuinely democratic.

In the European context different movements, networks and working groups with a multi-scale base of operations are trying to take a proactive approach with regard to these issues. These groups are very diverse and are being formed with different approaches and interests. Some have a more general approach and long-term focus trying to promote changes in behaviour and local relationships that enable society to cope with the socio-ecological crisis (local networks of de-growth as the well known "town in transition" initiatives). Other ones have been created in a more urgent and immediate way. They are constructed as networks of solidarity and mutual support to defend their participants against the institutional collapse in the context of crisis. Some groups are connected with other movements at the regional level, national or international and others are much more local nature.

However, despite of their differences in the motivations, composition or modes of organization, when different groups operate on the same territory, their interests, and often the specific individuals that are part of local movements and interaction platforms, overlap. This interaction gives rise to a territorial network of networks of collaboration and mutual learning.

Here emerges a new "awareness of place", the awareness of a community on how the place, understood as collective heritage, ensures the biological and social reproduction of the community itself (Magnaghi, 2010) as well as new forms of "territorial intelligence", production and management of knowledge and construction of the collective intelligence in the territory (Girardot, 2010). This new "awareness of place" and this "territorial intelligence" are associated to a new social economy, more diversified, which combines traditional formulas with other modes of informal economic organization. This new social economy becomes a tool for the social factory for the construction of new lifestyles and local new identities by promoting cooperation and the reconstruction of the links in all the scales to meet the socioecological challenges.

2. The co-construction of territorial intelligence.

In this context, one of major challenges for those involved, both from a research point of view as well as of the action, in these processes is to try to identify which kind of knowledge rises from local, formal and informal networks that are being currently developed on the ground and what capabilities it is necessary to create for these networks to empower and to work effectively in their territories.

An important part of the defies have to do to do with the idea of enhancing capabilities to enable the stakeholders to work together by building networks of trust (Senet, 2012): methods and tools to generate collaborative processes of analysis of their individual and collective problems; tools to build up common knowledge and languages allowing to deliberate effectively on the proposed solutions; design of frameworks of cooperation at all levels to establish and carry out the agreements; methods for a common sustainable management of shared resources and so on.

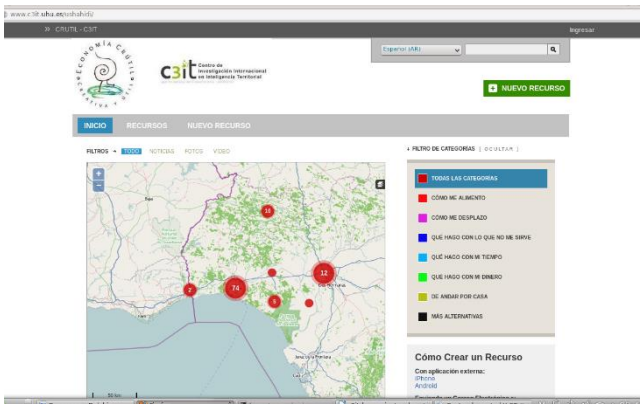
In a very general way, the question for those who work in the development of territorial intelligence is how to strengthen these processes of transformation and how to put the current potential of technological knowledge under the control of the citizenship. An important issue is how to put information and communication technologies (not just the digital ones, although these occupy a predominant role) at the service of production and communication of knowledge of actors who share projects in a territory. A question that emerges is how the co-construction of knowledge has an impact on their awareness and shared responsibility, as well as in the recovery of the common ties, especially in the areas of territorial proximity, in which the exchanges that determine the quality of life occur. In short, methods and tools of territorial intelligence seek to contribute to the construction of spaces of empowerment that enable innovative socio-ecological actors to identify their needs, to advocate their projects, to evaluate their actions, to make their speeches, to strengthen and amplify your voice, and with all this to increase their power of resistance, confrontation and proposition in the context of globalization. This is the

main purpose of this collaborative tool Map of the "Economy Crutil" that we present in this work.

3. Description of the “economía crútil” tool.

The map of the Economy "Crutil" (creative and useful) was created by a group of social activists in the city of Huelva (Spain, 165,000 inhabitants,). These activists were students in a graduate course on social entrepreneurship and solidarity of the University of Huelva, in collaboration with the C3IT (International Research Centre in territorial intelligence[45]) of this university.

One of the main objectives of the C3IT is the development, adaptation and dissemination of tools of mutualisation of the information in order to foster the territorial knowledge management and the development of collective intelligence in favour of sustainable development. The map of the Crutil Economy is a clear example of the potential of this type of processes.



The original idea was to develop a repertoire of local economic activities, creative and useful economic activities to illustrate clearly the existence of individual and collective endeavours pioneering a socially more supportive, more just and more respectful of the natural environment model of development.

In collaboration with the C3IT this initial repertory has been implemented in an interactive cartographic tool called Ushahidi (http://www.ushahidi.com/) that allows users to register a database of actors/actions, and at the same time create a temporary file and geospatial records. As explains Victor Miclovich, member of the Ushahidi team, in the presentation in your web is "a platform on which the user generate content". Citizens can send information using Apps, SMS, e-mail, Twitter or a form embedded in the web page. After, the information is geocoded and displayed via OpenLayers, a

[45] www.c3it.uhu.es

JavaScript library to load maps from Google Maps, Bing and OpenStreetMap Maps. Previously the data are reviewed and rated by managers before appear on the map. The software is open source and can be freely adapted.
The dynamics of the tool is that any person can upload information to the platform on any activity that is considered "crutil", however, for an initiative or project to get the Crutil Seal and to be included in the map it must meet the criteria of transparency and justice as well as one of the principles of ecological, local or minimum waste.
On the other hand, the information is organized in the face of those users, both those that include the information in the map, such as those that use it merely to obtain information, classified or looking for information on the basis of the basic activities that make up their lifestyles[46].

The influence on the global operators (the ideology of market and its main driver, consumerism, as well as educational systems, mass media and more recently social networks) to guide the lifestyles toward more sustainable modes are necessary, but are insufficient if we don't act on the local facilitators to enable the people really can change their behaviours in a meaningful way. Therefore, the local, the space of territorial proximity, becomes a space of crucial importance for the management of global tensions.

These elements are on the basis of the "Map of the Crutil Economy" whose creators are very aware that the way that adopt these styles of life -what people eat, drink, buy, conmmute, climatize their households - determine the metabolic profile of our societies. Hence, the aim of the map is to put the emphasis on the implementation of behaviors and actions to facilitate the change in these styles of life and this not only through actions to raise awareness, but, more importantly through the development of local facilitators (meeting spaces, data banks, mappings, action groups, …) to boost alternative lifestyles.

4. Results.

The ultimate aim of the tool is the generation and maintenance of networks of trust. It follows that as important is the information that contains the map as the reliability of those who validate the information to be included on it. That is the reason why after the elaboration of the initial repertoire of activities, both the promoter group of students and the members of C3IT agreed that the tool should be managed by a sufficiently representative organization in order the users were able to rely on the validity and reliability of the map.

[46] In Europe, for example, 70-80% of the environmental impact comes from food and households comsume, only the meat and dairy products account for 24 %. The consumption of heating and the domestic water represents 40% of the total energy spent. With regard to the modes of transport, the number of cars has increased 35% between 1990 and 2007. A third of the cars in the world are in the EU. On the other hand, 60% of the largest and 20% of school-age children are obese which is correlated with the development of coronary heart disease, which are the major cause of death in the EU (Breukers et al., 2012:2-4).

After a long process of transfer, finally the Network of Alternative Social and Solidary Economy of Huelva (NASSEH), which has been recently promoted by some of the designers of the tool, will be the responsible for validating the information that appears on the map. This has been seen to be absolutely essential to build confidence around the activities included in the map, because in the long term it is intended that this tool will promote not only the virtual interaction of the entities in the social networks, but also its physical interaction in the constitution of social markets, both from the individual (purchase-sale) and from the collective (cooperation, crowfunding) point of view.

Currently the map account with 192 resources, 85 in the city and the rest are distributed in other populations of the same province as well as in neighbouring provinces. We hope that in the short term the map properly to streamline both the process of collaborative knowledge construction around the territorial economy "Crutil" as well as the actions for its development.

Notwithstanding, the main limitation of this tool is that requires time for its technical maintenance and for the organization of the validation process. It is not completely free, which is a problem for the group's promoters that do not always have access to resources. Hence, the need for these processes to be articulate with entities and public research centres that can support them technically and organizationally.

5. Conclusions.

As it has been already mentioned, for the territories to have any guarantee of success when transiting toward more viable, more fair and more liveable socio-ecological models, the actors with decision-making capacity in the territory, in complicity with the citizenship (more than with transnational financial interests), and with other territories, will have to find ways to win decision space in the global framework. The social innovation to revolutionize the lifestyles and the adaptation of the territorial systems of governance have become key factors and for this reason it is necessary to work in the capacity building (knowledge and powers) of the involved actors.

The road is not free of difficulties; to translate the collaborative processes in the construction of knowledge into coordination and collaboration actions promoting social and ecological wellbeing is not an automatic process. It depends primarily on actor's determinations, on the generation of spaces for cooperation and good use of the available resources. Today we have more knowledge and more instruments than ever before. We have a responsibility to use them.

Bibliography.

- Breukers, S. et al (2012): "Sustainable Lyfestyles. Todays Facts and Tomorrow's Trends", Deliverable 1. del proyecto SPREAD, Sustainable Lifestiles 2050 (European Commission, 7th Rtd Framework Programm), disponible en [http://www.sustainable-lifestyles.eu]
- caENTI, Quality Letter of Action-Research favouring Territorial Governance of Sustainable Development, Deliverable n. 47, www.territorial-intelligence.eu/caenti, dec. 2007.
- Carpintero, O. y Riechmann, J. (2013): "Pensar la transición: enseñanzas y estrategias económico-ecológicas". Economía Crítica, Nº 16, Segundo Semestre, 45-107.
- Fischer-Kowalski M. y Haberl, H. (eds) (2007): *Socioecological transitions and global change. Trajectories of social metabolism and land use*. Edward Elgar, Cheltenham.
- Gallopín, G. C. (2006). "Linkages between vulnerability, resilience, and adaptive capacity" *Global Environmental Change*, Volume 16, Issue 3, August 2006, Pages 293–303.
- Girardot, J-J. (2010): "Inteligencia territorial y transición socio-ecológica", *Trabajo*, Num.23, pp. 11-24.
- Magnaghi, A. (2011): El Proyecto Local, Arquitectonics, Mind, Land&Society, UPC, BarcelonaTECH, Barcelona.
- Miedes B. (2009): "Territorial Intelligence: Towards A New Alliance Between Sciences and Society in Favour of Sustainable Development", RES *Ricerca E Sviluppo Per Le Politiche Social* IVol 1, pp-105-116.
- Senet, R. (2012): *Together: The Rituals, Pleasures, and Politics of Cooperation*, Yale University Press, New Haven.

Virtual space serving territory and interaction

Ács, P – Filó Cs. – Wilhelm Z.
University of Pécs (Hungary)

Background

The multidisciplinary Research Team on Territorial Development and Information & Communication Technologies has been studying the scientific collaboration in territorial development more than 10 years. The research integrated to number of international projects involving partners from the INTI (International Networks of Territorial Intelligence) networks.

This research has to be connected to the practice. The base of research and also the result is also relation to the organization of the society. In term of territory it is also true. The network is not only a network of researchers, but network of territorial actors. The INTISC (International Networks of Territorial Intelligence Service Cloud) register offer possibilities to join to an international network, to make transparent the research areas, and best practices connected to territory, territorial knowledge, intelligence territory. But this is not just a directory of people and projects, but a framework of access web based territory and project specific applications, methods and databases (Service Cloud). In the same way INTISC project describes how to, and promote to build and offer web based thematic services for the INTISC society. More data we have made available about our projects raises the opportunities finding partners involved in similar fields from wide world. Now we are working on a researcher and actor registration framework, and the authentication, and authorization service for specific applications accepting it.

The aim of project is facilitating the long-term sustainability and efficient operation of a research infrastructures, which will help to respond to grand challenges in science and society European level. In addition, the next generation of new research infrastructures can be identified through design studies.

Research infrastructures play an increasing role in the advancement of knowledge and technology and their exploitation. By offering high quality research services to users from different countries, by attracting young people to science and by networking facilities, research infrastructures help structuring the scientific community and play a key role in the construction of an efficient research and innovation environment. Because of their ability to assemble a 'critical mass' of people, knowledge and investment, they contribute to national, regional and European economic development. Research infrastructures are also key in helping Europe lead a global movement towards open, interconnected, data-driven and computer-intensive science and engineering.
European researchers need effective and convenient access to the best research infrastructures in order to conduct research for the advancement of knowledge and technology. The aim of this action is to bring together, integrate and open up key

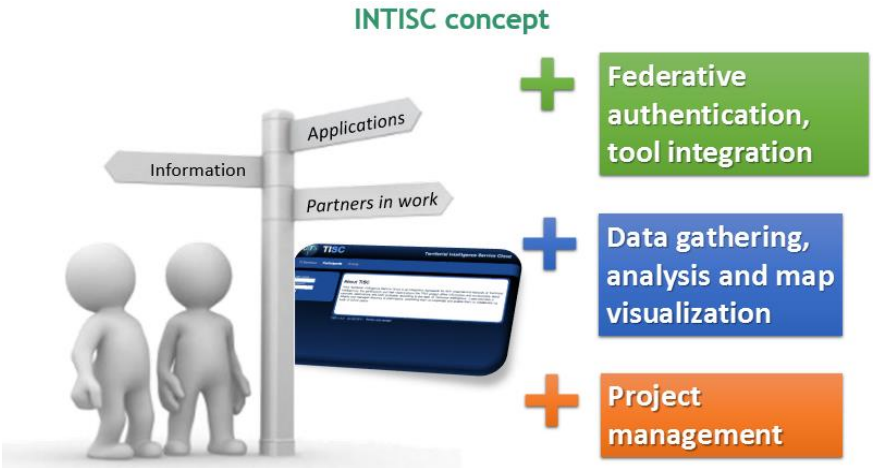
national and regional research infrastructures to all researchers, from both academia and industry, ensuring their optimal use and joint development.

This virtual environment may target any area of science and technology, especially interdisciplinary ones, including ICT and social sciences and humanities. Focusing on the ICT infrastructures needed for addressing the Societal Challenges is especially encouraged.

This regard, extend and spread the INTISC [International Network of Territorial Intelligence Service Cloud] (research and management) applications at the European level.

The multidisciplinary research group of University of Pécs has been organizing the scientific collaboration in territorial development for more years. The research integrated a number of international projects involving partners from the ENTI, INTI (International Network of Territorial Intelligence) networks. University of Pécs concentrates on gathering, collecting and registering and gaining best practices. The knowledge which was gained and cumulated in the research centers should and can be available and usable to be ready for real territorial action.

This concept strongly connects the actors of the territories and participants involved in the research. The wide variety of fields in research and action entail the international and interdisciplinary characteristics of the project. Geography, social-geography, communication studies, sociology, information technology are just the topmost of the fields emerged. The current project aims to realize and spread an IT application which can help integrate the existing applications in an appropriate level enabling higher responsiveness and collaboration in the involved research and action communities. In society level University of Pécs aims development of non-material resources: Positive filings, Self-esteem, Optimism, Resilience, Competence, Autonomy, Engagement, Purpose, Trust and Belonging. The project is based on the work and experience we have from the previous years of cooperation.



1. *Figure. Concept of INTISC application*

The territorial intelligence aims at providing tools made for, with and by territorial actors and other stakeholders that allow them to elaborate, to manage and to evaluate partnership and participative projects of territories sustainable development.

Territorial intelligence, as it is defined by our research group, should also be able to design forms of promotion of the available resources development that should even increase the historical-environmental heritage, but should also conform to a culture which main objective is to increase public awareness of immaterial needs, apprehended as transnational ones.

One of these aspects, in the post-industrial societies, concerns the development concept. Indeed, we cannot define it in a unique way as a form of industrial production at a large scale, as it was done until a recent moment. This new concept, and here is the change, implies the convergence of the different forms of non-material resources, and particularly of human resources. For example culture that chooses non material values, especially concerning the artistic heritage and landscapes, feed the tourism sectors multiple dimensions.

Nevertheless, we do not consider economic development is useless and it is not necessary to keep the organizations that had an impact on modernity, by making it integrate the importance of immaterial goods. However, we would like to underline the contribution of the immaterial resources to general development. In a scientific and formal way, we want to define the limit between sustainable development in a given socio-cultural area and the mere economic development.

The development of the cultural, traditional value resources available on a territory should allow selecting and promoting those that meet a heterogeneous people needs, through the global technology. At the same time, it allows the differences not becoming an obstacle to these needs affirmation, but on the contrary underlining a territory heritage.

Territorial intelligence should conciliate the post-material values with those of the industrial society culture, by supporting the territories resources development. This discipline recognizes the latter implicit qualities and uniqueness and makes their use attractive for the heterogeneous local societies.
This kind of environment, which main objective is to increase public awareness of development, should allow emancipating people from a vision only focused on available resources, by promoting the traditions that represent a territory peculiarities. Besides, if they consider the multiplicity of territories that make up a society, the latter should allow the expression of differences, which allow progressing.
As a consequence, this action has a particular importance among the territorial intelligence competences. It favours a global development of the specific culture, which underlines the differences between the external and endogenous elements of a territory and does not give in to exclusion and separation temptations.

Traditionally Territorial Intelligence has been fed on economics, geography, Information and Communication Sciences and Technologies (ICST) and knowledge management. The links with economic intelligence and the ICST are often quoted in the current definitions of territorial intelligence. The systems of territorial intelligence require the use of traditional processes of information broadcasting and of information technologies and communication by the means of Intranet or Internet sites, documentation, and geographical information systems and data analysis.

Why we do the development?

From the experience of the last decade in scientific cooperation it seems to be obvious, that one of the key factor which enables efficiency is the appropriate information technology support of collaboration. Our project is an interdisciplinary project. It includes not only the strictly technological aspect, but the cultural questions as well. In a multicultural, and spatially distributed environment we have to coordinate the collaborative actions. We has rather less occasion to have feedback compare to a usual everyday work. The people has to understand clearly what to do and why, what is the current state of the questions we are working on. The researchers and territorial actors has their own projects, which has collaborative aspects, but we have to face with that is something additional for them. When somebody getting involved an international collaboration there are a lot of questions, not always easy to answer:

- Who are the participants?
- What are their interests?
- Where they are?
- What is task to do?
- How it fits to the general principles of collaboration?
- What is the current state of problems we are discussing on?
- How can I contribute to collaboration currently from my own part?
- What are the decisions have been made?
- Is there any change in time scale of the project?
- Who to contact according to a specific problem? How can access the common information resources?
- How can I access the tools of cooperation and data processing?

These questions are arising regularly. We are working on a system aims to support the scientific collaboration process with information and application services. We are going to introduce it a specific field of research and action of Territorial Intelligence, but it could be useful experience for other fields as well. The INTISC application (International Network of Territorial Intelligence Service Cloud) is and it based but also cultural solution for participants. It is about maintaining a network of research and other institutions involved in intelligence territory. This support is tends form the Who is Who question to a support of common research data gathering.

International Network of Territorial Intelligence Service Cloud (INTISC) is like a crossway with signposts showing clearly what is available for use.

There are two important aspect of the system: Participants and applications accessible. It is a directory of participant persons (researchers and actors), institutions, researches and activity projects in one side. Otherwise, technically speaking it is also an authentication and authorisation framework. It means application offered for the community involved, could be registered and it is easy accessible for the users. This access is easy in technical terms, because of utilizing the authentication framework, but it is easy, because the INTISC (http://intisc.org) offer information about the certain offered services. It means the participants get informed about these possibilities and receives an easy way how to utilise them. For the partners offering services it is easy to join. They have to implement only the SAMIL based authentication and authorisation framework. Just after the successful implementation of the standard authentication schema, and the registration of the application in the INTISC database, the services will have immediate accessible for the participants of the community.

The INTISC system are going to support

- keep the network alive
- extension of the network
- collaboration in the network

Keep the network alive: Inside the INTISC application the update of descriptive data about the institutions and projects and individual participants are distributed. There are responsible representatives of institutions can keep the important data update regularly. In larger network it is not possible to keep track of every piece of information. The responsibility and possibility has to be distributed and balanced take into count that network is driven by common objectives, motives the participants shares.

Extension of the network

The decision making mechanism how the network could be extended with consensual agreement of the involved participants on who to invite or who's application has to be accepted. These are processes could be carried out with much less problem, with IT support. It is some kind of international workflow helps to extend smoothly. Another aspect of extension is to develop new connection between participants. Important part of the project we are working on to build mathematical models, to support finding possible eclipses of interest and study field especially assuming large number of involved institutions and projects. This is a scientific research involving strong mathematical competences.

Collaboration in the network

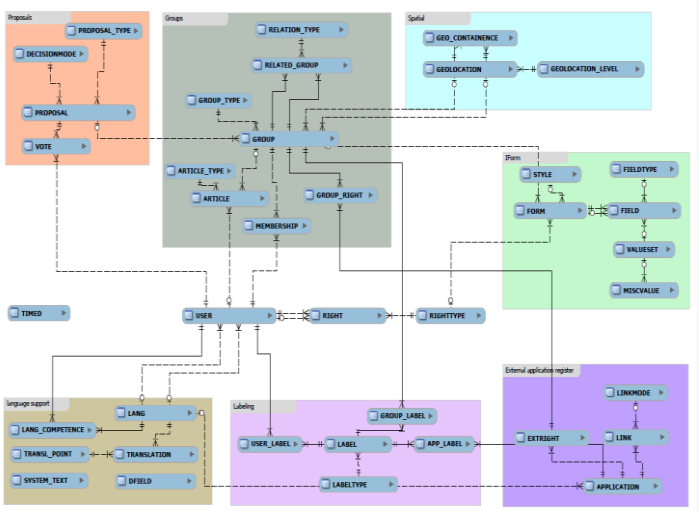
What are the applications participants can offer for the network? The collaboration could take place in very different fields from scientific research projects to synchronized actions unitising the best practices applied. That is why INTISC philosophy opens the authentication and authorisation platform for group of people in given case. There could be general purpose collaboration tools (for example: CooSpace provided and offered by a private company Dexter Ltd.) to specific scientific purpose tools (for example ePragma: offered by Université de Franche-Comté) The application supporting data gathering and

further analysis called (Catalyse Toolkit) developed and provided by the Université de Franche-Comté (Besançon), Territorial Intelligence Project Navigator (TIPRON) offered by University of Pecs (http://tipron.intisc.org/), ... and possibility is open.
Further plan for next years to develop special purpose project planning and management application. It could be focused for European joint projects, but should be relevant internationally as well.

Developing of register

The concrete application giving a platform for further application integration. First we are going to describe an important aspect: The computer based applications are not stands alone. As the ambiguity of the word application shows it, the computer application needs to have a culture of use, needs to have knowledge. How to use and what is it for? The recognition of the problem to be solved by the application is one of the key aspects of efficiency. In this article we are concentrating on the technical aspects of the development, but in other presentations and publication available we discussed it.

In the current phase of development we are working in a (SROP 4.2.2.C-11/1/KONV-2012-0005 Well being in information society) project at the University of Pécs. The targeted application, we are going to speak first about the core application called INTISC (International Network of Territorial Intelligence Service Cloud) the name describes a system going to be a signpost, the meeting point, for participants, institutions, projects, application involved in research and action of Territorial Intelligence. The application going to enable collaboration, research and action coordination, and research data processing via applications accepting the connection to the INTISC.

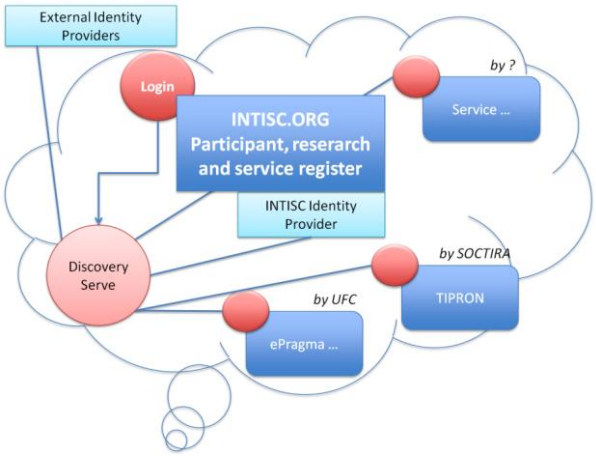


2. *figure: The data structure diagram shows the entities involved in the core application:*

We utilize an open standards for this project. The project application written in PhP, it's database background if in mySql. One of base part of the project is the directory of

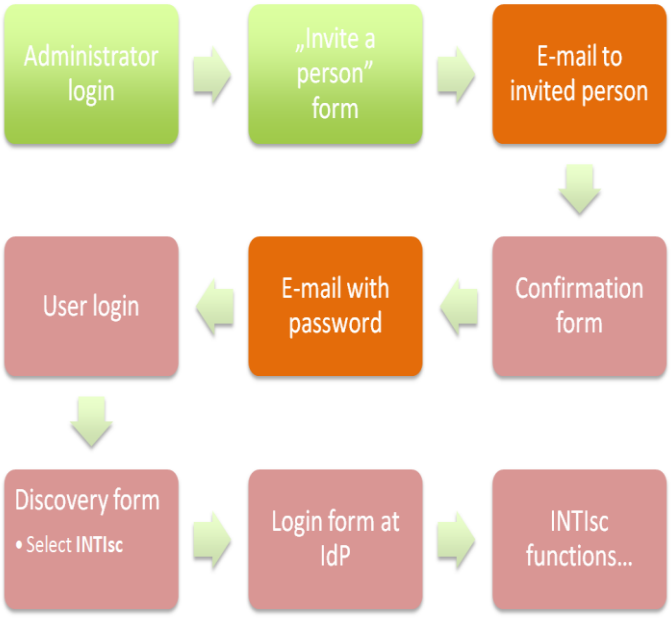
participants. In this article we are going to describe the technology chosen and used. The directory could contain institutional relation, project relations, grouping of participants and a lot of information about these entities. This is a basis of an authentication service offered for other application (services). Any application acception this authorization system can be connected. It means the application should implement, the login (sign in) protocol. This is beneficial for the service application, because it is standard protocol, and it can be used to for integration with other federations. It also provide single sign on (SSO) which makes it comfortable for the user.

Other part what we have to build authentication and authorization service. The authentication service (as identification of the user) it is could be partly based on federation but the system act as identity provider for a group of involved persons, having no other identity service provided for. The background for the INTISC identity service is an opneLDAP service. Authentication and authorization service as well, because we can offer information about user's project and institutional relations, group memberships, authorization process could take place based on. It is going to be provided via simple appropriate web services. The use for INTISC authorization services are optional integration possibilities based on standard federative authentication service identity.



3. *figure: The basic structure*

There is a INTISC directory in the middle connected to the INTISC identity provider service. We can login to the core system. The discovery service gives possibility to other systems to recognize and utilize the identity provider service. That provides possibilities for federative integrations. Currently we have a sample application called TIPRON (Territorial Intelligence PROject Navigator) showing an implementation example for using INTISC as identity provider. There are other application involved in the field of Territorial Intelligence for example ePragma provided by the Université de Franche-Comté – UFC is one of the important tools the integration is in close pans. It is very important to provide an open possibility for research institution, universities, organizations, public bodies to providing useful accessible applications to join.



4. *figure: The participants getting to the system*

Registering institutions (INTI)

On the next figure, it seems the access of user.

5. *figure: Invite key users (INTI)*

Invitation of a person starts with an invitation e-mail. The e-mail contains important information and a special link allowing the user to register.

The authentication integration

In the background of this integration has to have an appropriate agreement between partners. It is not enough to work on information technology driven questions, it is also crucial to build an acceptable collaboration logic to give real chance for agreements.

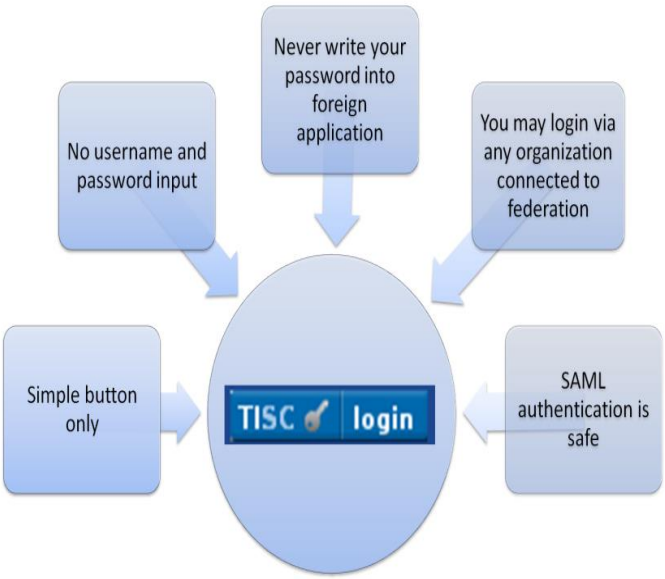
Technically the authentication integration has two steps:
-Implementation of SAML authentication in the service: It means putting a special piece of code appropriately to a service application. It is the easiest in commonly used PHP platform.


```php
<?php

require_once SAML_DIR.'/lib/_autoload.php';

$as = new SimpleSAML_Auth_Simple('default-sp');
if (array_key_exists('logout', $_REQUEST)) {
    $as->logout(SimpleSAML_Utilities::selfURLNoQuery());
}

if (array_key_exists('login', $_REQUEST)) {
    $as->requireAuth();
}

$isAuth = $as->isAuthenticated();
$attributes = $as->getAttributes();

...

?>
```

6. *figure: PHP platform*

Register the service in the cloud

Finally for authentication, a special button appears in the service application offering federation based authentication possibilities:



7. *figure: TISC login*

We enter the identification information (username, password) only on the page of identity provider chosen.

Summary

There are a lot of services available on the net, but it is as useful as we can let the involved researchers and actors to know about these applications and let them access these services. In practical sense "existence" means "accessibility" today. That is in the middle of INTISC system.

Websites: http://intisc.org, http://idp.intisc.org
OPEN Standard of the authentication system:
http://www.eduid.hu/hu
http://simplesamlphp.org/
https://www.oasis-open.org/standards
http://www.geant.net/service/eduGAIN/Pages/home.aspx

References:

ÁCS, P., 2008. Theoretical Approach of Network Communication and Collaboration in Research. In International Conference of Territorial Intelligence, Huelva 2007. Papers on territorial intelligence and governance, participative action-research and territorial development, Observatorio Local de Empleo, Huelva, 2008, p. 223-233. <URL: http://www.territorial-intelligence.eu/index.php/huelva07/Acs>
ÁCS, P., 2009. Theoretical approach of territorial intelligence and communication. In International Conference of Territorial Intelligence, Besançon 2008. Papers on Tools and methods of Territorial Intelligence, MSHE, Besançon, 2009. <URL: http://www.territorial-intelligence.eu/index.php/besancon08/Acs>
ÁCS, P. – FILÓ, CS., 2010. e-Support of researchactions. In Grand Ouestdays of Territorial Intelligence, IT-GO 2010. Papers on Territorial Intelligence and Socio-ecological Foresight, ENTI, Rennes.
ÁCS P. – BÉRES C. Z. – FILÓ CS, 2003. E-neighbourhood, azaz a hipertér lokális perspektívái. In Kultúra és Közösség 2003/1 p. 87-89
AHSAN, K. – GUNAWAN, I. 2010. Analysis of cost and schedule performance of international development projects. In. International Journal of Project Management 28 (2010) 68–78
CASTELLS, M. 1996. The Information Age: Economy, Society and Culture, Vol. I. The Rise of the Network Society, Blackwell, 1996.
CLELAND, D. I. – IRELAND, L. R. (ed.) 2008. Project Manager's Handbook, Applying Best Practices across Global Industries. Mc Graw Hill, New York
DIJK, J. 1999 [1991] The network society: social aspects of new media. [De Netwerkmaatschappij] Sage, London
FILÓ, CS. – ÁCS, P. 2006. New challenges of knowledge transfer in 21st century in Hungary. In: BERTACCHINI, Y. - MASSELOT, C. (ed.): Information, Savoirs, Décisions et Médiation (ISDM) n°26, Vecolloque TIC & TERRITOIRES, Besancon– Toulon, 186 Ko pp. 6-10.
FILÓ, CS. 2007. Knowledge–based development model. In: MOISE, A. – PASCARU PAG, M. (ed.): Region, Identity and Sustainable Development, <URL

http://www.territorial-intelligence.eu/telechargement/albaiulia2006/Alba06-Filo1.pdf>
FLEMMING, J 1998. Web Navigation, Design the User Experience. O'Reilly. Sebastopol.
Hand, Martin; Sandywell, Barry: 2002, E-topia as Cosmopolisor Citadel in Theory Culture & Society, Vol. 19 (1-2) SAGE, London
GIRARDOT, J.J. 2008. Evolution of the concept of territorial intelligence within the coordination action of the European network of territorial intelligence. In. ReS-Ricerca e Sviluppo per le politichesociali. Semestrale. NuovaSerie n. 1-2/2008 Numero special (English 11-29)
JOHNSON, D. B. 1999 [1991] A közösségi döntések elmélete [Original: Public Choice] Osiris, Budapest.
KUHN, T S. 1996 [1962]. The Structure of Scientific Revolutions. 3rd ed. Chicago and London, The University of Chicago Press
WILHELM, Z. ACS, P. FILO, Cs. BERES, Cs. UHERKOVITS, P [2013] A New Extension of Territorial Intelligence Service Cloud In: International Conference of Territorial Intelligence, "Territorial Intelligence, Socio-Ecological Transition and Resilience of the Territories"30-31 May 2013, Besançon (UFR SLHS, MSHE), Dijon (MSHD)

Information process and Territorial Intelligence

Saccheri, T.
Univeristy of Salerno (Italy)

Among territories knowledge and the territorial action it is necessary to insert an aspect linked to society and culture reflexive abilities, which are considered in a particular community context.

In this speech, the community does not mean one closed entity in an auto-referential way (ghetto), but it is represented as a subject that is able to auto-consider itself in a reflexive and relational way: only from this reflexivity springs the capacity to built subjectivity and action aiming at change.

This action will have necessarily the characters of *a critical action* as regards the pre-constituted assets by more traditional administrations and *a relational and co-planning action*, concerning more advanced administrations on the governance plan.

We could ask the question "what instruments can result advantageous in the second case"?
We are interested in the behaviour of individuals towards development as it developed over time, characterized, on the one hand, by strong elements of competitiveness and, on the other hand, by the emergence of social subjectivities more cooperation-oriented.

In this contradictory movement, in some ways loaded with social expectations, we must place the issue of territorial intelligence, which can not be reduced to mere strategies of economic interest, but must instead consider the territory as a hierarchical structure (managed by the politicians-specialists combo), but as a group of citizens. We therefore claim that we do not want to think of the territory as a structure "within" which to find the citizens; instead, the territory is "of" the citizens, it is made by them.

Generally speaking, each territory is regulated by administrative offices which, even on the area of economic and political planning, seldom provide widespread, clear, understandable, and, above all, shared information.

Individuals are carriers of the maximum of subjectivity and of a type of information that is directly dependent on their experience; however, this subjectivity must be related to and compared with that of the members of the group to which those individuals belong. The so-called "non-technical experts" have information qualitatively richer than those of individuals, as they are partly organized, arising from the position that the experts themselves hold within the community or derived from the profession they practice and the role they play.
Both of these social actors are involved in the production of the information coming from the organized groups within the population, and also of the further knowledge created

through discussion and the sharing of individual experiences and individual knowledge.

At the operational level, the course of action is hampered not only by the growing gap between the potential of design and the ability to systematize interventions, but also by other factors:
1) the lack of a common language and/or of effective communicaton between decision makers and technicians/experts, as well as between the latter and citizens;
2) the following problem of power management and decision-making;
3) the complexity of the various territorial realities.
Between "top down" communication flows, growing in strength, and "bottom up" ones, alas fading, it is more and more complex to push through the decisions taken by the apex of the system, whatever their nature. Communication channels able to promote the contribution of individuals or social groups appear increasingly jammed and full of obstacles (and sometimes it the same technical and political power to interrupt the flow of communication.

One of the first difficulties in the transition from an outlook oriented towards compensation intervention determined from above to a promotional outlook is the very need of in the methodology some *socially based variables:* rather than adhere to a specialized language, the inner language of the technical-political community, they must have the "duplicity" characteristic of the terminology shared between the community of experts and inexperienced people, between professional communities and "profane" communities.

Some of the most desirable questions are often excluded from the planning agenda:
- Who is going to identify the needs? How? Referred to whom and under what socio-spatial dimensions?
- Who is going to infer functions from needs?
- Who is going to monitor the effectiveness of the interventions emerging from this process, or to ensure the fact that the functions are correctly and accurately inferred?

Moreover, the answers are not necessarily simple, obvious, or natural.
- Needs can only be detected by local communities examining their own reality.
- The community itself, using for this purpose also the available technical expertise.
- Effectiveness control is carried out by the community itself through its representatives, who act as supervisors by means of tools and procedures supported by IT systems.

Therefore, a further element derives from these considerations: the need for "accountability," which can not be left out as not pertaining to the daily political or professional work. We mean that the information systems related to / concerning the various projects have to be constructed so that they are:
a) "relevant", i.e. that answer to the different parties involved for the decisions taken and the intermediate results, and

b) able to allow the parties to make other decisions concerning social choices and the allocation of actions and resources.
c)
Public opinion is not the result of mere information, in the same way that behaviours are not the outcome of mere scientific knowledge. Furthermore, public opinion is a set of opinions; instead, it springs from knowledge, beliefs, values, behavioural patterns acquired by and experienced within public awareness, and groups usually judge and adopt their behaviours according to such standards.

Therefore it is not so much a question of *persuading* or *feeding back* the results of research in the field, but rather of *analyzing*, of *discovering correlations*, of *agreeing upon* modes, even considering a feasibility that can only be assessed by those who take action. In order to do this, it is necessary to share objectives beyond the "well-being promotion" formula: «This requires parties and political institutions that respond more to the influences of organized social forces and movements and at the same time are keen to foster greater participation of the disadvantaged in public life, which can only be obtained by increasing social protection and welfare» [Touraine 1998: 281].
It also requires the overcoming of that attitude which perceives the expressions of subjectivity – often in conflict with established rules – as "inconvenience".
The principles underpinning the work of the INTI GDRI group[I] consider indispensable the immediate (i.e. from the very beginning of the action and research) involvement of stakeholders in the planning process envisaged by an ethical principle of planning and governance: this method requires an effort of knowledge and an ongoing exchange of experiences among the groups involved in the action of planning, sharing that looks at the reality of different observation points and has different kinds of expertise.
Only when these two modes and levels of knowledge come together we can implement governance as a product of a territorial intelligence and as an instrument of governance and the promotion of equity.

In CaENTI [Saccheri, Schmitz, 2009] we attempted to define the schema that should guide a participatory action:
6. residents play an active role throughout the execution of the project;
7. territorial actors know and can use the tools and research methods employed in the project;
8. training sessions are provided for the players who will make up a team of competent citizens;
9. subgroups have contributed to outline work plans defining intermediate and final objectives for the development of the project.
It is the people, thus, that bring real added value to the idea, its realization, its ownership by the local community and its permanence in time; and a working group for the detection of needs ensures that all the social sectors involved are adequately represented.

In this way, the topics covered during the works are shared by all the people involved and the topics identified take into account the weight of the participants and the interests they represent, so that the priorities on which to start working are shared by all the players based on three essential elements: participation, education, cohesion.

In our research *Tra resilienza e vulnerabilità* (between resilience and vulnerability, 2012) we are faced with an idea of community "in the singular", isolated and self-referential, while a development made from grassroots democracy and innovation with regard to the sustainability and future generations concerns a community "in the plural", freed from the constraints of self-confinement to a ghetto so typical of Southern Italy.

Which kind of approach can thus change the local context?
What are the obstacles to the strategies that can implement and increase the participation of stakeholders?
Our research shows that the constraints in a real process of local innovation can be described by the following three points:
there is no communication between the various areas of the territories;
there is no cooperation between the technical, political and social spheres;
it is difficult to eliminate the top-down approach that has governed the previous planning operations.
The question is if it is possible to imagine a different kind of communication, which is not pure information but it contains the seeds of a wider sharing of shared and truly democratic strategies.

Promoting and regulating social well-being

Ammaturo, N.
University of Salerno (Italy)

We present a brief report on an area now known as the "Land of fires" ("terra dei fuochi" in Italian), with its problems, emergencies and possible of solutions. The "Land of fires" occupies a vast area between the two main cities in the Campania region: Naples and Caserta. It comprises 57 municipalities and its population exceeded two million inhabitants.

This area is called "Land of fires" because of its being subjected to the spillage of toxic waste for the past thirty years; remarkably dangerous waste, with high levels of toxicity, sometimes buried in pits dug in considerable depth, and sometimes just a few feet below the surface of cultivated land.

More waste, consisting of polluting materials in large quantity, is deposited on the surface and burned, hence the name "Land of fires". The smoke releases into the atmosphere toxic substances such as dioxin. The buried refuse also includes radioactive waste from hospitals and chemical industries of various European countries, whose rate of pollution can last for centuries, and in some cases affects the aquifers.

Among the litter burned are tires and scraps of clothing, as well as other waste material. By means of the rain, dioxin is deposited on the land, once the pride of the agrarian economy of the area; it thus enter the animal food chain and is transmitted to humans through the consumption of milk and dairy products, such as the famous mozzarella cheese. The rate of respiratory diseases and the increase of cancers are the result of this environmental tragedy, which, together with the production stop in the entire agricultural sector, provides considerable damage to the already dismal economy of the area.

Through the testimonies of informers, once actors and performers of these death trades, it has been possible to identify the places where highly toxic substances were spilled. Only recently, the health system is providing data on the health of the population living in these areas, with particular attention to children.

In order to get a clear picture of the situation, it is enough to think that, while we talk, in these territories, which are now monitored by the police, there are still fires and spills going on. Thus, the higher the risk, the higher the costs for the "customers", and the Camorra runs the territory with extraordinary arrogance.

Experts have found that to relieve the area – at least that affected by radioactive waste pollution – it would be necessary to transfer elsewhere millions of cubic meters of highly contaminated soil, while other necessary work concerns measures of recovery of the seriously compromised natural resources.

In any case, the near future does not leave room for optimism.
In summary, this is the condition of the land of fires, whose density of population scores among the highest in Italy.

In this case, emergency management leaves little room for the promotion of social welfare. The only bright spot, and it is no small thing, relates to the awareness of the population. The inhabitants of the area recently expressed anger against the local institutions, claiming that they should have monitored the situation to avoid the incalculable damage on the ground perpetrated by the Camorra, whose elements penetrated into many public and private affairs.
The greater responsibility goes to the thirty-years-long indifference of the ruling class or of those who found silence to be more advantageous. We are talking about:

d) the landowners, who denounced abuses in their territories only after the scandal came out in the open;
e) the citizens, who looked on environmental violence, either in silence or unaware;
f) the series of clues often hidden by unprincipled observers;
g) a crooked territorial intelligence, willing to compromise;
h) the many people aware of the proceeding, who lacked the necessary social and environmental support to power their courage.

Years have passed since the publication of Roberto Saviano's book *Gomorrah*, in which he denounced the abuses of the Camorra on these territories, and the same Saviano warns the citizenship "against organized crime that first polluted, then thrived on the waste crisis and is now ready to build a new speculation on the recovery actions".

For many foreigners who know the Campania, the beauty of Naples and Caserta, along the coast of the peninsula of Sorrento, Amalfi, Ravello, Positano and Capri, a World Heritage Site, it seems an insult to common sense to be in an environment so devastated by human hands, a hoax, mixing third-world aspects with high-tech components, populism with the finest political theories, participation with marginalization.

These are the contradictions of an area that boasts extraordinary cultural roots.
In recent years, the group of European scholars who have been part of CaENTI investigated the sustainability of a development promoted by the territorial intelligence in various territories of the European countries chosen for their similarities. In the meetings it was possible to define the critical role of territorial intelligence in the

development the environment intended as a whole, particularly in the exploitation of the characteristic resources of the area. The research pinpointed territorial intelligence as a driver of development, whose sustainability is measured in terms of effectiveness and efficiency on both quantity and quality. The population is the key agent in defence of the resources available in the area, focusing on confirming the validity of the projects to promote, that have to be consistent and integrated in the socio-cultural system.

In the "Land of fires", the fact described above have highlighted the lack of care of the resident population with respect to its territory. Furthermore, to attribute the whole responsibility to the ruling class was an irresponsible act. Even when you delegate to policy-makers the care of your territory, in all countries of the world citizens have a direct relationship with their living environment, and the responsibility does not lie only on the political figures.

Every year in these areas, especially during the summer months, we spend millions of Euros to extinguish the fires regularly caused by arson in the nearby mountains; and even in these cases there is a suspicion of a criminal organization drawing ill-gotten gains from environmental disasters.

The population lives these phenomena with apprehension, however, at the same time, it seems as if someone had set a scheduling of such sinister and creepy events.
In many cases, it would be possible to put on the government's agenda – and carry on – the prevention of certain events. However, for many public leaders, it is better to ride the emergency. Whenever we are faced with phenomena affecting the public good, and therefore the population as a whole, citizens are absent.

Citizenship means awareness of the rights and duties of the citizen, who participates first-hand in the actions affecting the life of the area where he lives, expressed in various dimensions relating to politics, economy, culture intended as a combination of tradition and innovation, growth and development, religion and promotion of resources. In Naples and its province many ignore the meaning of citizenship and a lot of responsibility for this lies with the ruling class, regardless of their political flag, whose populist rule left the citizens in the squalor of ignorance and irresponsibility.

In our future research we try to give answers to a series of questions, addressing to the people directly involved:

- Why only in these territories of our country has been possible to create so great a harm?
- Why were the local authorities deaf to complaints from citizens?
- Why did foreign and North-Italian industries entrusted the Camorra with the disposal of toxic waste to be transferred in the rural South?
- Why did not this immense disaster become a national case?
- What is worse than a ban on the use of the water and the breathing of polluted air?

- Which future for the new generations?

These questions show how much the perpetrators of these criminal acts lack civil sense: they reside in this very area, in which they grew up and where they raise their children. But they have turned their territories in cemeteries, lands whose sterility prevents any form of life; they do not recognize any value outside of the gains acquired through a criminal act.
It is difficult to distinguish, in so mystified a social reality, the honest from the dishonest, the innocent from the guilty, and in the midst of so much confusion those middle classes that represent in many ways the driving force of development, because they are engaged in the fields of education, health, and social reproduction, must protect themselves from the threat of criminals often in collusion with politicians.

This awareness should not be an excuse to justify an accomodating immobility by the heads of territorial policies. We are referring to the territorial intelligence, those people who hold positions of authority, to which we can assign tasks of environmental reclamation and control on the more suspect population.
Territorial intelligence must triumph over the shackles of bureaucracy, alas by now become bureaucratization because deprived of a political rationalization: it must remove barriers in order to promote a development designed to enhance human, artistic, and environmental resources and to revitalize the area through the education of people. But criminal organizations and many silent citizens have turned their territories into cemeteries, lands whose sterility prevents any form of life; they are individuals who acknowledge no value besides the gains acquired either in a criminal activity, or by being accomplices to crimes.

The "Land of fires", however, has awakened political consciousness in much of the population, and many are willing to challenge any form of power that is exercised through violence, and not recognized by the rationality and effectiveness of the actions promotes.
They become agents of their own history, participate in the management of the public space by electing the territorial intelligence in charge of decisions addressed to public health, the monitoring of everything that happens in the area, the correct use of environmental resources, especially the well-being of the entire population, including the resident component of immigrants, often overlooked in the satisfaction of basic needs.

The match is yet to be played.
The population now knows about environmental risk, and it is not easy to live with the internally known culprits; the protections enjoyed by these criminals may not be enough in the face of popular protests and the monitoring actions the police carries on, with the order to stop any illegal action consumed in the environment.

It happens a few times in history that certain events compel the showdown: when you touch the vital interests, the future of the children and their hopes for a better life, people are willing to do anything and do not stop until they solve the problems.

Information retrieval and –management practices of Hungarian students in an international comparison

Egervári, D. – Sipos, A. M. – Varga, K.
University of Pécs (Hungary)

1. Introduction

Since the mid-20th century technology, economics and culture and their effect on each other have been rapidly changing. People need to adapt continuously to the changing environments, equipment, conditions and opportunities. The question is how this is to be done - how to face constant challenges, and what skills, abilities and competencies are needed to be able to progress in the information society and the digital world.

In the international literature Manuel Castells' Information Age trilogy[47] was the first work to emphasize the human factor while describing the digital foundations of the information society. Castells points out that the three pillars - technology, content and knowledge - are developed in a complex, collective and synchronized way. The paper discusses one of these components, knowledge, which has not received sufficient attention. The three pillars form a complex network of interdependencies. It would be a serious oversight, therefore, to ignore the other components and the complexity of the phenomenon.

In the theories of the information society the human factor has received attention mainly from an economic point of view, one that produces goods. However, while compiling the digital pillars, Castells considered the human factor to be the most crucial element. Humans are the primary influence in shaping the information society, since they are responsible for technological innovations, the direction they develop in, the fields in which they are utilized and the kinds of content that are produced or digitalized.

In Hungary, the digital pillars of information society have not been adequately considered as a complex entity, the structured foundation and the development of information literacy have not been achieved.[48] One reason for this is that the concept of information literacy still has not taken root. It is neither part of education policy, nor of normative documents in regard to public, higher and adult education. The complex foundation and the development of information literacy are not prioritized within the goals of public and higher education; therefore, information literacy has not had a chance

[47] Castells, M.: The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I-III. Blackwell, Cambridge (1996-1998)

[48] Koltay T., Varga K.: Conceptions, Ideas, What Else? Information Literacy in Hungary. In: Kurbanoğlu, S., Grassian, E., Mizrachi, D., Catts, R., Špiranec, S. (eds.) Worldwide Commonalities and Challenges in Information Literacy Research and Practice, pp. 625-631. Springer (2013)

to take a hold in educational practices. The first task is the complex interpretation of the concept of information literacy, which will allow for the term to become more prevalent, and it would also facilitate implementing it in practice.

Our attempt is to define opportunities and methods for the complex development of information literacy competencies based on the fact that competency-based education and training is one of the greatest challenges of the 21st century.

2. What is information literacy?

The meaning of the term literacy is constantly changing in different cultures. In English culture it has a strong and stable meaning, but for example in Eastern and Central Europe the term has a controversial life. In most European countries people don't like to mix the traditional literacies and cultures with the modern digital competencies. There are strong debates around the terms: literacy, competency, information, digital etc.

Information literacy is one of the most important key competencies of the 21st century knowledge societies. As long as we do not recognize that it is much more than just being able to use the digital technology, it will never fulfill its real mission. People in the information age need special weapons to be able to win in the war against information overload, and manipulation. The information literate person knows how to learn, how knowledge is organized, and how to find, organize and use information.[49]

Information literacy is not a new topic, especially in the field of library and information science, but in Eastern-European countries it is not really manifest in public education and higher education programs. Education policy makers are dealing only with the problems of digital literacy, and do not want to take into consideration, that it is necessary to have much broader information competencies in order to survive in the 21st century.[50]

Information literacy is a broad concept, consisting of different other literacies and competencies. We can only speak about information literacy, if these elements all are together at the same time.

Our good and bad experiences about the level of information literacy competencies, especially among young generations, force the communities to get exact indicators about the real situation. These indicators can help the countries to recognize the effects of their policy actions, and to get information about how their citizens are able to cope with the

[49] Association of College and Research Libraries: Information literacy competency standards for higher education, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf

[50] Partnership for 21st century skills: Framework for 21st century learning. http://p21.org/overview

challenges of the information and knowledge society.[51] Indicators we can get through different ways:

- Independent surveys.
- Use international surveys (e.g. PISA, PIRLS, TRAILS, SAILS etc.) in different levels of education.
- Combine international indicators and international surveys.
- Statistics related to creation/availability, distribution/supply, information reception, and use.

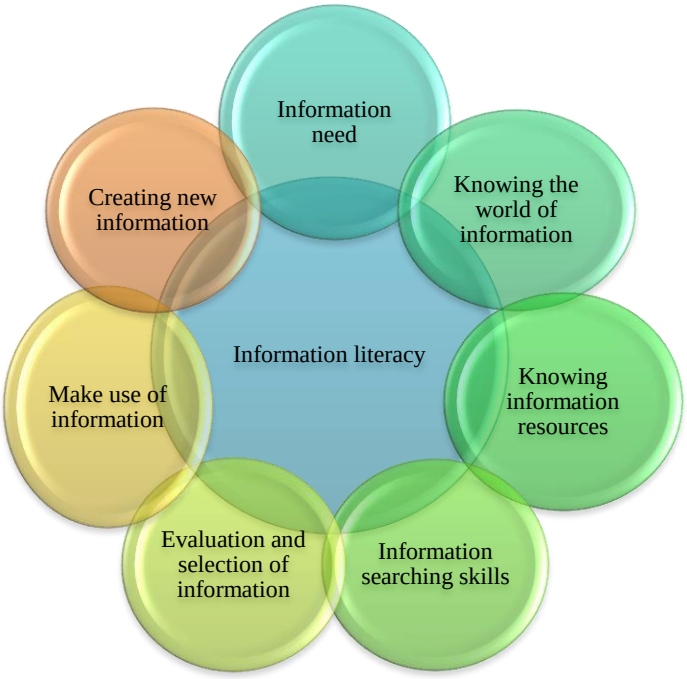


Fig. 1. The model of information literacy[52]

3. Research findings

The Institute of Library and Information Science at the University of Pécs made an online survey in 2014 (sponsored by the SROP 4.2.2.C-11/1/KONV-2012-005, Well-being in the Information Society project) about information competencies of university students all over Hungary. We wanted to know how students get information for their studies, what are their main resources, information seeking methods, how they select and evaluate the information. We got 2599 answers; our survey is not representative but significant.

[51] Towards media and information literacy indicators. Prepared by Susan Moeller, Ammu Joseph, Jesús Lau, Toni Carbo. Paris, Unesco, (2011). p. 16.

[52] Varga, K.: Az információtól a műveltségig. Az információs műveltség alapjai. L'Harmattan, Budapest (2013)

During their studies, students have a number of assignments that require competent literature searching and analysis. They like these assignments, and they do not feel any difficulty related to them. They also acquire substantial experience in making presentations, and have many opportunities to apply up to date digital technologies.
It is very interesting that the most frequent sources of information are not the books, or the media, or even the Internet, but the social relations, friends and colleagues. Media as information resource is not very much used by the young generation.

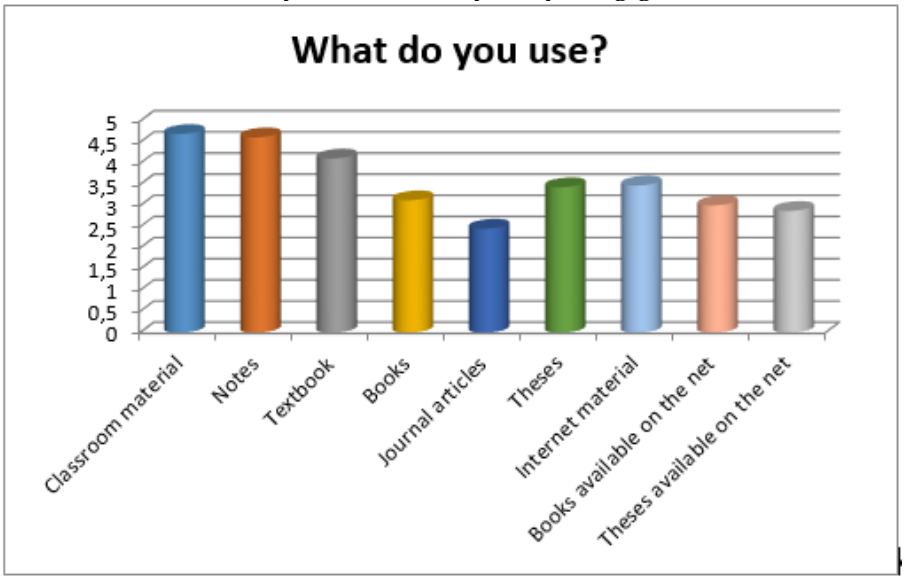


Fig. 2. What resources are used for the studies?

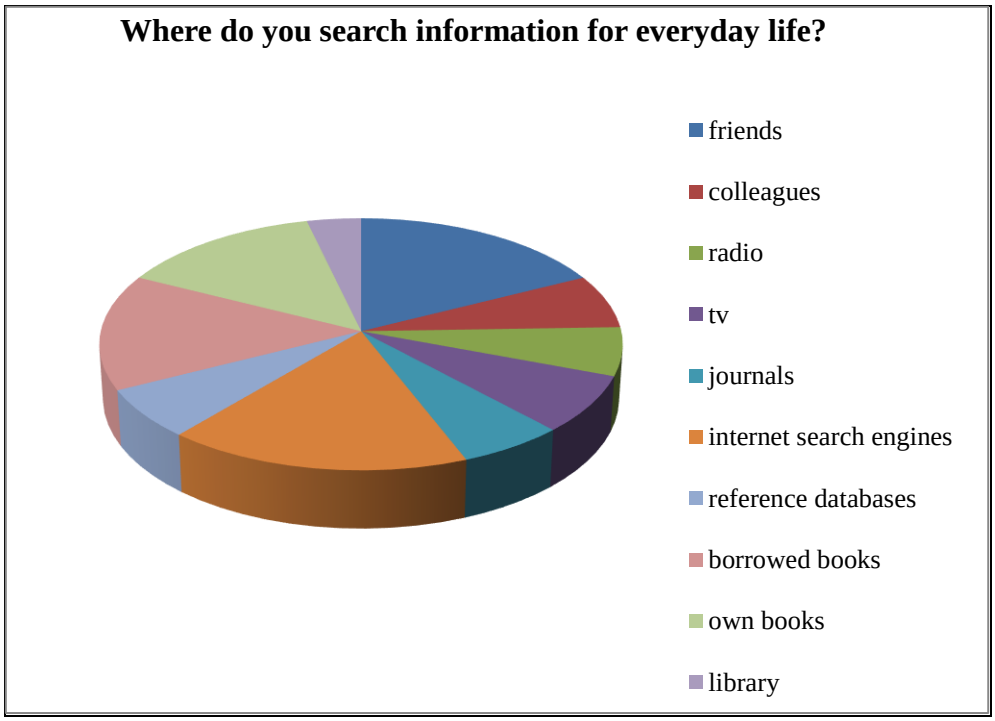


Fig. 3. Most frequent everyday information resources

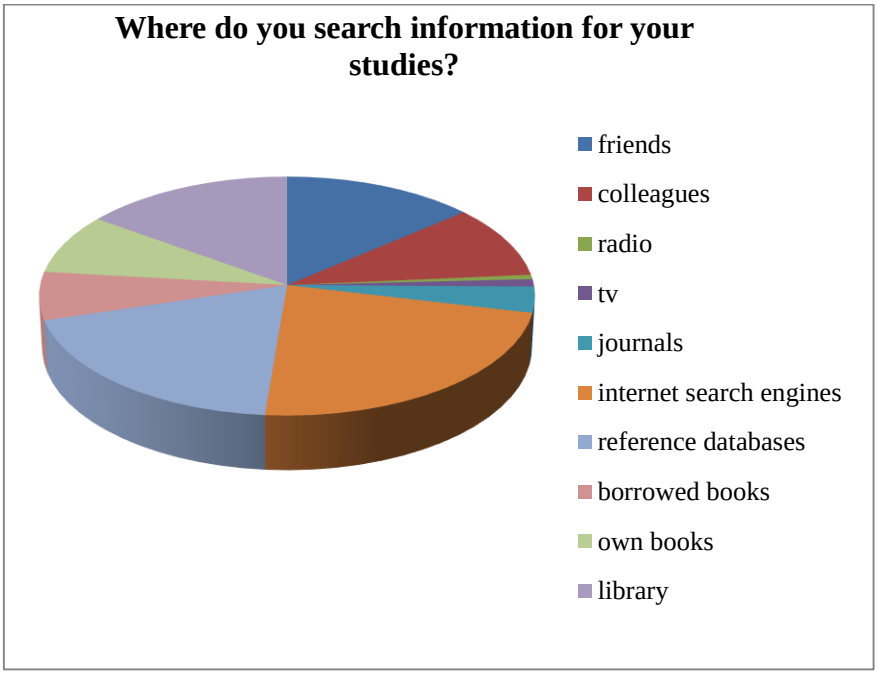


Fig. 4. Most frequent professional information resources

Students use the Internet mainly for social relations and learning, and less for getting political or economical informations.

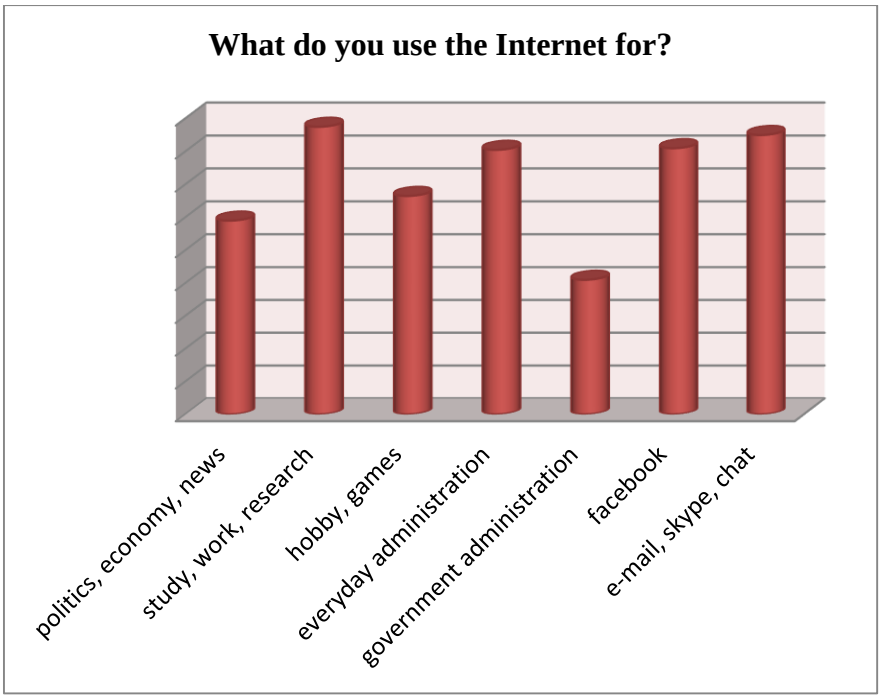


Fig. 5. The aims of Internet usage

The results show that there are big problems with the knowledge and competencies of our students. Only 1/3 of them apply information literacy competencies (e.g. search strategies) in their work. They have quite weak knowledge about professional information resources (databases), their main information resource is the internet, and the main information retrieval tool is Google. The complex competencies of information literacy are not known for them, and very often they ignore planning before an information solving problem.

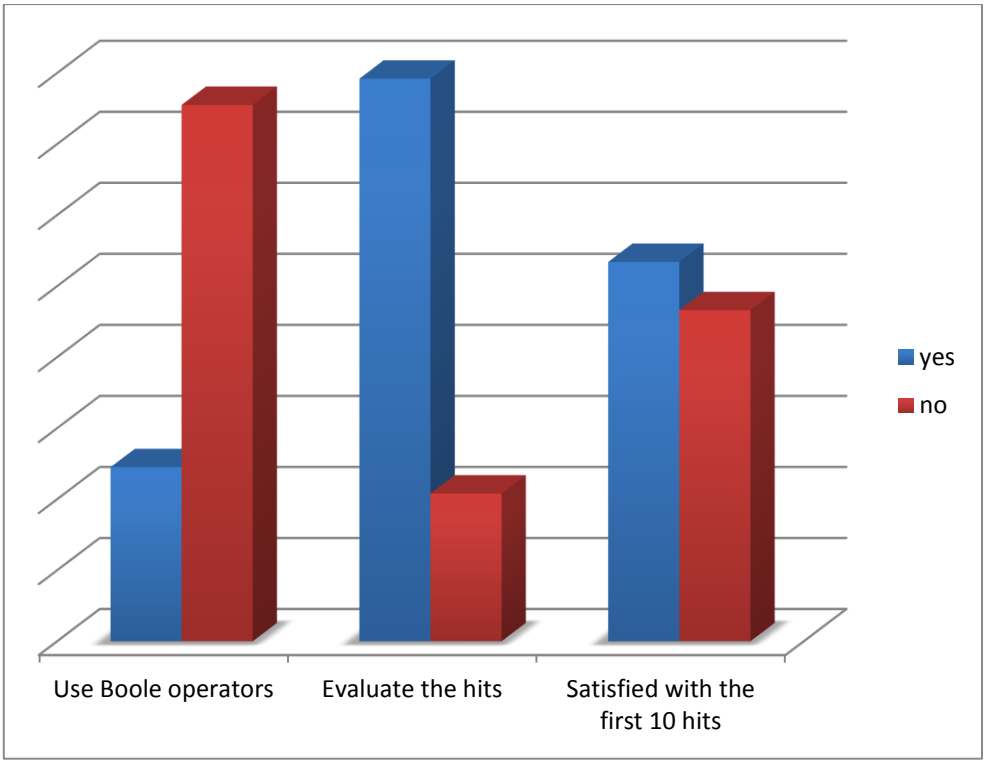


Fig. 6. Search strategies

Students have no bigger difficulties in defining a search question and strategy. However, about 20% of the respondents said they have problems with identifying relevant hits. It is difficult for 40% to determine, whether a web site is credible or not. It is also hard for them to convert the collected material into new information. This means that, despite the fact that they have opportunities for carrying out independent research, some very basic competencies that would enable them to accomplish these assignments in an efficient way, are missing.

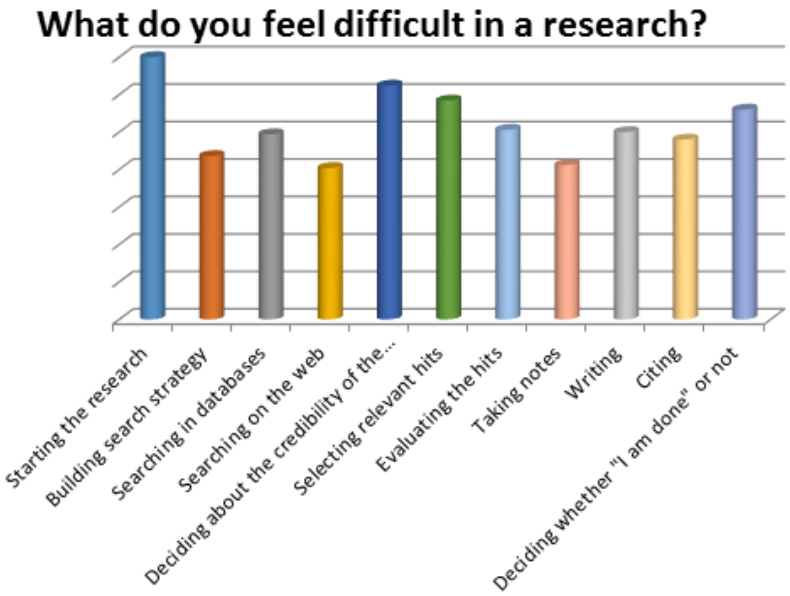


Fig. 7. Difficulties in a research process

For information seeking the majority of the students uses internet search engines, mainly Google. Library catalogues, encyclopaedias and lexicons are less frequently used. At the same time our students trust traditional information resources much more than the modern, digital resources. Hungarian students rarely consult government sites, and unfortunately they do not like to use research databases in order to solve study assignments.

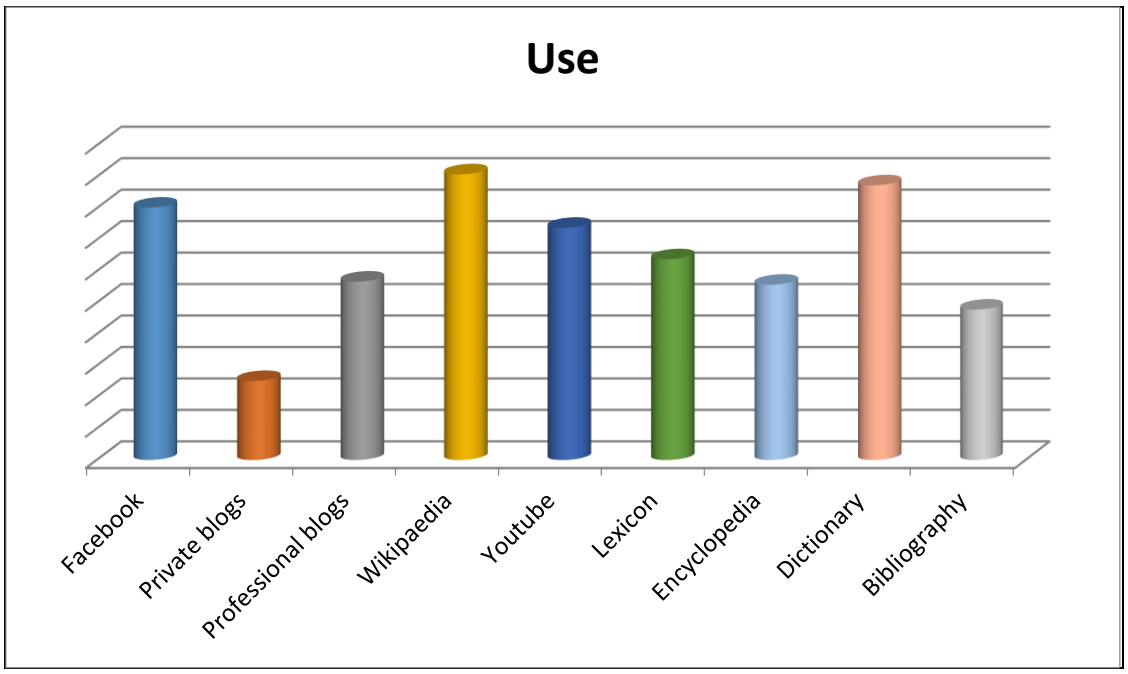


Fig. 8. Used resources

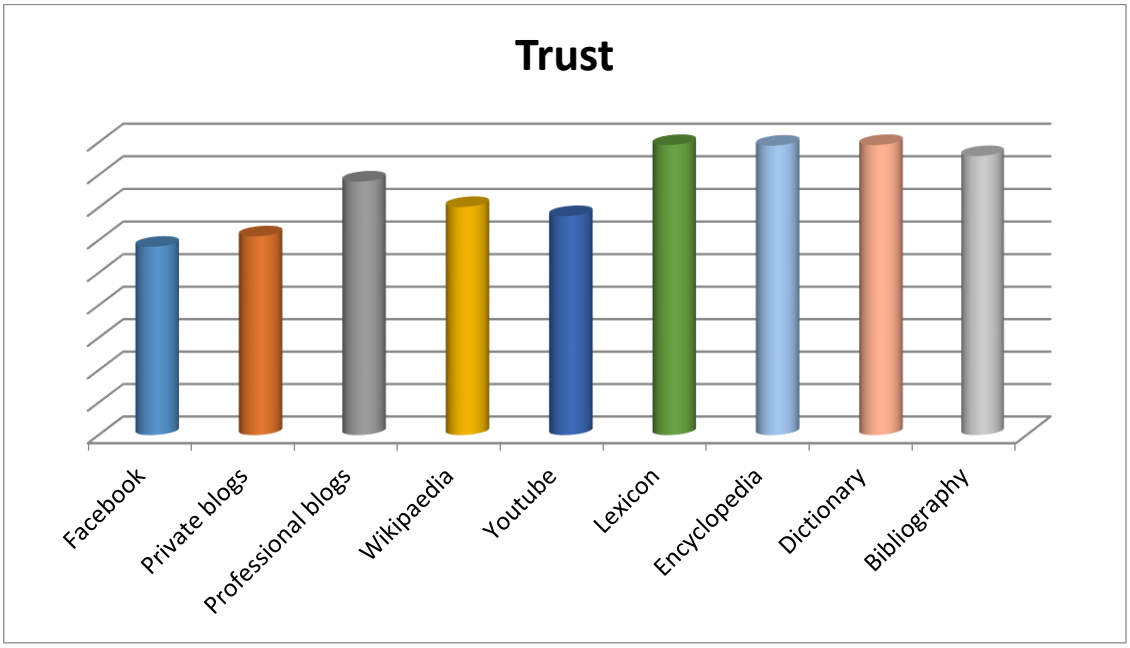


Fig. 9. Trusted resources

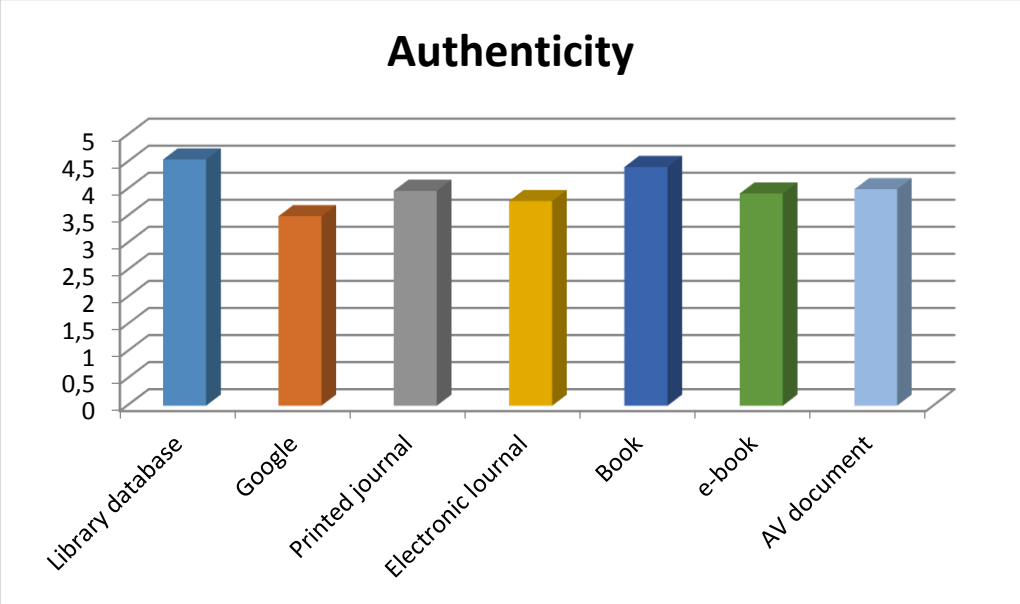


Fig. 10. Authenticity of the resources

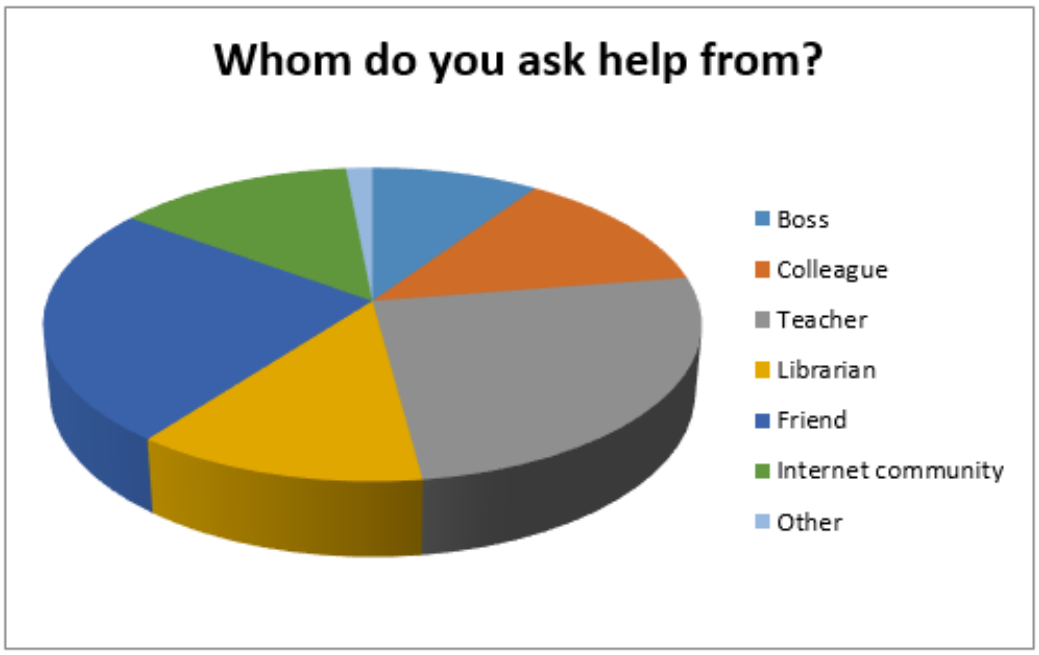


Fig. 10. Help

In the selection process freshness and reliability are major issues, the publisher or the existence of a bibliography is not important for them. Unfortunately, Hungarian students still have difficulties in using foreign languages, so one of the most important aspects is that the resource should be in Hungarian.

The results of this survey, compared with other surveys about students' information gathering methods[53], give some hints about information literacy in Hungary. Students all over the world like to choose the easiest ways to get information. Higher education institutions try to force students towards deep and reliable research methods, so they have to face several information seeking assignments. However, unfortunately Hungarian students are not well trained in gathering and selecting relevant information. In other words, their information literacy skills are limited.

Let us state that the roots of the problems are in public education. In the Hungarian National Core Curriculum there is no special focus on information literacy skills. Students in primary and secondary schools learn digital literacy and a little library literacy as a part of informatics (computing) courses, and there is a substantial emphasis on media literacy. Nonetheless, the holistic view of information literacy is not manifest

[53] McKiel, A., Dooley, J. (eds.): Changing library operations – Information literacy and E-resources: The Credo student survey. In: MLA, SLA Book Expo Issue-Pre-print, vol. 25. No. 2. April (2013) http://cdn.credoreference.com/images/PDFs/ATG_v25-2_McKiel_PrePrint.pdf
Head, A. J.: Project Information Literacy: What can be learned about the information seeking behaviour of today's college students? In: Association of College and Research Libraries (ACRL) Proceedings 2013, ALA, Chicago (2013) http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/Head_Project.pdf

in the curriculum or in everyday school practice. One of the reasons is that school teachers' information literacy skills often are below of that of their students'. This is one of the reasons why there is a strong and urgent need for reforms in teacher education.

6. Conclusions

The shortcomings and problems of research in information literacy in Hungary may lead to serious consequences. As information literacy is not considered a key competency, educational institutions are not addressing it as belonging to the core group of basic skills, and thus they do not spend resources on establishing and developing it. That is why it is crucial to place more focus on studying the Hungarian aspects of such theoretical questions and the methodology of teaching 21st century competencies.

There is an immediate need for a system of educational principles for information literacy, for curricula supporting the acquisition of sub-skills, and for these curricula to be integrated into the public education system. We have to provide an opportunity for students to acquire, practice and improve sub-skills of information literacy in a structured system. In addition, curricula provide special content and tasks relevant to each subject area. The intermediate-level information literacy acquired in public education can then be further developed and made specific within institutions of higher education. That is why close cooperation between educators and librarians in higher education is essential, which can only be effective if educators provide specific tasks and projects for students that require regular use of library resources and services.

Educational institutions as well as libraries are lacking in precise definitions of the roles and tasks, which would be essential for the complex development of this competency. What is needed is the availability of and access to the latest technology and the most modern infrastructure, along with a re-evaluated role and precise task definition for institutions of public and higher education and for libraries, as these are the places where establishing and developing 21st century competencies will need to be especially prioritized.

There is a special emphasis to be placed on the role of libraries, since they are the institutions that play a major role in the acquisition and development of information literacy. At the same time, this situation poses a serious challenge for libraries that they need to prepare for. There are international programs and projects that can help libraries in this endeavour.
All this could serve as a foundation for further research, pedagogical programs, and educational concepts, which in turn could contribute to the institutionalized foundation and development of information literacy. Information literacy as an attitude plays an important role for members of the information society acquiring other 21st century skills and competencies, which in turn result in life-long learning and the mitigation of the secondary digital divide.

References

1997. évi CXL. törvény a muzeális intézményekről, a nyilvános könyvtári ellátásról és a közművelődésről, http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=99700140.TV
Association of College and Research Libraries: Information literacy competency standards for higher education, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf
Castells, M.: The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I-III. Blackwell, Cambridge (1996-1998)
Egervári, D.: Development Opportunities for the Digital Pillars of Information Society. In: Lehocka, I., Szabó, T., Vargová, Z., Viczayivá, I. (eds.) Science for Education-Education for Science, 3rd International Conference, Nitra, 2013. pp. 99-111. Univerzita Konštantína Filozofa v Nitre Fakulta Stredoeuropskych Studii (2014)
Head, A. J.: Project Information Literacy: What can be learned about the information seeking behaviour of today's college students? In: Association of College and Research Libraries (ACRL) Proceedings 2013, ALA, Chicago (2013) http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/Head_Project.pdf
Koltay T., Varga K.: Conceptions, Ideas, What Else? Information Literacy in Hungary. In: Kurbanoğlu, S., Grassian, E., Mizrachi, D., Catts, R., Špiranec, S. (eds.) Worldwide Commonalities and Challenges in Information Literacy Research and Practice, pp. 625-631. Springer (2013)
Martin, A., Madigan, D. (eds.): Digital literacies for learning. Facet, London (2006)
McKiel, A., Dooley, J. (eds.): Changing library operations – Information literacy and E-resources: The Credo student survey. In: MLA, SLA Book Expo Issue-Pre-print, vol. 25. No. 2. April (2013) http://cdn.credoreference.com/images/PDFs/ATG_v25-2_McKiel_PrePrint.pdf
Partnership for 21st century skills: Framework for 21st century learning. http://p21.org/overview
Sipos, A. M.: Competency for Information Literacy. In: Lehocka, I., Szabó, T., Vargová, Z., Viczayivá, I. (eds.) Science for Education-Education for Science, 3rd International Conference, Nitra, 2013. pp.235-249. Univerzita Konštantína Filozofa v Nitre Fakulta Stredoeuropskych Studii (2014)
Sipos, A. M.: Információs kompetencia – realitás vagy frázis? Az információs kompetenciával kapcsolatos ismeretek oktatása a Pécsi Tudományegyetem informatikus könyvtáros képzésében. In: Herdon M. – Pethő A. (eds.) Informatika a felsőoktatásban. Debreceni Egyetem, Informatikai Kar, Debrecen (2008), http://www.agr.unideb.hu/if2008/kiadvany/papers/F73.pdf
Varga, K.: Az információtól a műveltségig. Az információs műveltség alapjai. L'Harmattan, Budapest (2013)
Wilson, C. et al.: Media and information literacy curriculum for teachers. UNESCO, Paris (2011)

The Institute of Library and Information Science of the University of Pécs conducted an online questionnaire research in 2014 about the information culture of the students of various Hungarian universities and colleges. A total of 2599 questionnaires were completed. The survey is not representative, but the large number of questionnaires can make it significant.
Information culture is the basis of lifelong learning, it is equally important in all scientific areas and all learning environments. With its help we can recognise when we need more information and how we can acquire it the most efficiently. It gives us the necessary skills to utilise the various sources of data, analyse and evaluate the retrieved information and gain the confidence to make decisions. Information culture makes the learning humans able to handle content, to make research, to become independent and to gain more control over their learning. It is indispensable to fulfil the tasks of higher education and to perform in a position requiring a diploma. Our research aims to investigate the information competences of Hungarian students in an international context.
During their studies students receive several tasks where they have to find information, quality literature, and form their own opinions based on that. The research shows that the students like these tasks and they provide no significant problems. They also had no problems in presenting the acquired information adequately with the creative use of digital technologies. Internationally true that almost every research starts on the web and often finishes there as well. The most important source of information is the open web; the primary tool used is Google or a different search engine.
National and international surveys both show that the biggest challenge for the next generation of scholars is the evaluation and selection if information. They do not possess the knowledge required to make informed decisions, and the biggest problem is that they do not feel the importance of this. It is not a question that the science-specific information culture of the students needs to be developed more efficiently.

New principles and instruments in the field of Data Protection Law

Kiss, A - Szőke, G. L.
National University of Public Service, University of Pécs (Hungary)

Background

The birth of data protection regulation in Europe was directly linked to technological developments – mainly to the impressive IT developments of the '70s and their application in public administration. These changes have challenged data protection law on every single day ever since.[54] One of the most important objectives of the data protection regulation is decreasing the information power of data controllers by providing limitations on processing of personal data and limitations of combining different databases.

Data protection Acts of the 70s, sometimes referred to as the first generation of data protection regulation, were enacted in a world where few data controllers (mostly government bodies and some major companies) used automated data processing technology, and where the general purpose was to limit the state's power by ensuring the transparency of the state's databases.[55]

In the 80s and 90s the world changed a lot – also from privacy risks' perspective. Various developments such as the spread of personal computers (PCs) and the wide usage of it by business sector imposed potential new risks. Later, from the middle of the 90s, the rapid expansion of Internet usage and the appearance of many online services set new challenges for regulators. The establishment of the "information society" became a political agenda in the European Union, and so documents were adopted in this field, all emphasizing the importance of privacy.[56] The legal regulation of personal data changed a lot in the 90's in order to face these challenges. The main result was the adoption of the European Data Protection Directive.[57]

During the last 10-15 years, there have been further significant social, economic and cultural changes which EU legislation has had to face and respond to, like web 2.0, cloud computing, ubiquitous computing, mobile data processing, new ways of profiling and Big Data.[58]

[54] Kiss – Szőke, 2015, 311-312.
[55] Jóri, 2005, 24-25.
[56] Kiss – Szőke, 2015, 314-315.
[57] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
[58] Kiss – Szőke, 2015, 316-317.

Besides these tendencies, the all-encompassing digital governance and surveillance methods grant for the state a ubiquitous monitoring option and a valuable database containing a wide range of personal data. In addition to the basic services, like online enrolment in higher education or filling in tax return, electronic government includes the use of interconnected databases, biometric identification systems (e.g. to issue personal documents), tracking systems such as RFID tags and geographical positioning systems, camera surveillance, and the collection of vast amounts of citizens' data through everyday transactions. These make governments one of the largest data processors (together with multinational companies) of the 21st century. On the other hand they have to face concern among citizens about the possibility of intrusive data collection, misuse or loss of personal information, and constant surveillance practices, just to mention some of the top barriers in the progress of electronic governance.[59]

European data protection reform

The legal framework currently in force in the EU Member States cannot provide adequate answers to the problems of mass data collection by the state and by companies, and the lack of transparency and efficiency in their processing activities, while the development of applicable information security (and identity management) is an ongoing challenge.[60]

Currently, the European data protection law is undergoing a long awaited revision. One of the most important aims of the reform is to react appropriately to the latest technological developments (like Big Data) and to the related social changes once more.

In our research, we have investigated, whether the development of data protection will be or could be able to face the challenges of the mentioned technological changes or not. As for the key outcome, it seems that a new philosophical approach is needed in the regulation. The core element of this should aim at the effective protection of individual privacy, even if subjects' privacy awareness is low, or no steps are taken by them in order to be protected. In other words, there is an elementary need for ensuring a form of »background protection«. Therefore our research team has analyzed the key elements of the Proposal for a new European Data Protection Regulation[61] in details, to decide whether it could fit into this new approach. We have concluded, that the Proposal for the GDPR is more relevant than a simple fine-tuning of existing legislation and the focus is clearly shifting to the issues of »what the data controllers shall do«, from the question of »what the data subject has the right to«."[62]

[59] Jacobi et. al., 2013, 14.
[60] Kiss, 2014, 267.
[61] European Parliament, "European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, GDPR)" (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)), http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0212 [30.06.2014.]
[62] Kiss – Szőke, 2015, 329.

Law and Technology

Theories of Lawrence Lessig argue that the Code (that means both regulation and computer program lines) has a major relevance in cyberspace. This Code, which includes software, hardware, and the entire online infrastructure, as a general rule defines cyberspace, innovations and also the permitted and forbidden behaviours.[63] However, European data controllers are obliged to process personal data in line with not just the Code, but all principles of the EU Directive on data protection and will have to face a number of new duties under the framework of GDPR.[64] Therefore the use of technologies and principles which foster the legitimate processing of data could effectively reduce the costs of meeting the obligations and the chance of being sanctioned for illegal data processing activities.[65]
For that reason, the second main goal of our research was to sum up how technological developments change the regulation of personal data processing, and what are the new principles of the GDPR that aims to regulate the issue of privacy invasive technologies and projects. The principles of Privacy by Design and Privacy by Default, including one of their key instruments, Privacy Enhancing Technologies (PETs) were the subjects of many scientific research papers of the team.

Article 23 of the GDPR impose the principle of Privacy by Design (PbD) stating that „Where required, mandatory measures may be adopted to ensure that categories of goods or services are designed and have default settings meeting the requirements of this Regulation relating to the protection of individuals with regard to the processing of personal data." PbD is the new corporate hotness affecting IT systems, business practices, and networked infrastructure.[66] It is the concept of embedding privacy into the design specification of various technologies, and also a set of rules laying down ideas of taking a proactive approach contrary to the Surveillance by Design phenomenon, appearing in the US in the 90's.[67] Although PbD was evolved originally in the technology area, its scope expanded to business practices and physical design due to the works of Ann Cavoukian.[68] This approach is applicable for all types of data processing, especially for processing sensitive data, such as health or financial data, and includes the determination of the means for data processing and at the time of the processing itself.

The meaning of Privacy by Default is similar to PbD, however needs even more attention on the data controllers' side – to protect privacy in cases when there are multiple choices to select the level of protection, but the user remains passive. In this case the default setting shall provide the highest level of protection (e.g. Do Not Track settings in Internet browsers).

[63] Lessig, 2006, 5.
[64] Balogh et al., 2014a,
[65] Rubinstein, 2012, 1411.
[66] Hill, 2011.
[67] Böröcz, 2014a, 280-281.
[68] Cf. Cavoukian, 2009.

Privacy Enhancing Technologies refer to a "system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system."[69] It is more than just a field of research in IT security, as the key goal is to enforce legal privacy principles.[70]

Although there is an overlap in the aims of PbD and PETs, the latter are clearly engineering approaches which focus on the positive potential of technology, on tools used to maintain anonymity, confidentiality, or control over personal information,[71] whilst PbD is a broader concept comprising several elements[72] to balance technologies with a framework highlighting the process and their fundamental components.[73]

On one hand side, in various projects, where the processing of personal data is necessary and have a legitimate purpose, the introduction of PETs can serve as a technological and institutional background to enforce the principle of data minimization, both at communicational and at application level (e.g. by providing anonymous communication, securing online transactions). On the other hand, some PETs are also tools for the creation and analysis of machine readable privacy policies, therefore enhancing transparency and legitimacy in data processing, facilitating users to exercise their right of informational self-determination.[74] PETs can be deployed at both ends, by controllers and data subjects, therefore with the combination of these solutions privacy enhancing identity management can be introduced for users and data controllers, tools that allow users to negotiate privacy policies with service providers.[75] However, PETs have to be developed directly to face the challenges of the invasive technologies' development, and to be more understandable for the everyday user.[76]

Data security, in close relation to privacy, is a crucial issue of data processing. A series of legal, organizational, and technical safeguards, appropriate to the sensitivity of the information are needed in order to ensure that data maintained are remaining confidential, integral and available only to authorized persons. Laws governing authorization, encryption of data and procedures to protect files are enacted differently in the EU Member States, implementing the general rules of Article 16 and 17 of the Directive, while there is a lack of cooperation between IT specialist and policy makers. The level of data security can be enhanced either by introducing obligatory compliance of data controllers to standards (such as ISO or Cobit),[77] or by the application of new

[69] van Blarkom – Borking – Olk, 2003, 33.
[70] „Data security technologies are PETs if they are used to enhance privacy. But, it should be noted that they can be used in inherently privacy-invasive application, in which case they cannot properly be counted as PETs." London Economics, 2010, ix.
[71] Rubinstein, 2012, 1412.
[72] Cf. Cavoukian, 2011.
[73] Rost – Bock, 2011, 1.
[74] Kolter, 2009, 31.
[75] Leenes – Schallaböck – Hansen, 2008, 17.
[76] Kolter, 2009, 2.
[77] Szádeczky, 2013, 153.

PETs, that also increase the level of protection of personal information by setting the protection of personal data as 'default' in different services, as a result making the use of technology one of the key elements of a suggested new European legislation.[78]

The Article 30 (1) of the GDPR also imposes the obligation on data controllers to take security measures in accordance with "taking into account the results of a data protection impact assessment". Contrary to the current regulation, data controllers and processors should adopt (written) security policies to comply with the new provisions of Article 30 (1a) of the GDPR.

Data Protection Impact Assessment

One of the most important novelties of the future data protection law is data protection impact assessment. First of all, the GDPR lays down for all data controllers the obligation to carry out a risk analysis of the potential impact of the intended data processing. If specific risks are likely to be presented by the data processing, the controllers shall also carry out data protection impact assessment and periodical compliance review. The Proposal lists the circumstances of data processing operations which are likely to present specific risks; e.g. processing of more than 5,000 data subjects' personal data, or processing special categories of personal data (sensitive data), or profiling, if it has legal effects on data subjects, automated monitoring of publicly accessible areas on a large scale (like CCTV systems), etc. For that reason the obligation to carry out data protection impact assessment concerns a well-defined, but somewhat wide range of data controllers.[79]

Carrying out a Privacy Impact Assessment is far not an easy task to do. In order to help out the data controllers in our research, we summarized the most important element of a PIA methodology,[80] based on the newest European tendencies and research results,[81] and on actually working best practices of the US, Canada and Australia. In these countries privacy impact assessment is mostly used for the data processing operations of the public sector's body and of the health care system.[82]

Generally it can be said, that a "privacy impact assessment (PIA) is a process for assessing the impacts on privacy of a project, policy, program, service, product, or other initiative (hereinafter: project) and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimize the negative impacts. The concept of PIA has been known since the mid '90s and has become progressively more common. The growing interest in PIA is caused by the robust development of privacy-invasive tools."[83]

[78] Kiss, 2013, 116.
[79] Kiss – Szőke, 2015, 321.
[80] Balogh – Böröcz – Kiss – Polyák – Szőke, 2014
[81] De Hert – Kloza, – Wright, 2012
[82] Böröcz, 2014b, 104.
[83] De Hert – Kloza – Wright, 2012, 5.

The main important elements of a PIA process are:

- Deciding whether PIA is necessary at all;
- Choosing the organization who carries out the privacy impact assessment;
- Project description;
- Pointing out potential privacy implications;
- Consultation with the stakeholders;
- Risk management (including assessment and mitigation);
- Legal compliance check;
- Drawing up recommendations;
- Reporting;
- Implementation of recommendations (including the justification of the non-implemented recommendations);
- (Periodical) compliance review [84]

In our research the characteristics and key elements of all of these steps have been summarized, taking into account the planned provisions of the GDPR. As a result of the project a detailed questionnaire with 29 exact questions have been worked out, which may be applied in various PIA processes. Thus, these results are widely applicable for data processing activities of data controllers.

Summary

Our research in the past two years has focused on the new tendencies and new legal institutions in the field of data protection law. We have analyzed the European data protection reform, including the most important output of this process, the Proposal for new Data Protection Regulation. We have concluded, that the Proposal for the is more relevant than a simple fine-tuning of existing legislation and the focus is clearly shifting from the rights of data subjects to the compliance duties of the data controllers. We have also shown the increasing role of the technology as a regulatory means, and the emerging of a new principle calls "Privacy by Design". "Setting strict rules for data controllers on applying technologies for personal data processing fits the concept of a paradigm shift; it can be seen as a change in balancing responsibilities from the data subjects' informational self-determination towards an automatic protection."[85] Finally we've put the focus of the research activity on the new legal instrument of "privacy impact assessment", and have summarized the most important steps to carry out such process. These results are widely applicable for data processing activities of data controllers and offer an actual help for them to increase the level of data protection, which totally fits to our original research goals.

[84] Balogh – Böröcz – Kiss – Polyák – Szőke, 2014, 80., De Hert – Kloza – Wright, 2012, 27-32.
[85] Kiss-Szőke, 2015, 12.

References

Balogh Zsolt György – Kiss Attila – Polyák Gábor – Szádeczky Tamás – Szőke Gergely László (2014a): Technológia a jog szolgálatában? Kísérletek az adatvédelem területén. Pro Futuro - A Jövő Nemzedékek Joga 4 (1), pp. 33-46.
Balogh Zsolt György – Böröcz István – Kiss Attila – Polyák Gábor − Szőke Gergely László (2014b): Az adatvédelmi hatásvizsgálat módszertana, Médiakutató, 2014/4, pp. 77-92.
Böröcz István (2014a): The tools of privacy protection at the dawn of a new era. Privacy by Design in theory and in practice. In: Alexander Balthasar, Hendrik Hansen, Balázs Kőnig, Robert Müller-Török, Johannes Pichler (eds.): Central and Eastern European eGov Days 2014 - eGovernment: Driver or Stumbling Block for European Integration. Austrian Computer Society, Wien, pp. 269-282.
Böröcz István (2014b): A Privacy Impact Assessment forrásai, *Infokommunikáció és jog 59,* 2014/3, pp. 103-110.
Cavoukian, Ann (2009): Privacy by Design. …Take the challenge. Information and Privacy Commissioner of Ontario, available at: http://www.privacybydesign.ca/content/uploads/2010/03/PrivacybyDesignBook.pdf [13/11/2014]
Cavoukian, Ann (2011), "Privacy by Design, The 7 Foundational Principles," available at: http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf [13/11/2014]
De Hert, Paul – Kloza, Dariusz – Wright, David (2012): Recommendations for a privacy impact assessment framework for the European Union. http://piafproject.eu/ref/PIAF_D3_final.pdf [2013.07.17.]
Hill, Kashmir: Why 'Privacy by Design' is the New Corporate Hotness, 2011, Available at: http://www.forbes.com/sites/kashmirhill/2011/07/28/why-privacy-by-design-is-the-new-corporate-hotness/ [13/11/2014]
Jacobi, Anders. – Jensen, Mikkel Lund – Kool, Linda – Munnichs, Geert – Weber, Arnd (2013): Security of eGovernment Systems, Science and Technology Options Assessment, European Union, Brussels, http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/513510/IPOL-JOIN_ET%282013%29513510_EN.pdf [31.01.2014]
Jóri, András (2005): Adatvédelmi kézikönyv, Osiris, Budapest.
Kiss Attila (2013): A privátszférát erősítő technológiák, *Infokommunikáció és jog 56*, 2013/3, pp. 113-120.
Kiss, Attila (2014): The European data protection legal framework in relation to e-Government. In: Alexander Balthasar, Hendrik Hansen, Balázs Kőnig, Robert Müller-Török, Johannes Pichler (eds.): Central and Eastern European eGov Days 2014 - eGovernment: Driver or Stumbling Block for European Integration. Austrian Computer Society, Wien, pp. 253-268.
Kiss, Attila – Szőke, Gergely László (2015): Evolution or Revolution? Steps Forward to a New Generation of Data Protection Regulation, In: Gutwirth – Leenes – De Hert (eds.): Reforming European Data Protection Law, Springer, pp. 311-332.

Kolter, Jan P. (2009): User-Centric Privacy. A Usable and Provider-Independent Privacy Infrastructure. Dissertation, University of Regensburg, available at: http://www.ics.uci.edu/~kobsa/phds/kolter.pdf [13/11/2014]
Leenes, Ronald - Schallaböck, Jan - Hansen, Marit (2008): PRIME white paper. Third and final version, available at: https://www.prime-project.eu/prime_products/whitepaper/PRIME-Whitepaper-V3.pdf [13/11/2014]
Lessig, Lawrence (2006): Code, Version 2.0, Basic Books, New York. http://codev2.cc/download+remix/Lessig-Codev2.pdf [13/11/2014]
London Economics (2010): Study on the economic benefits of privacy-enhancing technologies (PETs). Final Report to The European Commission DG Justice, Freedom and Security, available at: http://ec.europa.eu/justice/policies/privacy/docs/studies/final_report_pets_16_07_10_en.pdf [13/11/2014]
Rost, Martin – Bock, Kirsten (2011): "Privacy by Design and the New Protection Goals," available at: https://www.european-privacy-seal.eu/results/articles/BockRost-PbD-DPG-en.pdf [13/11/2014]
Rubinstein, Ira S. (2012): "Regulating Privacy by Design." *Berkeley Technology Law Journal* 26, pp. 1410–1456.
Szádeczky, Tamás (2013): Az IT biztonság szabályozásának konfliktusa, *Infokommunikáció és jog 56*, 2013/3, pp. 151-155.
Van Blarkom, G. W. − Borking, J. J. − Olk, J. G. E. (eds.)(2003): Handbook of Privacy and Privacy-Enhancing Technologies. The case of Intelligent Software Agents. College bescherming persoonsgegevens, The Hague.

Data Protection Law in the age of Big Data

Böröcz, I - Kiss, A.
University of Pécs, National University of Public Service (Hungary)

Due to the evolution and spreading of informational and communicational technologies the legal protection of personal information and privacy has become one of the most important and rapidly developing topics of the 21st century. This paper will guide us through the latest threats of online data processing, especially behavioural advertising and the Big Data with its – both structured and unstructured – data collecting and data evaluating techniques, including the privacy aspects of Big Data. During the project these methods and technologies were researched on the basis that we accept their presence in the cyberspace without arguing for or against them, but searching for the possible ways to ensure our privacy meanwhile their application by private companies.

The nature of data in the age of cloud computing

Nowadays personal data can be interpreted as the „*new oil of the internet and the new currency of the digital world*",[86] a raw material for organisations to produce a new form of economical value.[87] Data is not a static, obsolete thing anymore that is only important for a single purpose, while it is recyclable so as may serve as a resource of new services based on correct attitude. Due to the technological innovations society tries to react and learn how to use the new tools which make life easier. This led us to an unending process where data has to be generated from everything in order to evaluate them and create services, which upgrades the uncomfortable or ineffective tools in everyday life, such as a navigation system, a traffic rationalising algorithm or a personalised product range in a webshop. The new analytical methods and evaluating techniques are more accurate than ever before, but they generate numerous negative side-effects on the data subjects as well. The project team had the aim to prove that in the age of Big Data privacy, as a fundamental right of the persons, and the protection personal data still could be granted by new (or redefined) principles of data protection.

Characteristics of Big Data

The Article 29 Data Protection Working Party described the Big Data phenomenon as " *a great number of data processing operations, some of which are already well-identified, while others are still unclear and many more are expected to be developed in the near future."*[88] The

[86] Kuneva, Meglena (2009): Keynote speech. Roundtable on Online Data Collection, Targeting and Profiling. Brussels, 31 March 2009. http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm [22/11/2014]
[87] Mayer-Schönberger, Viktor – Cukier, Kenneth (2014): Big Data, HVG Könyvek, Budapest, p. 13.
[88] Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU, WP 221, 2014, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf [18/11/2014]

definition is quite inaccurate, however it could serve as a common core for the more than forty existing approaches of Big Data.[89] The term was mentioned first by NASA scientist in 1997,[90] and appeared in Oxford English Dictionary only in 2013, which clearly shows its novelty and the rapid changes in this field. A commonly used definition is the „4V" given by Gartner Inc., an information technology research and advisory company, stating that business and IT leaders must focus on information volume, variety, velocity and veracity in order to manage Big Data. It can be argued therefore that it's not merely the size and amount of data that counts, but the tools being used or the insights that can be drawn from a dataset[91]

In the 21st century organisations started to create, collect and store an incredible amount data, despite the lack of purpose and capability to process it.[92] Leaded by IBM, Google and Yahoo, new data mining processes were developed in the past few years (e.g. Apache Hadoop[93], Infosphere BigInsights[94], MapReduce[95]), as the constant growth of data analytics evoked a change: from quantitative change to qualitative change. With Big Data analytics the large amount of data can be reused in different context and provide an improved result in every field of life.

The process of Big Data analytics can be separated into three stages: collecting data, analyzing and utilising the outcomes.

The source of data may vary, however IBM's survey showed, that the most commonly analysed data sources are transaction data, log data, event data and emails.[96] With the outcomes Big Data scenarios appear in many industries:

- Multi-channel customer sentiment and experience analysis;
- Detection of life-threatening conditions at hospitals in time to intervene;
- Prediction of weather patterns to plan optimal wind turbine usage, and optimise capital expenditure on asset placement;

[89] Dutcher, Jenna (2014): What Is Big Data? http://datascience.berkeley.edu/what-is-big-data/ [18/11/2014]
[90] Press, Gil (2014): 12 Big Data Definitions: What's Yours? http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/ [18/11/2014]
[91] Dutcher, Jenna (2014): What Is Big Data? http://datascience.berkeley.edu/what-is-big-data/ [18/11/2014]
[92] Szőke, Gergely László (2013): Az adatvédelem szabályozásának történeti áttekintése *Infokommunikáció és jog 56*, 2013/3, p. 110.
[93] Apache Hadoop is an open-source software framework for distributed storage and distributed processing of Big Data on clusters of commodity hardware, created in 2005. Read more at: http://hortonworks.com/hadoop-tutorial/hello-world-an-introduction-to-hadoop-hcatalog-hive-and-pig/ [21/11/2014]
[94] InfoSphere BigInsights is built on the Apache Hadoop software framework, the open source technology for reliably managing large volumes of data. Read more at: http://www-01.ibm.com/software/data/infosphere/biginsights/index.html [21/11/2014]
[95] MapReduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster, created in 2004. Read more at: http://research.google.com/archive/mapreduce.html [21/11/2014]
[96] IBM Institute for Business Value: Analytics: The real-world use of big data, 2013, p 11., available at: http://www-03.ibm.com/systems/hu/resources/the_real_word_use_of_big_data.pdf [21/11/2014]

- Make risk decisions based on real-time transactional data;
- Identification of criminals and threats from disparate video, audio, and data feeds, and much more territories data analysis is useful.[97]

The two main concepts of collecting data are select before collect and collect before select.[98] With select before collect data controllers decide prior to the collection what type of (personal) data they want to process, and collection just starts afterwards. Collect before select covers the opposite approach: "*In the old, data-is-scarce model, companies had to decide what to collect first, and then collect it, but with the new, data-is abundant model, we collect first and ask questions later.*"[99] The latter is an economically beneficial method – it is much easier to create profiles about the users and provide personalised content for them.

Big Data as a threat to privacy

The European Union set a baseline common level of privacy in the Member States in 1995 with Directive 95/46/EC.[100] The directive reinforced national laws and introduced a range of new rights applying for both electronic and manual processing of personal data. Its key concept was to provide enforceability to data protection, therefore the rights of the data subjects were given explicitly to be implemented to national regulation, and data protection agencies had to be set up in all countries of the Union to enforce those rights.[101] Since the Lisbon Treaty entered into force in 2009, the protection of personal data was recognized as a fundamental right,[102] so European citizens have a right to data protection, even in the absence of specific rules.[103]

The common principles in European data protection laws require that personal data shall be obtained fairly and lawfully; used only for the original specified purpose;

[97] Viale, Joel (2011): BAO & Big Data Overview Applied to Real-time Campaign. p. 6., avalable at: https://www.ibm.com/developerworks/community/files/form/anonymous/api/library/d210b4b1-0df8-4b7a-9e7a-f22a8a958b01/document/b6bdf989-12b3-44a6-82a9-e7ac8805f99a/media/Big%20Data%20%26%20Real-time%20Campaign_GSE%20-%20External.pdf [24/11/2014]

[98] Hijmans, Hielke (2010): Recent developments in data protection at European Union level, p. 222., available at: http://link.springer.com/content/pdf/10.1007%2Fs12027-010-0166-8.pdf [23/11/2014]

[99] Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It, 2012, available at: http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/ [21/11/2014]

[100] Directive 95/46/EC of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (hereinafter: Data Protection Directive).

[101] Öqvist, K. L. (2009): *Virtual shadows: Your privacy in the information society.* (Swindon, British Computer Society, p. 113.

[102] Article 16 of the Treaty on the Functioning of the European Union

[103] http://epic.org/privacy/intl/lisbon_treaty.html [20/11/2014]

adequate, relevant and not excessive to purpose; accurate and up to date; accessible to the data subject; kept secure; and destroyed after its purpose is completed.[104]

The key problem with Big Data is that analyzing processes may affect a wide range of subjects, as in most cases data is created and shared by users/customers for their own purposes, not necessarily in connection to the goals or needs of the Big Data analyst. Surveys have shown that almost half of the organisations try to use Big Data to provide customer-centric outcomes even from the anonym or non-personal data collected. After these are collected and analysed by Big Data methods at the end of the process the result will most likely contain valuable information on the costumers.[105] These profiles are used in different fields, for example behavioural advertising, where only personalised advertisements are shown to the user which are close to his field of interest. Personalisation increases user experience, but also the chances that the targeted groups will click on a personalised advertisement and buy goods therefore realizing profit on the side of data collecting organizations. This is also a brief description how Google AdWords work.
With these profiles plans, future behaviour of the users can be predicted violating their free will and dignity.[106] According to Székely, the users think they still control the information they share and receive in the cyberspace, but in reality they will see the world only behind a filter bubble produced by the results of the above mentioned analysis.[107] Before users got behind a filter, they had the right to define who they are, what they like, but after behavioural analysis they are defined by a profile. They are told who they are and what they like.[108]

However, the value of more than 500 million European citizens' data makes it worth for companies collecting and using them (e.g. Google's total revenue from advertising was more than 50 billion USD in 2013),[109] even unlawfully for their business purposes without the consent or the notification of the data subjects.[110] This is the processing method why data protection has to intervene and prevent unauthorised data processing.

[104] Ramage, S. (2007): *Privacy – Law of Civil Liberties*. p. 7., available at: http://books.google.hu/books?isbn=0595449018 [20/11/2014]
[105] IBM Institute for Business Value (2013): Analytics: The real-world use of big data, p 7., available at: http://www-03.ibm.com/systems/hu/resources/the_real_word_use_of_big_data.pdf [21/11/2014]
[106] Balogh, Zsolt György (2004): Adatvédelem és információszabadság, Fundamentum, Issue 4., p. 56.
[107] Székely, Iván (2013): Jog ahhoz, hogy elfelejtsenek és töröljenek. Információs társadalom, Issue 3-4., p. 11.
[108] Richards, Neil M. – King, Jonathan H. (2013): Three Paradoxes of Big Data. Stanford Law review Online, Issue 41, pp. 43-44., available at: http://www.stanfordlawreview.org/sites/default/files/online/topics/66_StanLRevOnline_41_RichardsKing.pdf [24/11/2014]
[109] https://investor.google.com/financial/2013/tables.html [25/11/2014]
[110] http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf (2012.10.28.) p. 5.

On the other hand, privacy-centric thinking is essential for an organisation, since this can keep users' trust and *„without it, technologies cannot move forward"*.[111] With Privacy by Policy approach data controllers trust the organisation, while with Privacy by Architecture approach they trust the system. Trust (of the policy) may be increased by the data controllers by auditing and certifying their data processing operations.[112]

Data protection legislation regarding Big Data

The Proposal for a General Data Protection Regulation[113] aims to reflect to the threats of Big Data to European citizens' private life by laying down general rules concerning profiling.[114] Article 4 Section (3a) of GDPR defines that profiling means *„any form of automated processing of personal data intended to evaluate certain personal aspects relating to a natural person or to analyse or predict in particular that natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour"*.
The definition and further provisions of the Regulation seem to be clear, however the applicability of the Regulation (or the 95/46/EC Data Protection Directive) to many forms of Big Data is not unambiguous.

Article 2 (1) highlights that GDPR "*applies to the processing of personal data wholly or partly by automated means, irrespective of the method of processing.*" Users, however leave mostly digital footprints that are not regarded as personal data, but by recombining these footprints Big Data analysts can gain personal information. Cavoukian also argued that de-identified data can be reused by third parties, and nothing ensures that even the correct de-identification – combined with quasi-identifiers, will prevent the data subject to be identified again.[115] An often cited study from Sweeney proves that 87% of the U.S. population can be identified just by using three types of data: gender, date of birth and postal code of the subjects.[116] This rate is incredibly high, but without a name, a picture, or any kind of additional quasi-identifier, it is still not known who the exact person these

[111] Hill, Kashmir (2011): Why 'Privacy by Design' is the New Corporate Hotness http://www.forbes.com/sites/kashmirhill/2011/07/28/why-privacy-by-design-is-the-new-corporate-hotness/ [25/11/2014]
[112] The planned new General Data Protection Regulation aims to promote the auditing and certification schemes by various new provisions. Cf. Szőke, Gergely László (2014): Az önszabályozás, audit és tanúsítás lehetőségei és korlátai az adatvédelem területén, *Infokommunikáció és jog 57*, p. 17.
[113] European Parliament, "European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, GDPR)" (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)), http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0212 [10/11/2014]
[114] Article 20 of GDPR
[115] Cavoukian, Ann (2014): Big data innovation, Setting the Record Straight: De-identification Does Work, Ontario, p. 9. available at: http://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identification_ITIF1.pdf [30/10/2014]
[116] Sweeny, Latanya (2000): Uniqueness of Simple Demographics in the U.S. Population, Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, available at: http://dataprivacylab.org/projects/identifiability/paper1.pdf [22/11/2014]

data refer is, but databases established after the spreading of Big Data methods will overcome this 'problem'.

Due to that direct and indirect identification is a subject of vivid discussions, and it is also a relevant issue of the Directive currently in force. In case of direct identification data controller is able to restore the connection between data and the stakeholder with its own tools, without an exceptional effort. Indirect identification means the connection has to be theoretically replaceable. A person is identifiable by any process the data controller would reasonably use.[117] The main criterion of indirect identification is to create unique combinations. Although single identifiers may not allow identifying someone, but combined with other identifiers the new information is able to distinguish a person from others.[118] This new information becomes personal data.

The intention of the data controller is also significant. If the data controller is intended to identify the person, the data will count as personal data, but if not, the technical measures of prevention become very important.

If personal data were removed properly the organisation will not process personal information, meaning it will not be a subject of the Directive.[119] Various de-identification techniques were developed in order to achieve this goal, such as obfuscation, spatial and temporal aggregation, replacing direct identifiers with irreversible pseudonyms, or removing uncommonly high values in the dataset (top-coding).[120]

By using Big Data algorithms it is possible not just to re-identify anonym data, but also to create personal related information from raw data without any intention. Based on the '5V' and the theory of indirect identification every data can be considered as personal data. The Hungarian Data Protection Authority stated that the processing of encrypted data is out of the scope of the Hungarian data protection regulation if it does not affect the stakeholder.[121] If we take the invert meaning of this opinion and encrypted data would have an effect on a stakeholder, it would become personal data. This expands the scope of personal data enormously and paralyses data protection law. Big Data highlighted several unclear elements of data protection law, but in the past few years new principles and solutions appeared which are able to give a feasible answer to the current questions.

[117] Data Protection Directive recital (26)
[118] Opinion 4/2007 of the Article 29 Data Protection Working Party on the concept of personal data, WP136, 2007, p 13., available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf [05/11/2014]
[119] WP136. p. 17.
[120] Cavoukian, Ann (2014): Big data Innovation, Setting the Record Straight: De-identification Does Work, Ontario, 2014, p 10., available at: http://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identification_ITIF1.pdf [30/10/2014]
[121] NAIH-2512-2/2012/V Decision of the Hungarian National Authority for Data Protection and Freedom of Information, available at: http://www.naih.hu/files/2512_V_2012-2.pdf [11/22/2014]

Possible solutions under the GDPR

The GDPR emphasises that by the application of both old and new principles of European data protection, such as purpose limitation, data minimisation, and transparency, Privacy by Design or Privacy Impact Assessment the liabilities of the data controllers can be tightened, therefore limiting the unlawful usage of Big Data methods. Data Protection Impact Assessment was introduced to GDPR in order to form a lifecycle data protection management system for all privacy invasive projects, and assess the impacts on the rights and freedoms of the data subjects.[122]

By introducing the principles of Privacy by Design and Privacy by Default to data protection regulation, organisations will be guided to take a proactive approach to privacy, and that will affect IT systems, business practices, and networked infrastructure.[123]

As a conclusion we have to highlight that the new principles and tools of data protection that could be introduced in order to protect users and consumers from Big Data exploiting methods will enter into our legal systems only after the Council of the European Union and the European Parliament manage to forge a consensus on the goals and the exact wording of GDPR, which could be achieved only in the long run.

[122] Article 32a Section 2 of GDPR.
[123] Cavoukian, Ann (2014): Privacy Engineering: Proactively Embedding Privacy, by Design, Ontario, p 3., available at: http://www.privacybydesign.ca/content/uploads/2014/01/pbd-priv-engineering.pdf [19/11/2014]

On the Moral Foundations of Democracy

Boros, J – Jenei, S – Bretter, Z. – Kruzslicz, F. – Zuh, D – Horváth – Kocsis, L – Pete, K. – Di Blasio, B. – Hornyák, M.
University of Pécs (Hungary)

Introduction

This research was focused mainly on the moral concepts which genereally occur in the laws of a democratic society, and investigated the issue of moral foundations of it. Our question was: whether the laws of a democratic society can prove that it is founded in morality, or at least that a certain set of laws can underpin the efforts of the lawmakers to constitute and enact laws on moral principles.[124] Our research admitted the premise that the political frame of an Info-communicational Society (ICS) should be democracy only and vice versa: in the modern age democracy can function only as a certain kind of info-communicational community. Democratic societies are societies of trust respectively safety, and are built upon the continuous, transparent, rational, effective and scientific forms of communication, which all use the outcomes of recent developments in information technology. The free flow of information and the basic openness of a democratic infrastructure is the first condition of research and innovation, subsquently of trust and safety. The comprehensive theory and the wider range of practices of democratic society is inconceivable without informatical approaches, and even more ICS is an inomissible constituent of social issue management. This means in addition that ICS must be submitted to a rigorous legislature to prevent the abuse of power through controlling and manipulating information.

Friedrich A. Hayek warns us in his remarkable classic *The Road to Serfdom,*[125] that moral foundations oftenly lead to a different kinds of servility. The core of legality is not in the content of a specific regulation or law, which would determine the actors how to behave, jow to react in a certain situation. Laws should only mark formal crutches, or more succintly: points of orientation towards a more synthetic conception of (social) justice. Formal law is – in Hayek's view – synonym for justice. He emphasized, that "it may even be said that for the Rule of Law to be effective it is more important that there should be a rule applied always without exceptions, than what this rule is. Often the content of the rule is indeed of minor importance, provided the same rule is universally enforced."[126] A formal law is fully rational, and this is sufficient to be just and rightful. It is also true, that a modern dictatorial system usually passes laws which are charged with (sometimes emphatically stressed) moral words – an even more as the laws of democracy. In a further study the researchers should turn their attention to the ways democratic legislature transforms moral principles in a highly rational vocabulary.

[124] For the historical background of this thesis see section no. 6 of present study.
[125] Friedrich A. Hayek: *The Road to Serfdom*. Text and Documents. University of Chicago Press, 2007 [first published in 1944].
[126] Ibid. 83.

Constituting, transmitting and processing data has reached a higher gear, which continuously influences society as a whole: the moods of thinking at the individual level, the everyday rituals of social actors, circles of friends, public institutions, scientific research, everyday jurisprudence. The basic task is to follow the development of democracies on the field of traditional moods of communication from the early ancient times of the greek polis to the Declaration of Independence of the United States and to contrast it with all the recent advances in information technology and the restructured levels of societies after WW II. A comprehensive research should unfold the following three problems: 1.) which social changes have occured due to the new media of communication?; 2.) is democracy weakened of reinforced by the fast transmission of information?; 3.) should we see a menace to democracy in the new techniques of communication, and if so, how could we organize and set out the adequate self-defense? To support the research marked by principles mentioned above we constituted a seminal data base of moral concepts through the use of IT data mining methods, and submitted it to analysis.

2. The Democracy Database: a specific goal

Our database could be a serious help in comprehending the basic rules of an infocommunicational society that will be able to present itself as a democracy. Such data bases could find their way not only to the main curricula of higher education, but could also catalyze the legislative practice.

Our research group has developed solid interest in answering the following question: did Hungarian political legislature consider moral principles in the enactment of laws in the period between 1990–2013, namely in the last 24 years. Are the laws of our present democracy founded in morality? This undertaking is philosophical, therefore our practical orientation had to be completed with the theoretical background of the material analyzed.

The philosophical hypothesis against which our study is to be situated is that democratic legistrature is inconceivable without a basis made up of moral principles, a basis that has to be respected in the process of enforcing these laws. The existence of moral principles is our strong philosophical conviction, which has to be investigated. We suppose that there are certain social moral principles, or to put it differenty, that every society has a „Moral Codex". But we are at the horns of the dilemma, whether the legislators would respect and codify this in their law-making practices, respectively would their moral law-making activity exert a certain influence on wider categories of society. If legal moralism is to characterize the will of legislators, then how strong this will really is? What the content of legal moralism is concerned, we have to differentiate between two specific tendencies which mirror the influence of two strong social philosophies: *indiviualism* and resp. *collectivism* in legislative subject matters, where the latter explicitly tends to strenghten the force of the so-called „Moral Codex" and to put it in concrete legislative fomulas. Legal moralism is much more well-set in a collectivist-based communities, which by no means leads to a collectivist monopoly of moral issues.

Individualists also accept the existence of a Moral Codex, they just they don't want to compell legislators to transplant it to the realm of everyday jurisprudence.

The proper functioning of a democracy is secured by the authority of laws, consequently democratic functioning should respect those moral principles which have found their way to the text of concrete legislative products and assure the social setting and the welfare of those who take part in. But this is neither guarantee for the permanent application of moral principles in law-making actions, nor the warranty seal which prevent a moral set-up to be distorted in a specific factual case of applying laws. Moral foundations of a democracy thus could be investigated on a narrower field, on the practice of law-making. This is the very area, we are approaching now.
Our project is based on already existing data, we had not a single task of law-making or working on the provisory laws, which have to be summitted and adopted. Existing information then had been investigated through IT-data mining programs, which had to be adequate to our case. Data mining as we see it, is a certain form of text mining, which raises an immediate question: how could it prove that democracy is *sufficiently* grounded in moral principles? There is not doubt that has been cast upon the idea, that democracy *should* and *ought to* be morally founded, but the issue of its range and strength can't be clarified not even after the research was completed. But we can show up some useful hints on the character of moral foundations.

The first step was to found the adequate corpus of laws, namely the proper set of data which will have a kind of mathematical-statistical analysis prepared for further investigation, and which will have pinned down the main points of hermeneutical activity. The text basis consisted of Hungarian laws passed between 1947 and 2012. We innitially wanted to analyze both the specific text of the laws (content of the proper paragraphs), and their full justification. These are both essential to the extent of morality in these laws, to the extent of moral principled used by the law-makers. It is true that *just* the content of laws is not sufficient to excavate their moral springs, but in the mathematical analysis we concentrated firstly on the laws themselves to draw the first order conclusions. We have crafted a four-way divide of categories: Hungarian laws could be categorized in four types: of *autochtone or home* initiative, of *EU-law-harmonization*, of domesticating and proclaiming the *consequences of an international treaty or engagement* through the civil code and finally the set of *the Constitution and the pivotal laws adherring to that* (we were considering both the Constitution of 1990 and 2011). Then finally we used an other standpoint: we categorized laws after the parlamentary or electoral period in which they were formulated and passed. Legislative periods are crucial to characterize the *style* of law-making (or at least the elements of such styles) in a certain period of time, and definitely more comprehensive then the category or type the laws they belong to.

3. Preparation and preliminary results

After setting the goals of our researh and discussing the principled questions we prepared the texts for informatical processing. The task was simple, to read as much laws as possible and to define a set of moral expressions prevalent in them therefore to

constitute a special golossary which help us to investigate the final prevalence index of these syntagms. Rading 42 Hungarian laws we knew the basic moral concepts they purport (128 moral concepts or concepts with moral connotations), and then we generated the list of concepts for further analysis. Afterwards we stipulated a list of 141 laws to be contrasted with the glossary just constituted and this lead to determine the specific morality index of these laws (we analyzed the occurence of these 128 concepts in the overall word count of the relevant texts). The highest rate of moral phrasing was reached by the law no. CXI of 2011 on the ombudsman or public advocate reaching up to 13.82 %.

At this stage we have taken two problems into consideration: 1. Laws they use moral concepts, but the cosequent and pretty frequent use of them is not a proof of the relevant subject matter, that they are built on moral foundations – the less pretentios thesis, that these laws pursue moral goals is nevertheless plausible; 2. Some moral (or seeming to be moral) words are as much dependent on the context of their usage, that we have to deliber the terms of text mining techniques (and the applied statistical skills) very carefully. These problem still remain to be solved.

Let just take the sneak preview of a further case study, which underscores the above mentioned problem. In most of the cases laws contain moral words, but there are serious exceptions such as the rules of the road or the highway code. Notwithstanding with our first impressions, these exceptions show up the limits of our survey, because some of the traffic rules are perfectly apt to signal the case of moral law-making. A classic example is the issue of the safety belt. At a superficial glance it is a sheer practical case, but making compulsory to fasten it was contrary to some libertanian moral principles, which put a serious weight on personal responsability and stated that rules are not made up to save the individual from him/herself. Seatbelt fastening regulations can invoke utilitarian arguments too: after a watchful calculation of the health and car insurance acquittances, respectively the costs of personal (sometimes lethal) damages, we can infer, that the utilitarian interest of all societies is to prevent car accidents and to sink collateral damages to the minimum. But these considerations could be formulated without using a single moral word in the very technical texts we read.

The use of moral words in Hungarian legislature shows afterall a very mixed picture: from 1990 their prevalence is much more excessive, but the year 1998 was even more salient. The wider qualitative investigation then led to the following results:

4. Qualitative survey

In our investigation we evaluated the proportion of moral words (respectively which could be morally contextualized) in the text of specific laws. We tried to prove or refute the following two theses:

T_1: The proportion of the used etical/moral words is changing periodically.
T_2: Examinig morals words, their prevalence and context of use benefits the scientific analysis of a certain age in legislature.

The first technical step was the choose of the target and of the right samples. We entered two rounds of gaining samples (we'll unravel the causes of this self-correction below). To gain samples we downloaded the online assets of laws. The downloaded files were then cleared of the non-usable font-sequences (e.g. the HTML-terms), then we executed the NLTK text mining algorythm of the Python language on this raw stock of files. From the preliminary data we substracted the list of those words, which occur frequently in the Hungarian text of laws (this is called: finding the „stop-words"): like conjunctions and articles. Then we formed a matrix, where one row consisted of the documents analyzed and the other of the word which are relevant for us. The elemts of the matrix show the frequency of the word in the documents (this is a TDM-matrix). For better handling we transponed the TDM-matrices in the analyses. In our research we used those words as means of control (evidently leaving out of consideration the results themselves), that we preliminarily defined as „connected to the realm of morality".

The first sampling took as a starting point the 1949 Constitution of Hungary (a Charta of constituting a new state) and followed the laws till 2012 (the year just before the start of our research). The choose of the laws was random. These texts were donwloaded from a free-access online sources. The first analyses they stated that however there are some words that randomly occur from one year to another such as „loyalty" (from figure No. 1), but „proud", „morals" and „prostrate" just act in the same way.

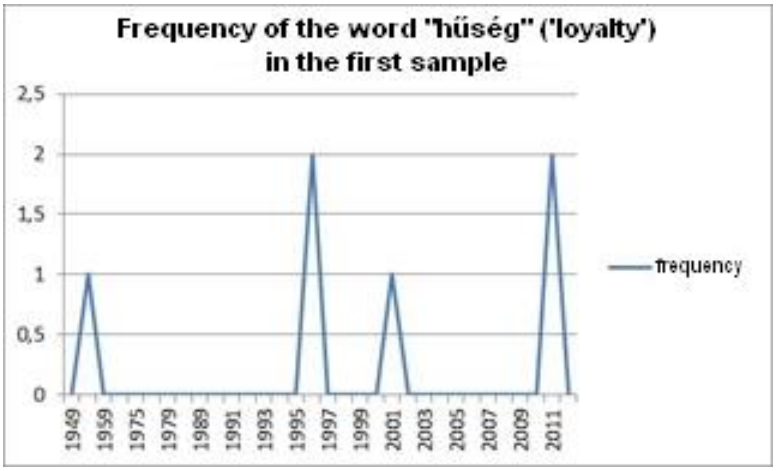


Fig. No.1

Notwithstanding with all these facts, some concepts, words reflecting a moral vocabulary practically occured only after the change of the political system in 1989, or when occuring in isolated cases before 1989, it was used then regularly the 1990s.

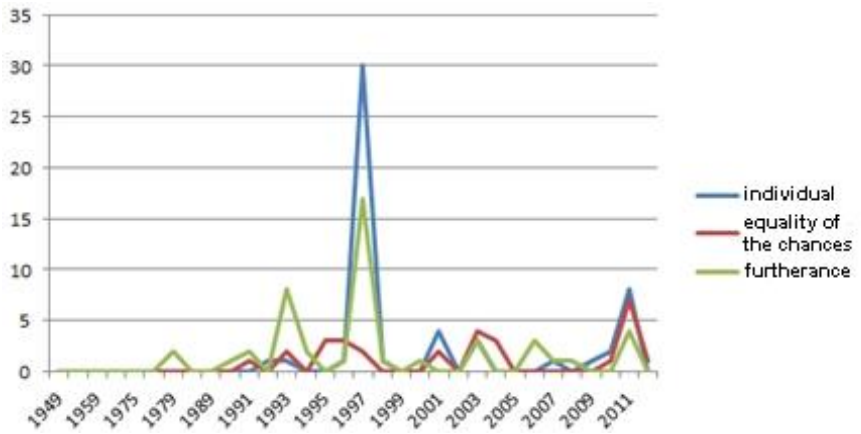


Fig No. 2

Dealing with the first set of data motivated us to look for a diferrent source of samples, because we left some points out of consideration.
We preceived it only posteriorly, that the texts of laws are accessible to us in their last and updated operative form. In cases of the Constitution/Fundamental Law, Penal and Civil Codes, which were modified in the last decades almost a hundred times each, our current samles were certainly biased by the guise of actuality.
It was also a problematic point, that we sampled a series of law-like orders decreed befor 1989 by the communist gevernment without the de facto acceptance of the general assembly. These orders could be interpreted in a historical investigation, but not in a legal mood of speaking, were we want to study the morailty of a democratic system.

We have trated these issues in the following way: we gained a different set of samples to redemedy the flaws of past researches. Then contacted Wolters Kluwer Ltd. to access the full database of Hungarian legislature, in order to call down all the historic tokens of currently effective laws. We followed the strategy of pinning *nine* consecutive periods and respectively pairing them with the form that it fit at the period investigated. The legislative texts were exclusively laws, law-like orders were rigorously excluded (even if on a certain point and in certain areas law-like orders was running parallelly to the specific law passed by a parliament). Laws were then original texts, not modifications of existing ones. These laws were the following:

a. Constitution/Fundamental Law
b. The law of the order of penal processes
c. Penal Code
d. Family Code
e. The law on the right of assembly
f. National defence code
g. National education code
h. Labour code
i. Civil code
j. The law on the order of civil lawsuit
k. The law on copyright

l. The law on strike
m. The law on social security
n. The law on local self-governments
o. The law on religious freedom

The dates marked here represent the end of a legislative period. These are mostly elective periods which were completed by marking two significant events (in italics):

- *23 October 1989*: *proclamation of the Third Hungarian Republic*
- 23 May 1990: accession to office of the Antall-government
- 15 July 1994: accession to office of the Horn-government
- 27 August 1998: accession to office of the first Orbán-government
- 27 May 2002: accession to office of the Medgyessy-government
- *1 May 2004: Hungary joining the European Union*
- 9 June 2006: accession to office of the second Gyurcsány-government
- 29 May 2010: accession to office of the second Orbán-government
- 6 June 2014: accession to office of the third Orbán-government

The whole surveyed period is depicted in the following diagram:

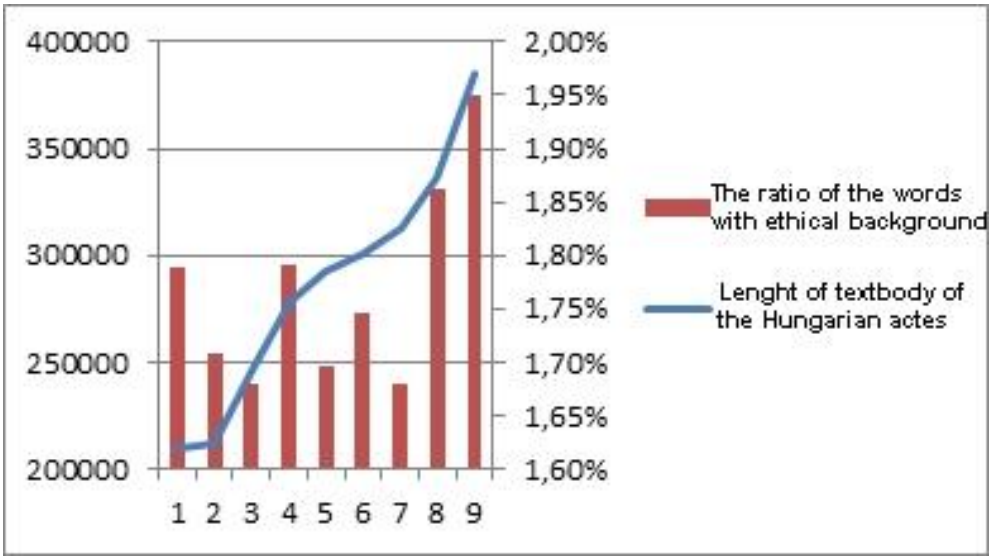


Fig. No.3

It is worth noticing that the total count of words in laws from different periods is facing a slow but strictly monotonous growth (between the first and the nineth period has increased by 83,44%), but the distribution of moral terms, words and expressions in them is very unbalanced and divergent. A far reaching conclusion can't actually be drawn. Starting from 1990, we can diagnose, that the left-wing goverments pursued in their law-making strategies a kind of *Weltanschauung*-neutrality, and they sunk the proportion of ethic-moral words (in the 2nd, 5th and 7th period), but after joining the EU it became higher one more time (and totally ignoring the political pedigree of a government officials) – see the 6th period.

After their prevalence we can divide the analyzed words in five categories:

1) The words of marginal importance, the frequency of which is between 0 and 10 unities in a given period, without any substantive tendency. These are „protection", „fair", „negation", „love", „suffering", „to suffer", „poor", „rejuveniation", „preservance" (it appeared only in the penultimate period), „common good", „christianity", „wealth", „proud", „prostrated", „require", „value", „equality of chance", „diligence", „enlargement", „faith", „promise", „truth", „fidelity".

2) The words with a slightly grown importance were the following (those, that occur more then 10 times in a certain period and they are subjected to a slow multiplication as time passes evidently not reaching the proportion of 83,44 %.): „do". „gain", „liberty", „damage", „responsibility", „to miss".

3) The words with a slightly sinking importance: their prevalence is marked by a receding tendence: like those of „labour" (1st periоad at 37, in the 9th at 12); „step" (1st period at 20, 9th period at 14); „obligation" (1st period at 112, 9th period at 80) and finally „family" (1st period at 342, 9th period at 158).

4) The words with a grossly rising importance: the occurence of which is higher than 10 in a given period and is continuously growing – with marginal set-backs – and the growth exceeds 83,44 %. These words are: „privity", „respect", „help", order", „nation", „furtherance", „dignity", „preserving", „obliged", „restriction", „harm", „sincerity", „law", „release", „extant" „imputable", „responsible" „reasonable", „interest", „to omit", „collaboration", „safety" and „to give".

5) Words with indecisive tendencies: no significant differencies in prevalence between the 1st and 9th period, but in a sigle period they show up a serious salience (in an either negative or positive way): „injury" (from 15 sunk back to just 14, but in the 3rd period was at 17 unities); „negligence" (grew from 44 to 47, but in the 7th period took a value of 52), „morals" („oscillating" between 23 and 24, took more then once the value of 19).

Afterall we can conclude, that it is useless too characterize one period with the prevalence of moral words and concepts, because the absolute plevalence notes of a certain concept can in itself be swagging („liberty" came from 136 to reach 239, „dignity" from 6 to 419). These are – to put it better – tendencies not character traits of a certain conceptual word. A better perspective would be to analyze *in concreto* each elective period which should constitute the topic of further discussions.

5. Further steps

The question still prevails: does a method of text mining provide us the necessary equipment to examine the moral foundations of democracy? Does it prove that Hungarian legislature is morally well- or less-endowed? It is thus true, that the style of law-making is showing differces regarding the four-year elective periods after 1990? The relation of Hungarian legislature to moral principles and moral keywords was quite

different under socialist-liberal and centre-right-conservative governments and this could give further investigations a new and fascinating twist. Last but not least its actuality is on a very high level too. How were the styles of law-making practiced by different governing parties and how were they respecting the moral principles of constituting a democratic society? The answers should be in the lime-light of not just political scientists but of philosphers, sociologists and specialists in law, not the mention the experts of contemporary history. The challange is becoming in this way interdicplinary.

6. Outlook on the two meanings of the expression „the moral foundations of democracy"

Fot the purpose of present outlook on the possibilities of further investigation we have chosen a paradigmatic case in examining the relation of law-making and morality. Our case is not only historically laden but also has a clear tangency to present Hungarian legislative reality. This further analyis should be taken into account as a wider inquiry concerning all elective periods of post-communist Hungary. We speak about the law on the status of local self-governments. Moral foudation has here two distinct meanings: a traditional one (A.) and a second one (B.), the adaptation of A to the Hungarian reality, which is narrowly tied to our starting hypothesis:

A.) If we assume, that every society has a certain „texture", which is an unwritten, tacit „Moral Codex" holdig its parts firmly together and saving it form disintegration, then we'll face the vital question, whether laws have the task to protect this „texture", or to formulate it unequivocally: do they have to compel the actors to obey and the sinners and transgressors to accept their authority? In which measure have laws to be content with technical formulations and to in which to codify the principles given by the „Codex"?

The above mentioned assumption and the debate which it triggered was first initiated by Sir Patrick Devlin in his *The Enforcement of Morals*, a lecture from 1958.[127] Devlins position then was commented by H. L. A. Hart in various writings and the relevant literature has now reached the point where it couldn't be easily reviewed. This debate – we have to mention – is still on today. Many scholars have pointed out that Devlin's conservative arguments are very solid and sound, that is why we have to do a serious effort to deliver a scientifically relevant refutation of them. But anyway, Devlin's approach underscored that we have to reckon with two basic issues: (a) democracy theory has an immediate connection to the enforcement of moral principles (Devlin was certaily not speaking about the morally founded rights, just only about the enforcement of moral principles in political action, where these principles come to being as an initiative of civil communities); (b) and he sturdily proposed to put an unconventional

[127] Sir Patrick Devlin held the lecture on *The Enforcement of the Moral* at the British Academy as the second Maccabean lecture. It is till today *the* classic example of conservative foundation of laws, and remains also the most disputed, controversial text of its kind. See also the first annotated edition the lecture got by with a critical introduction of the author at Oxford UP, 1959.

accent on the definition of the concept which dominates the actual discussion. This concept was and has been since Devlin's proposal the one of *public morality*.[128]

The moral foundations of democracy was the indirect but not collateral research field of Alexis de Tocqueville, so we reach back to his formulations to grasp the topic we are speaking about. When Tocqueville described the pecularities of a democracy in full functioning (the American political structure), he also made a distinction between the „habits of the heart" and the „habits of the mind" which constitute together the broader field of the morals (mores, *Mœrs).* This could also be translated with an informal artlessness and naivete' as „morality" in a very general sense, but it is much more correct, if we say that it covers the sphere of political culture. All nations have a specific political culture, a specific bundle of feelings, habitus, cultural and social behavioral patterns, a realm of ideas that culminate in forging institutions and laws, which mirror not only political culture, are not only part of that, but they are also shaping social morality.

Tocqueville had observed in 19th century United States, that social activity, respectively the community-shapig ability of personal freedom (as a peculiar form of morality) manisfested itself in the towns and villages (more precisely in the very forms of civil self-organization, the self-government of the settlements) with an enthralling creative force. As he formulates it somewhere:

„The laws they influence morality, as morality influences laws. Where this reciprocity fails unrest and revolution disrupts society. In New England legislature created towns. It was completed through the variety of behavioral patterns which created a veritable town spirit. Towns are the centre of passions and interests, where true and tenacious, enduring activity reigns."[129]

In modern Hungary the first democratic local self-government system was basicly a Tocquville-ian enterprise (see law LXV from 1990); it was a seminal attempt to establish 3000 self-governmental unities (settlements, which are in a methodical sense: micro-states). Many traits of this law were to project the frame of civil sel-organization (like the model of local „school-chairs", where the parents, teachers and self-goverment functionaries where jointly in charge of the local forms of instruction). But afterwards and opposed to the Tocqueville-ian model these developments couldn't have been fulfilled – and this is our second hypothesis about the local institutions of morality, which has to be developed in an upcoming research: Hungarian towns have lost the sight of a full-fledged local self-government, which is crucial in forming a „social statality" having some serious differences in comparison to „political statility" a

[128] See: Ronald Dworkin: „Lord Devlin and the Reinforcement of Morals", in: Yale University. *Faculty Scholarship Series. Paper 3611.* (1966)

[129] This passage is a note from the manuscript of Alexis de Tocqueville's *Democracy in America*. He wrote the book in 1835 after nearly five years of empirical research between 1831-35. His text became one of the classics of political science, an ineludible piece of 19th century scholarship. See the annotated edition of the work: University of Chicago Press, 2000.

centralized system of counter-subsidiarity. The relation of laws to morality (and general political culture) took a different accent which influenced the reciprocity of support demanded by Tocqueville himself. Therefore Hungarian settlements became progressivey parts of a statal hierarchy and of statal politics. This situation nests the danger, that civilian communities loose their possibilities of real self-goverment and will exit the field of constituting social morality.

Further research can shed new light on this issue by analyzing the process of shaping the law on local self-governments. If the conclusion will emerge, that in the process of Hungarian legislation, the use of moral words after 2010 is running higher then before, when the former legislature tended to occupy a position which is neutral of common *Weltanschauung*-types, then it can rewrite the whole agenda of morality in democratic laws. Hence the force of self-governmental independency was decreasing in Hungary after 2010, we should take the educated guess that the simple use of moral words is not catalyzing morality. The use of them won't make a society based on moral principles. Moralizing in law-making (undertaking and deepening a world view in phrasing laws) is not a guarantee for the moral foundation of legislature.

The public library: an arena for an enlightened and rational public sphere or an arena for individual experiences conflicts between the library field and politicians? A Scandinavian case study.

Audunson, R,
Oslo and Akershus University College (Norway)

1. Introduction[1]

On January 1, 2014, a revised law on public libraries came into force in Norway Also Sweden got a new library law from that date. Both laws underline the libraries' role as arenas for democracy and the public sphere. The revised mission statement in the Norwegian law states that public libraries shall be independent meeting places and arenas for public discourse. Like in the previous mission statement, libraries are still obliged to promote "enlightenment, education and other cultural activities". The mission statement of the new Swedish library law states that all publicly financed libraries – public libraries, university and college libraries as well as school libraries – shall contribute in the development of the democratic society by promoting knowledge and the free formation of public opinion. Swedish publicly financed libraries are also supposed to promote the position of literature in society and the interest in education, research, cultural activities and that which in German is called Bildung and for which it is difficult to find an English term[1]. The mission statement of the Finnish library law adopted in 1998 contains a formulation stating the Finnish public libraries shall contribute in promoting civic skills. Also in the Finnish values related to knowledge and enlightenment are central, as they are in the Danish library law.

With the newly revised/new library laws in Norway and Sweden as a point of departure, it is interesting to discuss the relationship between the perspectives on public libraries, which can be identified with the legislators, and new trends in the field of public librarianship. In this article, the following two questions will be discussed:

A. A public sphere characterized by that which in §100 of the Norwegian constitution is called and informed and enlightened public discourse, is based on an ideal of rationality. In the open, public discourse, the participants have committed themselves to the power of the best argument in order to reach optimal solutions to common problems. By confronting different points of view and different, but knowledge based perceptions of reality, we can gradually approach truth and optimal policies. By comprehensive access to information and enlightenment the empowered citizen can make rational choices. Such ideals of rationality are obviously at the bottom of Nordic library laws with their weight on enlightenment and libraries as arenas for democracy, public discourse and civic skills. However, the library field in the Nordic counties as well as internationally has to an increasing degree oriented itself towards experiences. From enlightenment to experiences[1] has been a slogan. (Carlsson, 2013; Hvenegaard Rasmussen, Jochumsen & Skot Hansen, 2011). Do we have a conflict between new trends

in the field of librarianship and the politicians' perceptions of what a public library is and what legitimizes the use of scarce public funds on libraries? This issue will be discussed in the first part of the article.

B. The second part of the article will present and discuss results from a survey undertaken by heads of local government libraries in Norway on their attitude towards the amendment in the library law's mission statement focusing upon the libraries' role as independent meeting places and arenas for public discourse.

2. Literature and theoretical background

Public libraries as instruments for promoting civic skills is not a new idea. George Ticknor, one of the Boston Brahmins who in 1850 established that which is regarded to be the first modern public library, seems to have been driven at least partly by fear that the immigrants coming in from Europe were "not fitted to understand our free institutions or to be entrusted with the political power given by universal suffrage". (Harris, 1978). By developing civic skills, libraries could prevent social unrest and the lower classes from being led astray by radical politicians and agitators. According to Harris, the growth of totalitarianism in the form of Fascism, Nazism and Stalinism in the 1930s gave public libraries a new role as institutions promoting democracy. "The American public library had become a bureaucracy, a social institution without a purpose – except perhaps to preserve itself. Nearly two decades would pass before the actions of Hitler and Mussolini would provide the public library with a new and viable justification for its existence". (Harris, op.cit, p. 50). Through nonpartisanship and through securing access to all sides of the most salient and acute social and political issues, libraries could become effective partisans of democracy.

The tradition dominating Western public libraries from the thirties up until now has been termed "The grand tradition" (Jones, 1978; Muddiman, 1990). Within this tradition, public libraries are seen as liberal and humane institutions defending the open society. The revised mission statements in the Norwegian and Swedish library laws are well within this tradition.
This dimension of the role of public libraries can be understood within the framework of Habermas' theory of the public sphere, i.e. a sphere between and independent from the state, the market and the private sphere, characterized by undistorted communication and the free formation of (public) opinion. The public sphere is dependent upon institutions where citizens can come together and communicate in an undistorted manner or which can be channels of communication and mediation, for example newspapers, public service broadcasting, universities living up to the norms of academic freedom, community organizations etc. The public sphere is in principle open to everyone, although everyone is not guaranteed to have their viewpoints published in edited and moderated channels, for example a newspaper.

Are public libraries a part of the public sphere? Local governments own them. In many countries, all the Nordic countries included, Parliaments have adopted library laws and governmental agencies have been set up to oversee that these laws are complied with. If

the public sphere is defined as a sphere independent of state and government, that might be an argument against regarding libraries as a part of the public. However, we do have state owned institutions, which definitely belong to the public sphere, e.g. public service broadcasting and universities. Public libraries as independent and neutral providers of knowledge and information for citizens, open to everyone and with no particular ideological and political axe to grind other empowering citizens by giving them access to as broad and varied a range of information as possible, are institutions belonging to the public sphere[1]. The British sociologist Frank Webster (2006, p.) argues, "The public library is arguably the nearest thing we have to an achieved public sphere". He obviously means that the public library contributes vitally in laying the basis for a Habermasian public sphere by making information accessible for the general public and by doing this with an independent professional staff without a political agenda. Librarians are, as stated above, partisans of democracy by being nonpartisan. Extending the library's role from being an informational source for the empowered and active citizen into also being the arena where discourse and debate take place might seem as a new idea. Already in the 1930s, however, Oslo Public Library – the Deichmanske public library – organized a club called Young Deichman where youngsters between 14 and 17 years of age (alas, only boys were admitted into the club in those years) were trained in giving public speeches. One more recent example from Denmark is an initiative from the year 2000 in the city of Kolding called Kolding future forum. Every month the library invited a lecturer of national reputation within a given field. A local expert commented upon his or her lecture. Then followed a public debate, partly by the audience being physically present in the lecture hall, partly in the following weeks on a web based discussion forum. The Danish researcher Dorte Skot Hansen concludes, based on the Kolding project, that the library can provide a framework for a sustainable public sphere in local communities, but that this presupposes an open agenda, as in traditional collection development and other library services. (Skot Hansen, 2002).

Traditionally the following institutions have been vital in keeping up a sustainable public sphere:

- Media such as newspapers, radio and television, which partly have been channels through which authorities can communicate with and inform the citizenship, but which but which have been forums scrutinizing and discussing governmental policies critically.
- Civic organization – political parties, trade unions and other kinds of voluntary organizations, which can mobilize political interest and participation, be forums for political discourse, integration of political viewpoints and interests, for example in the form of programs and with a capacity of mediating between the local level, regional level and national level.
- Institutions like universities and libraries with a considerable intellectual autonomy – autonomy from the market as well as from their governmental owners – and without links to specific interest groups, political or religious ideologies and agendas.

Political scientists have expressed concern that developments have broken the links between the local, regional and national level. Membership in political parties has

decreased dramatically. Political parties tend to be dominated by elite of professional politicians. When people engage in political and social activities they do so often in the form of campaigns focusing upon one issue only (for example campaign to stop the closing down of a s school, defend the rights of animals, for a clean environment etc). The mediating and integrating function of for example political parties with a mass basis and links from local branches via regional level up to national level has been weakened. Digitization has accelerated this process of weakened mediation between small-scale level (the local community) and the large-scale level. (Tranvik&Selle, 2003).

Traditionally, mass media such as newspapers, radio and television have been channels which government has used to inform the citizens, and which citizens have used to keep themselves informed. At the same time, media have ben arenas for critical journalism and critical debate. Due to technological developments, the mediating function of traditional media is under pressure. Governmental bodies inform more and more directly via the web. Social media is used to an increasing degree and communication is individualized. The proportion of the Norwegian population using traditional media such as printed newspapers, radio and television a daily basis has been dramatically reduced over the last two decades. According to Nydén this might lead to the media focusing unilaterally *either* on infotainment *or* on the critical dimension. In the first case, the result will be a citizenship, which is not adequately informed via traditional media and which therefore becomes dependent upon the government's own version of reality, communicated via the web and social media. In the second case, the result might be a one sided focus on wrongs and negative aspects related to governmental policies. In the end, the result might be that "the citizens' sense of community and unity will diminish along with their involvement in the democratic process". (Johansson 2004, p. 49).

Developments such as these might be the reason for the increased concern with libraries as meeting places with a capacity for promoting democracy and public discourse. It is probably the reason behind the new and revised mission statements in Norwegian and Swedish library legislation, but the trend is international. Willingham (2008) has described how American public libraries have strong traditions in developing civic skills and functioning public spheres in US communities and the central role this dimension of the public library tradition has had in the presidential programs of the presidents of the American Library Association since 2000.

3. Enlightenment, knowledge provision for an informed citizenship or experience versus libraries as arenas for individual experiences

As stated above, there is a trend in modern public librarianship of supplanting enlightenment and knowledge with experience in the second meaning of that word – the individual, emotion filled experience. Is that trend in conflict with the way politicians perceive the role of libraries? It was clearly expressed in the strategy of the Swedish Malmø city library: "The Darling library in the world – your life, your dreams, your library". (Carlsson, op.cit, p.). Here the focus is upon the individual and his or her dreams, whereas the mission statement of the new Swedish library law focuses upon democracy, knowledge and interest in research – values which seem to be quite far away

from the dreaming individual. In the Royal proposition to the Norwegian parliament in connection with the revision of the mission statement in the library law, the government states that the mission statement, which we have had since 1972, focusing upon enlightenment, knowledge and education, overall has served us well.

Earlier research indicates that both local government politicians and members of Parliament in Norway perceives the promotion of quality reading, based on the literary and cultural canon, as the main role of public libraries. In a survey comprising a representative sample of the adult population in Norway, a representative sample of local government politicians and all professionally educated librarians in Norwegian public libraries, the respondents were asked to indicate the main reason legitimizing the use of scarce public funds on public libraries. Among the politicians, the largest group pointed at the promotion of quality reading based on the canon as the main reason for using scarce funds on public libraries. A substantial group, approximately 20 per cent, regarded the library's democratic role as the main reason for legitimizing the allocation of funds to the library. Among the librarians, the democratic role of public libraries received the highest score, whereas the citizens tended to regard the instrumental usefulness of libraries in relation to education, work and everyday life as the most important reason. (Audunson, 2001). In 2005 qualitative interviews with members of the committee for culture and education in the Norwegian parliament was undertaken. (Audunson, 2005). The results confirmed the impression that politician perceive the promotion of quality reading as the main role of libraries.

Carlsson, when analyzing the experience-oriented strategy of Malmø city library, argues that the proponents of the new strategy do not reject the traditional role based on knowledge and enlightenment. It is more a question of adaption of the traditional role to new circumstances, she argues. According to Carlsson, the strategy does not cut itself "*completely*" (our italics) loose from the traditional role of libraries of promoting democracy and education. The problem, however, was that important agents in the environment perceived the strategy as a break with these traditions.

The role perceptions reflected in the policy documents preparing the ground for the Norwegian law revision adopted by the parliament in June 2013, are complex. In the report to the Parliament on public libraries from 2009 (Ministry of Culture, 2008-2009), the government states that the legitimacy of public libraries is rooted in making knowledge and education accessible for all. Legitimacy, thus, is rooted in the enlightenment mission. At the same time, public libraries shall represent and promote a set of common values, i.e. a set of values, which, in spite of pluralism and multiculturalism, make us perceive that we are one common society, not a number of parallel societies. Mediating a set of common values is related to that which in German is called Bildung.

However, the promotion of knowledge, education and common values must, according to the report, go hand in hand with "accidental meetings with the unexpected" and the library visit shall stimulate phantasy and interpersonal social processes related to Accidental meetings with the unexpected is a formulation closely related to culture,

knowledge and learning. References to meetings with the unexpected and phantasy are closely related to the experience-oriented library.

Also in the Swedish policy documents, the social mission of public libraries is rooted in rationality and knowledge. The basis of general support for libraries is the conviction that transparency and an informed citizenry strengthen democratic processes and participation. Publicly funded libraries transmit knowledge and create the preconditions for a public discourse where citizens freely can discuss common issues and exchange opinions. (Governmental proposition 2101/13:147, p. 13). Here, the weight is upon knowledge, the empowered citizen and community more than on individual experiences. The proposition also underlines that the mediation of *knowledge* is that which unites and constitutes a joint platform for different kind of libraries, for example public libraries, school libraries and university and college libraries.

The new Swedish mission statement also states that libraries shall promote "Bildung" – the integrated and overall forming of humans, taking into regard knowledge and culture as well as values. The proposition explicitly states that focusing only upon education would be too instrumental and too narrow. The formative role is not defined in detail. The proposition refers to the necessity of supporting individuals striving to develop as integrated human beings. When something is former, however, it necessarily implies forming according to a pre-existing pattern, for example a body of values or knowledge, a literary canon etc. The forming agent, for example a librarian or a teacher takes such pre-existing patterns as their point of departure in their forming activity.

In sum, then, although the governmental policy documents contain some elements which are not necessarily in conflict with the experience-oriented trend – the Norwegian policy documents more than the Swedish – the legitimacy is still primarily rooted in the ideals of rational discourse, knowledge. The experience-oriented trend might easily collide with strongly institutionalized perceptions of what a library is and, therefore, the roots of the legitimacy of public libraries. (Evjen & Audunson, 2008). Carlsson (op.cit) convincingly documents how the efforts to implement an experience oriented strategy in Malmö, Sweden, collided strongly against such institutionalized perceptions in important part of the environment.

4. How do municipal librarians perceive the libraries' role as arenas for public discourse

In this section, we will briefly discuss how Norwegian municipal librarians perceive the role as arenas for public discourse.

The librarians were asked what kind of programs and activities they would give priority to when adapting to the new role of being an arena for public discourse. The alternatives were:

- Give people access to knowledge and information they need to be active citizens.

- Develop civic skills so that people can involve themselves in traditional as well as digital channels for public discourse
- Arrange physical meetings in the library
- Develop digital arenas

A clear majority hold the traditional task of giving people access to knowledge and information and to develop the library as an arena for physical meetings and debates as the two most important. Almost 90 per cent perceive physical meetings and debates as important or very important, whereas 78 per cent hold the more traditional role of mediating information and knowledge as important or very important. Very few regard the development of civic skills – one of the tasks in the mission statement of the Finnish library law as important. Norwegian librarians seem, however, to be more in line with the new Swedish law, which states that libraries shall promote democracy and the free building of opinion by giving promoting knowledge.

There are practically no correlations between attitudes to which issues libraries should focus upon when striving to develop into arenas for public discourse, and background variables such as holding a LIS degree or not and size of municipality.

When asked what kind of competencies

The librarians were also asked how they perceive their role when programming for public discourse: Are they similar to the independent editor of a newspaper or is their role more to be a provider of physical space, where other agents are responsible for the content. A clear majority see themselves as editors. The proportion who are of this opinion, however, increases almost linear with increasing number of inhabitants in the municipality; this is shown in the table below.

	Below 5000	5000 and more
Independent editor	29	66
Information officer	11	15
Provider of space/premises	60	20

Table 1 – perceived role of librarian and size of municipality

In the smallest communities, a majority of 60 per cent perceive their role primarily as a provider of premises, whereas in the municipalities with more than 5000 inhabitants 66 per cent perceive the librarians' new role as that of an independent editor. In the largest municipalities with 50000 inhabitants or more, 87 per cent perceive their role as that of an independent editor.

A bivariate cross table with professional library education as independent variable also indicates a significant relationship between having a bachelor or master degree in LIS and perceived role. They with a degree in LIS seem to perceive the role as that of an editor more frequently than those without such a degree. This correlation, however, did

not hold when we performed a multivariate test. As shown in the table below, the difference between small and larger municipalities is not affected by the librarian having professional education or not.

5000 inhabitants or more		Less than 5000 inhabitants	
LIS degree	No LIS degree	LIS degree	No LIS degree
66	57	29	28

Table 2. Proportion perceiving the librarians' role as that of an editor depending on size of municipality and LIS degree

5. Conclusion

Summing up, there seems to be a potential conflict between the trend of experience-oriented libraries prevalent in the public library field and the dominating library perceptions of the responsible politicians. Politicians seem to base the legitimacy of public libraries on rationality, knowledge, democracy, education and the promotion of common social values as a basis for a common public discourse – the policy documents analyzed in this article do not so much refer to individual experiences. The survey among Norwegian librarians indicate that they are more in line with the politicians in giving priority to the promotion of knowledge. The high proportion, almost 80 per cent, holding the promotion of access to knowledge necessary for democratic knowledge when adapting to the new role might serve as an argument for that. The even higher priority given to organizing events and meetings and the high proportion giving responding that there is a need for increased competencies in event planning might, on the other hand, indicate a leaning towards the experience oriented library. More research is needed to elicit in which way librarians want to take their libraries as meeting places and arenas for public discourse.

Librarians are local government employees who have been hierarchically subordinated to the upper echelons of the local government's administrative and political structure. In their capacity as moderators of public discourse, however, a clear majority feel that they must have the same independence as that of an editor. That is a fundamental change from traditional roles, which it will be interesting to follow how public librarians adapt to. That is also a topic for further research.

References

Lov om folkebibliotek. (1985).
Arnestad, G. (2013). Vil den nye §1 føre til filialdød. *Bok og Bibliotek*(6).
Audunson, R. (2001). Folkebibliotekenes rolle i en digital framtid: Publikums, bibliotekarenes og bibliotekarenes bilder. Audunson og Windfeld Lund (red). *Det siviliserte informasjonssamfunn*. Oslo, Fagbokforlaget, s. 206 – 224.
Audunson, R. (2005). How do Politicians and Central Decision Makers View Public Libraries – the Case of Norway. *IFLA Journal*, 31 (2), pp. 174 – 182
Carlsson, H. (2013). Den nya stadens bibliotek. Om teknik, förnuft och känsla i gestaltningen av kunnskaps- och upplevelsesstadens folkebibliotek. (The library of the

new city: On technology, rationality and emotions in the creation of the new city's library). Universitetet I Lund, doctoral dissertation.
Det kongelige kulturdepartement. (2013). *Endringer i lov om folkebibliotek*. Proposisjon til Stortinger. Prop. 135 L (2012-2013). Oslo
Evjen, S. & Audunson, R. (2008). The complex library: Do the public's attitudes represent a barrier to institutional change in public libraries? *New Library World*, 110 (3/4),pp. 161 - 174
Jones, K.H. (1978). Towards a reinterpretation of public library purpose.*Totterdell, B. (ed). Public library purpose,* pp. 122-135 London, Clive Bingley
Harris, M. (1978). The purpose of the American public library. *Totterdell, B. (ed). Public library purpose,* pp. 39-53. London, Clive Bingley
Holmquist, T. (2014, 10 April). Utfordrer biblioteksjefene. *Kommunal Rapport*.
Johansson, V. (2004). Public Libraries as Democratic Intermediaries. Some examples from Sweden. *New Library World*, 105 (1), p. 47 – 59
Kulturdepartementet /Ministry of Culture (2009). *Bibliotek – kunnskapsallmenning, møtestad og kulturarena I ei digital tid. (Libraries – commons of knowledge, meeting places and cultural arenas in a digital time*. Stortingsmelding nr. 23, 2008-2009
Muddiman, D. (1990). Towards a definition of public library purpose. *Public Library Journal,* (4), pp. 5 90-94
Nydén, M. (2000).Myndigheter, *Internet och Integritet*. Stockholm, Styrelsen för psykologiskt försvar
Regeringens proposition 2012/13:147. *Ny bibliotekslag*. Stockholm 18. April 2013
Skot-Hansen, D. (2002). Viewpoint: The public library in the service of civic society. *Scandinavian Public Library Quarterly, 35*(3).
Tranvik, T.&Selle,). (2007). The rise and fall of popular mass movements. Organizational change and globalization – the Norwegian case. *Acta Sociologica*, 50 (1), pp. 57-70
Webster, F. (2002) *Theories of the Information Society*. Routledge, London

The use of information. The needs, skills, difficulties and preferences related to the use of diverse sources of information among young people.

Zybert, E. B.
University of Warsaw (Poland)

The 21st century is called "the century of the knowledge society". Its fundamental characteristics and functions are: generating, storing, transferring, collecting and using information[130]. The success in running a business or involvement in any other activity lies in the ability to adapt to change and to keep ahead of it[131]. At the same time, communication, scientific, cultural and technological transformations, as well as introducing new communication technologies result in changes in the information environment and in their stronger influence on every aspect of people's lives. According to Digital Agenda for Europe Scoreboard Report, by 2015, 90% of jobs will require at least basic digital skills, which means that in order to find employment in the future, young people already have to learn how to use their digital competences effectively[132]. Therefore, people's education and competences, including information competences, become crucial values in the contemporary information civilisation[133]. Moreover, deficiencies in this respect may lead to social exclusion.

In order to prevent social exclusion and create conditions for full participation in the information society, people need to develop necessary skills from an early age. This may be achieved through promoting information literacy, which involves development of information competences aimed at acquiring a habit of constant knowledge acquisition through the active and conscious use of various sources of information. Emphasis is placed not only on the use of the new technologies, but also on providing knowledge of traditional information sources and modern databases as well as on independence in the use of scientific information, which will enable students to achieve better educational results and pursue lifelong learning.

The key element of information competences is the ability to identify the need of information, determine the needs in this respect and formulate questions, as well as knowledge of effective strategies needed to search and locate potential information

---

[130] M. Nowina Konopka: Istota i rozwój społeczeństwa informacyjnego. In: Społeczeństwo informacyjne: istota, rozwój, *wyzwanie*. Warsaw 2006, pp. 19-21. As cited in: E. Klich: *"Latarnicy" społeczności lokalnej na przykładzie Kiełczowa w województwie dolnośląskim- kształcenie kompetencji informacyjnych osób 50+*. In: *Czas przemian- czas wyzwań. Rola bibliotek i ośrodków informacji w procesie kształtowani kompetencji współczesnego człowieka*. J. Jasiewicz, E. B. Zybert (eds.) A project implemented in cooperation with Orange Foundation. Warsaw: Wydawnictwo SBP, 2014 p. 132

[131] *Kompetencje zawodowe.* https://www.biurokarier.umk.pl/kompetencje-zawodowe

[132] *Społeczeństwo informacyjne w liczbach*. V. Szymanek (ed.), Ministry of Administration and Digitization, Warsaw 2013. As cited in: E. Krupa, D. Drabik: *Skąd się biorą cyfrowi humaniści? - Nowe technologie w projektach społecznych.* In: *Czas przemian - czas wyzwań. ...* op.cit. p. 21

[133] Ibidem

sources, the ability to assess their credibility, competent interpreting of information and assessment of its value, selecting information and the effective use of obtained information to broaden existing knowledge. In the process of achieving literacy, it is also important to help users in developing desired behaviour related to the use of information, including its use in an ethical and lawful way[134].

The aim of the research presented here was to diagnose the state of knowledge about information needs, the ability to search for and use information, as well as preferences and sceptical approach in the use of information among young people who enter the labour market or who have already started working.

The research was conducted in May and June 2014. Altogether, it included 250 extramural students of:

- a state higher education institution (state HEI): 3rd year of BA studies, subject: scientific information (130 people, i.e. 52%) and 1st year of MA studies, subject: scientific information (30 people, i.e. 12%);
- a private higher education institution (private HEI): 3rd year of BA studies, subject: other than scientific information (altogether 90 people, i.e. 36%).

The researchers used a questionnaire developed at the University of Pécs (Pécsi Tudományegyetem), which provided a possibility to compare the obtained results. The questionnaire consisted of 46 questions.

The first 7 questions related to information about respondents, i.e. their age, place of residence, sex, education, occupation and its areas, as well as income. The group consisted of 180 women (72% of all respondents) and 70 men (28% of all respondents).

All respondents were less than 50 years of age. 12 people (4.8%) were 35-50 years old; 100 people (40%) were 25-34 years old and 138 people (55.2%) were 22-23 years old. As regards the place of residence, 80% (200 people) of respondents lived in cities, suburban areas and small towns (this group included 64% students of the state HEI), whereas 20% of respondents (50 people) lived in the country (among them 32 students of the state HEI, i.e. 64%). The vast majority of respondents (220 people, i.e. 88%) completed their secondary education (BA students), whereas 30 people (12%) completed higher vocational studies (BA level).

Since only extramural students were involved in the research, 98% of them (245 people) had already started working. Only 5 people (2%) did not work at the time of the research. Occupation areas varied significantly, i.e. the results were as follows: administration (91 people, i.e. 37.1%); education (75 people, i.e. 30.6%); trade (18 people, i.e. 7.3%); other (undefined) sectors (17 people, i.e. 6.9%), tourism (16 people, i.e. 6.5%), mass media (15 people, i.e. 6.1%); human resources (10 people, i.e. 4.01%), transport (8 people, i.e. 3.26%). Thus, the research involved representatives of various areas of life and economy.

The monthly income of the respondents fell within the national average. The income of 140 people (57.1%) ranged from PLN 2501 to 3500, while the income of 75 people (30.6%) ranged from PLN 1601 to 2500; 30 people (12.2%) earned a different (undefined) amount of money. According to the results, the income of the respondents did not influence

[134] K.Borawska-Kalbarczyk: *Internet jako miejsce uczenia się a kompetencje informacyjne uczniów*. http://www.up.krakow.pl/ktime/ref2005/borawska.pdf.

either the frequency of using various sources of information or the ability to search for and use it.
The results concerning ownership of modern information tools were not surprising. Since it was a check-all-that-apply question, it was possible to assume that all respondents had access to these tools. 235 people (94%) owned a desktop computer, 160 people (64%) owned a laptop, 147 people (58.8%) owned a tablet computer, 205 people (82%) owned a smartphone and 240 people (96%) had Internet access. 69 people (27.6%) owned an e-book reader. Therefore, it can be concluded that the respondents who owned the above-mentioned tools were also able to use them.

The respondents also gave interesting answers concerning books they had at home. 180 people (72%) owned a collection of 100-500 volumes; 30 people (12%) owned a collection of 500-1000 volumes, 10 people (4%) owned a collection of 1000-2500 volumes and 30 people (12%) owned less than 100 volumes. Although it was difficult to determine the type of these collections, it seems significant that all respondents had books at home.
Several questions concerned the use of books and Internet. The results showed that all respondents made use of these sources of information. However, although 100% of respondents claimed that they use computers, Internet and traditional sources of information, the frequency of their use varied significantly. 50 people (20%; among them were students of the state HEI, of both BA and MA studies) used books every day, 170 people (68% of all respondents) used books several times a week, whereas 30 people (12%) claimed they used books several times a month. As far as Internet was concerned, most of respondents used it regularly (the frequency was described as: constantly, when I'm at work, several times a day, every day). Only 100 people (40%) stated clearly that they used Internet several times a week. 180 people (72%) spent several hours using Internet, 50 people (20%) spent 1 hour and 20 people (8%) spent less than 1 hour. Usually the respondents used Internet at home: such answer was selected by 180 people (72%); 150 people (60%) used Internet at work and 80 people (32%) used it in HEI. 130 people (60%) used Internet somewhere else, e.g. in libraries or Internet cafés. Since it was a check-all-that-apply question, it can be concluded that the respondents make use of the Internet access on various occasions.

As far as the use of websites was concerned, the respondents most often gave the following answers: email (100% of respondents); news (150 people, i.e. 60%); science (150 people, i.e. 60%); Skype (140 people, i.e. 56%); work (100 people, i.e. 40%); social networks (80 people, i.e. 32%); hobby, games (50 people, i.e. 20%); everyday information, e.g. public transport timetables, opening hours, recipes, etc. (40 people, i.e. 16%). The respondents also used the following Web 2.0 sites: Facebook (180 people, i.e. 72%), private blogs (20 people, i.e. 8%), professional blogs (50 people, i.e. 20%), Wikipedia (230 people, i.e. 92%), YouTube (60 people, i.e. 24%).
Some of the questions related to the use of libraries. 90 people (36%, among them 80 people, i.e. 87.7%, studied scientific information) claimed that they were members of at least one library. 220 people (88%) claimed that they used libraries, while 30 people (12%) claimed the contrary. As in the previous case, the latter group consisted of the private HEI students. The answers to the question concerning the frequency of using libraries

were as follows: several times a month (150 people, i.e. 60%), several times a week (50 people, i.e. 20%), every day (20 people, i.e. 8%), seldom (30 people, i.e. 12%).
Card catalogues were used by 100 people (40%), while 150 people (60%) did not use them. The group of the Internet catalogue users consisted of 200 people (80%); 50 people (20%) did not use them. Regardless of the fact that Internet catalogues seem an understandable alternative to card catalogues, it would be interesting to find out how the respondents searched for necessary materials in cases when Internet catalogues were not available – after all, as many as 60% of respondents claimed that they did not use the card ones.

The answers to the question concerning the aim of using libraries were as follows: research, studying and work (230 people, i.e. 92%); reading news, e.g. about politics and economics (10 people, i.e. 4%); hobby, games (20 people, i.e. 8%); everyday information (20 people, i.e. 8%); government sites (20 people, i.e. 8%); social networks (30 people, i.e. 12%); e-mail, Skype, chat (50 people, i.e. 20%). While it was understandable that the respondents used electronic resources (social networks, chat, Skype, e-mail, etc.) in libraries, in case of other answers (various news, everyday information, research, work, studying, etc.) it was not clear whether the respondents used traditional or electronic resources. The library may have been indicated simply as a place of searching for information, regardless of the sources.
The next questions related to the following skills:

- searching for information and using available sources (including selection of adequate search methods, sceptical approach to information sources, using application and tools that facilitate the information search process);
- creating own contents that combine various tools and forms of presentation and creating unique contents out of elements found in other sources.

After comparing the answers concerning the aim of using websites with the ones concerning the aim of using libraries, it could be clearly concluded that most of the respondents (92%, i.e. 230 people) preferred to use library sources in order to work, study or conduct a research. These results applied to both printed and Internet materials. 60% of respondents (150 people as regards studying) and 40% of respondents (100 people as regards work) used only Internet sites, i.e. this answer related only to electronic sources.
The answers concerning the use of the published information materials in order to work may confirm the above-mentioned result: lexicons were used by 130 people (52%), encyclopaedias were used by 200 people (80%), dictionaries were used by 220 people (88%) and bibliographies were used by 150 people (60%). These results were directly reflected in the answer to the question "Where do you find information necessary to work and study?", which indicated that 150 people (60%) found such information in libraries, 70 people (28%) found it in magazines and newspapers, 130 people (52%) found it on the radio and TV, 90 people (36%) heard it from their friends and 50 people (20%) used web portals and social networks.

As regards information concerning everyday matters, the library sources were used much less often, i.e. only by 30 people (12%). Such information was searched for mainly in web portals (by 180 people, i.e. 72%), on the radio and TV (by 150 people, i.e. 60%), heard from friends (by 100 people, i.e. 40%) and searched for in traditional magazines (by 40 people, i.e. 16%).

The answers to the direct question "What kind of sources do you prefer to use: traditional or electronic ones?" showed that traditional sources were used by 130 people (52%) and the electronic ones were used by 120 people (48%). It is necessary to emphasise that nearly all students of the state HEI (subject: scientific information) preferred using traditional sources.

As far as the frequency of using various information sources was concerned, it is necessary to mention the following results: 130 people (52%) used library catalogues very often, 80 people (32%) used them often and 40 people (16%) used them not very often. Google was used very often by 180 people (72%), often by 50 people (20%) and seldom by 20 people (8%). Online magazines were used very often by 130 people (52%), often by 80 people (32%) and seldom by 40 people (16%). Online audiovisual documents were used very often by 50 people (20%), often by 70 people (28%), not very often by 100 people (40%), seldom by 20 people (8%) and very seldom by 10 people (4%). Electronic books were used very often by 20 people (8%), often by 80 people (32%) and not very often by 150 people (60%). As far as the use of traditional sources was concerned, the results were as follows: printed books were used very often by 90 people (36%), often by 130 people (52%) and not very often by 30 people (12%). Similar results were obtained in the case of the use of traditional magazines, which were used very often by 80 people (32%), often by 120 people (48%) and not very often by 50 people (20%). Audiovisual materials in the library were never used by 50 people (20%), seldom by 30 people (12%), not very often by 90 people (36%); often by 50 people (20%) and very often by 30 people (12%).

According to the results, the respondents needed and searched for information and spent a lot of time on collecting it. 30 people (12%) spent less than 1 hour a day on this activity; 40 people (16%) spent up to 2.5 hours; 50 people (20%) spent up to 5 hours and 130 people (52%) indicated that they did it constantly at work.

The research was also aimed at establishing whether the respondents had any online information search strategies. Usually they searched for information using search engines, such as Google (as much as 180 people, i.e. 72%); professional sites were used by 30 people (12%); both subject gateways and bookmarks were used by 20 people (8%). Boolean operators were never used by 180 people (72%; it can be assumed that these respondents did not know how to use them); limiters were used by 190 people (76%) and special settings were used by 230 people (92%). Only 100 people (40%) checked whether information they were searching for was correct and 200 people (80%) selected it randomly. The indicators of correctness of the choice were as follows: becoming familiar with the text (200 people, i.e. 80%), the title of the text (100 people, i.e. 40%), keywords (60 people, i.e. 24%) and the name of the website (40 people, i.e. 16%). The above-mentioned results seem surprising if we take into account the answers concerning satisfaction with the search results on the first pages. 80 people (32%) described them as always satisfying; 125 people (50%) described them as often satisfying; 15 people (6%)

described them as occasionally satisfying; 30 people (12%) described them as sometimes satisfying; no one was dissatisfied with the search results on the first page.
It is interesting to analyse the assessment of reliability of information sources used by the respondents, which may indicate certain consciousness of the users in this respect. Facebook reliability was estimated between 3 and 5 (1 = very low, 5 = very high). Among those who used the source (180 people, i.e. 72%), 10 respondents (4%) estimated it at 5; 90 respondents (36%) at 4 and 80 respondents (32%) at 3. Private blog users estimated their reliability as follows: 20 people (8%) gave them 4, 5 people (2%) gave them 5 and 15 people (6%) gave them 3. Professional blogs were used by the respondents more often (by 50 people, i.e. 20%). They were estimated at 4 by 30 people (12%) and at 3 by 20 people (8%). YouTube was used by 60 people (24%) and its reliability was estimated as follows: 4 (10 people, i.e. 4%), 3 (40 people, i.e. 16%) and 2 (10 people, i.e. 4%).
Most respondents (230 people, i.e. 92%) used Wikipedia and regarded it as the most reliable source: 80 people (34.8%) estimated it at 5, i.e. they had no objections; 120 people (52.2%) estimated it at 4 and 30 people (13%) at 3. The reliability of information found on Google was estimated at 5 (very high) by 30 people (12%), at 4 by 80 people (32%), at 3 by 110 people (44%) and at 2 (very low) by 30 people (12%). Online magazines were estimated as very reliable by 120 people (48%); 100 people (40%) estimated them at 4 and 30 people (12%) at 3. Online audiovisual documents were considered unreliable by 30 people (12%); 150 people (60%) estimated them at 3, 50 people (20%) at 4 and 20 people (8%) at 5. E-books were considered very reliable by 50 people (20%); 180 people (72%) estimated them as reliable and 20 people (8%) described them as not very reliable (3).
Reliability of information found in library catalogues was estimated at 5 (very high) by as much as 200 people (80%); 50 people (20%) estimated it at 4.
Also traditional sources were regarded much more reliable than online documents: as far as printed books were concerned, 150 people (60%) estimated them at 5, 80 people (32%) at 4 and 20 people (8%) at 3; printed magazines were estimated as very reliable by 90 people (36%), while 140 people (56%) estimated them as reliable and 20 people (8%) as not very reliable. Audiovisual library documents were considered very reliable by 150 people (60%); 80 people (32%) estimated them as reliable, while 20 people (8%) estimated them as not very reliable. It can be assumed that a publication available in the library may have been regarded by the respondents as reliable because of the fact that a librarian must have approved it beforehand.
Also printed library materials that were used for work and studying were considered reliable. Among 130 people (52%) who described themselves as lexicon users, 100 people (77%) considered this source very reliable, while 30 people (23%) considered it reliable. As much as 85% paper encyclopaedia users (200 people, i.e. 80%) considered them very reliable and 30 people (15%) estimated them as reliable. Dictionaries, which were used by 220 people (88%), were estimated at 5 by 210 respondents (95.5%), while 10 people (4.5%) estimated them at 4; bibliographies, which were used by 150 people (60%), were considered very reliable by 140 people (93.3%) and reliable by 10 people (6.7%).
Respondents who had problems in finding necessary information (180 people, i.e. 72%) turned for help to the librarian. This can be considered as a positive approach towards the role of the librarian as an intermediary in finding, assessing and using obtained information. The option „ask a librarian" was used, on the other hand, by only 50 people (20%), which meant that the respondents preferred a direct contact with the librarian.

150 respondents (60%) turned for help to the Internet community, 50 respondents (20%) to their friends and 30 respondents (12%) to other people.

The researchers were also interested in finding out which sources were used by the respondents while preparing for the exam and what the frequency of their use was. The obtained results showed that most respondents (200 people, i.e. 80%) used own notes and the Internet; 150 people (60%) used both course notes and coursebooks; 100 people (40%) used their classmates' notes. As regards articles, 150 people (60%) claimed that they used them seldom and very seldom. Audiovisual documents were used with the average frequency (i.e. 3) by 200 people (80%); classmates' notes were also used with the average frequency by 130 people (52%).

The research was also aimed at determining to what extent the respondents were able to be active creators of their own knowledge – an ability that requires complex cognitive skills, interpretation, reorganisation, transformation and generalisation of obtained information. As was indicated above, these processes comprise the basis of information competence development. Therefore, one of the questions in the questionnaire concerned respondents' awareness of search problems, transforming information into knowledge and self-editing of the content. Most respondents indicated that the actions listed in the questionnaire were difficult for them. Only a small percentage of the respondents (below 20%) did not claim that it was difficult for them. According to the respondents, the following questions were problematic for them: starting to write the text (180 people, i.e. 72%), defining the subject of the text (170 people, i.e. 68%); narrowing the subject of the text (150 people, i.e. 60%); formulating search queries (180 people, i.e. 72%); creating a search strategy (170 people, i.e. 68%), taking decisions (180 people, i.e. 72%); finding average information in the Internet (180 people, i.e. 72%). The respondents found difficult also such things as writing itself (200 people, i.e. 80%); formulating thoughts found in the source by themselves (170 people, i.e. 68%); assessing whether the work was done well (150 people, i.e. 60%).
The respondents also had difficulties in listing quoting rules (200 people, i.e. 80%); estimating when it is necessary to distinguish quotations (150 people = 60%) and whether the way of using a source is a plagiarism (140 people, i.e. 56%). The respondents also found in difficult to: find sources in the Internet (200 people, i.e. 80%); estimate reliability of a website (180 people, i.e. 72%); identify information sources in libraries (180 people, i.e. 72%); find recent materials (170 people, i.e. 68%); assess sources (200 people, i.e. 80%); take notes (180 people, i.e. 72%); integrate the materials found (120 people, i.e. 48%). Most respondents estimated as not very difficult integration of the materials found (100 people, i.e. 40%) and becoming familiar with the materials found (170 people, i.e. 68%).
Lack of skills in transforming information into content was earlier confirmed in the question that concerned active creation of the Internet content. Very low results obtained here may suggest or confirm problems faced by the respondents in this respect.
People who took part in the research could be described as passive information users. Their role in creating the Internet content was insignificant. 20 people (8%) took part in creating Wikipedia entries; 40 people (16%) created blogs; 20 people (8%) commented

information found in the Internet; 40 people (16%) tagged information, 50 people (20%) were engaged in the Internet activity in other ways.

As regards the trainings on the use of the Internet, libraries and their resources, which (as it can be assumed) included elements of developing information competences, 160 people (64% of all respondents) took part in them; 90 people (36%) did not participate in such trainings. These results can be explained by the fact that in the state HEI, especially in the one that was included in the present research, all students who start their BA or MA studies have to take part in the obligatory library training, as well as in the training on the use of library catalogues and databases.

Considering that the respondents had problems related to the knowledge of the quoting rules, estimating where to indicate that the text was not written by themselves and therefore obeying adequate rules and regulations, they were asked a question whether they were familiar with the term "plagiarism". The obtained results showed that most of the respondents were familiar with the term (240 people, i.e. 96%, answered yes). They described plagiarism as "copying someone else's work, claiming credit for someone else's work; using materials without presenting sources; accurate copying of the content; using someone else's words and thoughts to present them as their own; putting their own signature to someone else's work; lack of quoting".

The answer to the next questions was, however, surprising. Only 100 respondents (40%) observed the quoting rules, while as much as 150 people (60%) consciously broke law in this respect.

The last question in the questionnaire was aimed at determining whether the respondents were able to name the skills that were mentioned in the questionnaire. 100 people (40%) knew the term information literacy, which is used to describe such forms of education that provide adequate knowledge and competences that guarantee full satisfaction and the use of necessary information. However, it was difficult for them to give a precise definition. Most often, the respondents claimed that it was "lack of information competences; we are not able to find necessary information; education aimed at gaining, using and promoting knowledge and information; that a person does not know any computer programs; the ability (or obtaining an ability) to search for, evaluate and use information; that a person knows what to do in the Internet, that a person knows how to search for information; having adequate skills and competences as regards widely understood information; the ability to use information sources, the ability to search for and select them; education aimed at preparing students for the effective use of information, assessment of its reliability and rules of its correct (lawful) use at work; development of such skills as information search and analysis; the ability to use various sources of (traditional and online) information".

The research helped to discover mutual relations between the knowledge of the term information literacy and the HEI type. Nearly all respondents (95 people, 95% of the group) who knew the term studied in the state HEI (subject: scientific information). The term was unknown to all private HEI students (90 people, 60% of the group) and to 60 state HEI students (40% of the group). It is important to mention that those respondents who were familiar with the term information literacy were more likely to indicate that they observed the quoting rules.

The fundamental results of the research indicate that traditional documents have their everlasting value, especially as regards science and research. Internet sources, on the other hand, are more often used for everyday needs and for entertainment, as well as a supplement to information found in traditional sources. It is also important to mention that people who attend classes aimed at developing their information skills become more conscious and responsible information recipients. Therefore, the importance of such classes seems to be justified; among other things, they help to develop consciousness and effectiveness in the lawful use of materials found in various sources.