THE IMPLEMENTED HUMAN INTERPRETER AS A DATABASE

Gábor Alberti and Márton Károly
Department of Linguistics, University of Pécs, Pécs, Hungary

Keywords: Discourse semantics, Interpretation, Discourse analysis, Logical programming.

Abstract: In this paper we continue to publish the results of our work with ℜeALIS, a new "post-Montagovian" discourse semantic theory, demonstrating its functioning on a few classical semantic problems. We retain mathematical exactness while simultaneously applying cognitive paradigm. Previously we wrote study programs, for testing purposes, now we are building a (lexical) Prolog fact database. Although we are implementing the grammatical analysis, too, it is important to note here that the whole process of grammatical (phonological, morphological and syntactic) analysis is practically included in only one (σ) of the four basic functions while the other three describe various parts of semantics.

1 INTRODUCTION

The long-term goal of the ℜeALIS project is automated discourse analysis, preferably including effective information extraction, text summarization, machine translation (e.g. the translation of European legal text) and analyzing NL queries.

To do all this, we provide the ℜeALIS theory of discourse semantics (Alberti et al., 2010b) as a simultaneous extension of SDRT (Asher–Lascarides 2003) and LDRT (Alberti, 2000), which can be regarded as a compromise between the total rejection of formalization by cognitive semanticists and the different kinds of higher order intensional logic applied by formal semanticists, say, Pollard (2007) or Kamp et al. (2011). Here, we provide a basic data structure based on Prolog facts on which the implementation of ℜeALIS will be based.

2 ℜEALIS

2.1 ℜeALIS, Reciprocal and Lifelong Interpretation System

We hereby summarize the main concepts of ℜeALIS which were introduced in Alberti et al. 2010b as a "post-Montagovian" (Kampian) theory concerning the formal interpretation of sentences constituting coherent discourses (see also Asher–Lascarides 2003). ℜeALIS has a LIfelong model of lexical, interpersonal ("REciprocal") and cultural knowledge of interpreters, too. "LIfelong" also means that the DRS-like structures of ℜeALIS are continuously being built, being able to reach an arbitrary degree of complexity – much like the structures of "world knowledge" in the human mind itself.

ℜeALIS reconciles three objectives of formal semantics: the exact formal basis itself (Montague's Thesis), compositionality (postulating the existence of a homomorphism from syntax to semantics) and, most importantly, ℜeALIS' own "discourse representationalism". The main difference between Kampian (and its extensions) and ℜeALIS' DRT is that the outer world, the states of the interpreters' mind and the information structure of discourses (which are also stored in the interpreter's mind, becoming part of it) are described by a unified model (see detailed def. in Alberti 2011a:139-149). To do this, the *infons* of Seligman–Moss (1997:245) are used to describe information structure from the view of the external world and the vertical hierarchy of the so-called *worldlets* which are parts of the *internal (mental) world.* Shortly, the *information state* of any interpreter is depicted by ℜeALIS' own DRS boxes. A part of the external world is projected into them so that each interpreter's unique perception and knowledge about the world – or parts of these – is represented by an internal world with multiple worldlets. Embedded worldlets are used to describe the hierarchy of *fictiveness* and interpersonal knowledge (see Figure 4), too.

The principal theoretical difference between the Montagovian semantics, the Kampian DRT and ℜeALIS is shown here:

Alberti G. and Károly M..
THE IMPLEMENTED HUMAN INTERPRETER AS A DATABASE.
DOI: 10.5220/0003639903790385
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2011), pages 379-385
ISBN: 978-989-8425-80-5

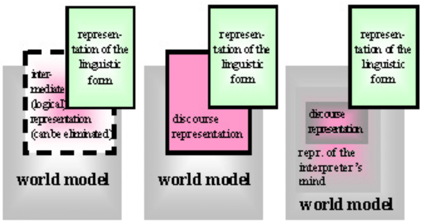

Figure 1: Montagovian model, (S)DRT and ℜeALIS.

2.2 Explanation of ℜeALIS' Definition

To provide sufficient theoretical background to our database, we give a short explanation of ℜeALIS' world model (Alberti, 2011a:139-149), see also pp. 151-179 for the detailed definition of its dynamic and static interpretation), providing some additional information to Alberti et al. 2010b. These structures were (partially) implemented as data structures (as shown in Section 4), the four *internal functions* form the starting point of a future ℜeALIS program.

U and the External World. ℜeALIS' world model is $\Re = \langle U, \mathbf{W_0}, W\rangle$ where $|U|=\aleph_0$, $\mathbf{W_0} = \langle U_0, \mathbf{T}, \mathbf{S}, \mathbf{I}, \mathbf{D}, \Omega, \mathbf{A}\rangle$, $W \subseteq I\times T_m \rightarrow \langle U[i], \sigma[i,t]^{\Pi}, \alpha[i,t]^{\Psi}, \lambda[i,t]^{\Lambda}, \kappa[i,t]^{K}\rangle$; $\mathbf{T}=\langle T,\Theta\rangle$, $\mathbf{S}=\langle S,\Xi\rangle$, $\mathbf{I}=\langle I,Y\rangle$, $\mathbf{D}=\langle D,\Delta\rangle$. First of all we note that all sets are *finite* or *countable*, e.g. time and space are based on $\mathbf{Q}$. Secondly, U_0 (and U) also contains elements that belong to the structure of ℜeALIS itself – because they, too, can be referred. The set of possible time intervals (T) is isomorphic with the set of $\mathbf{Q}$'s intervals). Spatial entities ($S \cong \mathbf{Q}^3$), interpreters (I) and physical linguistic signs (D) are also included.

The relations over T, S and I (marked by Θ, Ξ and Y) are arbitrary, but Δ must contain Dis ($\subset$ D as a *unary* relation, containing the *discourses* as linguistic entities) and Morph ($\subset D^2$ where $\langle d', d\rangle \in$ Morph if d' is a *morph* (in the linguistic sense) of the discourse d. The set of morphs of any discourse must be linearly ordered. The set of *core relations* ($\Omega \subset T\times U_0$*) of the external world must have a compulsory PERCEIVE element used by the infon (Seligman–Moss 1997) ι ($\in\mathbf{A}$) = ⟨PERCEIVE, t($\in$T), i($\in$I), j($\in$I), d($\in$D), s($\in$S)⟩.

U[i] and the Internal Worlds. ℜeALIS' interpretation actually defines the interpretation of discourse d in ι relative to the external world $\mathbf{W_0}$ and the internal world W[i,t], the latter representing the *information state* of interpreter i at the moment t. We also suppose that $\forall i\exists! t',t''\ t'\leq t''\ |\ t<t' \vee t>t'' \rightarrow$ σ[i,t], α[i,t], λ[i,t], κ[i,t] (i's *internal functions*) are empty with t' and t" being the *time of birth* and *death* of interpreter i (~"lifelong" property).

The elements of U[i], where $\forall i\forall j\ i,j\in I,\ i\neq j \rightarrow U[i]\cap U[j]=\varnothing$, are called *referents*. The referents may be *anchored* to each other (reversibly; not necessarily *identified* because identifying requires *accommodation*). Any relations between the referents are defined by the internal functions.

The two arguments of the **eventuality function** $\sigma[i,t]^{\Pi}$: $\Pi\times U[i] \rightarrow U[i]$ are the *eventual label* and the *eventual referent*. The label Π, is an ordered pair, too. A few examples of σ are: σ[i,t] (⟨Pred,π⟩,e)=p (resulting a *predicate referent*); σ[i,t] (⟨Temp,τ⟩,e)=t (resulting t which is the *temporal referent*); $\sigma[i,t](\langle Arg,\psi_k\rangle,e)=r_k$ (resulting an *argument* as a referent) etc. The first component of Π is actually a linguistic category, the second one represents a class within it. Here, π is a (linguistic) category of event structure, τ describes the relation between speech time, referred time and the event structure (e.g. ▼InRes describes the English present perfect tense, ▼ meaning $t_{ref}=t_{speech}$, and t_{ref} is *in* the *res*ult stage of e), while ψ_k marks either the grammatical case, or, according to Alberti–Kilián 2010, a generalized thematic role in a *polarized chain of influence.*

In Figure 4, a different (*simplified, timeless, eventuality-based*) notation is used. For example, e_1': p_{kill} r_{witch} r_{cow} means $\sigma(\langle Pred,\pi\rangle,e_1') = p_{kill}$ (*functional* notation); $\sigma(\langle Arg,\psi_1\rangle,e_1')=r_{witch}$ and $\sigma(\langle Arg,\psi_2\rangle, e_1')=r_{cow}$ respectively, where π describes the event structure of *to kill* (no preparation/result phase, only cumulative phase: *John was killed in a few seconds.*), ψ_1 is "Agent", ψ_2 is "Patient" according to Alberti–Kilián 2010, see Section 4.

The **anchoring function** works in a similar way: an example is $\alpha(\langle Ant,\langle Top,Gen\rangle\rangle,r_{she})=r_{witch}$. The first element of Ψ is the type of anchoring: <u>Arg</u>ument, <u>Pre</u>dicate, <u>Adj</u>unct, <u>Ant</u>ecedent, <u>Out</u>wards (meaning that the result of α is not a referent [$\notin\cup_i U[i]$] but an external entity of U_0). The second element of Ψ is an ordered n-tuple representing the (language dependent) factors that legitimize the anchoring: <u>Num</u>ber, <u>Gen</u>der, <u>Top</u>ic retaining, <u>Hum</u>anity etc.

The **level function** marks the worldlet which contains the referent. Only *fictive* referents are projected by λ, the root referents are not (but they can be anchored "outwards" by means of α instead). In general, $r'=\lambda(\langle\mu_k,\tau_k,i_k,\pi_k\rangle,r)$ where $\mu_k\in$ {⇖ [level-changing feature, linguistic *subordination*],⬀ [level-keeping, linguistic *co-ordination*]}×{., ?, ![modal markers ~*kinds of sentences*]}×{supp, cons, bel[ieve]$_n$, int[ent]$_n$, des[ire]$_n$, dream, see, hear, elab[orate], exp, nar[rative], back, conj, disj etc. [modal markers marking the *source of information*]} (not all combinations of μ_k are [linguistically] possible). $\tau_k \in T_m$ is the *temporal* component of the level label, i_k is the referent of the *direct owner* of r where the transitive closure of $\alpha(i_k)$ anchors i_k to a certain $i \in I$ interpreter. Finally, $\pi_k \in \{+, 0, -\}$ is the *polarity* of the level label (positive, neutral, negative), marking *true, don't know* and *false.*

The level label system is defined in a such way that one can reach the root referent i* by repeatedly applying λ: $r'' = \lambda(\langle \mu_{k-1}, \tau_{k-1}, i_{k-1}, \pi_{k-1} \rangle, r')$ … $r^{(s)} = \lambda(\langle \mu_{k-s+1}, \tau_{k-s+1}, i_{k-s+1}, \pi_{k-s+1} \rangle, r^{(s-1)})$… $r^{(k)} = \lambda(\langle \mu_1, \tau_1, i_1, \pi_1 \rangle, r^{(k-1)}) = i^*$ where $\alpha(\langle Out, \langle \ldots \rangle \rangle, i^*) = i$. The χ series of the Λ labels used in this formula is called the *worldlet index* of r (χ is *empty* if r is a root referent) and the set of referents with a certain χ worldlet index is called a *worldlet*. For the *root worldlet* of any i interpreter, χ is empty. This results in Kampian DRS boxes with SDRT-like (Asher–Lascarides 2003) rhetorical information (Figure 4).

The **cursor function** is $\kappa[i,t]: K \rightarrow U[i]$. K is a finite set of pre-defined cursor labels and must contain the following elements (among others):

- κ[i,t](Now/Here/Ego)=t/s/i, α(⟨Out,⟨…⟩⟩, t/s/i)=t/s(∈S)/i
- κ[i,t](ThisWay)=e (speech situation)
- κ[i,t](Then/There/Eve)=t'/s'/e'(the *referred time*/spatial entity/eventuality)
- κ[i,t](You)=j

The temporary states of these four internal functions above an interpreter's internal universe serve as her "agent model" in the process of (static and dynamic) interpretation.

The *interpretation* of any "perceived" discourse can be defined in our model relative to an external world $\mathbf{W_0}$ and internal world W[i,t]. (summary: (Alberti et al. 2011b)), (details: (Alberti 2011a)). Since the data structure that we intend to describe is a mapping of the actual world model, and most importantly, its four internal functions, we do not go into the depths of the mathematical definition of the interpretation. Indeed, we summarize the process of the actual discourse analysis later on.

2.3 Illustration of the Apparatus of ℜeALIS

Let us consider a sentence (1a) with two meanings. The figure in the center in (1d) shows the lexical items which have been retrieved by the words of the sentence (→eventuality function σ). The lexical items (except for those of *the piano*) are shown as eventualities more or less in the style of DRT. The formulae in (1b-c) express the difference between the collective and the distributive reading also in DRT-style. The two figures with arrows show two different anchorings (α) of referents belonging to arguments and eventualities to each other. And what is defined by these two anchorings exactly results in the two formulae in (1b-c).

Example 1: Readings and anchoring function α.

(1a) *The boys lifted the piano.*

(1b) COLLECTIVE READING:

$[_{e4}[_{e5=e41}[_{e2=e51}$ element r_{21} $r_{22}] \rightarrow [_{e1=e52}$ boy $r_{21}]] \wedge [_{e42=e3}$ lifted r_{22} $r_{32}]]$

'if referent r_{21} is an element of group r_{22}, it is a boy (hence, r_{22} is a group of boys); and *this group* r_{22} lifted the piano (r_{32})'

(1c) DISTRIBUTIVE READING:

$[_{e5}[_{e2=e51}$ element r_{21} $r_{22}] \rightarrow [_{e4=e52}[_{e1=e41}$ boy $r_{21}] \wedge [_{e42=e3}$ lifted r_{21} $r_{32}]]$

'if referent r_{21} is an element of group r_{22}, it is a boy who lifted the piano (r_{32}) (hence, r_{22} is a group of boys, **each** of them lifted the piano)'

(1d) α anchoring in the case of the **collective** reading (🡧) / the **distributive** reading (🡮):

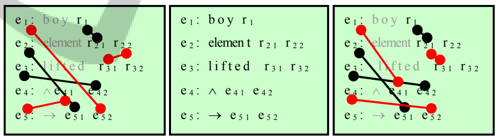

Figure 2: The two possible readings of (1a). The lines mark the anchoring function α.

3 WHERE ARE THE POSSIBLE WORLDS?

3.1 The Granularity Problem as a Further Argument for ℜeALIS

Let our starting point be Pollard's criticism on the mainstream Kripke/Montague-inspired *possible-worlds* (PW) semantics: it is "a framework known to have dubious foundations" (Pollard, 2007:1) because of the *granularity* problem (2), among other deficiencies (3); and hence "the idea of taking worlds as a primitive of semantic theory is a serious misstep" (Pollard, 2007:33). Another stubborn problem in mainstream NL semantics concerns distinct *accessibility* of certain referents in logically equivalent sentences (4). In this area a promising solution is offered by DRT, but at the cost (see (5)) of introducing an extra level of representation, that of *discourse structure* (4c). Nor does (the Higher Order Intensional Dynamic Logic of) current DRT exceed the *PW* approach criticized above; although – we claim – there would be an obvious way of using (gigantic) DRSs as lifelong representations of interpreters' *information states* and embedded DRS boxes (consisting of propositions) as PW-like ilks. Following this way we have elaborated a "ReAL" interpretation system (6a), which provides straightforward solutions (6b-f) to (2-5), including even the *Hob-Nob* problem (4d).

The essence of the *granularity* problem is that having the same reference / meaning (see (2)) is not a sufficient condition to allow replacement of one name / sentence for another in a larger expression. The other problem (Pollard, 2007:30-31) lies in "the standard view, [viz.] *reference* is compositional"; that is why "Frege had to resort to claiming that utterances of sentences in certain contexts [e.g. S3 in (3)] had the customary *sense* as the reference," which requires "sleight of hand."

Example 2: The Granularity Problem.

(2) The ancients realized that [Hesperus was Hesperus]$_{S1}$ / [Hesperus was Phosphorus]$_{S2}$.

Example 3: The problem of using the customary *sense* as the *reference* (see fn36, Pollard 2007:31).

(3) [(Justin Timberlake knows that) [Paris Hilton believes [snow is white]$_{S3}$]$_{S2}$]$_{S1}$.

We follow Pollard in assuming that "worlds are constructed from propositions ..., and not the other way around" (Pollard, 2007:34), but intend to work out this approach in a DRT-based framework in order to account for phenomena concerning *referent accessibility* (4a-d), at the same time. We claim, however, that our system is devoid of DRT's "extra level" problem (5).

Example 4: The problem of referent accessibility.

(4a) [[A delegate arrived.]$_{S1}$ *She* registered.]$_{Discourse1}$ / [[It is not the case that every delegate failed to arrive.]$_{S2}$ **She* registered.]$_{Discourse2}$

(4b) S1 and S2 are logically equivalent: $\exists x\Phi \Leftrightarrow \neg\forall x\neg\Phi$

(4c) The representation of Discourse1/ Discourse2 in DRT: x "enclosed" is not accessible to y

(4d) Hob believes that a witch has killed Cob's cow and Nob thinks that *she* has blighted Bob's sow.

(5) **An Argument against DRT:** DRS is an illegitimate extra level of representation between syntactic structure and the model of world in the course of interpretation (Groenendijk–Stokhof, 1991).

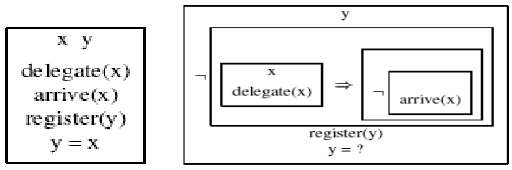

Figure 3: Illustration to (4a).

3.2 ℜeALIS as an Epistemic Multi-agent System

ℜeALIS is based on the idea that (gigantic) DRS-like structures, set W, are suitable for serving as *lifelong* representations of *information states* of interpreters, i.e. the "agents" getting information. Furthermore, as stated in Alberti et al. 2010b, both *static* evalutation (Tru) and *dynamic interpretation* (Dyn) can be carried out. *Labeled* tree systems of worldlets can serve as parts of the world model (6f): the interpretation of *modal* sentences is to be based on certain worldlets instead of W (6c).

Summary 6: ReALIS as an epistemic multi-agent system and its features.

(6a) $\mathfrak{R} = \langle \mathbf{W_o}, \mathbf{W}, \mathbf{Dyn}, \mathbf{Tru} \rangle$

(6b) →(2): co-anchoring (α(r')=α(r")) does not necessarily imply the identity of referents (r' and r"); *identifying* (and *concluding*, whose amount *does* matter) requires accommodation

(6c) →(3): S1 is *true* if S3's eventuality referent can be found in an appropriately labeled *worldlet* of JT (containing JT's thoughts on PH's beliefs)

(6e) →(4a-b): certain referents may be "enclosed" in the worldlet structure

(6f) →(4d): see Figure 4; what provides more freedom in constructing worldlet structures in ℜeALIS is that DRT's box structure depends on logical factors whereas the worldlet structure is affected by *pragmatic* factors (as well)

(6g) →(5): DRS-like representations form no "extra level" but turn into parts of the ℜeALIS world model (enriched with descriptions of interpreters themselves)

The content of worldlets can be enriched by (chiefly pragmatics-dependent) *accommodation*, too: *identifying referents* (6b, e) or drawing *logical inferences* can be bound to pragmatic conditions.

Let the last figure serve as an illustration. The representation to the left expresses the pure semantic content of (4d). According to this, Hob's witch is not yet accessible to the referent of *she*. But then we can consider some pragmatic factor, discussed by Zeevat (2005:549) as follows: "Hob may have told Nob about his belief [→repr. to the right], [or] there may be a rumour in the village about a witch that has played a causal role in the formation of Hob's and Nob's beliefs [→repr. in the middle]..."

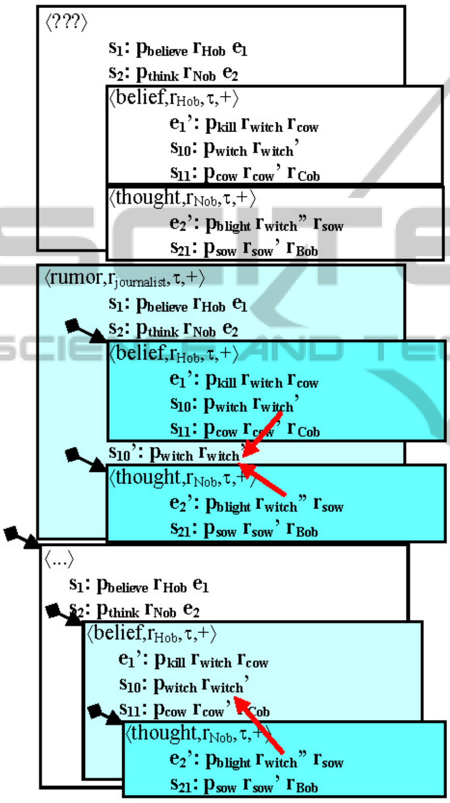

Figure 4: The Hob-Nob sentences in the ℜeALIS model.

4 IMPLEMENTATION

4.1 Basic Principles of the Data Structure and the Function σ

In ℜeALIS, the decomposition of syntax plays a crucial role (Alberti, 2011b). Syntactic and semantic information are stored *together* in the core lexicon or produced by the lexical rules (Alberti et al., 2010a), according to our 'totally lexicalist' approach. To handle all these data, we use a Prolog engine which is also able to handle **non-determinism** (see also Figure 2: the sentence can be disambiguated after having some more sentences of the discourse). For now, theoretical consistency in the implementation of the ℜeALIS model is more important – but in the long run, the theory itself may be refined to improve the speed of our program and/or reduce its computational complexity: as the human brain has its own limits revealed by psycholinguists, our program does not need to surpass them either. (One such limit is the depth of discourse analysis.)

Here we summarize the practical Prolog equivalents of our concepts as parts of a (future) Prolog fact database. In the program itself we excessively use `assert` and `retract` to manage it. We note here that for now, we are working primarily on Hungarian (a highly agglutinative language with free word order) and English.

- **Input:** Step 1: morphologically tagged SL text; Step 2: untagged SL text
- **Output:** discourse representation, Step 3: make bidirectional use of the program possible (for e.g. machine translation).
- **Logic:** linguistic/semantic data → Prolog facts (Horn clauses)
  - modal operators → λ-labels
  - disjunction → different λ-labels
- **Reversing basic Prolog mechanism:** linguistic data = **facts**, derivation of possible semantics = **rules** (forward chaining)
- **Lexicon:** Prolog facts (later SQL), totally lexicalist approach (Alberti et al. 2010a-b)
- **Referents:** variables/dynamic facts, values: instances of any class (unique entities are also regarded as classes, e.g. Paris Hilton)
- **Problem:** implementing second-order logic with first-order tools
  - **reification:** predicates as data
  - use of extra-logical methods to extract predicates if needed

Predicate Referents can only be handled by reification. An example of a predicate template:

```
semrolefeat(111,'Agent').
semrolefeat(112,'Patient').
semrolefeat(113,'Beneficiary').
ref(20,p,'R111^R112^R113^give(R111,
R112,R113)').
```

The verb *to give* is reified and stored as data (string) but we can rebuild it by using extra-logical techniques. The roles of `Agent`, `Patient` and `Beneficiary` keep their linguistic meanings. In the case of active transitive verbs, `Agent` becomes subject. A Hungarian exception: *Mari*$_{Mary}$ *meg*$_{Perf}$-

híz$_{\text{grow-fat}}$-*ott*$_{\text{Past}}$. (*Mary has grown fat.*) The intransitive verb *meghízik* has only one argument, the subject, which is `Patient`.

Eventual Referents can be regarded as instantiated predicate referents: *Peter gave Mary the book. And **that** makes me angry.*

Analysis. We suppose that or input is already morphologically analyzed and POS-tagged. So we can (Step 1) assign potential referents to all nouns (r), verbs (e), adverbs (e.g. t), adjectives (e, p, r) and pronouns (typeless). Then (Step 2) we can determine σ-relations by doing syntactical analysis. Here are some facts that should be asserted during the analysis of the two sentences:

```
ref(...,p,...). %see give above
ref(11,r,'Peter'). ref(12,r,'Mary').
ref(13,r,'book').
%--calculate evref here
ref(21,e,'give(11,12,13)').
ref(22,x,'that')].%type is not set
...
alpha(22,21). %see below
ref(23,e,make_angry(22,EGO)).
```

Conclusion: `make_angry(21,EGO).`

Temporal Referents are handled by time intervals: *a week ago* is described by `date(past, (0,0,1,0,0,0,0),(0,0,0,1,0,0,0))`. The first 1 is in the field `weeks` while the second one marks the 'precision' of the linguistic expression – about *one day* in the English language. `now` is regarded as a special temporal referent.

The **words** and their default morphology and parts of their syntax and semantics are stored in the core lexicon. This is basically the same as stated in (Alberti et al., 2003) and (Alberti–Kilián, 2010) but each feature (POS, type of morpheme etc.) is described by a different fact. Lexical units search for each other by *offering and demanding* features.

```
lexfeatval(301,3,'Nom-Subj').
lexfeat('Bob',301,+7).
```

This means that the 'Bob' *offers* a grammatical case of 'NOM-Subject' with rank +7. The same applies to POS, referentiality/definiteness etc. If morphological inflection takes place, the default features are overridden: in Hungarian, *Bob-ot*$_{\text{Acc}}$ will get a new *`lexfeat('Bob',302,+1)`* feature, which has a "stronger" rank of +1.

Now we take a look into the "demand" part of the database. Verbs search for their arguments:

```
semdemand('GIVE','Agent',1).
%1 would be 2,3 etc. for more args
synsem('GIVE','en','give').
synsem('GIVE','hu','ad'). %etc.
%Note: syndemand is for English
syndemand('give','Noun',1,+2).
syndemand('give','Nom-Subj',1,+2).
syndemand('give','Nei-3',1,+2).
%1 is the same as in semdemand
%strings become ID's in the real DB
```

As it is shown, the semantic field GIVE has several language-dependent syntactical representations. Here we only show the facts for the agent (subject) of the English verb *give* (*ad* in Hungarian). Its patient and beneficiary can be represented in a similar way. `Nei-3` is a neighborhood rank of –3 meaning that Argument 1 must take place *before* the verb with rank 3 but this requirement can be overridden with any other fact with rank e.g. +1.

4.2 The Anchoring Function α

The function α *anchors* referents to each other. Anchored referents are supposed to refer to the same thing but anchoring is not necessarily permanent. The relations behind α must be defined by a background ontology: the legitimizing factors are often extra-grammatical, based only on semantic categories (e.g. when a *parrot* is mentioned as a *bird*). In subsection 4.1 (`alpha(22,21)`) this is not the case: *that* can have a function of referring to the eventuality – formally, $\alpha(\langle Ant,\langle Eve\rangle\rangle,r_{22})=e_{21}$. However, `alpha` is restricted to mark antecedents (Ant) in its present form without labels.

Multiple grammatical and semantic factors are involved in α and not all analyses are correct. So we plan to use a cost/weight metric for α to measure discourse coherence. If it surpasses a certain limit (which is set by the user), the discourse is considered incoherent and ill-formed. (The same applies for tolerance to grammatical errors and σ.)

4.3 The Level Function λ

λ assigns all referents to the specified worldlet(s) of ℜeALIS and describes **mood** and **rhetorical** relations. As defined, all referents must carry all level labels – as shown here:

Example 5: Storing the level labels of referents.

(6) *If only* I$_{\text{ego}}$ had$_{\text{e2}}$ a car$_{\text{r1}}$! Sue$_{\text{r3}}$ *would* drive$_{\text{e5}}$ it$_{\text{r4}}$, too. But now I$_{\text{ego}}$ only *have*$_{\text{e6}}$ a motorcycle$_{\text{r7}}$.

```
ref(1,r,'car').
ref(2,e,'have(EGO,1)').
ref(3,r,'Sue'). ref(4,r,'it').
ref(5,e,'drive(3,4)').
```

```
ref(6,r,'motorcycle').
ref(7,e,'have(EGO,6)'). alpha(5,2).
lambda(1,[[sub,now,des,0,101,+1]]).
lambda(2,[[sub,now,des,0,101,+1]]).
lambda(3,[]).%Sue is a real entity
lambda(4,[[sub,now,des,0,101,+1]]).
lambda(5,[[sub,now,des,0,101,+1]]).
lambda(6,[]). lambda(7,[]).
```

The second argument of `lambda` is the above-mentioned list of level labels. Each level label has six parameters according to the definition: *co/sub* marks how the actual worldlet is related to the next one in the label chain, the first number (0 by default) marks the level of belief etc. (where applicable), 101 is a placeholder for the interpreter's ID, and, finally, the polarity is marked by +1/0/–1 (believed etc. to be *true*, *"don't know"* or *false*).

λ is set inherently by verbs (e.g. *to believe, to think, to desire*), adjectives (*alleged*) and some other words (e.g. negative words) and morphemes (modal markers).

5 CONCLUSIONS

A full implementation of ℜeALIS is yet to come but our progress and partial results are going to be published continuously. We demonstrated the functioning of the ℜeALIS model on a few classical semantic problems, arguing that the cognitive paradigm does not necessarily exclude mathematical exactness. By now, we fixed most of the data formats of the lexicon and database.

Although we plan to use external ontologies and/or dictionaries, their integration into the ℜeALIS software is only the first step. Since ℜeALIS is a **lifelong** [self reference] interpretation system, our database is designed to **build itself (**by `assert`ions) when analyzing discourses. So the values of the three base functions – σ, α and λ (and the cursor function κ whose exact functioning is yet to be determined) form an integrate part of the database even if the relations discussed here are analyzed on-the-fly.

ACKNOWLEDGEMENTS

We are grateful to SROP-4.2.1.B-10/2/KONV/2010/ KONV-2010-0002 (Developing Competitiveness of Universities in the Southern Transdanubian Region) for their contribution to our costs at KEOD 2011.

REFERENCES

Alberti, G. 2000. Lifelong Discourse Representation Structures. *Gothenburg Papers in Computational Linguistics 00-5*, 13-20.

Alberti, G., Balogh, K., Kleiber, J., Viszket, A. 2003. Total Lexicalism and GASGrammars: A Direct Way to Semantics. In Gelbukh, A. ed.: *Proceedings of CICLing 2003*, Springer, Mexico City. 37-48.

Alberti, G. 2011a. *ReALIS* [in Hungarian]. Academic Press, Budapest. To appear.

Alberti, G. 2011b. ReALIS, or the decomposition of syntax [in Hungarian]. In: *General Studies of Linguistics.* Academic Press, Budapest. To appear.

Alberti, G., Károly, M., Kleiber, J., 2010a. From Sentences to Scope Relations and Backward. In Sharp, B., Zock, M. eds.: *Natural Language Processing and Cognitive Science. Proceedings of NLPCS 2010.* SciTePress. Funchal, Madeira. 100-111.

Alberti, G., Károly, M., Kleiber, J., 2010b. The ReALIS model of human interpreters and its application in computational linguistics. In Cordeiro, J., Virvou, M., Shiskov, B. eds.: *Proceedings of the 5th International Conference on Software and Data Technologies Vol. 2.* SciTePress. Funchal, Madeira. 468-474.

Alberti, G., Kilián, I., 2010. Polarized chains of influence instead of argument structure lists – or the function σ of ℜeALIS [in Hungarian]. In *Proceedings of the 7th Hungarian Conference on Computational Linguistics.* Univ. of Szeged. Szeged. 113-126.

Asher, N., Lascarides, A., 2003. *Logics of Conversation*, Cambridge University Press. Cambridge.

Dekker, P., 1999. Coreference and Representationalism. In von Heusinger, K., Egli, U. eds.: *Reference and Anaphoric Relations*. Kluwer. Dordrecht. 287-310.

Groenendijk, J., Stokhof, M., 1991. Dynamic Predicate Logic. In *Linguistics and Philosophy Vol. 14.* Kluwer. Dordrecht. 39-100.

Kamp, H., van Genabith, J., Reyle, U., 2011. Discourse Representation Theory. In *Handbook of Philosophical Logic, Vol. 15*. Springer. Heidelberg. 125-394.

Pollard, C., 2007. Hyperintensions. *ESSLLI 2007.* http://www.cs.tcd.ie/esslli2007.

Reyle, U., 1993. Dealing with Ambiguities by Underspecification. *Semantics* 10, 123-179.

Selingman, J., Moss, L. S., 1997. Situation Theory. In van Benthem, J., ter Meulen, A. eds.: *Handbook of Logic and Language*. MIT Press. Cambridge. 239-309.

Szabolcsi, A. ed., 1997. *Ways of Scope Talking.* Kluwer. Dordrecht.

Zeevat, H. 2005. Overlaying Contexts of Interpretation. In Mayer, E., Bary, C., Huitink, J. eds.: *Proceedings of SuB 9.* Radboud University Nijmegen. Nijmegen. 538-552.