Temporal Aspects of Knowledge and Information

Marie Duží[1], Aleš Horák[2], and Pavel Materna[3]

[1] VŠB–Technical University of Ostrava
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
marie.duzi@vsb.cz
[2] Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
hales@fi.muni.cz
[3] Institute of Philosophy of the Academy of Sciences of the Czech Republic, v.v.i.
Jilská 1, 110 00 Prague 1, Czech Republic

Abstract. The authors of the paper have proposed a collaborative research project that should start in 2010 and proceed in three successive years. The goal of the proposed project is to develop tools and mechanisms for computer-aided natural-language analysis, knowledge management and reasoning. In a broader sense, the main objective is to propose a new solution to natural language analysis and reasoning based on the procedural semantics of Transparent Intensional Logic (TIL). Primarily, our research is classified in the area of philosophical logic. Subsidiary works then fall under theoretical linguistics and computer science. Our research is focused on the temporal, modal and epistemic aspects of knowledge representation and reasoning. The main problems under our scrutiny are: TIL proof calculus for hyperintensional, partial, typed lambda calculus; analysis of tenses and temporal logic; analysis of epistemic verbs and events; analysis of anaphora references and topic-focus articulation. Concerning practical applications, we develop functional programming language TIL-Script, the computational variant of TIL. The main result we aim to achieve is computer-aided analysis of natural-language texts based on web ontologies like WordNet, FrameNet/VerbNet, VerbaLex and large text corpora.

Key words: Transparent intensional logic; TIL; temporal knowledge

1 State-of-the-art and objectives of the project

In the age of computers there is a new challenge that might be characterised as the problem of human-computer communication, or rather *human communication via computers*. Currently, communication *via* computers is mostly computer driven. Humans must be able to formulate quasi-logical queries, and the answer provided by a computer is far from being comprehensible and natural-language like. The World Wide Web (www) media became a prevailing source of *information* and *knowledge*. The most common format of web documents, HTML, focuses on the *form* of knowledge representation rather than its content.

Petr Sojka, Aleš Horák (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2009, pp. 39–47, 2009.

While this format standardisation was undoubtedly a necessary and important step, the *semantic content* of a HTML document is what we should focus on now. Otherwise the human-computer communication and web-information processing are limited and do not serve the goal.

An improvement of this situation comes with the Semantic Web.[4] Semantic Web is a vision of web data defined and linked in such a way that it can be processed and understood by machines, not just by humans. Though the key idea of Semantic Web, i.e. meta-data, is an old one, still only a negligible portion of today's web pages make use of meta-data techniques. The semantic web approach currently offers well-defined static features, e.g. ontologies, but the inference mechanisms and *temporal features* are in need of scrutiny. The evolution of the web itself shows that the vision of Semantic Web cannot be realised unless intelligent automatic procedures are developed in order to analyse and transform natural-language knowledge into a Semantic Web representation and *vice versa*. The goal of our project is to fill this gap.

The goal of the proposed project is to develop tools and mechanisms for a computer-aided natural-language analysis, knowledge management and reasoning. In a broader sense, the main objective is to propose *a new solution to natural language analysis and reasoning*.

To this end we will make use of the expressive system of *Transparent Intensional Logic* (TIL).[5] The logical system we vote for makes it possible to explicitly formally specify and process all kinds of knowledge including difficult natural-language phenomena such as temporal relations, modalities, personal attitudes to other knowledge, and anaphoric references. This is due to the *procedural semantics* of TIL. From the theoretical point of view, much has been done in the area of logical analysis of natural language, as the results of the previous projects testify. Yet the problem of *tenses* and generally *temporal logic* has been rather neglected till now in TIL-like analyses, though due to numerous applications there is a pressing need for such analyses. The only TIL paper that dealt with the analysis of tenses was a pioneer paper of Pavel Tichý [12]. Thus we will focus on *temporal aspects* of knowledge representation and reasoning. Another area under scrutiny is the analysis of *context-dependent ambiguities* such as topic-focus articulation and anaphora.[6] Here we will make use of the results of *linguistic analysis*, because theoretical linguistics and logic must collaborate and walk hand in hand.

The character of the proposed project is *interdisciplinary*. It combines *logical analysis of natural language* with *theoretical linguistics* and *computer science.*

1.1 Natural Language Analysis and Knowledge representation

Natural language analysis has a long tradition, beginning with Frege, Chomsky, Carnap, Montague and many others. However, it became a hot topic as soon as the first real-world computer applications came into being. The possibility of an automatic "intelligent" processing of the huge amount of texts that people

[4] See [1,6]. [5] See, for instance, [8,9,14,15,7]. [6] See [2].

have produced is a great contribution to the way people can make use of computers. Intelligent processing techniques usually analyse natural language sentences on several consecutive levels, to name at least the most important of these, namely morphological, syntactic and semantic levels. Morphological and syntactic natural language analysis is currently well-developed and described in numerous papers. Thus we may say that the automatic techniques dealing with morphology and syntax produce plausible results (at least regarding frequently used languages). However, this is not the case of *semantic analysis* where the results are far from being satisfactory. Logical tools that are broadly applied are mostly based on the first-order logic (FOL) paradigm. FOL tools have been primarily developed for mathematics, where they are successfully applied. But natural language is too rich to be encoded by a first-order formalism.

In the last decades numerous methods of natural-language analysis and knowledge representation have been developed. The approaches range from encoding knowledge in the form of some strictly specified procedures in a programming-like language through semantic networks and frames up to statistical and probabilistic methods. However, none of these approaches is universal and fine-grained enough to make it possible to specify all the semantically salient features of natural language expressions. Moreover, there is a pressing need for a unique knowledge representation framework. *The goal of the project is to demonstrate that TIL can serve as such a fine-grained universal framework.*

Concerning temporal aspects of knowledge, we have to take into account that sentences in the present, past and future tenses have different truth-conditions. This fact has been observed by numerous logicians, and many variants of so-called *temporal logic* have been developed. These formal systems are mostly viewed as a special case of modal logic interpreted by means of Kripkean possible-world semantics. Another view of temporal logic is motivated by the logical analysis of natural language. This approach develops temporal logic as a formalization of linguistic conventions with respect to tenses, and it has been applied in particular by Arthur Prior who developed the axiomatic system of so-called *tense logic*. Prior introduced operators like P (*weak past*, corresponding to the modal operator $\Diamond$) and H (*strong past*, corresponding to $\Box$), whose intended meanings are:

$P(q)$ – "It has at some time been the case that q".

$H(q)$ – "It has always been the case that q".

Subsequently, systems of temporal logic have been further developed by computer scientists, notably Zohar Manna and Amir Pnueli, and widely used for formal verification of programs and for encoding temporal knowledge within artificial intelligence. Despite the great applicability of particular variants of tense logic in the semantics of programming languages, the systems just mentioned suffer a drawback when applied to the semantics of natural language. The drawback is their inability to adequately analyse sentences indicating a 'point of reference' referring to the interval when the sentence was or will be true. Such sentences come attached with a *presupposition* under which a sentence is true or false. To illustrate the problem, consider the sentences

"Tom is sick".

"Tom has been sick".

"Tom was sick throughout the year 2008".

"The Mayor of Dunedin was sick throughout the year 2008".

The first two sentences do not come with a presupposition. They ascribe to Tom the property of being sick and of having been sick, respectively. They are true or false according as Tom has the relevant property. However, the truth-conditions of the third sentence depend not only on whether Tom has the property of being sick throughout the year 2008, but also on the time at which the sentence is evaluated. If T is the time of evaluation, then the truth-conditions are specified as follows:

(a) *If* $T \leq$ December 31, 2008, 24:00, *then* ***no value*** *(fail), else (b)*;
(b) *If* the whole year 2008 precedes T (i.e., $T >$ December 31, 2008, 24:00), *then if* Tom was sick at all times during 2008, *then* the value ***True***, *else* the value ***False***.

The fourth sentence is ambiguous. Its ambiguity is pivoted on the difference between *de dicto* and *de re*. On its *de re* reading, the truth-conditions are identical to (a), (b), provided there is a Mayor of Dunedin. If the office of Mayor of Dunedin is vacant, the sentence has *no* truth-value. The *de dicto* reading of the sentence requires us to evaluate the proposition *in the past* according as the office was occupied throughout the year 2008, and according as the Mayor *had been* sick throughout the year 2008. Thus it is irrelevant whether there is a Mayor at the present time. Our analysis must respect these truth-conditions. Tichý's solution in [12] is difficult to understand, because Tichý applies the *singulariser* function to a singleton typed as containing a truth-value in order to make the set fail to deliver a truth-value in case the associated presupposition is not satisfied. Tichý's analysis is analogous to what the computer scientist would call an *imperative* rather than *declarative* analysis. The downside to an imperative analysis is that it may conceal flaws that rear their head only when the analysis is applied to extreme situations. Yet there is an elegant alternative that makes use of the 'if-then-else' connective. There has been much dispute over the semantics of 'if-then-else' in the logic of computer science, and it is often said that *if-then-else* is a non-strict function that does not behave in compliance with the compositionality principle. For instance, the application of *if-then-else* to a condition P and two formulae F_1 and F_2 is not improper (by failing to produce a product) even when F_2 is improper whenever P is true. However, there is no cogent reason to settle for non-strictness. The *procedural* semantics of TIL enables us to specify a strict definition of *if-then-else* that meets the compositionality constraint. The definition of "If P then F_1, else F_2" is a procedure that decomposes into two phases. First, on the basis of the condition P, select one of F_1, F_2 as the procedure to be executed. Second, execute the selected procedure. Thus, for instance, if P is true then F_1 is executed rather than F_2, and the possible improperness of F_2 does not matter.

1.2 Inference

There are a huge amount of systems of inference developed within contemporary logic. Beginning with the systems of 'classical logic', like general resolution method or Gentzen-like systems, there are inference systems based on the so-called 'non-classical logics'.What is important from our viewpoint is that many particular logics build up their apparatus dependently on a particular problem so that the notions necessary for solving such a problem are introduced *ad hoc*, even if they could be defined in terms of a more general system. Such are the 'of-logics' ("logic of events", "epistemic logics" (i.e. logic of knowledge), "logic of questions", "deontic logic" (i.e., logic of norms) etc.).

TIL offers a unique approach, which helps to find the *meaning-driven* logical tools for solving particular problems on the basis of a homogeneous general conception. Thus in comparison with logics based on a distinct philosophy TIL can demonstrate that its philosophy yields greater expressivity. Logics that do not recognize hyper-intensions (such as procedural constructions, i.e., the great majority of logics) do not adequately solve for example inferences in the milieu of propositional attitudes. This is also the case of the most important rival logic, viz. Montague's IL.[7] There are several shortcomings of Montague's IL. Summarising briefly. First, IL fails to satisfy Church-Rosser property. The reason typically cited for the logic of IL displaying such a deviant behaviour is that the logic has been designed to reflect opacity phenomena of natural language. But this is actually a serious deficiency caused by *ad hoc* introducing the operators $^\wedge$, $^\vee$, F and P that are equipped with hidden 'ghost' variables ranging over possible worlds and times. TIL explicit intensionalisation and temporalisation makes it possible to uniquely adhere to anti-contextualism. Second, the functions of IL are restricted to total functions. But we need to work with partial functions, unless we are content with an unmanageable explosion of domains. Third, functions typically have more than one argument. Usually we are told that these $n-$ary functions can be *represented* by unary composite functions. Schönfinkel in [11] observed that there is a one-to-one isomorphic correspondence between $n-$ary functions and certain unary composite functions.However, this isomorphism breaks down when *properly* partial functions are involved, as Tichý showed in [13], see [15, pp. 467–8]. A final objection to IL is that it fails to accommodate hyperintensionality, as indeed any formal logic interpreted set-theoretically is bound to fail to. Only when we embrace an algorithmic/procedural semantics are we able to handle structured, hyperintensional meanings of natural-language expressions. And we argue that any theory of natural-language analysis needs a hyperintensional (preferably procedural) semantics in order to render synonymy in natural language accurately, as well as to adequately analyse hyperintensional (*de dicto/de re*) attitude reports, the phenomena of anaphora, tenses, presuppositions, and other hard cases.

The central notion of logical semantics, viz. *entailment*, is exactly defined in terms of constructions. Moreover, structured meanings known as TIL constructions make it possible to precisely distinguish between analytical and

[7] See [10].

logical validity of sentences as well as of arguments, and examine the *analytical content* of these. Thus the 'scandal of deduction', as Hinttikka formulates the failure of logical deduction to explain the utility of a valid argument, is easily solved away.[8]

2 Methodology and project planning

The project work will run in parallel in three cooperating and partly overlapping groups:

1. *TIL theoretical backgrounds* group
2. *TIL logical analysis* group
3. *TIL inference machine* group

The group of TIL theoretical backgrounds will concentrate on extensions and specifications of the current TIL theory for the purposes of the project. Summary of the milestones and topics of the project plan is stated in Table 2.

2.1 Logical Analysis of Natural Languages

The main goal of the part of logical analysis is *natural-language text processing*. We aim at the development of automatic facilities to process real-world texts in natural languages (in particular English and Czech) in order to obtain a machine-readable form of the formal representation of sentences. This logical representation will then serve as an initial stock of knowledge from which inferable knowledge will be computed and deduced by the TIL Inference Machine.

Our method of analysis consists of three steps:

(a) *Type-theoretical analysis*, i.e., assigning types to all the objects talked about by the analysed sentence, i.e. denoted by semantically self-contained subexpressions of the analysed expression.
(b) *Synthesis*, i.e., combining the constructions of the objects *ad* (a) in order to construct the entity denoted by the whole expression.
(c) *Type-theoretical checking*, i.e., checking the type-theoretical consistency of the resulting analysis.

Within the project we will advance the fundamentals of TIL theory in particular with respect to temporal, modal and epistemic aspects. We will develop logical modules as semantic upgrades of the currently implemented tool for the full-fledged syntactic analysis of natural languages called *synt*.[9] This tool has been developed by the NLP Centre of the Faculty of Informatics, Masaryk University in Brno within the group of A. Horák, and it has been thoroughly tested on the samples of Czech and English texts. The *synt* system will be further enhanced by the new techniques and algorithms defined by the *theoretical backgrounds* group.

[8] Cf. Cohen-Nagel's 'paradox of inference'. [9] See [4,5].

Table 1. Summary of the project planning.

Milestones/topics	*Theoretical results*	*Applications*	
		Logical analysis	Inference
1st **year (2010)** Simple sentences in present, past and future tenses	– specification of TIL calculus – reference corpus of TIL constructions – type classification of basic lexicon tokens	*Computer-aided analysis* of simple sentences in past, present and future tense containing selected verbs	In the scope of FOL (enriched by explicit intensionality and temporality)
2nd **year (2011)** Complex sentences in present, past and future tenses	– analysis of events – analysis of grammatical tenses – type classification of attitudes	*Computer-aided automatic analysis* of relative time-related subordinate sentences	In the scope of classical typed λ-calculus
3rd **year (2012)** Context dependencies	– inference with background and common-sense knowledge – Analysis of sentence context and discourse	*Computer-aided analysis* of complex sentences with temporal events including direct speech	TIL inference machine including partiality and hyperintensional features

The implemented system will be tested on real texts from various domains. One specific domain we have at our disposal is the domain of medicine texts. They are available at Masaryk Memorial Cancer Institute in Brno, where the system will be used for advanced question-answering techniques in cooperation with the Masaryk University NLP Centre and the Digital Enterprise Research Institute, Ireland. We will pay attention to the multilingual features of the systems. Particular subtasks of the proposed project include constructions of logical lexicons for natural language based on previous works on the subject. These lexicons include WordNet semantic networks, verb frame lexicons, rules for building the meaning representation including meta-knowledge (knowledge about knowledge).

2.2 TIL Inference Machine – Theory and Development

From the formal point of view, TIL is a partial, hyper-intensional typed lambda calculus. Pavel Tichý in [13] specified inference rules only for the pre-1988 simpler version of TIL that was based on the simple theory of types. In order to deal with agents' (*de dicto/de re*) hyper-intensional attitudes in an appropriate way, we need to work in the full-fledged TIL, i.e., with the post-1988 version based on the ramified theory of types. In this version constructions are full-fledged objects *sui generic*. They can be not only used to identify an object, but also mentioned as objects of predication. Thus we need and have to develop and specify the TIL *inference machine* in full. From the theoretical point of view, such a calculus specification has been presented in [3].

Concerning applications, the computational variant of TIL, viz. the functional programming language *TIL-Script*, is currently being developed. *TIL-Script* is a FIPA compliant language. The development of TIL-Script as well as ontology languages is still a work in progress. Currently we combine traditional tools and languages like OWL (Ontology Web Language), logic programming inference tools (Prolog) and FOL proof calculi (Gentzen system and natural deduction) with the full-fledged features of TIL-Script by building transcription bridges. Thus the implementation of TIL-Script inference machine proceeds in stages. In the first stage we implemented the subset of language corresponding to the expressive power of Horn clauses. Then we extend it to the full FOL inference machine. The next stage is to implement the inference machine for the subset of classical λ-calculi, and finally, the hyper-intensional features and partiality are to be taken into account.

3 Conclusions

We have described the aims and objectives of the on-coming collaborative research project of VŠB-Technical University of Ostrava, Masaryk University in Brno and the Institute of Philosophy of the Academy of Sciences of the Czech Republic, Prague. The project will concentrate on further investigations and development of both the theory and applications of the Transparent Intensional Logic (TIL) with the focus on temporal aspects of knowledge and language. We believe that this project will be another successful step on the way to full natural language semantics by means of automatic computer-based logical systems.

Acknowledgements This work has been partly supported by the Ministry of Education of CR within the Center of basic research LC536 and in the National Research Programme II project 2C06009, by the Czech Science Foundation under the project 401/07/0451 (Semantization of pragmatics), and by the internal grant agency of FEECS VSB-TU Ostrava – IGA 22/2009 (Modeling, simulation and verification of software processes).

References

1. Berners-Lee, T., Hendler, J. and Lassila, O. (2001): The semantic web. *Scientific American*, May 2001.
2. Duží, M. (2009): Topic-Focus Articulation from the Semantic Point of View. In *CICLing 2009*. Ed. Gelbukh Alexander, Berlin Heidelberg: Springer-Verlag LNCS, vol. 5449, pp. 220–232, 2009.
3. Duží, M., Jespersen, B., Materna, P.: *Procedural Semantics for Hyperintensional Logic – Foundations and Applications of Transparent Intensional Logic*. Series Logic, Epistemology and the Unity of Science. Springer, Berlin, forthcoming.
4. Horák, A. (2002): *The Normal Translation Algorithm in Transparent Intensional Logic for Czech*. Ph.D. thesis, Faculty of Informatics, Masaryk University, Brno, 2002.
5. Horák, A., Kadlec, V. (2005): New Meta-grammar Constructs in Czech Language Parser synt. *Lecture Notes in Artificial Intelligence*, Berlin: Springer-Verlag, 3658/2005, 1, pp. 85–92, 2005.
6. Horrocks, I. and Patel-Schneider, P.F. (2003): Three Theses of Representation in the Semantic Web. WWW2003, May 20–24, Budapest, Hungary, `http://www2003.org/cdrom/papers/refereed/p050/p50-horrocks.html`.
7. Jespersen, B. (2005): 'Explicit Intensionalisation, Anti-Acatualism, and How Smith's Murderer might Not Have Murdered Smith'. *Dialectica* Vol. 59, No. 3, 285–314.
8. Materna, P. (1998): *Concepts and Objects*, Acta Philosophica Fennica, vol 63. The Philosophical Society of Finland, Helsinki.
9. Materna, P.(2004): *Conceptual Systems.* Logos Verlag, Berlin.
10. Montague, R. (1974): 'The proper treatment of quantification in ordinary English', in: *Formal Philosophy: Selected papers of Richard Montague,* Thomason, R. (ed.), New Haven: Yale University Press, pp. 247–170.
11. Schönfinkel, M. (1924): 'Über die Bausteine der mathematischen Logik', *Mathematische Annalen*, vol. 92, pp. 305–316.
12. Tichý, P. (1980): 'The logic of temporal discourse', *Linguistics and Philosophy*, vol. 3, pp. 343–369, reprinted in Tichý (2004), pp. 373–69.
13. Tichý, P. (1982): 'Foundations of partial type theory', *Reports on Mathematical Logic*, vol. 14, pp. 52–72, reprinted in: Tichý (2004), pp. 467–80.
14. Tichý, P. (1988): *The Foundations of Frege's Logic*, Berlin, New York: De Gruyter.
15. Tichý, P. (2004): *Collected Papers in Logic and Philosophy*, V. Svoboda, B. Jespersen, C. Cheyne (eds.), Prague: Filosofia, Czech Academy of Sciences, and Dunedin: University of Otago Press.