Interpretability Logic

Mladen Vuković, Department of Mathematics University of Zagreb, Croatia

This is an overview a study of interpretability logic in Zagreb for the last twenty years: a brief history and some planes for further research. The idea of treating a provability predicate as a modal operator goes back to Gödel. The same idea was taken up later by Kripke and Montague, but only in the mid–seventies was the correct choice of axioms, based on Löb's theorem, seriously considered by several logicians independently: G. Boolos, D. de Jongh, R. Magari, G. Sambin and R. Solovay. The system **GL** (Gödel, Löb) is a modal propositional logic. R. Solovay 1976. proved arithmetical completeness of modal system **GL**. Many theories have the same provability logic - **GL**. It means that the provability logic **GL** cannot distinguish some properties, as e.g. finite axiomatizability, reflexivity, etc. Some logicians considered modal representations of other arithmetical properties, for example interpretability, Π_n-conservativity, interpolability ... Roughly, a theory S interprets a theory T if there is a natural way of translating the language of S into the language of T in such a way that the translations of all the axioms of T become provable in S. We write $S \geq T$ if this is the case. A derived notion is that of relative interpretability over a base theory T. Let A and B be arithmetical sentences. We say that A interprets B over T if $T + A \geq T + B$.

Modal logics for relative interpretability were first studied by P. Hájek (1981) and V. Švejdar (1983). A. Visser (1990) introduced the binary modal logic **IL** (interpretability logic). The interpretability logic **IL** results from the provability logic **GL**, by adding the binary modal operator $\rhd$. The language of the interpretability logic contains propositional letters $p_0,\ p_1, \ldots$, the logical connectives $\wedge,\ \vee, \rightarrow$ and $\neg$, and the unary modal operator $\Box$ and the binary modal operator $\rhd$. The axioms of the interpretability logic **IL** are: all tautologies of the propositional calculus, $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$, $\Box A \rightarrow \Box\Box A$, $\Box(\Box A \rightarrow A) \rightarrow \Box A$, $\Box(A \rightarrow B) \rightarrow (A \rhd B)$, $(A \rhd B\ \wedge B \rhd C) \rightarrow (A \rhd C)$, $((A \rhd C) \wedge (B \rhd C)) \rightarrow ((A \vee B) \rhd C)$, $(A \rhd B) \rightarrow (\Diamond A \rightarrow \Diamond B)$, and $\Diamond A \rhd A$, where $\Diamond$ stands for $\neg\Box\neg$ and $\rhd$ has the same priority as $\rightarrow$. The deduction rules of **IL** are modus ponens and necessitation. Arithmetical semantics of interpretability logic is based on the fact that each sufficiently strong theory S has arithmetical formulas $Pr(x)$ and $Int(x, y)$. Formula $Pr(x)$ expressing that "x is provable in S" (i.e. formula with Gödel number x is provable in S). Formula $Int(x, y)$ expressing that "$S + x$ interprets $S + y$." An arithmetical interpretation is a function $*$ from modal formulas into arithmetical sentences preserving Boolean connectives and satisfying $(\Box A)^* = Pr(\ulcorner A^* \urcorner)$ and $(A \rhd B)^* = Int(\ulcorner A^* \urcorner, \ulcorner B^* \urcorner)$ ($\ulcorner A^* \urcorner$ denote Gödel number of formula A^*). The system **IL** is natural from the modal point of view, but arithmetically incomplete. Various extensions of **IL**are obtained by adding some new axioms. These new axioms are called the principles of interpretability. We denote by **ILX** the system obtained by adding a principle X to the system **IL**. System **ILM** is the interpretability logic of

Peano arithmetic. The arithmetical completeness of system **ILM** is proved in [1]. Visser (in [8]) proved the arithmetical completeness of the system **ILP**.

There are several kinds of semantics for the interpretability logic. The basic semantics is Veltman models. D. de Jongh and F. Veltman proved the completeness of **IL** w.r.t. Veltman models (see [5]). We think that there are two main reasons for other semantics. First, the proofs of arithmetical completeness of interpretability logic are very complex. Second, the characteristic classes Veltman frames of some principles of interpretability are equal. Generalized Veltman models were defined by de Jongh. We use generalized Veltman models in [12] to prove independence between principles of interpretability. A question is which kind of connection exists between generalized Veltman models and general Kripke models.

If we want to study a correspondence between Kripke models K and K' we consider an isomorphism or an elementarily equivalence. If we want to study "weaker" correspondence we can consider a bisimulation. Van Benthem defined bisimulations of Kripke models. Visser in [8] defined a notion of bismulation between two Veltman models. We defined a notion of bisimulation between two generalized Veltman models in [13], and proved Hennessy–Milner theorem for generalized Veltman semantics. We study various kinds of bisimulations of Veltman models in [11]. In [10] bisimulation quotients of generalized Veltman models are considered. We proved in [14] that there is a bisimulation between Veltman model and generalized Veltman model. The existence of a bisimulation in general setting is an open problem.

P. Hájek and V. Švejdar in 1990. determined normal forms for the system **ILF**. The existence of the normal forms for system **IL** is an open problem. In [3] are determined normal forms in **IL** for some special classes of formulas.

The correspondence theory is the systematic study of the relationship between modal and classical logic. Bisimulations and the standard translation are two of the tools we need to understand modal expressivity. Van Benthem's characterization theorem (cf. [7]) shows that modal languages are the bisimulation invariant fragment of first–order languages, and it is established by classical methods of first-order model theory. The preservation theorems (cf. [6]) characterise a correspondence between semantic conditions of a class of models and logical formulas, too. However, the preservation property is usually much less significant than the corresponding expressive completeness property that any formula satisfying the semantic invariance condition is equivalent to one of the restricted syntactic form. D. Janin and I. Walukiewicz prove that a formula of monadic second–order logic is invariant under bisimulations if, and only if, it is logically equivalent to a formula of the μ–calculus. E. Rosen prove that the characterization theorem holds even in restriction to finite structures. A. Dawar and M. Otto in [4] investigate ramifications of van Benthem's characterization theorem for specific classes of Kripke structures. They study in particular Kripke modal classes defined through conditions on the underlying frames. Classical model theoretic arguments as saturated models and ultrafilter extensions do not apply to many of the most interesting classes. In the proofs the game–based analysis is used. V. Čačić and D. Vrgoč defined in [2] a bisimulation game between Veltman models and they proved the basic properties. We are interested in corresponding characterizations of modal fragments of first–order formula over Veltman models. The main problem when we prove van Benthem's theorem for interpretability logic is the existence of saturated Veltman model. We considered ultraproduct of Veltman models in [15].

References

[1] A. Berarducci, *The Interpretability Logic of Peano Arithmetic*, Journal of Symbolic Logic, 55(1990), 1059-1089

[2] V. Čačić, D. Vrgoč, *A Note on Bisimulation and Modal Equivalence in Provability Logic and Interpretability Logic*, Studia Logica 101(2013), 31–44

[3] V. Čačić, M. Vuković, *A note on normal forms for closed fragment of system IL*, Mathematical Communications, 17(2012), 195–204

[4] A. Dawar, M. Otto, *Modal characterization theorems over special classes of frames*, Annals of Pure and Applied Logic 161(2009), 1–42

[5] D. de Jongh, F. Veltman, *Provability Logics for Relative Interpretability*, In: *Mathematical Logic*, (P. P. Petkov, Ed.), Proceedings of the 1988 Heyting Conference, Plenum Press, New York, 1990, 31–42

[6] T. Perkov, M. Vuković, *Some characterization and preservation theorems in modal logic*, Annals of Pure and Applied Logic 163(2012), 1928-1939

[7] J. van Benthem, *Modal Logic and Classical Logic*, Bibliopolis, Napoli, 1983.

[8] A. Visser, Interpretability logic, In: P. P. Petkov (ed.), Mathematical Logic, Proceedings of the 1988 Heyting Conference, Plenum Press, New York, 1990, 175–210

[9] A. Visser, *An overview of interpretability logic*, In: K. Marcus (ed.) et al., Advances in modal logic. Vol. 1. Selected papers from the 1st international workshop (AiML'96), Berlin, Germany, October 1996, Stanford, CA: CSLI Publications, CSLI Lect. Notes. 87(1998), 307–359

[10] D. Vrgoč, M. Vuković, *Bisimulations and bisimulation quotients of generalized Veltman models*, Logic Journal of the IGPL, 18(2010), 870–880

[11] D. Vrgoč, M. Vuković, *Bismulation quotients of Veltman models*, Reports on Mathematical Logic, 46(2011), 59–73

[12] M. Vuković, *The principles of interpretability*, Notre Dame Journal of Formal Logic, 40(1999), 227–235

[13] M. Vuković, *Hennessy–Milner theorem for interpretability logic*, Bulletin of the Section of Logic, 34(2005), 195–201

[14] M. Vuković, *Bisimulations between generalized Veltman modles and Veltman models*, Mathematical Logic Quarterly, 54(2008), 368–373

[15] M. Vuković, *A note on ultraproducts of Veltman models*, Glasnik matematički, 46(2011), 7–10