

# Characterization and Definability in Modal First-Order Fragments

Facundo Carreiro\*

Departamento de Computación, FCEyN  
Universidad de Buenos Aires

Characterization results for modal logics below first order identifies them as a fragment of first order logic. These results are very important to analyze the expressive power of a given logic. The first work in this direction was done by van Benthem [12] who used bisimulations to characterize the basic modal logic as the bisimulation invariant fragment of first order logic.

There exists a huge number of modal logics, to enumerate some of them: temporal logic [7], propositional dynamic logic [8], description logic [3], sub-boolean modal logics [11], etc. Each of these logics was specially designed to fit a particular purpose, that is one of the jobs modal logics do best.

There is no single definition of bisimulation, each of the aforementioned logics has its own definition to match its particular semantics. It is interesting to analyze to which fragment of first order logic each of those logics correspond. Analogue results to van Benthem’s characterization hold for many other logics besides basic modal logic but, as the basic elements have changed (the bisimulation definition, the logic, etc.), there comes the need to re-prove the result. It is well known that the proofs of those results for several modal logics have, somehow, the same ‘taste’.

Definability results identify the properties that a class of models should satisfy in order to have a formula or a set of formulas whose models are exactly the class we are trying to define.<sup>1</sup> This question had previously been stated and answered for classical first order logic [6], basic modal logic [4] and many others [11, 10]. As with characterization results, each logic has its own different proof although they share most of the key ideas.

In this article we set up a framework which will aid us to prove generalized results for modal logics which lay (in terms of expressivity) below first order logic. We will focus on the following two axes.

**Arbitrary modal logic.** We want to obtain characterization and definability results which hold for an arbitrary modal logic. When we say ‘arbitrary’ we mean any modal logic with conjunction and disjunction (interpreted as usual) which is interpreted over extensions of Kripke models. Observe that this includes sub-boolean logics. We also want to be able to derive results no matter what the simulation or bisimulation for these logics looks like. It will later become clear that this last generalization comes with a great price to pay: we know nothing about the structural properties involved in this notion.

**Relativization classes of models.** When talking about definability we can think of *relative definability* as follows: Is the class  $K \subseteq C$  definable with by formula (or set) *given* that we only consider models within the class  $C$ ? In practice, depending on the domain of application, it is common to work with restricted classes of models such as finite models, tree models, acyclic models, etc. We want to know whether these restrictions give us extra information and turn classes that were previously undefinable into definable classes. A relativized version of the definability theorem should aid us in this quest.

---

\*Partially supported by CONICET (Argentina).

<sup>1</sup>Here we study definability of classes of models; definability of classes of frames will not be analyzed in this work.

# 1 Basics

We note  $\mathcal{L}$  and  $\mathcal{F}$  as the source and target languages respectively. The source language is an extension of the language  $\mathfrak{P} = \langle (p_i)_{i \in \mathbb{N}}, \wedge, \vee, \top, \perp \rangle$ . The target language is a (countable) first-order language which may contain equality.  $\text{FORM}(\mathfrak{A})$  is the set of formulas of the language  $\mathfrak{A}$  and  $\text{FORM}(\mathfrak{F}^1)$  is the subset with at most one free variable ‘ $x$ ’.

**Definition 1.1** (Models). We define  $\text{MODS}(\mathcal{L})$  to be the class of all models of the source logic and  $\text{MODS}(\mathcal{F})$  to be the class of all models of the target first-order logic. We define the class of pointed models for the source logic as  $\text{PMODS}(\mathcal{L}) = \{ \langle \mathcal{M}, w \rangle : \mathcal{M} \in \text{MODS}(\mathcal{L}) \text{ and } w \in |\mathcal{M}| \}$ . Similarly, in the target logic  $\mathcal{F}$ , we use  $\mathcal{M}^f, g \models \varphi$  to note that a formula  $\varphi$  is true in the model  $\mathcal{M}^f$  under the valuation (or assignment)  $g$ .<sup>2</sup> We define the class of pointed models for the target logic as  $\text{PMODS}(\mathcal{F}) = \{ \langle \mathcal{M}^f, g \rangle : \mathcal{M}^f \in \text{MODS}(\mathcal{F}) \text{ and } g : \{x\} \rightarrow |\mathcal{M}^f| \}$ .

During this article we will deal with source logics which are at most as expressive as first order logic. If  $\mathcal{L}$  is less or equally expressive than  $\mathcal{F}$  we should be able to express in  $\mathcal{F}$  everything that is expressible in  $\mathcal{L}$ . We can think of models as ‘information bearers’, they represent some information in a way that is compatible with some specific logic. We need to define some way to ‘look at’ this information from different perspectives, one compatible with  $\mathcal{L}$  and other compatible with  $\mathcal{F}$ .

**Definition 1.2** (Truth preserving translations). A function  $\text{Tf}_x : \text{FORM}(\mathcal{L}) \rightarrow \text{FORM}(\mathcal{F}^1)$  is a *formula translation* if it preserves conjunctions and disjunctions. Given a class of models  $K \subseteq \text{PMODS}(\mathcal{F})$ , a *model translation* is a bijective function  $T : \text{PMODS}(\mathcal{L}) \rightarrow K$ . A pair of translations  $(\text{Tf}_x, T)$  is said to be *truth-preserving* if for all  $\varphi \in \text{FORM}(\mathcal{L})$  and all  $\langle \mathcal{M}, w \rangle \in \text{PMODS}(\mathcal{L})$  we have that  $\mathcal{M}, w \models \varphi$  iff  $T(\mathcal{M}, w) \models \text{Tf}_x(\varphi)$ .

As an abuse of notation we use  $T(\mathcal{M})$  when we are not interested in the associated assignment. Let  $\langle \mathcal{M}^f, g \rangle, \langle \mathcal{N}^f, h \rangle \in \text{PMODS}(\mathcal{F})$ . We write  $\mathcal{M}^f, g \sqsubseteq_{\mathcal{F}} \mathcal{N}^f, h$  to mean that for every first order formula  $\alpha(x)$ ; If  $\mathcal{M}^f, g \models \alpha(x)$  then  $\mathcal{N}^f, h \models \alpha(x)$ . We write  $\mathcal{M}^f, g \equiv_{\mathcal{F}} \mathcal{N}^f, h$  when  $\mathcal{M}^f, g \sqsubseteq_{\mathcal{F}} \mathcal{N}^f, h$  and  $\mathcal{N}^f, h \sqsubseteq_{\mathcal{F}} \mathcal{M}^f, g$ . This notation extends analogously to  $\text{PMODS}(\mathcal{L})$ .

From now on we fix  $(\text{Tf}_x, T_K)$  as our pair of *truth-preserving* translations for the rest of the article. We will also want to translate formulas from  $\mathcal{L}$  to  $\mathcal{F}$  and then go back to  $\mathcal{L}$ -formulas. *A priori*, as we are not requiring  $\text{Tf}_x$  to be injective this could lead to a problem.

We make the following claim: For any  $\alpha, \beta$  such that  $\text{Tf}_x(\alpha) = \text{Tf}_x(\beta)$  we have  $\models_{\mathcal{L}} \alpha \leftrightarrow \beta$ . To simplify the proofs in this article, and without loss of generality, we will work with the set of formulas divided by the relation of  $\mathcal{L}$ -equivalence. All of our proofs also work with the original set of formulas but they would require excessive detours and justifications. In this setting we will be working up to formula equivalence and we will assume that our formula translation  $\text{Tf}_x$  is injective.

We will not assume any structural property for the notion of  $\mathcal{L}$ -simulation. We only require it to preserve the formulas which are true in a given model. This is something that any reasonable model equivalence notion should respect.

**Definition 1.3** ( $\mathcal{L}$ -simulation). Given two  $\mathcal{L}$  models  $\mathcal{M}$  and  $\mathcal{N}$  we define an  $\mathcal{L}$ -simulation to be a *non-empty* relation  $Z \subseteq \text{PMODS}(\mathcal{L}) \times \text{PMODS}(\mathcal{L})$  such that if  $\langle \mathcal{M}, w \rangle Z \langle \mathcal{N}, v \rangle$  then  $\mathcal{M}, w \sqsubseteq_{\mathcal{L}} \mathcal{N}, v$ .

We write  $\mathcal{M}, w \rightrightarrows_{\mathcal{L}} \mathcal{N}, v$  to indicate that there exists an  $\mathcal{L}$ -simulation between  $w$  and  $v$ . In general  $\mathcal{M}, w \sqsubseteq_{\mathcal{L}} \mathcal{N}, v$  does not imply  $\mathcal{M}, w \rightrightarrows_{\mathcal{L}} \mathcal{N}, v$  but we will be interested in some special classes of models which have this property. It is called the Hennessy-Milner Property [4].

<sup>2</sup>Observe that in this case  $g$  is a valuation and not a point of the domain.

## 2 Main Results

In the following definition we make explicit the requirements for the Characterization and Definability theorems in this section to hold for an arbitrary logic  $\mathfrak{L}$  and with respect to a class of models  $\mathbf{K}$ .

**Definition 2.1** (Adequate pair). A logic  $\mathfrak{L}$  and a class of models  $\mathbf{K} \subseteq \text{PMODS}(\mathfrak{F})$  is said to be an *adequate pair* if they fulfill the following requirements:

1.  $\mathbf{K}$  is closed under ultraproducts.<sup>3</sup>
2. There exist truth-preserving translations  $\text{Tf}_x, \text{T}_\mathbf{K}$ .
3. There exists an  $\mathfrak{L}$ -simulation notion.
4. The class of  $\omega$ -saturated  $\mathfrak{L}$ -models [6] has the Hennessy-Milner property w.r.t.  $\mathfrak{L}$ -simulations.<sup>4</sup>

In general, points one to three can be easily checked by just ‘looking at’ the logic  $\mathfrak{L}$ . The strongest requirement is the last one and will be all we would need to prove to automatically derive the results.

### 2.1 Characterization

We restate the classical notion of (bi)simulation invariance in terms of  $\mathfrak{L}$  and  $\mathbf{K}$ .

**Definition 2.2** ( $\mathfrak{L}$ -simulation  $\mathbf{K}$ -invariance). Let  $\langle \mathfrak{L}, \mathbf{K} \rangle$  be an adequate pair. A formula  $\alpha(x)$  of  $\mathfrak{F}^1$  is  *$\mathbf{K}$ -invariant for  $\mathfrak{L}$ -simulations* if for all  $\mathfrak{L}$ -models  $\mathcal{M}, \mathcal{N}$  and  $w \in |\mathcal{M}|, v \in |\mathcal{N}|$ : If  $\mathcal{M}, w \rightrightarrows_{\mathfrak{L}} \mathcal{N}, v$  and  $\text{T}(\mathcal{M}, w) \models \alpha(x)$  then  $\text{T}(\mathcal{N}, v) \models \alpha(x)$ .

We write  $\alpha \models_{\mathbf{K}} \beta$  to mean that the entailment is valid in the class  $\mathbf{K}$ . We say that  $\alpha$  and  $\beta$  are  $\mathbf{K}$ -equivalent when  $\models_{\mathbf{K}} \alpha \leftrightarrow \beta$ . Our main result regarding characterization is the following:

**Theorem 2.3** (Characterization). *Given an adequate pair  $\langle \mathfrak{L}, \mathbf{K} \rangle$  then A formula  $\alpha(x)$  of  $\mathfrak{F}^1$  is  $\mathbf{K}$ -equivalent to the translation of an  $\mathfrak{L}$ -formula iff  $\alpha(x)$  is  $\mathbf{K}$ -invariant for  $\mathfrak{L}$ -simulations.*

*Sketch of the proof.* Left-to-right direction is a direct consequence of the invariance of  $\mathfrak{L}$ -formulas over  $\mathfrak{L}$ -simulations (inherent to the definition of  $\mathfrak{L}$ -simulation). To prove the converse consider the set

$$\text{SLC}(\alpha) = \{\text{Tf}_x(\varphi) : \varphi \text{ is an } \mathfrak{L}\text{-formula and } \alpha(x) \models_{\mathbf{K}} \text{Tf}_x(\varphi)\}.$$

One can show, using a compactness argument, that if  $\text{SLC}(\alpha) \models_{\mathbf{K}} \alpha(x)$  then  $\alpha(x)$  is  $\mathbf{K}$ -equivalent to the translation of an  $\mathfrak{L}$ -formula. Hence, it all boils down to proving that result. Suppose that  $\text{T}(\mathcal{M}, w) \models \text{SLC}(\alpha)$ . We have to show that  $\text{T}(\mathcal{M}, w) \models \alpha(x)$ . For this proof we will use the following notation: we add a subscript  $f$  to the first-order translations of  $\mathfrak{L}$  models, a superscript  $+$  to first-order saturated models and a superscript  $*$  to modal saturated models. The proof goes as follows:

Define the set  $\Sigma(x) = \{\neg \text{Tf}_x(\varphi) : \varphi \text{ is an } \mathfrak{L}\text{-formula and } \mathcal{M}, w \not\models \varphi\} \cup \{\alpha(x)\}$ . Using a compactness argument we show that there is a model  $\mathcal{N}_f, g_v$  for  $\Sigma(x)$ . Its modal counterpart satisfies  $\mathcal{N}, v \sqsubseteq_{\mathfrak{L}} \mathcal{M}, w$ . We would like to transfer the validity of  $\alpha(x)$  from  $\mathcal{N}_f, g_v$  to  $\mathcal{M}_f, g_w$ .

Using the ultrapower construction we obtain, for  $\mathcal{M}_f, g_w$  and  $\mathcal{N}_f, g_v$ ,  $\omega$ -saturated extensions  $\mathcal{M}_f^+, g_w^+$  and  $\mathcal{N}_f^+, g_v^+$  which are elementary equivalent to their originators [9]. Observe that this implies that  $\mathcal{M}^*, w^* \equiv_{\mathfrak{L}} \mathcal{N}^*, v^*$  and  $\mathcal{N}_f^+ \models \alpha(x)[v^+]$ .

As  $\mathcal{N}^*, v^* \sqsubseteq_{\mathfrak{L}} \mathcal{M}^*, w^*$  and they are  $\omega$ -saturated, using point number 4 in Definition 2.1 we have that  $\mathcal{N}^*, v^* \rightrightarrows_{\mathfrak{L}} \mathcal{M}^*, w^*$ . Finally, as  $\mathcal{N}_f^+ \models \alpha(x)[v^+]$  and  $\alpha(x)$  is invariant under  $\mathfrak{L}$ -simulations we can be sure that  $\mathcal{M}_f^+ \models \alpha(x)[w^+]$ . As  $\mathcal{M}_f^+$  has the same first order theory as its originator we conclude that  $\mathcal{M}_f \models \alpha(x)[w]$ .  $\square$

<sup>3</sup>A very informative survey on ultraproducts and ultrapowers can be found in [9].

<sup>4</sup>We say that an  $\mathfrak{L}$ -model  $\mathcal{M}$  is  $\omega$ -saturated if and only if  $\text{T}(\mathcal{M})$  is.

## 2.2 Definability

In this section we will use *pointed models* for a smoother proof. Recall the notion of definability:

**Definition 2.4** (Definability). A class  $M \subseteq \text{PMODS}(\mathfrak{L})$  is said to be *definable by a set of formulas* if there exists a set  $\Gamma$  of  $\mathfrak{L}$ -formulas such that  $\langle \mathcal{M}, w \rangle \in M$  if and only if  $\mathcal{M}, w \models \Gamma$ .

We have two main results concerning definability. The first one considers *arbitrary sets* of formulas:

**Theorem 2.5** (Definability by a set). *Given an adequate pair  $\langle \mathfrak{L}, K \rangle$  and a class of pointed models  $M \subseteq \text{PMODS}(\mathfrak{L})$ , the following are equivalent*

- a)  $M$  is definable by a set of  $\mathfrak{L}$ -formulas.
- b)  $M$  is closed under  $\mathfrak{L}$ -simulations,  $T(M)$  under ultraproducts and  $T(\overline{M})$  under ultrapowers.

*Sketch of the proof.* From a to b suppose that  $M$  is defined by the set  $\Gamma$  of  $\mathfrak{L}$ -formulas and there is a model  $\langle \mathcal{M}, w \rangle \in M$  such that  $\mathcal{M}, w \rightrightarrows \mathcal{N}, v$  for some pointed model  $\mathcal{N}, v$ . As  $\langle \mathcal{M}, w \rangle \in M$  it must occur that  $\mathcal{M}, w \models \Gamma$ . By simulation preservation we have  $\mathcal{N}, v \models \Gamma$  therefore  $\langle \mathcal{N}, v \rangle \in M$ . Therefore  $M$  is closed under  $\mathfrak{L}$ -simulations.

To see that  $T(M)$  is closed under ultraproducts we use [6, Theorem 4.1.9]: If  $\mathcal{M}_i^f \models \Gamma$  for all  $i$  then  $\prod_D \mathcal{M}_i^f \models \Gamma$  therefore  $\prod_D \mathcal{M}_i^f \in T(M)$ . To see that  $T(\overline{M})$  is closed under ultrapowers we use that an ultrapower is elementary equivalent to the original model [6, Corollary 4.1.10].

From b to a we proceed as follows: Define the set  $\Gamma = \text{Th}(M)$ , i.e. the  $\mathfrak{L}$ -formulas which are valid in the class  $M$ . Trivially  $M \models \Gamma$ , we still have to show that if  $\mathcal{M}, w \models \Gamma$  then  $\langle \mathcal{M}, w \rangle \in M$ . We define the following set

$$\text{NTh}^w(x) = \{\neg \text{Th}_x(\varphi) : \varphi \text{ is an } \mathfrak{L}\text{-formula and } \mathcal{M}, w \not\models \varphi\}$$

and using a compactness argument we can see that it is satisfiable in  $T(M)$ . Therefore there is a model  $\langle \mathcal{N}, v \rangle \in M$  such that  $T(\mathcal{N}, v) \models \text{NTh}^w(x)$ . Observe that these models satisfy  $\mathcal{N}, v \sqsubseteq_{\mathfrak{L}} \mathcal{M}, w$ .

Suppose that  $\langle \mathcal{M}, w \rangle \in \overline{M}$ , using an argument similar to that in Theorem 2.3 we can conclude that there exist  $\omega$ -saturated extensions  $\langle \mathcal{N}^*, v^* \rangle \in M$  and  $\langle \mathcal{M}^*, w^* \rangle \in \overline{M}$  such that  $\mathcal{N}^*, v^* \rightrightarrows_{\mathfrak{L}} \mathcal{M}^*, w^*$ . As  $M$  is closed under  $\mathfrak{L}$ -simulations then  $\langle \mathcal{M}, w \rangle \in M$ . Absurd, therefore  $\langle \mathcal{M}, w \rangle \in M$ .  $\square$

Our second result considers classes of models definable by a *single* formula:

**Theorem 2.6** (Definability by a single formula). *Given an adequate pair  $\langle \mathfrak{L}, K \rangle$ , and a class of models  $M \subseteq \text{MODS}(\mathfrak{L})$ , the following are equivalent*

- a)  $M$  is definable by a single  $\mathfrak{L}$ -formula.
- b)  $M$  is closed under  $\mathfrak{L}$ -simulations and both  $T(M)$  and  $T(\overline{M})$  are closed under ultraproducts.

*Ideas involved in the proof.* The proof of this theorem uses theorems 2.5 and 2.3 in combination with a relativized version of the definability theorem for first-order logic [5, Theorem A.2]. The full proof can be found in [5].  $\square$

## 3 Conclusions and further work

The advantages of the general framework presented in this work is twofold. On the one hand, we can give new and unifying proofs of Characterization and Definability for logics where these properties are well-known to be true (e.g. see [5] for hybrid logics). On the other we can address the properties of Characterization and Definability for logics that have not yet been investigated so far. In both cases, it is only needed to check that a logic meets the requirements (Definition 2.1) to automatically derive the desired results. In this section we mention an application to memory logics and discuss a research line which looks very promising to give a general proof for a large family of modal logics.

**Memory Logics.** Memory logics are a novel family of modal logics introduced in [2]. They allow to model dynamic behavior through explicit memory operators that *change* the evaluating Kripke structure. It is important to note that although the authors give a bisimulation definition for this family, they do not prove any Characterization or Definability results. In [5] we prove that the class of  $\omega$ -saturated memory models have the Hennessy-Milner property for all the logics in the family and conclude the desired Characterization and Definability results.

**Coinductive models.** The ‘Basic Temporal Logic’ is a modal logic with two modalities  $F$  and  $P$ . The classical perspective on this logic interprets it over Kripke models  $\langle W, R, V \rangle$  such that:

$$\mathcal{M}, w \models F\varphi \text{ iff } \exists v. wRv \text{ and } \mathcal{M}, v \models \varphi \qquad \mathcal{M}, w \models P\varphi \text{ iff } \exists v. vRw \text{ and } \mathcal{M}, v \models \varphi$$

Observe that  $F$  can be thought as a normal ‘diamond’ over the relation  $R$  but that is not possible with  $P$ . An alternative interpretation is as follows. Interpret the logic over Kripke models which are tuples  $\langle W, R_1, R_2, V \rangle$  where  $R_1 = R_2^{-1}$ . With this restriction both modalities  $F$  and  $P$  turn to be simple diamonds over  $R_1$  and  $R_2$  respectively. The key point is that, in this setting, the right model equivalence notion for this logic is *exactly* the same bisimulation notion as in basic modal logic.

In [1] this idea is generalized for (almost) any modality which can be defined with the pattern of the diamond operator ( $\forall\exists$ ). From our perspective, the most important point of their work is that, by restricting the class of models, we get a *unique* notion of model equivalence for every logic that fits in their framework.

To the moment, there is no direct way to prove Characterization and Definability results using the framework developed in [1]. The problem laid in the restriction applied to the class of models: there is no classical proof which takes this kind of restrictions into account. With the results developed in this thesis it should be possible to prove a more general result using their framework.

## References

- [1] Coinductive models and normal forms for modal logics. *Journal of Applied Logic*, 2010. To appear.
- [2] C. Areces, D. Figueira, S. Figueira, and S. Mera. Expressive power and decidability for memory logics. In *Logic, Language, Information and Computation*, volume 5110 of *Lecture Notes in Computer Science*, pages 56–68, Edinburgh, Scotland, 2008. Springer Berlin / Heidelberg. Proceedings of WoLLIC 2008.
- [3] F. Baader, D. Calvanese, and D. M. D. N. P. Patel-Schneider. *The description logic handbook*. Cambridge University Press New York, NY, USA, 2007.
- [4] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.
- [5] F. Carreiro. Characterization and definability in modal first-order fragments. Master’s thesis, Department of Computer Science, FCEyN, University of Buenos Aires, Mar. 2010. <http://www.glyc.dc.uba.ar/facu/papers/tesislicFC.pdf>.
- [6] C. C. Chang and H. J. Keisler. *Model Theory, Volume 73 (Studies in Logic and the Foundations of Mathematics)*. Elsevier Science B.V., Amsterdam, The Netherlands, 1973.
- [7] E. Emerson. Temporal and modal logic. *Handbook of theoretical computer science*, 1990.
- [8] M. Fischer and R. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, 1979.
- [9] H. J. Keisler. The ultraproduct construction. In *Proceedings of the Ultramath Conference*, Pisa, Italy, 2008.
- [10] N. Kurtonina and M. de Rijke. Bisimulations for temporal logic. *Journal of Logic, Language and Information*, 6:403–425, 1997.
- [11] N. Kurtonina and M. de Rijke. Simulating without negation. *Journal of Logic and Computation*, 7:503–524, 1997.
- [12] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, Universiteit van Amsterdam, Instituut voor Logica en Grondslagenonderzoek van Exacte Wetenschappen, 1976.