Reasoning about comparative similarity in ontologies: a first step

Régis Alenda[1], Nicola Olivetti[1], and Camilla Schwind[2]

[1] LSIS - UMR CNRS 6168
Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen ,
13397 MARSEILLE CEDEX 20
regis.alenda@lsis.org et nicola.olivetti@univ-cezanne.fr

[2] LIF, CNRS
Faculté des Sciences de Luminy, Université de la Méditerranée
163 Avenue de Luminy, Case 901 13288 Marseille Cedex 9, France,
camilla.schwind@lif.univ-mrs.fr

Résumé : Nous étudions la logique de *similarité comparative des concepts* $\mathcal{CSL}$, introduite par Sheremet, Tishkovsky, Wolter et Zakharyaschev en 2005, et permettant de formuler des assertions qualitatives sur la similarité entre des concepts, de la forme "A est plus similaire à B qu'à C". Dans cet article, nous présentons une extension de ce langage avec les *nominals*. Nous donnons aussi une méthode de preuve à tableaux permettant de tester la satifiabilité d'une formule par rapport à une base de connaissance. Ce travail est une première étape pour une utilisation de cette logique dans la représentation et le raisonnement sur des ontologies.

Mots-clés : logiques modales, procédures à tableaux, logiques de description, similarité comparative des concepts

1 Introduction

The logics of comparative concept similarity $\mathcal{CSL}$ have been introduced in Sheremet *et al.* (2005) to capture a form of qualitative comparison between concept instances. In these logics we can express assertions or judgments of the form : "Renault Clio is more similar to Peugeot 207 than to WW Golf", "Marseilles is more similar to Barcelona than to Naples"' ; arguing about music we may wish to express (arguably) "‘Mozart (meaning his music) is more similar to Händel than to any composer of the romantic age"'. These logics may find an application in ontology languages, whose logical base is provided by Description Logics (DL), allowing concept definitions based on proximity/similarity measures. For instance (Sheremet *et al.* (2005)), the color "Reddish" may be defined as a color which is more similar to a prototypical "Red" than to any other color (in some color model as RGB). The aim is to dispose of a language in which logical classification provided by standard DL is integrated with classification mechanisms based on calculation of proximity measures. The latter is typical for instance of domains like bio-informatics or linguistics.

But as the last examples show, comparative similarity assertions are not necessarily the outcome of an objective similarity measure, but they may just express arbitrary agents' belief (or a combination thereof). In this respect the logic of concept similarity is not far from languages for handling agent's preferences.

The logic $\mathcal{CSL}$ contains concepts defined by boolean operators and a single binary modal connective$\leftleftarrows$ expressing comparative similarity [1]. In this paper we consider an extension $\mathcal{CSLO}$ of $\mathcal{CSL}$ which contains also nominals, i.e. individual names, the latter can be used for individual assertions as well as in concept definitions. In this language the above examples can be encoded as follows :

(1) $\mathit{reddish} \equiv \{\mathit{red}\} \leftleftarrows \{\mathit{green}, \ldots, \mathit{black}\}$,
(2) $\mathit{Clio} \sqsubseteq (\mathit{Peugeot207} \leftleftarrows \mathit{Golf})$,
(3) $(\{\mathit{barcelona}\} \leftleftarrows \{\mathit{naples}\})(\mathit{marseilles})$,
(4) $(\{\mathit{haendel}\} \leftleftarrows \mathit{Romantic})(\mathit{mozart})$.

Developing the last example further, a (simplifed) knowledge base about classical music may contain some other "knowledge"' or beliefs about composers, for example : "Mozart is more similar to romantic composers than to contemporary composers", "Haendel is a Baroque music composer", "Baroque composers are closer to romantic composers than to contemporary composers" ; moreover the "modern composers" are defined as romantic or contemporary composers :

(5) $(\mathit{Romantic} \leftleftarrows \mathit{Contemporary})(\mathit{mozart})$,
(6) $\mathit{Baroque}(\mathit{haendel})$,
(7) $\mathit{Baroque} \sqsubseteq (\mathit{Romantic} \leftleftarrows \mathit{Contemporary})$,
(8) $\mathit{Modern} \equiv \mathit{Romantic} \sqcup \mathit{Contemporary}$,
(9) $\{\mathit{mozart}, \mathit{beethoven}, \mathit{scarlatti}, \mathit{haydn}\} \sqsubseteq \mathit{Classic}$.

No matter how plausible are the above claims from a musicological point of view, from (4)–(9) a system supporting the logic CSL should be able to derive the following information, whose interpretation is left to the reader :

$(\mathit{Baroque} \leftleftarrows \mathit{Romantic})(\mathit{mozart})$
$(\mathit{Baroque} \leftleftarrows \mathit{Modern})(\mathit{mozart})$

and obviously $(\mathit{Romantic} \leftleftarrows \mathit{Contemporary})(\mathit{haendel})$.

The semantics of $\mathcal{CSL}$ (and of $\mathcal{CSLO}$) is defined in terms of distance spaces, that is to say structures equipped by a distance function d, whose properties may vary according to the logic under consideration. In this setting, the evaluation of $A \leftleftarrows B$ can be informally stated as follows : x satisfies $A \leftleftarrows B$ iff $d(x, A) < d(x, B)$ meaning that the object x is an instance of the concept $A \leftleftarrows B$ (i.e. it belongs to things that are more similar to A than to B) if x is strictly closer to A-objects than to B-objects according to distance function d, where the distance of an object to a set of objects is defined as the *infimum* of the distances to each object in the set. In Sheremet *et al.* (2005, 2008); Kurucz *et al.* (2005); Sheremet *et al.* (2007), the authors have investigated the logic

[1] In a more general setting, the language might contain several $\leftleftarrows_{\mathrm{Feature}}$ where each $\mathrm{Feature}$ corresponds to a specific distance function d_{Feature} measuring the similarity of objects with respect to one $\mathrm{Feature}$ (size, price, power, taste, color...).

$\mathcal{CSL}$ with respect to different classes of distance models, see Sheremet *et al.* (2008) for a survey of results about decidability, complexity, expressivity, and axiomatisation. Remarkably it is shown that $\mathcal{CSL}$ is undecidable over subspaces of the reals.

In the present work, following Alenda *et al.* (2009a), we restrict our attention to so-called *minspaces*, that is spaces where the infimum of a set of distances is actually their *minimum*. The minspace property entails the restriction to spaces where the distance function is discrete. This requirement does not seem incompatible with the purpose of representing qualitative similarity comparisons, whereas it might be less reasonable for applications of $\mathcal{CSL}$ to spatial reasoning.

Since the logic in itself can only express a qualitative comparisons and not numerical values, it is no hard to see that the semantics can be reformulated in terms of preferential structures, that is to say possible-world structures equipped by a family of strict partial (pre)-orders $y \prec_x z$ indexed on objects x (see Lewis (1973); Stalnaker (1968)), whose intended meaning is that x is more similar to y than to z. It turns out that the semantics over minspaces is equivalent to the semantics over preferential structures satisfying the well-known principle of Limit Assumption according to which the set of minimal elements of a non-empty set always exists.

In Alenda *et al.* (2009a,b) we have provided an axiomatization of $\mathcal{CSL}$ for a propositional language augmented by the concept similarity operator. Moreover we have proposed the first practical decision procedure for this logic in the form of tableaux calculus. In this paper we extend and adapt our tableaux procedure to $\mathcal{CSLO}$, this extension allows us to handle knowledge bases comprising as usual a TBOX containing concepts definitions (possibly involving nominals) and an ABOX containing individuals assertions. We present this extension as a first step to integrate concept similarity with significant description logics dialects at the base of ontologies languages.

2 The Logic $\mathcal{CSLO}$

We first introduce the syntax and the semantics of logic $\mathcal{CSLO}$.

Definition 1
The alphabet of $\mathcal{CSLO}$ contains a finite set of concept names $A_i \in \mathcal{C}_{nom}$ and a finite set of individual names $a_i \in \mathcal{A}_{ind}$. The language $\mathcal{L}_{\mathcal{CSLO}}$ contains all concepts formed out from concept names and singletons by applying boolean connectives and the $\leftleftarrows$ operator :

$$C \ ::= \ \bot \mid \top \mid A_i \mid \{a_i\} \mid \neg C \mid C \sqcap C \mid C \leftleftarrows C$$

where $A_i \in \mathcal{C}_{nom}$ and $a_i \in \mathcal{A}_{ind}$. Dealing with enumerative concepts, we write $\{a_1, \ldots, a_n\}$ in place of $\{a_1\} \sqcup \ldots \sqcup \{a_1\}$. An inclusion axiom has the form $C \sqsubseteq D$, where C, D are concepts. A TBox $\mathcal{T}$ is a finite set of inclusion axioms. An individual assertion has the form $C(a)$, where C is a concept and a is an individual name. An ABox $\mathcal{A}$ is a finite set of individual assertions. A knowledge base KB is a pair $(\mathcal{T}, \mathcal{A})$. By an assertion F we mean either an inclusion $C \sqsubseteq D$ or an individual assertion $C(a)$.

As for $\mathcal{CSL}$ in Sheremet *et al.* (2005) , the semantics of $\mathcal{CSLO}$ can be defined in terms of models based on distance spaces.

$\mathcal{CSLO}$ is a logic of pure qualitative comparisons, this motivates an alternative semantics where the distance function is replaced by a family of comparisons relations, one for each object. We call this semantics *preferential* semantics, the same as the semantics of conditional logics (Nute (1980); Lewis (1973)). Preferential structures are equipped by a family of strict pre-orders. We may interpret this relations as expressing a comparative similarity between objects. For three objects, $x \prec_w y$ states that w is more similar to x than to y.

The preferential semantics in itself is more general than distance model semantics. However, if we assume the additional conditions here below, it turns out that these two are equivalent. The conditions are :

Definition 2
We will say that a preferential relation $\prec_w$ over Δ :

(i) *is* modular *iff* $\forall x, y, z \in \Delta, (x \prec_w y) \rightarrow (z \prec_w y \vee x \prec_w z)$.

(ii) *is* centered *iff* $\forall x \in \Delta, x = w \vee w \prec_w x$.

(iii) *satisfies the* Limit Assumption *iff* $\forall X \subseteq \Delta, X \neq \emptyset \rightarrow \min_{\prec_w}(X) \neq \emptyset$,[2] *where* $\min_{\prec_w}(X) = \{y \in X \mid \forall z(z \prec_w y \rightarrow z \notin X)\}$.

Modularity is strongly related to the fact that the preferential relations represent distance comparisons. Centering states that w is the *unique* minimal element for its preferential relation $\prec_w$. The Limit Assumption states that each non-empty set has at least one minimal element with respect to any preferential relation (i.e it does not contain an infinitely descending chain), and corresponds to the minspace property.

Definition 3 ($\mathcal{CSLO}$-preferential model)
A $\mathcal{CSL}$-preferential model is a triple $\mathcal{M} = (\Delta, (\prec_w)_{w\in\Delta}, .^{\mathcal{M}})$ where :
- *Δ is a non-empty set of* objects *(or* possible worlds*).*
- *$(\prec_w)_{w\in\Delta}$ is a family of* preferential relation, *each one being modular, centered, and satisfying the* limit assumption.
- *$.^{\mathcal{M}}$ is the evaluation function defined as usual on the boolean operator, and as follow for the $\leftleftarrows$ operator :*

$$(A \leftleftarrows B)^{\mathcal{M}} = \left\{w \in \Delta \middle| \exists x \in A^{\mathcal{M}} \text{ such that } \forall y \in B^{\mathcal{M}},\ x \prec_w y\right\} .$$

All semantic notions regarding knowledge bases (satisfiability and entailment) are exactly the same as above.

Distance models and preferential models provide an equivalent semantics of $\mathcal{CSLO}$. More precisely, we say that a $\mathcal{CSLO}$-preferential model $\mathcal{I}$ and a $\mathcal{CSLO}$-distance minspace model $\mathcal{J}$ are *equivalent* iff they are based on the same set Δ, and for all formulas $A \in \mathcal{L}_{\mathcal{CSL}}$, $A^{\mathcal{I}} = A^{\mathcal{J}}$.

[2] We note that the Limit Assumption implies that the preferential relation is asymmetric. On the other hand, on a finite set, asymmetry implies Limit Assumption. Modularity and asymmetry imply that the relation is also transitive and irreflexive.

Theorem 1 (see Alenda *et al.* (2009b))

1. *For each $\mathcal{CSLO}$-distance minspace model, there is an equivalent $\mathcal{CSLO}$-preferential model.*
2. *For each $\mathcal{CSLO}$-preferential model, there is an equivalent $\mathcal{CSLO}$-distance minspace model.*

An axiomatization of $\mathcal{CSL}$, that is to say $\mathcal{CSLO}$ without nominals is provided in Alenda *et al.* (2009a).

We extend the semantics to knowledge bases in the usual way : (i) we say that a model $\mathcal{M}$ satisfies a concept C if $C^M \neq \emptyset$; moreover (ii) M satisfies an inclusion $C \sqsubseteq D$ if $C^{\mathcal{M}} \subseteq D^{\mathcal{M}}$ and that (iii) $\mathcal{M}$ satisfies an individual assertion $C(a)$ if $a^{\mathcal{M}} \in C^M$. We say that $\mathcal{M}$ satisfies a knowledge base KB (or that M is a model of KB) if $\mathcal{M}$ satisfies every assertion (either inclusion or individual assertion) contained in the KB, we denote this by $\mathcal{M} \models KB$. We say that a knowledge base KB is *satisfiable* if there exists a model $\mathcal{M} \models KB$. We finally define that KB *entails* an assertion F ($KB \models_{\mathcal{CSLO}} F$) if every model of KB satisfies F.

Entailment can be easily reduced to (un)satisfiability, as we have :

$KB \models_{\mathcal{CSLO}} C(a)$ iff $KB \cup \{\neg C(a)\}$ is unsatisfiable and
$KB \models_{\mathcal{CSLO}} C \sqsubseteq D$ iff $KB \cup \{C(a'), \neg D(a')\}$ is unsatisfiable,

where a' is an individual name not occurring in KB.

It is worth noticing that in $\mathcal{CSLO}$, checking (un)satisfiability of a knowledge base can be reduced to checking the (un)satisfiability of a *single* concept. To this aim, observe first that for any concept C and model $\mathcal{M}$ with domain Δ, $(\neg C \leftleftarrows \bot)^{\mathcal{M}} = \emptyset$ or $(\neg C \leftleftarrows \bot)^{\mathcal{M}} = \Delta$. Thus $\neg C \leftleftarrows \bot$ behaves as a modal operator $\Box C$ (in modal logic S5). By this fact we can translate a whole knowledge base KB into a single concept : just replace every inclusion $C \sqsubseteq D$ by the concept $(C \sqcap \neg D) \leftleftarrows \bot$ and every individual assertion $C(a_i)$ by $\{a_i\} \sqcap C$, finally let C_{KB} be the conjunction of the so-obtained concepts. We obtain :

Proposition 1
Let KB be knowledge base, then there is a $\mathcal{CSLO}$ concept C_{KB} in the language of KB such that for any model $\mathcal{M}$:

$\mathcal{M} \models KB$ *iff* $C^{\mathcal{M}}_{KB} \neq \emptyset$.

Moreover, the size of C_{KB} is linear in the size of KB.

3 The Tableaux Calculus for $\mathcal{CSLO}$

In this section, we present a tableau calculus for $\mathcal{CSL}$, this calculus provides a decision procedure for this logic. This calculus is an extension of the calculus given in Alenda *et al.* (2009a) with additional rules for nominals and the knowledge base KB. We identify a tableau with a set of sets of formulas $\Gamma_1, \ldots, \Gamma_n$. Each Γ_i is called a *tableau set*[3]. It will make use of labels to represent objects of the domain. Labels are either introduced during the proof or they correspond to nominals of the language $\mathcal{L}_{\mathcal{CSLO}}$.

[3] A tableau set corresponds to a *branch* in a tableau-as-tree representation.

Let us consider formulas $(A \leftleftarrows B)$ and $\neg(A \leftleftarrows B)$ under preferential semantics. We have :

$$w \in (A \leftleftarrows B)^{\mathcal{M}} \text{ iff } \exists x(x \in A^{\mathcal{M}} \wedge \forall z(z \in B^{\mathcal{M}} \rightarrow x \prec_w z))$$

In minspace models, the right part is equivalent to :

$$w \in (A \leftleftarrows B)^{\mathcal{M}} \text{ iff } \exists u \in A^{\mathcal{M}} \text{ and } \forall y(y \in B^{\mathcal{M}} \rightarrow \exists x(x \in A^{\mathcal{M}} \wedge x \prec_w y))$$

We now introduce a pseudo-modality $\Box_w$ indexed on objects :

$$x \in (\Box_w A)^{\mathcal{M}} \text{ iff } \forall y(y \prec_w x \rightarrow y \in A^{\mathcal{M}})$$

Its meaning is that $x \in (\Box_w A)^{\mathcal{M}}$ iff A holds in all worlds preferred to x with respect to $\prec_w$. Observe that we have then the equivalence :

Claim 1
$w \in (A \leftleftarrows B)^{\mathcal{M}}$ iff $A^{\mathcal{M}} \neq \emptyset$ and $\forall y(y \notin B^{\mathcal{M}}$ or $y \in (\neg\Box_w\neg A)^{\mathcal{M}})$.

This equivalence will be used to decompose $\leftleftarrows$-formulas in an analytic way.

In addition to the tableaux rules for $\mathcal{CSL}$ Alenda *et al.* (2009a), we have the following rules for nominals and KB :

$$(T\mathcal{O}) \quad \frac{\Gamma[x : \{a_1, \ldots a_n\}]}{\Gamma[x/a_1] \mid \Gamma[x/a_2] \ldots \mid \Gamma[x/a_n]} \qquad (F\mathcal{O}) \quad \frac{\Gamma, x : \neg\{a_1, \ldots a_n\}}{\Gamma, x : \neg\{a_1\},\ x : \neg\{a_2\}, \ldots x : \neg\{a_n\}}$$

$$(Unfold) \quad \frac{\Gamma}{\Gamma, y : \neg C \sqcup D}\ (C \sqsubseteq D \in KB) \qquad (Ind) \quad \frac{\Gamma}{\Gamma, b : A}\ (A(b) \in KB)$$

The tableau rules make also use of a universal modality $\Box$ (and its negation). The language of tableaux comprises the following kind of formulas : $x : A, x : (\neg)\Box\neg A, x : (\neg)\Box_y\neg A, x <_y z$, where x, y, z are labels and A is a $\mathcal{CSL}$-formula. The meaning of $x : A$ is the obvious one : $x \in A^{\mathcal{M}}$. The reading of the rules is the following : we apply a rule

$$\frac{\Gamma[E_1, \ldots, E_k]}{\Gamma_1 \mid \ldots \mid \Gamma_n}$$

to a tableau set Γ if each formula E_k is in Γ. We then replace Γ with any tableau set $\Gamma_1, \ldots, \Gamma_n$. As usual, we let Γ, A stand for for $\Gamma \cup \{A\}$, where A is a tableau formula. Figure 1 shows all the tableaux rules for $\mathcal{CSLO}$.

Let us comment on the rules which are not immediately obvious. The rule for $(T \leftleftarrows)$ encodes directly the semantics by virtue of claim 1. However in the negative case the rule is split in two : if x satisfies $\neg(A \leftleftarrows B)$, either A is empty, or there must be an $y \in B$ such that there is no $z \prec_x y$ satisfying A ; if x satisfies B then x itself fulfills this condition, i.e. we could take $y = x$, since x is $\prec_x$-minimal (by centering). On the other hand, if x does not satisfies B, then x cannot satisfy A either (otherwise x would satisfy $A \leftleftarrows B$) and there must be an y as described above. This case analysis with respect to x is performed by the (F1$\leftleftarrows$) rule, whereas the creation y for the latter case is performed

$$(T\sqcap)\quad \frac{\Gamma[x : A \sqcap B]}{\Gamma,\ x : A,\ x : B} \qquad (F\sqcap)\quad \frac{\Gamma[x : \neg(A \sqcap B)]}{\Gamma,\ x : \neg A \mid \Gamma,\ x : \neg B}$$

$$(NEG)\quad \frac{\Gamma[x : \neg\neg A]}{\Gamma, x : A} \qquad (F1 \leftleftarrows)\quad \frac{\Gamma[x : \neg(A \leftleftarrows B)]}{\Gamma, x : \Box\neg A \mid \Gamma, x : B \mid \Gamma, x : \neg A,\ x : \neg B}$$

$$(T \leftleftarrows)(*)\quad \frac{\Gamma[x : A \leftleftarrows B]}{\Gamma, x : \neg\Box\neg A,\ y : \neg B \mid \Gamma, y : B, y : \neg\Box_x\neg A} \qquad (F2 \leftleftarrows)(**)\quad \frac{\Gamma[x : \neg(A \leftleftarrows B),\ x : \neg A,\ x : \neg B]}{\Gamma,\ y : B,\ y : \Box_x\neg A}$$

$$(F1\Box_x)\quad \frac{\Gamma[z : \neg\Box_x\neg A]}{\Gamma, x : \neg A \mid \Gamma, x : A}$$

$$(T\Box_x)(*)\quad \frac{\Gamma[z : \Box_x\neg A, y <_x z]}{\Gamma,\ y : \neg A,\ y : \Box_x\neg A} \qquad (F2\Box_x)(**)\quad \frac{\Gamma[z : \neg\Box_x\neg A,\ x : \neg A]}{\Gamma,\ y <_x z,\ y : A,\ y : \Box_x\neg A}$$

$$(T\Box)(*)\quad \frac{\Gamma[x : \Box\neg A]}{\Gamma,\ y : \neg A, y : \Box\neg A} \qquad (F\Box)(**)\quad \frac{\Gamma[x : \neg\Box\neg A]}{\Gamma,\ y : A}$$

$$(T\mathcal{O})\quad \frac{\Gamma[x : \{a_1, \ldots a_n\}]}{\Gamma[x/a_1] \mid \Gamma[x/a_2] \ldots \mid \Gamma[x/a_n]} \qquad (F\mathcal{O})\quad \frac{\Gamma, x : \neg\{a_1, \ldots a_n\}}{\Gamma, x : \neg\{a_1\},\ x : \neg\{a_2\}, \ldots x : \neg\{a_n\}}$$

$$(Unfold)(*)\quad \frac{\Gamma}{\Gamma, y : \neg C \sqcup D}\ (C \sqsubseteq D \in KB) \qquad (Ind)(****)\quad \frac{\Gamma}{\Gamma, b : A}\ (A(b) \in KB)$$

$$(Mod)(*)\quad \frac{\Gamma[z <_x u]}{\Gamma, z <_x y \mid \Gamma, y <_x u} \qquad (Cent)(***)\quad \frac{\Gamma}{\Gamma, x <_x y \mid \Gamma[x/y]}$$

(*) y is a label occurring in Γ. (**) y is a new label not occurring in Γ. (***) x and y are two distinct labels occurring in Γ. (****) b is a nominal name.

FIG. 1 – Tableau rules for $\mathcal{CSL}$.

by ($F2 \leftleftarrows$). We have a similar situation for the (F$\Box_x$) rule : let z satisfy $\neg\Box_x\neg A$, then there must be an $y \prec_x z$ satisfying A ; but if x satisfies A we can take $x = y$, since $x \prec_x z$ (by centering). If x does not satisfy A then we must create a suitable y and this is the task of the (F2$\Box_x$) rule. Observe that the rule does not simply create a $y \prec_x z$ satisfying A but it creates a *minimal* one. The rule is similar to the (F$\Box$) rule in modal logic GL (Gödel-Löb modal logic of arithmetic provability) Boolos (1993) and it is enforced by the Limit Assumption. This formulation of the rules for (F$\leftleftarrows$) and for (F$\Box_x$) prevents the unnecessary creation of new objects whenever the existence of the objects required by the rules is assured by centering. The rule for $(T\mathcal{O})$ encodes the substitution of x by either of $\{a_1, \ldots a_n\}$ within Γ, whereas $(F\mathcal{O})$ states the converse : x cannot be substituted by either of $\{a_1, \ldots a_n\}$. The rule $(Unflod)$ takes care of the inclusion axioms in the KB while the rule (Ind) ensures that if an assertion $A(b)$ is in the KB, a formula $b : A$ is added to the tableau. The rule $(Cent)$ is of a special kind : it has no premises (ie. it can always be applied) and generates two tableau sets : one with $\Gamma \cup \{x <_x y\}$, where x and y are two distinct labels occurring in Γ), and one where we replace x by y in Γ, i.e. where we identify the two labels.

Definition 4 (Closed set, closed tableau)
A tableau set Γ is closed *if one of the four following conditions hold : (i) $x : A \in \Gamma$ and $x : \neg A \in \Gamma$, for any formula A, or $x : \bot \in \Gamma$. (ii) $y <_x z$ and $z <_x y$ are in Γ. (iii) $x : \neg\Box_x A \in \Gamma$. (iv) $a : \neg\{a\} \in \Gamma$. (v) Under the unit names assumption UNA : $a : \{b\} \in \Gamma$, for different a and b. A $\mathcal{CSLO}$-tableau is closed if every tableau set is closed.*

In order to prove soundness and completeness of the tableaux rules, we introduce the notion of satisfiability of a tableau set by a model.

Given a tableau set Γ, we denote by Lab_Γ the set of labels occurring in Γ and by $\mathcal{O}_\Gamma$ the set of nominals occurring in Γ.

Definition 5 ($\mathcal{CSLO}$-mapping, satisfiable tableau set)
Let $\mathcal{M} = (\Delta, (\prec_w)_{w\in\Delta}, .^{\mathcal{M}})$ be a preferential model, and Γ a tableau set. A $\mathcal{CSLO}$-mapping from Γ to $\mathcal{M}$ is a function $f : \mathrm{Lab}_\Gamma \cup \mathcal{O}_\Gamma \longrightarrow \Delta$ satisfying the following condition : for every $y <_x z \in \Gamma,$ we have $f(y) \prec_{f(x)} f(z)$ in $\mathcal{M}$.

Given a tableau set Γ, a $\mathcal{CSLO}$-preferential model $\mathcal{M}$, and a $\mathcal{CSLO}$-mapping f from Γ to $\mathcal{M}$, we say that Γ is satisfiable under f in $\mathcal{M}$ if $x : A \in \Gamma$ implies $f(x) \in A^{\mathcal{M}}$. A tableau set Γ is satisfiable if it is satisfiable in some $\mathcal{CSLO}$-preferential model $\mathcal{M}$ under some $\mathcal{CSLO}$-mapping f. A $\mathcal{CSLO}$-tableau is satisfiable if at least one of its sets is satisfiable.

We can show that our tableau calculus is sound and complete with respect to the preferential semantics.

Theorem 2 (Soundness of the calculus)
A formula $A \in \mathcal{L}_{\mathcal{CSLO}}$ is satisfiable wrt. KB if any tableau started by $x : A$ is open.

The proof of the soundness is standard, and goes in the same lines than the one for $\mathcal{CSL}$ in Alenda *et al.* (2009b). Checking that our additional rules preserve satisfiability is easy.

Theorem 3 (Completeness of the calculus)
If a tableau started by $x : A$ is open, then A is satisfiable wrt. KB.

The proof of completeness is standard : we show that given any saturated open teableau set we can build a canonical model $\mathcal{M}_\Gamma$ such that Γ is satisfiable in $\mathcal{M}_\Gamma$ (and so $\mathcal{M}_\Gamma \models A$).

Saturated tableaux are defined as in Alenda *et al.* (2009b), the additional conditions for nominals and KB are :

$(T\mathcal{O})$ There is no $x \in \mathrm{Lab}_\Gamma$ and $a \in \mathcal{O}_\Gamma$ such that $x : \{a\} \in \Gamma$

$(F\mathcal{O})$ If $x : \neg\{a_1, \dots a_n\} \in \Gamma$ then $\forall i,\ 1 \leq i \leq n, x : \neg\{a_i\} \in \Gamma$.

$(Unfold)$ If an assertion $C \sqsubseteq D \in KB$ then for all label x occuring in Γ, $x : \neg C \sqcup D \in \Gamma$.

(Ind) If an assertion $A(b) \in KB$, then $b : A \in \Gamma$.

The following lemma shows that the preference relations $<_x$ satisfies the Limit Assumption for an open tableau set.

Lemma 1 (See Alenda *et al.* (2009b))
Let Γ *be an open tableau set containing only a finite number of positive* $\leftleftarrows$*-formulas* $x : A_0 \leftleftarrows B_0$, $x : A_1 \leftleftarrows B_1$, $x : A_2 \leftleftarrows B_2$, ..., $x : A_{n-1} \leftleftarrows B_{n-1}$. *Then* Γ *does not contain any infinite descending chain of labels* $y_1 <_x y_0$, $y_2 <_x y_1$, ..., $y_{i+1} <_x y_i, \ldots$.

Given an open tableau set Γ, we define a canonical model $\mathcal{M}_\Gamma = \langle \Delta, (\prec_w)_{w \in \Delta}, .^{\mathcal{M}_\Gamma} \rangle$ as follows :

- $\Delta = \mathrm{Lab}_\Gamma \cup \mathcal{O}_\Gamma$ and $y \prec_x z$ iff $y <_x z \in \Gamma$.
- For all propositional variables $V_i \in \mathcal{V}_p$, $V_i^{\mathcal{M}_\Gamma} = \{x \mid x : V_i \in \Gamma\}$

$\mathcal{M}_\Gamma$ is indeed a $\mathcal{CSLO}$-model, as each preferential relation is centered, modular, and satisfies the limit assumption. The first two came from the rules $(Cent)$ and (Mod), and we have the latter by lemma 1.

We now show that Γ is satisfiable in $\mathcal{M}_\Gamma$ under the trivial identity mapping, i.e for all formula $C \in \mathcal{L}_{\mathcal{CSL}}$: (i) if $x : C \in \Gamma$, then $x \in C^{\mathcal{M}_\Gamma}$. (ii) if $x : \neg C \in \Gamma$, then $x \in (\neg C)^{\mathcal{M}_\Gamma}$.

The proof is by induction on the complexity $cp(C)$ of a formula C and goes along the same lines as for $\mathcal{CSL}$ (the proofs when $C = \{a_i, \ldots, a_n\}$ or its negation are straightforward). We can also easily show that $\mathcal{M}_\Gamma$ satisfies each assertion of the KB (because of the corresponding saturation condition), which ensures the satisfiability of the formula with respect to KB.

Due to the interplay beween dynamic rules (which creates new labels) and static ones (which can add new formulas to these new labels, to which the dynamic rules could also be applied, and so on.), the calculus can lead to infinite tableau set. However, it can be made terminating by using the same blocking (or *loop-checking*) conditions than the ones used in Alenda *et al.* (2009b). The proof goes in the same way, as our new rules for nominals and KB cannot loop by themselves, and do not interfere with the original rules for $\mathcal{CSL}$ to create such loops.

4 Conclusion - Further Research

In this paper, we extended the logic $\mathcal{CSL}$ with nominals, and provided a descision procedure for checking satisfiability of a formula with respect to a knowledge base, in description logics style. This is a first step toward integration with this kind of logics, and thus toward integration with ontologies representation and reasonning systems.

There are number of issues to explore in future research. As the calculus for $\mathcal{CSL}$ presented in Alenda *et al.* (2009a) is not optimal, the calculus presented here is not guaranteed to have an optimal complexity, and thus might be improved (by using *caching* *techniques* for example). Another issue is to consider integration of the concept similarity operator with significant description logics. Finally, one interesting issue is to check if the logic $\mathcal{CSL}$ (and its extension) could be useful to represent agent's preference. The link with other logics for preference (or making use of preferentials relations) can also be investigated.

Références

ALENDA R., OLIVETTI N. & SCHWIND C. (2009a). Comparative concept similarity over minspaces : Axiomatisation and tableaux calculus. In M. GIESE & A. WAALER, Eds., *TABLEAUX 2009*, volume 5607 of *Lecture Notes in Computer Science*, p. 17–31 : Springer.

ALENDA R., OLIVETTI N. & SCHWIND C. (2009b). *Comparative concept similarity over Minspaces : Axiomatisation and Tableaux Calculus*. Rapport interne, `http ://www.lsis.org/olivetti/TR09/CCS_AOS09.pdf`.

BOOLOS G. (1993). *The Logic of Provability*. Cambridge University Press.

KURUCZ A., WOLTER F. & ZAKHARYASCHEV M. (2005). Modal logics for metric spaces : Open problems. In S. N. ARTËMOV, H. BARRINGER, A. S. D'AVILA GARCEZ, L. C. LAMB & J. WOODS, Eds., *We Will Show Them ! (2)*, p. 193–108 : College Publications.

LEWIS D. (1973). *Counterfactuals*. Basil Blackwell Ltd.

NUTE D. (1980). *Topics in Conditional Logic*. Dordrecht : Reidel Publishing Company.

SHEREMET M., TISHKOVSKY D., WOLTER F. & ZAKHARYASCHEV M. (2005). Comparative similarity, tree automata, and diophantine equations. In G. SUTCLIFFE & A. VORONKOV, Eds., *LPAR 2005*, volume 3835 of *Lecture Notes in Computer Science*, p. 651–665 : Springer.

SHEREMET M., TISHKOVSKY D., WOLTER F. & ZAKHARYASCHEV M. (2007). A logic for concepts and similarity. *J. Log. Comput.*, **17**(3), 415–452.

SHEREMET M., WOLTER F. & ZAKHARYASCHEV M. (2008). A modal logic framework for reasoning about comparative distances and topology. submitted.

STALNAKER R. (1968). A theory of conditionals. *In N. Rescher (ed.), Studies in Logical Theory, American Philosophical Quarterly, Monograph Series no.2, Blackwell, Oxford*, p. 98–112.