Implicit Representation of Bigranular Rules for Multigranular Data

Stephen J. Hegner[1] and M. Andrea Rodríguez[2]

[1] DBMS Research of New Hampshire
PO Box, 2153, New London, NH 03257, USA
dbmsnh@gmx.com
[2] Millennium Institute for Foundational Research on Data
Departamento Ingeniería Informática y Ciencias de la Computación
Edmundo Larenas 219, Universidad de Concepción
4070409 Concepción, Chile
andrea@udec.cl

Abstract. Domains for spatial and temporal data are often multigranular in nature, possessing a natural order structure defined by spatial inclusion and time-interval inclusion, respectively. This order structure induces lattice-like (partial) operations, such as join, which in turn lead to *join rules*, in which a single domain element (granule) is asserted to be equal to, or contained in, the join of a set of such granules. In general, the efficient representation of such *join rules* is a difficult problem. However, there is a very effective representation in the case that the rule is *bigranular*; i.e., all of the joined elements belong to the same granularity, and, in addition, complete information about the (non)disjointness of all granules involved is known. The details of that representation form the focus of the paper.

1 Introduction

In a multigranular attribute, the domain elements are related by order-like and even lattice-like operations, leading to a much richer family of integrity constraints than is found in the traditional monogranular setting. The ideas are best illustrated via example. Let $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$ be the schema in which the spatial attribute A_{Plc} identifies certain geographical areas of Chile, the temporal attribute A_{Tim} identifies intervals of time, and the thematic attribute B_{Bth} has numerical values representing the number of births. A tuple of the form $\langle p, t, b\rangle$ denotes that in the region defined by p, for the time interval defined by t, the number of births was b. An example instance for this schema is shown in Fig. 1. Think of the two tables of that figure to be part of a single relation; the division is for expository reasons, as well as to conserve space. In that instance, for domain elements (called *granules*) of A_{Plc}, the suffix *_prv* identifies the name as that of a province, *_rgn* identifies a region, *_cmn* identifies a county, while *_urb* identifies a metropolitan area. For A_{Tim}, *Y2017Qx* denotes quarter *x* of year 2017, while *Y2017* represents the entire year. Such a multigranular schema and

$\underline{A_{\mathsf{Plc}}}$	$\underline{A_{\mathsf{Tim}}}$	B_{Bth}
Los_Lagos_rgn	*Y2017Q1*	b_1
Osorno_prv	*Y2017Q1*	b_2
Llanquihue_prv	*Y2017Q1*	b_3
Chiloé_prv	*Y2017Q1*	b_4
Palena_prv	*Y2017Q1*	b_5
Puerto_Montt_cmn	*Y2017Q1*	b_6
Puerto_Varas_cmn	*Y2017Q1*	b_7
Gran_Puerto_Montt_urb	*Y2017Q1*	b_8

$\underline{A_{\mathsf{Plc}}}$	$\underline{A_{\mathsf{Tim}}}$	B_{Bth}
BíoBío_rgn	*Y2017*	b'_1
BíoBío_rgn	*Y2017Q1*	b'_2
BíoBío_rgn	*Y2017Q2*	b'_3
BíoBío_rgn	*Y2017Q3*	b'_4
BíoBío_rgn	*Y2017Q4*	b'_5

Fig. 1. Multigranular relational instance

instance may arise, for example, when data of varying granularities of space and time are integrated, into a single schema, with respect to the same *thematic attribute* (here B_{Bth}).

It is clear that the ordinary functional dependency (FD) $\{A_{\mathsf{Plc}}, A_{\mathsf{Tim}}\} \rightarrow B_{\mathsf{Bth}}$ is expected to hold. However, there are also several other natural dependencies, induced by the structure of the multigranular domains. Each of the four listed provinces is contained in the region Los Lagos, expressed formally as *Osorno_prv* $\sqsubseteq$ *Los_Lagos_rgn*, *Llanquihue_prv* $\sqsubseteq$ *Los_Lagos_rgn*, *Chiloé_prv* $\sqsubseteq$ *Los_Lagos_rgn*, and *Palena_prv* $\sqsubseteq$ *Los_Lagos_rgn*. Similarly, both counties, as well as the metropolitan area of Gran Puerto Montt, are contained in the province Llanquihue; *Puerto_Montt_cmn* $\sqsubseteq$ *Llanquihue_prv*, *Puerto_Varas_cmn* $\sqsubseteq$ *Llanquihue_prv*, and *Gran_Puerto_Montt_urb* $\sqsubseteq$ *Llanquihue_prv*. For the temporal domain, each of the quarters of 2017 is contained in the entire year: *Y2017Qx* $\sqsubseteq$ *Y2017* for $x \in \{1, 2, 3, 4\}$. Since the number of births is monotonic with respect to region size and time-interval size, these conditions in turn lead to the constraints $b_i \leq b_1$ for $i \in \{2, 3, 4, 5\}$, $b_i \leq b_3$ for $i \in \{6, 7, 8\}$, and $b'_i \leq b'_1$ for $i \in \{2, 3, 4, 5\}$.

More is true, however. The region Los Lagos is composed exactly of the four provinces listed, without any overlap, written as the *disjoint-join equality rule* (r-LLr) below.

$$\mathit{Los\_Lagos\_rgn} = \bigsqcup^{\perp} \{\mathit{Osorno\_prv}, \mathit{Llanquihue\_prv}, \mathit{Chilo\acute{e}\_prv}, \mathit{Palena\_prv}\} \quad \text{(r-LLr)}$$

Specifically, the symbol $\bigsqcup$ means that the four provinces cover the region completely, while the embedded $\perp$ means that the join is *disjoint*; that is, that the regions do not overlap. This leads to the *spatial aggregation constraint* $\sum_{i=2}^{5} b_i = b_1$. Additionally, the metropolitan area of Gran Puerto Montt lies entirely within the combined areas of the counties Puerto Montt and Puerto Varas, leading to the *disjoint-join subsumption rule* (r-Llp) shown below, and consequently the spatial aggregation constraint $b_8 \leq b_6 + b_7$.

$$\mathit{Gran\_Puerto\_Montt\_urb} \sqsubseteq \bigsqcup^{\perp} \{\mathit{Puerto\_Montt\_cmn}, \mathit{Puerto\_Varas\_cmn}\} \quad \text{(r-Llp)}$$

Such aggregation constraints arise in the same fashion for temporal multigranular attributes, such as A_{Tim}. For example, the disjoint-join equality rule (r-YQ2017) shown below holds, leading to the *temporal aggregation constraint* $\sum_{i=2}^{5} b'_i = b'_1$.

$$\mathit{Y2017} = \bigsqcup^{\perp} \{\mathit{Y2017Q1}, \mathit{Y2017Q2}, \mathit{Y2017Q3}, \mathit{Y2017Q4}\} \qquad \text{(r-YQ2017)}$$

Aggregation constraints arising from join rules, as illustrated by the examples above, are instances of *TMCDs* or *thematic multigranular comparison dependencies*, which are developed in detail in [8], including a notion of *tolerance* which replaces absolute equality with an approximate one (to account for differences arising from rounding and measurement errors). In order to enforce such TMCDs, it is first of all essential to know which ones hold. This, in turn, requires a means to determine which disjoint-join rules hold. Although a formal semantics and inference mechanism for such rules is developed in [8], it is quite resource expensive to enforce all TMCDs by identifying the associated join rules via direct inference. The focus of this paper is the development of a compact and efficient representation for certain types of join rules which occur frequently in practice.

Key to these results are the observation that the granules of a multigranular attribute may be partitioned naturally into so-called *granularities* (hence the term *multigranular*) of *disjoint* members, as illustrated in Fig. 2 for both space and time. Arrows of the form $G_1 \longrightarrow\!\!\!< G_2$ represent the basic refinement order

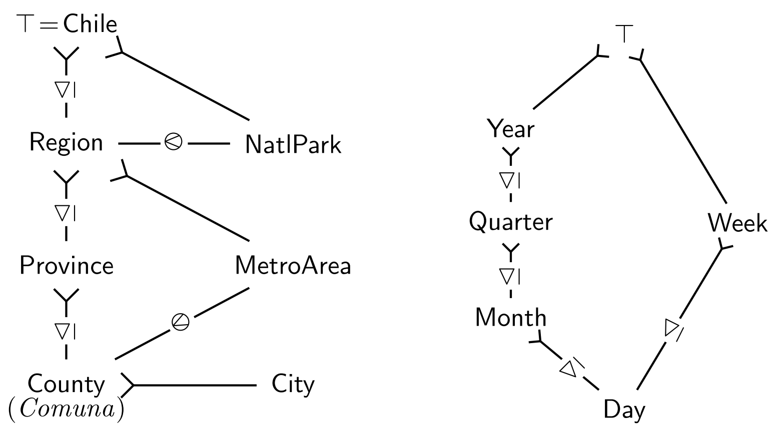

Fig. 2. Granularity hierarchies for Chile and for time

of granularities, in the sense that for every granule g_1 of granularity G_1 there is a granule g_2 of granularity G_2 with $g_1 \sqsubseteq g_2$. Inline, this typically written $G_1 \leq G_2$. Thus, every county is contained in a (unique) province, every province is contained in a (unique) region, and every region is contained in Chile. Similarly, every metropolitan area is contained in a region, (although not necessarily in a single province.)

In support of the representation of rules, there are two additional binary relations on granularities which are of fundamental importance, *equality join order*, denoted $\trianglelefteq$, and *subsumption join order*, denoted $\otimes$. $G_1 \trianglelefteq G_2$ holds just in case every granule g_2 of granularity G_2 is the (necessarily disjoint) join of some granules of granularity G_1; i.e., if $g_2 = \biguplus S$ holds for some finite set S of granules of G_2. As can be seen in Fig. 2, with the symbol $\trianglelefteq$ embedded in a line indicating that this relation holds between the granularities which it connects, this condition characterizes many practical situations. As a concrete example, $\mathsf{Province} \trianglelefteq \mathsf{Region}$, with (r-LLp) a specific instance of a join rule arising from it. Similarly, for the time hierarchy, (r-YQ2017) is a specific instance of a rule arising from $\mathsf{Quarter} \trianglelefteq \mathsf{Year}$.

The main result of this paper regarding $\trianglelefteq$ may be summarized as follows. Let $\mathsf{NRel}_{\langle G_1, G_2 \rangle}$ denote the relation which identifies pairs $\langle g_1, g_2 \rangle$ of granules from $\langle G_1, G_2 \rangle$ (i.e., with g_1 of granularity G_1 and g_2 of granularity G_2) which are not disjoint. Then, it must be the case that $S = \{g_2 \mid \langle g_1, g_2 \rangle \in \mathsf{NRel}_{\langle G_1, G_2 \rangle}\}$; in other words, S must be exactly the set of all granules of g_2 which are not disjoint from g_1. As a specific example, to identify those provinces which lie in $\textsf{\textit{Los\_Lagos\_rgn}}$, it is only necessary to retrieve $\{g \mid \langle \textsf{\textit{Los\_Lagos\_rgn}}, g \rangle \in \mathsf{NRel}_{\langle \mathsf{Region}, \mathsf{Province} \rangle}\}$; no complex inference procedure is necessary. In assessing this solution, it must be remembered that knowledge about granules, including subsumption, disjointness, and join, is specified via statements. There is the possibility that a given assertion is *unresolvable*; i.e., it is not possible to establish that it is true or it is false. (See Summary 2.7 for details.) What is remarkable about this result is that no such unresolvability can occur for $\langle G_1, G_2 \rangle$ disjointness. For $G_1 \trianglelefteq G_2$ to hold, it must be the case that for any pair $\langle g_1, g_2 \rangle$ of granules of $\langle G_1, G_2 \rangle$, it is the case that the disjointness of $\langle g_1, g_2 \rangle$ is resolvable.

This idea applies also, subject to an additional condition, when subsumption replaces equality. $G_1 \otimes G_2$ holds just in case every granule of G_1 is subsumed by the join of some granules in G_2; i.e., if $g_2 \sqsubseteq \biguplus S$ holds for some finite set S of granules of G_2. This is illustrated in particular by rule (r-Llp), as an instance of $\mathsf{County} \otimes \mathsf{MetroArea}$. Of course, $G_1 \trianglelefteq G_2$ always implies $G_1 \otimes G_2$, but this example shows that the converse need not hold. The additional condition which must be imposed is that the join be *resolved minimal*, meaning that if any element is removed from the join set, the assertion becomes resolvably false. In other words, both $\textsf{\textit{Gran\_Puerto\_Montt\_urb}} \not\sqsubseteq \textsf{\textit{Puerto\_Montt\_cmn}}$ and $\textsf{\textit{Gran\_Puerto\_Montt\_urb}} \not\sqsubseteq \textsf{\textit{Puerto\_Varas\_cmn}}$ must follow from the rules. In this case, to determine the counties in which $\textsf{\textit{Gran\_Puerto\_Montt\_urb}}$ lies, it is only necessary to retrieve $\{g \mid \langle \textsf{\textit{Gran\_Puerto\_Montt\_urb}}, g \rangle \in \mathsf{NRel}_{\langle \mathsf{County}, \mathsf{MetroArea} \rangle}\}$.

To clarify the terminology, a join rule $g = \biguplus S$ is *bigranular* if every granule in S is of the same granularity G_2. (Since granules of the same granularity are disjoint, it must be the case that the granularity G_1 of g is different from that of the members of S, hence the term bigranular.) Thus, any rule arising from the application of a condition of the form $G_1 \trianglelefteq G_2$ or $G_1 \circledcirc G_2$ is necessarily bigranular.

The representations developed above are termed *implicit*, since a rule of the form $g = \biguplus S$ or $g \sqsubseteq \biguplus S$ is represented by a way to recover S from the appropriate $\mathsf{NRel}_{\langle -,- \rangle}$. In the remainder of this paper, the details of how and why this method of representing of join rules works are developed.

The paper is organized as follows. Section 2 provides necessary details of the multigranular framework developed in [8]. Section 3 develops the general ideas of minimality for join rules, while Sec. 4 contains the main results of the paper on the representation of bigranular join rules. Finally, Sec. 5 contains conclusions and further directions.

2 Multigranular Attributes and Their Semantics

The results of this paper are based upon the formal model of multigranular attributes, as developed in [8]. It is thus appropriate to begin with a summary of that framework. Although [7] covers similar material, it is of a preliminary nature, so the reader is always referred to [8] for clarification of details. For terminology and notation regarding logic, consult [11], while for issues surrounding order structures, including posets, see [3]. For basic concepts surrounding the relational model, see [9].

Notation 2.1 (Special mathematical notation). $X_1 \subsetneq X_2$ (resp. $X_1 \subseteq_f X_2$ denotes that X_1 is a proper (resp. finite) subset of X_2. The cardinality of the set X is denoted $\mathsf{Card}(X)$.

Overview 2.2 (Constrained granulated attribute schemata). In the ordinary relational model with SQL used for data definition, several attributes may use the same data type. For example, two distinct attributes may be declared to be of the same type `VARCHAR(10)`. Similarly, in the multigranular model, several distinct attributes may be declared to be of the same type. Such a type is called a *constrained granulated attribute schema*, or *CGAS*, and is a triple $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle)$ in which $\mathbf{Glty}\langle\mathfrak{S}\rangle$ is a poset of *granularities* and $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$ is a *granule assignment*, both elaborated in Summary 2.3 below, while $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle$ is a unified set of constraints, elaborated in Summary 2.5 below.

Summary 2.3 (Granularities and granules). A *granularity poset* for the CGAS $\mathfrak{S}$ is an upper-bounded poset $\mathbf{Glty}\langle\mathfrak{S}\rangle = (\mathsf{Glty}\langle\mathfrak{S}\rangle, \leq_{\mathsf{Glty}\langle\mathfrak{S}\rangle}, \top_{\mathsf{Glty}\langle\mathfrak{S}\rangle})$; that is, it is poset with a greatest element $\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}$. The two diagrams of Fig. 2 represent the specific granularity posets for $\mathfrak{S}$ replaced by $\mathfrak{C}$ and $\mathfrak{T}$, respectively, with $G_1 \leq_{\mathsf{Glty}\langle\mathfrak{C}\rangle} G_2$ (resp. $G_1 \leq_{\mathsf{Glty}\langle\mathfrak{T}\rangle} G_2$) iff there is an arrow of the form $G_1 \multimap G_2$ in the associated diagram. In that which follows, $\mathfrak{S}$ will be used to represent a general CGAS, while $\mathfrak{C}$ (for Chile) and $\mathfrak{T}$ (for time) will be used to represent, respectively, the spatial and the temporal schema whose granularities are depicted in Fig. 2.

A *granule assignment* $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle = (\mathbf{Gnle}\langle\mathfrak{S}\rangle, \Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle)$ for $\mathfrak{S}$ extends the idea of a domain assignment for an ordinary relational attribute, in the sense that it assigns (with one exception) every granule to a granularity. $\mathbf{Gnle}\langle\mathfrak{S}\rangle = (\mathsf{Granules}\langle\mathfrak{S}\rangle, \bar{\sqsubseteq}_{\mathfrak{S}}, \top_{\mathfrak{S}}, \bot_{\mathfrak{S}})$ is the (bounded) *granule preorder*, while $\Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle = \{\mathsf{Granules}\langle\mathfrak{S}|G\rangle \mid G \in \mathsf{Glty}\langle\mathfrak{S}\rangle\}$ is a partition of $\mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle = \mathsf{Granules}\langle\mathfrak{S}\rangle \setminus \{\bot_{\mathfrak{S}}\}$ that identifies which granules are assigned to which granularities. The bottom granule $\bot_{\mathfrak{S}}$ (the least element of the preorder $\mathbf{Gnle}\langle\mathfrak{S}\rangle$) is not a member of $\mathsf{Granules}\langle\mathfrak{S}|G\rangle$ for any granularity G, while the top granule $\top_{\mathfrak{S}}$ (the greatest element of the preorder $\mathbf{Gnle}\langle\mathfrak{S}\rangle$) lies in $\mathsf{Granules}\langle\mathfrak{S}|\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}\rangle$.

The orders of granularities and granules are closely related. Specifically, for granularities G_1 and G_2, $G_1 \leq_{\mathsf{Glty}\langle\mathfrak{S}\rangle} G_2$ iff for every $g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$, there is a $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$ with the property that $g_1 \bar{\sqsubseteq}_{\mathfrak{S}} g_2$. Since $\mathbf{Gnle}\langle\mathfrak{S}\rangle$ is only a preorder, distinct granules may be equivalent, in the sense that $g_1 \bar{\sqsubseteq}_{\mathfrak{S}} g_2 \bar{\sqsubseteq}_{\mathfrak{S}} g_1$. Write $[g_1]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}$ to denote the equivalence class of g_1; thus, with g_1, g_2 as above, $g_2 \in [g_1]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}$ and $[g_1]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle} = [g_2]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}$. To avoid problems, the special notation $g_1 \overset{\text{id}}{=} g_2$ will be used to mean that g_1 and g_2 are the same granule, with the meaning of $g_1 = g_2$ deferred until Summary 2.5, when semantics are discussed. With this in mind, further conditions may be stated. First of all, the top granularity $\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}$ is the only one which may contain equivalent but not identical granules. It contains the top granule $\top_{\mathfrak{S}}$ (the greatest element of the poset $\mathbf{Gnle}\langle\mathfrak{S}\rangle$), as well as any granule equivalent to it. For example, in the CGAS $\mathfrak{C}$, $[\top_{\mathfrak{C}}]_{\mathsf{Gnle}\langle\mathfrak{C}\rangle} = [\mathsf{Chile}]_{\mathsf{Gnle}\langle\mathfrak{C}\rangle}$ (see Fig. 2). Otherwise, non-identical granules of the same granularity may not be equivalent, and they furthermore must have the bottom granule as GLB (greatest lower bound). More precisely, if g_1 and g_2 are of the same non-$\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}$ granularity, and $g_1 \overset{\text{id}}{\neq} g_2$, then both $([g_1]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle} \neq [g_2]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle})$ and $(\mathsf{GLB}_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}\langle\{g_1, g_2\}\rangle = \bot_{\mathfrak{S}})$ hold.

Summary 2.4 (Semantics of granules). A *granule structure* $\sigma = \sigma = (\mathsf{Dom}\langle\sigma\rangle, \mathsf{GnletoDom}_{\sigma})$ for the granule assignment $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$ provides set-based semantics. $\mathsf{Dom}\langle\sigma\rangle$ is a (not necessarily finite) set, called the *domain* of σ, and $\mathsf{GnletoDom}_{\sigma} : \mathsf{Granules}\langle\mathfrak{S}\rangle \to \mathbf{2}^{\mathsf{Dom}\langle\sigma\rangle}$ is a function which assigns to each granule a subset of the domain. In this assignment, granule subsumption translates to set inclusion ($g_1 \bar{\sqsubseteq}_{\mathfrak{S}} g_2$ implies $\mathsf{GnletoDom}_{\sigma}(g_1) \subseteq \mathsf{GnletoDom}_{\sigma}(g_2)$), granule disjointness translates to empty intersection (if g_1 and g_2 are of the same granularity with $g_1 \overset{\text{id}}{\neq} g_2$, then $\mathsf{GnletoDom}_{\sigma}(g_1) \cap \mathsf{GnletoDom}_{\sigma}(g_2) = \emptyset$); equivalent granules have identical semantics $((\mathsf{GnletoDom}_{\sigma}(g_1) = \mathsf{GnletoDom}_{\sigma}(g_2)) \Leftrightarrow [g_1]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle} = [g_2]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle})$; and the bottom granule maps to the empty set $(\mathsf{GnletoDom}_{\mathfrak{S}}(\bot_{\mathfrak{S}}) = \emptyset)$.

As already mentioned in Sec. 1, for a spatial attribute such as $\mathfrak{C}$, a natural granular structure might be σ_{Chile}, the subset of the real plane $\mathbb{R} \times \mathbb{R}$ representing Chile, with $\mathsf{GnletoDom}_{\sigma_{\mathsf{Chile}}}(g)$ exactly the geographic region corresponding to granule g. While such a structure is mathematically correct, it involves an enormous amount of detail, much more than is necessary in many cases. It is for this reason that the semantics of a multigranular attribute is modelled not by a single granular structure, but rather by any such structure which satisfies the constraint, or rules, of the schema, as defined in Summary 2.5 below. For a more complete explanation, see [8, 3.6].

Summary 2.5 (Rules). In [8, Sec. 3], general constraints for GGASs and their semantics are developed extensively. In this paper, only those constraint types which are used in the theory developed here are sketched.

The *primitive basic rules* over the CGAS $\mathfrak{S}$, denoted, $\mathsf{PrBaRules}\langle\mathfrak{S}\rangle$ are of the following two forms.

(pjrule-i) A *subsumption join rule* is of the form $(g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S)$ for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. The *elemental subsumption rule* $(g_1 \sqsubseteq_{\mathfrak{S}} g_2)$, with $g_1, g_2 \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$, is shorthand for $(g_1 \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} \{g_2\})$.

(psrule-ii) A *basic disjointness rule* is of the form $(\bigsqcap_{\mathfrak{S}} \{g_1, g_2\} = \perp_{\mathfrak{S}})$ for $g_1, g_2 \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ and $[g_1]_{\mathfrak{S}} \neq [g_2]_{\mathfrak{S}}$.

Extending the notion of semantics of Summary 2.4 to $\mathsf{PrBaRules}\langle\mathfrak{S}\rangle$, a granule structure σ for $\mathfrak{S}$ is a *model* of the subsumption rule $(g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S)$ if $\mathsf{GnletoDom}_\sigma(g) \subseteq \bigcup_{s\in S} \mathsf{GnletoDom}_\sigma(s)$, while σ is *model* of the basic disjointness rule $(\bigsqcap_{\mathfrak{S}} \{g_1, g_2\} = \perp_{\mathfrak{S}})$ if $\mathsf{GnletoDom}_\sigma(g_1) \cap \mathsf{GnletoDom}_\sigma(g_2) = \emptyset$. For $\Phi \subseteq \mathsf{PrBaRules}\langle\mathfrak{S}\rangle$, $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle$ denotes the collection of all models of Φ.

For any CGAS $\mathfrak{S}$, the *built-in rules* $\mathsf{BuiltInRules}\langle\mathfrak{S}\rangle$ are those which are satisfied by every granular structure σ for $\mathfrak{S}$. These include the subsumption rule $(g_1 \sqsubseteq_{\mathfrak{S}} g_2)$ whenever $g_1 \mathrel{\bar{\sqsubseteq}}_{\mathfrak{S}} g_2$ holds,[3] as well as $\bigsqcap_{\mathfrak{S}} \{g_1, g_2\} = \perp_{\mathfrak{S}}$ whenever $g_1 \overset{\mathsf{id}}{\neq} g_2$ are of the same granularity.

A *complex rule* is a conjunction of primitive basic rules. Write $\mathsf{Conjuncts}\langle\varphi\rangle$ to denote the set of conjuncts of the complex rule φ. Thus, if $\varphi = \varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_k$, then $\mathsf{Conjuncts}\langle\varphi\rangle = \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$. The most important kind of complex rules are the *complex join rules*:

(cjrule-i) An *equality join rule* is of the form $(g = \bigsqcup_{\mathfrak{S}} S)$, for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. Its definition in terms of primitive basic rules is
$$\mathsf{Conjuncts}_{\mathfrak{S}}\langle(g = \textstyle\bigsqcup_{\mathfrak{S}} S)\rangle = \{(g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S)\} \cup \{(g_i \sqsubseteq_{\mathfrak{S}} g) \mid g_i \in S\}.$$

(cjrule-ii) A *disjoint-join subsumption rule*, written as $(g \sqsubseteq_{\mathfrak{S}} \biguplus_{\mathfrak{S}} S)$ for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$, is defined in terms of primitive basic join rules as
$$\mathsf{Conjuncts}_{\mathfrak{S}}\langle(g \sqsubseteq_{\mathfrak{S}} \textstyle\biguplus_{\mathfrak{S}} S)\rangle = \mathsf{Conjuncts}\langle(g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S)\rangle \cup \{(\bigsqcap_{\mathfrak{S}} \{g_1, g_2\} = \perp_{\mathfrak{S}}) \mid g_i, g_j \in S \text{ and } g_i \overset{\mathsf{id}}{\neq} g_2\}.$$

(cjrule-iii) A *disjoint-join equality rule*, written as $(g = \biguplus_{\mathfrak{S}} S)$ for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. is defined in terms of primitive basic join rules as
$$\mathsf{Conjuncts}_{\mathfrak{S}}\langle(g = \textstyle\biguplus_{\mathfrak{S}} S)\rangle = \mathsf{Conjuncts}_{\mathfrak{S}}\langle(g = \bigsqcup_{\mathfrak{S}} S)\rangle \cup \mathsf{Conjuncts}_{\mathfrak{S}}\langle(g \sqsubseteq_{\mathfrak{S}} \biguplus_{\mathfrak{S}} S)\rangle.$$

For convenience, a complex rule will be represented by its set of conjuncts. Thus, every complex rule is a regarded as a finite nonempty set of primitive basic rules.

[3] $\bar{\sqsubseteq}_{\mathfrak{S}}$ is the granule preorder defined in the granule assignment $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$ (see Summary 2.3) while $\sqsubseteq_{\mathfrak{S}}$ is the general subsumption relation used to define rules. For $g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, it is always the case that $g_1 \mathrel{\bar{\sqsubseteq}}_{\mathfrak{S}} g_2$ implies $(g_1 \sqsubseteq_{\mathfrak{S}} g_2)$). The converse is not required to hold, although in practice it usually does.

For simplicity, the example rules in Sec. 1 were presented without qualifying subscripts on the operators. Using the notation for specific granular attributes introduced in Summary 2.3, for example, rule (r-Llp) should be written more properly as $\mathsf{\mathit{Gran\_Puerto\_Montt\_urb}} \sqsubseteq_{\mathfrak{C}} \bigsqcup^{\perp}_{\mathfrak{C}}\{\mathsf{\mathit{Puerto\_Montt\_cmn}}, \mathsf{\mathit{Puerto\_Varas\_cmn}}\}$. It is assumed that the reader will add these qualifying symbols, as necessary.

Summary 2.6 (Negation of rules). It is also necessary to work with negations of primitive basic rules over the CGAS $\mathfrak{S}$; the most important example is negation of disjointness; for $g_1, g_2 \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$, write $(\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} \neq \perp_{\mathfrak{S}})$ to mean $\neg(\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \perp_{\mathfrak{S}})$. Similarly, $(g_1 \not\sqsubseteq_{\mathfrak{S}} g_2)$ means $\neg(g_1 \sqsubseteq_{\mathfrak{S}} g_2)$ and $(g_1 \not\sqsubseteq_{\mathfrak{S}} S)$ means $\neg(g_1 \sqsubseteq_{\mathfrak{S}} S)$. The set of all negations of primitive basic rules is denoted $\mathsf{NegPrBaRules}\langle\mathfrak{S}\rangle$. The granule structure σ is a model of $\psi = \neg\varphi \in \mathsf{NegPrBaRules}\langle\mathfrak{S}\rangle$, iff it is not a model of φ; i.e., $\mathsf{Models}_{\mathfrak{S}}\langle\psi\rangle$ is the collection of all granule structures which do not lie in $\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle$.

For $\Phi, \Phi' \subseteq \mathsf{PrBaRules}\langle\mathfrak{S}\rangle$, define $\mathsf{Not}\langle\Phi\rangle = \{(\neg\varphi) \mid \varphi \in \Phi\}$. Thus, $\mathsf{NegPrBaRules}\langle\mathfrak{S}\rangle = \mathsf{Not}\langle\mathsf{PrBaRules}\langle\mathfrak{S}\rangle\rangle$.

Finally, it is convenient to combine positive and negated rules into one set. Define $\mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle = \mathsf{PrBaRules}\langle\mathfrak{S}\rangle \cup \mathsf{NegPrBaRules}\langle\mathfrak{S}\rangle$. For $\Phi \subseteq \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$, $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle = \bigcap\{\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle \mid \varphi \in \Phi\}$.

Summary 2.7 (Satisfiability and Resolvability). Continuing with $\mathfrak{S}$ a CGAS, for $\varphi \in \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$ and $\Phi \subseteq \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$, define semantic entailment $\Phi \models_{\mathfrak{S}} \varphi$ to mean that $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle \subseteq \mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle$, and for $\Phi' \subseteq \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$, $\Phi \models_{\mathfrak{S}} \Phi'$ to mean that $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle \subseteq \mathsf{Models}_{\mathfrak{S}}\langle\Phi'\rangle$. In other words, Φ imposes stronger constraints than does Φ'. φ (resp. Φ) is *satisfiable* (or *consistent*) if it has a model; i.e., $\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle \neq \emptyset$ (resp. $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle \neq \emptyset$).

Let $\Phi \subseteq \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$ and $\varphi \in \mathsf{PrBaRules}\langle\mathfrak{S}\rangle$. Say that φ is *resolvable* from Φ, written $\Phi \models^{\pm}_{\mathfrak{S}} \varphi$, if one of $\Phi \models_{\mathfrak{S}} \varphi$ or else $\Phi \models_{\mathfrak{S}} \neg\varphi$ holds. In other words, the truth value of φ is determined by Φ; either φ is true in every model of Φ, or else φ is false in every model of φ.

The set $\mathsf{PrBaRules}\langle\mathfrak{S}\rangle$ has the property of admitting *Armstrong models* [6], in the precise sense that for any consistent $\Phi \subseteq \mathsf{PrBaRules}\langle\mathfrak{S}\rangle$, there is a model which satisfies only those members of Φ. This means that members of $\mathsf{NegPrBaRules}\langle\mathfrak{S}\rangle$ whose negations are not entailed by Φ may be added to Φ in any combination while retaining satisfiability. See [8, 3.15-3.20] for details.

Finally, $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \subseteq \mathsf{AllPrBaRules}\langle\mathfrak{S}\rangle$ is a consistent set of rules, representing the set of constraints of $\mathfrak{S}$, as first identified in Overview 2.2. In [8] this set is represented as a pair $\langle\mathsf{Constr}(\mathfrak{S}), \mathsf{cwa}\langle\mathfrak{S}\rangle\rangle$, with $\mathsf{Constr}(\mathfrak{S})$ the positive constraints and $\mathsf{cwa}\langle\mathfrak{S}\rangle$ those to be negated; $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle = \mathsf{Constr}(\mathfrak{S}) \cup \mathsf{Not}\langle\mathsf{cwa}\langle\mathfrak{S}\rangle\rangle$ provides the equivalence of notation.

3 Minimality of Join Rules

Roughly, a join rule is minimal if removing any of the joined granules results in a rule which is no longer a consequence of the constraints. In this section, this idea of minimality is developed formally.

Context 3.1 (CGAS). Unless stated specifically to the contrary, for the remainder of this paper, let $\mathfrak{S} = (\mathsf{\mathbf{Glty}}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle)$ denote an arbitrary CGAS.

Notation 3.2 (Components of join rules). There are four variants of join rule over $\mathfrak{S}$, identified in (pjrule-i) and (cjrule-i)–(cjrule-iii) of Summary 2.5, collectively denoted $\mathsf{JRules}\langle\mathfrak{S}\rangle$. A join rule over $\mathfrak{S}$ is thus a statement of the form $(g \circledast \bigsqcup^{?} S)$ with $\circledast \in \{=, \sqsubseteq_{\mathfrak{S}}\}$, and $\bigsqcup^{?} \in \{\bigsqcup_{\mathfrak{S}}, \biguplus_{\mathfrak{S}}\}$, for $g \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$, and $S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ nonempty. Using terminology borrowed from logic, g is called the *head* of the rule while S is called the *body*, denoted by $\mathsf{Head}\langle\varphi\rangle$ and $\mathsf{Body}\langle\varphi\rangle$, respectively, for $\varphi \in \mathsf{JRules}\langle\mathfrak{S}\rangle$. In addition, $\mathsf{CompOp}\langle\varphi\rangle \in \{=, \sqsubseteq_{\mathfrak{S}}\}$ denotes the operator of the rule, and $\mathsf{JoinOp}\langle\varphi\rangle \in \{\bigsqcup_{\mathfrak{S}}, \biguplus_{\mathfrak{S}}\}$ denotes the join operation of the rule. In other words, $\mathsf{CompOp}\langle\varphi\rangle$ is just $\circledast$ and $\mathsf{JoinOp}\langle\varphi\rangle$ is just $\bigsqcup^{?}_{\mathfrak{S}}$, as defined above. The new notation is introduced in order to be able to parameterize these items in terms of the underlying rule φ. Thus, φ may be written, somewhat cryptically, as $(\mathsf{Head}\langle\varphi\rangle\ \mathsf{CompOp}\langle\varphi\rangle\ \mathsf{JoinOp}\langle\varphi\rangle\ \mathsf{Body}\langle\varphi\rangle)$.

Definition 3.3 (Primitive reduction and minimality of join rules). The *primitive reduction* of $\varphi \in \mathsf{JRules}\langle\mathfrak{S}\rangle$ by $Z \subseteq \mathsf{Body}\langle\varphi\rangle$, denoted $\mathsf{PrReduct}\langle\varphi : Z\rangle$, is obtained by removing the members of Z from $\mathsf{Body}\langle\varphi\rangle$, and by replacing, if necessary, equality with subsumption as the comparison operator. Formally, $\mathsf{PrReduct}\langle\varphi : Z\rangle$ is the rule $\varphi' \in \mathsf{JRules}\langle\mathfrak{S}\rangle$ with $\mathsf{Body}\langle\varphi'\rangle = \mathsf{Body}\langle\varphi\rangle \setminus Z$ and $\mathsf{JoinOp}\langle\varphi\rangle = \bigsqcup_{\mathfrak{S}}$, while $\mathsf{Head}\langle\varphi'\rangle$ and $\mathsf{CompOp}\langle\varphi'\rangle$, remain unchanged from φ. If $\mathsf{Body}\langle\varphi'\rangle$ is a proper subset of $\mathsf{Body}\langle\varphi\rangle$; i.e., $\mathsf{Body}\langle\varphi'\rangle \subsetneq \mathsf{Head}\langle\varphi\rangle$, then φ' is called a *proper primitive reduction* of φ. For example, letting φ be the rule (r-LLr) of Sec. 1, with $Z = \{\mathit{Osorno\_prv}, \mathit{Chiloé\_prv}\}$,

$$\mathsf{PrReduct}\langle\varphi : Z\rangle = (\mathit{Los\_Lagos\_rgn} \sqsubseteq_{\mathfrak{C}} \textstyle\bigsqcup_{\mathfrak{C}}\{\mathit{Llanquihue\_prv}, \mathit{Palena\_prv}\}).$$

$\varphi \in \mathsf{JRules}\langle\mathfrak{S}\rangle$ is *minimal* (for $\mathfrak{S}$) if for no proper primitive reduction φ' of φ is it the case that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \varphi'$. More formally, φ is minimal if for no nonempty $Z \subseteq \mathsf{Body}\langle\varphi\rangle$ is it the case that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \mathsf{PrReduct}\langle\varphi : Z\rangle$. In other words, if any nonempty subset of the body is removed, the resulting rule is no longer a consequence of $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle$. φ is *resolved minimal* (for $\mathfrak{S}$) if for every nonempty $Z \subseteq \mathsf{Body}\langle\varphi\rangle$ it is the case that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \neg\mathsf{PrReduct}_{\mathfrak{S}}\langle\varphi : Z\rangle$. Put another way, if any element of the body is removed, and the comparison operator is replaced by subsumption, the rule becomes false. If φ is minimal but not resolved minimal, then it is called *unresolved minimal*. Both forms of minimality may be characterized by the removal of single elements from the body. Define the *primitive reduction set* of φ, denoted $\mathsf{RedSet}\langle\varphi\rangle$, to be

$$\{\mathsf{PrReduct}_{\mathfrak{S}}\langle\varphi : \{h\}\rangle \mid h \in \mathsf{Body}\langle\varphi\rangle\} \text{ if } \mathsf{Card}(\mathsf{Body}\langle\varphi\rangle) \geq 2,$$

and to be $\emptyset$ otherwise. For example, letting φ again be (r-LLr),

$$\begin{aligned}\mathsf{RedSet}\langle\varphi\rangle = \{&(\mathit{Los\_Lagos\_rgn} \sqsubseteq_{\mathfrak{C}} \textstyle\bigsqcup_{\mathfrak{C}}\{\mathit{Osorno\_prv}, \mathit{Llanquihue\_prv}, \mathit{Chiloé\_prv}\}),\\ &(\mathit{Los\_Lagos\_rgn} \sqsubseteq_{\mathfrak{C}} \textstyle\bigsqcup_{\mathfrak{C}}\{\mathit{Osorno\_prv}, \mathit{Llanquihue\_prv}, \mathit{Palena\_prv}\}),\\ &(\mathit{Los\_Lagos\_rgn} \sqsubseteq_{\mathfrak{C}} \textstyle\bigsqcup_{\mathfrak{C}}\{\mathit{Osorno\_prv}, \mathit{Chiloé\_prv}, \mathit{Palena\_prv}\}),\\ &(\mathit{Los\_Lagos\_rgn} \sqsubseteq_{\mathfrak{C}} \textstyle\bigsqcup_{\mathfrak{C}}\{\mathit{Llanquihue\_prv}, \mathit{Chiloé\_prv}, \mathit{Palena\_prv}\})\}.\end{aligned}$$

For φ to be minimal, no element of $\mathsf{RedSet}\langle\varphi\rangle$ may be implied by the constraints, while to be resolved minimal, the negation of every such element must be so implied. This is formalized by the following, whose proof is immediate.

Observation 3.4 (Removing single elements suffices). *Let* $\varphi \in \mathsf{JRules}\langle\mathfrak{S}\rangle$ *with* $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \varphi$.

(a) φ *is minimal iff for no* $\psi \in \mathsf{RedSet}\langle\varphi\rangle$ *does* $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \psi$ *hold.*

(b) φ *is resolved minimal iff* $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \mathsf{Not}\langle\mathsf{RedSet}\langle\varphi\rangle\rangle$. □

Proposition 3.5 (Disjoint equality join implies resolved minimality). *A disjoint equality join rule* φ *for which* $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \varphi$ *is resolved minimal.*

Proof. Writing φ as $(g = \biguplus_{\mathfrak{S}} S)$, according to Summary 2.5, it has the representation $\mathsf{Conjuncts}_{\mathfrak{S}}\langle\varphi\rangle =$

$(g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S) \cup \{(s \sqsubseteq_{\mathfrak{S}} g) \mid s \in S\} \cup \{(\bigsqcap_{\mathfrak{S}} \{s, s'\} = \bot_{\mathfrak{S}}) \mid s, s' \in S \text{ and } s \overset{\text{id}}{\neq} s'\}$

in terms of primitive basic rules. Now, let $\sigma \in \mathsf{Models}_{\mathfrak{S}}\langle\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle\rangle$ and choose any $s \in S$. Since $\sigma(s) \neq \emptyset$, $\sigma(s) \cap \sigma(s') = \emptyset$ for all $s' \in S \setminus \{s\}$, and $\sigma(g) = \bigcup\{\sigma(s'') \mid s'' \in S\}$, it follows that $\sigma(g) \subsetneq \bigcup\{s'' \in S \mid s'' \overset{\text{id}}{\neq} s\}$. Since σ is an arbitrary model of $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle$, it follows that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \neg(g \sqsubseteq_{\mathfrak{S}} S \setminus \{s\}) = \neg\mathsf{PrReduct}_{\mathfrak{S}}\langle\varphi : \{s\}\rangle$. Finally, since s is arbitrary, the proof follows from Observation 3.4(b). □

Discussion 3.6 (Subsumption join and minimal rules). In view of Proposition 3.5, (r-LLr) is automatically resolved minimal. This is clear, since if any of the provinces are removed from the body, the subsumption will fail. However, this idea does not extend to subsumption join. For example, any metropolitan area of Chile lies within the join of all counties; e.g.,

$$(\mathit{Gran\_Puerto\_Montt\_urb} \sqsubseteq_{\mathfrak{C}} \biguplus_{\mathfrak{C}} \mathsf{Granules}\langle\mathfrak{C}|\mathsf{County}\rangle).$$

This rule is not even unresolved minimal; there are only two counties with which Gran Puerto Montt is not disjoint. Thus, resolved minimality must be asserted explicitly for a rule such as (r-Llp) of Sec. 1.

Definition 3.7 (Resolved-minimal join rules). For any $\varphi \in \mathsf{JRules}\langle\mathfrak{S}\rangle$, define $\mathsf{RMinSet}\langle\varphi\rangle = \mathsf{Not}\langle\mathsf{RedSet}\langle\varphi\rangle\rangle$, and define the *resolved minimization* of φ to be $\mathsf{ResMin}\langle\varphi\rangle = \mathsf{Conjuncts}_{\mathfrak{S}}\langle\varphi\rangle \cup \mathsf{RMinSet}\langle\varphi\rangle$. In light of Observation 3.4(b), $\mathsf{RMinSet}\langle\varphi\rangle$ consists of exactly those constraints necessary to make φ a resolved minimal join rule. For φ set to (r-Llp) of Sec. 1,

$$\begin{aligned}\mathsf{ResMin}\langle\varphi\rangle = \{&\neg(\mathit{Gran\_Puerto\_Montt\_urb} \sqsubseteq_{\mathfrak{C}} \mathit{Puerto\_Montt\_cmn}),\\ &\neg(\mathit{Gran\_Puerto\_Montt\_urb} \sqsubseteq_{\mathfrak{C}} \mathit{Puerto\_Varas\_cmn})\}\end{aligned}$$

Just as the basic join symbol $\bigsqcup_{\mathfrak{S}}$ is embellished with $\bot$ to yield $\biguplus_{\mathfrak{S}}$ to indicate disjoint join, it is also useful to embellish the symbol to indicate resolved minimal joins. More precisely, for any type of join rule φ identified in Notation 3.2, replacing $\bigsqcup_{\mathfrak{S}}$ by $\overset{\text{rmin}}{\bigsqcup}_{\mathfrak{S}}$, or $\biguplus_{\mathfrak{S}}$ by $\overset{\text{rmin}}{\biguplus}_{\mathfrak{S}}$, denotes its resolved minimization. For this paper, the concrete case of interest is the *resolved-minimal disjoint subsumption join rule* $(g \sqsubseteq_{\mathfrak{S}} \overset{\text{rmin}}{\biguplus}_{\mathfrak{S}} S)$, shorthand for $\mathsf{Conjuncts}_{\mathfrak{S}}\langle(g \sqsubseteq_{\mathfrak{S}} \biguplus_{\mathfrak{S}} S)\rangle \cup \mathsf{RMinSet}\langle(g \sqsubseteq_{\mathfrak{S}} \biguplus_{\mathfrak{S}} S)\rangle$. Formally, the *resolved-minimal disjoint equality join rule* $(g = \overset{\text{rmin}}{\biguplus}_{\mathfrak{S}} S)$, shorthand for $\mathsf{Conjuncts}_{\mathfrak{S}}\langle(g = \biguplus_{\mathfrak{S}} S)\rangle \cup \mathsf{RMinSet}\langle(g = \biguplus_{\mathfrak{S}} S)\rangle$, is also used, but in view of Proposition 3.5, every disjoint equality join rule is resolved minimal, so the property is redundant. The set of all rules which are of one of these resolved forms is called the *resolved minimal join rules*, denoted $\mathsf{RMJRules}\langle\mathfrak{S}\rangle$.

$\varphi \in \mathsf{RMJRules}\langle\mathfrak{S}\rangle$ has $\mathsf{JoinOp}\langle\varphi\rangle \in \{\overset{\mathsf{rmin}}{\bigsqcup}_{\mathfrak{S}}, \overset{\mathsf{rmin}}{\dot{\bigsqcup}}_{\mathfrak{S}}\}$ but is otherwise syntactically identical to a rule in $\mathsf{JRules}\langle\mathfrak{S}\rangle$. As a concrete example, to express that it is resolved minimal, (r-Llp) may be rewritten as

$$\mathit{Gran\_Puerto\_Montt\_urb} \sqsubseteq_{\mathfrak{C}} \overset{\mathsf{rmin}}{\dot{\bigsqcup}}_{\mathfrak{C}} \{\mathit{Puerto\_Montt\_cmn}, \mathit{Puerto\_Varas\_cmn}\} \quad \text{(r-Llp}')$$

4 Bigranular Join Rules and Their Representation

In this section, the main results of the paper, on the implicit representation of multigranular join rules, are developed.

Definition 4.1 (Granularity pairs). A *granularity pair* over $\mathfrak{S}$ is an ordered pair $\langle G_1, G_2\rangle \in \mathsf{Glty}\langle\mathfrak{S}\rangle \times \mathsf{Glty}\langle\mathfrak{S}\rangle$ with $G_1 \neq G_2$.

Context 4.2 (Granularity names and granularity pairs). For the remainder of this section, unless stated specifically to the contrary, let $G_1, G_2, G_3 \in \mathsf{Glty}\langle\mathfrak{S}\rangle$. In particular, $\langle G_1, G_2\rangle$ and $\langle G_2, G_3\rangle$ are granularity pairs.

Definition 4.3 (Join-order properties of granularity pairs). The notions of equality-join order and subsumption-join order, introduced informally in Sec. 1, are formalized as follows.

(ej-ord) $\langle G_1, G_2\rangle$ has the *equality-join order property*, written $G_1 \trianglelefteq_{\mathfrak{S}} G_2$, if
$(\forall g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle)(\exists S \subseteq_f \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle)$
$(\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_2 = \bigsqcup_{\mathfrak{S}} S))$.

(sj-ord) $\langle G_1, G_2\rangle$ has the *subsumption-join order property*, written $G_1 \otimes_{\mathfrak{S}} G_2$, if
$(\forall g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle)(\exists S \subseteq_f \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle)$
$(\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_2 \sqsubseteq_{\mathfrak{S}} \overset{\mathsf{rmin}}{\bigsqcup}_{\mathfrak{S}} S))$.

While the join in these rules is not explicitly disjoint, in applications to bigranular rules (Definition 4.6), it will always be disjoint (Proposition 4.7).

Observation 4.4 (Equality join implies subsumption join). *If $G_1 \trianglelefteq_{\mathfrak{S}} G_2$ holds, then so too does $G_1 \otimes_{\mathfrak{S}} G_2$.*

Proof. Equality is a special case of subsumption, and equality join is always minimal (Proposition 3.5). □

Definition 4.5 (Biresolvability and equiresolvability). In order to characterize these order properties in terms of simpler ones, several new notions are essential. Local resolvability (for disjointness, subsumption, or both) characterizes resolvability at a fixed $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$, while full resolvability characterizes the corresponding property for all such g_2. Formally, given $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$, the pair $\langle G_1, G_2\rangle$ is *locally disjointness resolvable* (resp. *locally subsumption resolvable*) at g_2 if for every $g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$, $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models^{\pm}_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \bot_{\mathfrak{S}})$ (resp. $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models^{\pm}_{\mathfrak{S}} (g_1 \sqsubseteq_{\mathfrak{S}} g_2)$). If $\langle G_1, G_2\rangle$ is locally disjointness resolvable (resp. locally subsumption resolvable) for every $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$, then it is called *fully disjointness resolvable* (resp. *fully subsumption resolvable*). Call $\langle G_1, G_2\rangle$ *locally biresolvable* at g_2 (resp. *fully biresolvable*) if it is both locally disjointness resolvable and locally subsumption resolvable at g_2 (resp. both fully disjointness resolvable and fully subsumption resolvable).

The pair $\langle G_1, G_2\rangle$ is equiresolvable if subsumption and nondisjointness resolve equivalently. More formally, $\langle G_1, G_2\rangle$ is *equiresolvable* at g_2 if, for every $g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$, $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_1 \sqsubseteq_{\mathfrak{S}} g_2)$ holds iff $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} \neq \bot_{\mathfrak{S}})$ holds; and $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_1 \not\sqsubseteq_{\mathfrak{S}} g_2)$ holds iff $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \bot_{\mathfrak{S}})$ holds. Call $\langle G_1, G_2\rangle$ *fully equiresolvable* if it is equiresolvable at each $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$.

Definition 4.6 (Bigranular join rules). A join rule φ is of type $\langle G_1, G_2\rangle$ if $\mathsf{Head}\langle\varphi\rangle \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$ and $\mathsf{Body}\langle\varphi\rangle \subseteq \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$. Such a rule is also called *bigranular*.

Proposition 4.7 (Bigranular implies disjoint). *If a join rule φ is bigranular, then it is disjoint; i.e.,* $\mathsf{JoinOp}\langle\varphi\rangle \in \{\overset{\bot}{\bigsqcup}_{\mathfrak{S}}, \overset{\mathsf{rmin}}{\overset{\bot}{\bigsqcup}}_{\mathfrak{S}}\}$.

Proof. Distinct granules of the same granularity are disjoint; in particular, the granules of $\mathsf{Body}\langle\varphi\rangle$ have that property. □

The main characterization result for resolved minimality, in its most general form, is presented next.

Proposition 4.8 (Characterization of resolved minimality). *Let φ be a minimal join rule of type $\langle G_1, G_2\rangle$ with the property that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \varphi$. The following three conditions are then equivalent.*

(a) *$\langle G_1, G_2\rangle$ is locally disjointness resolvable at* $\mathsf{Head}\langle\varphi\rangle$.

(b) *φ is resolved minimal.*

(c) $\mathsf{Body}\langle\varphi\rangle =$
$$\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})\}.$$

Proof. (a) ⇒ (c): Regardless of whether or not (a) holds,
$\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})\} \subseteq \mathsf{Body}\langle\varphi\rangle$, since distinct elements of $\mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$ must be disjoint. If (a) holds, then every $g_1' \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \setminus$
$$\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})\}$$
must have the property that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1', \mathsf{Head}\langle\varphi\rangle\} = \bot_{\mathfrak{S}})$, by the very definition of local disjoint resolvability. Clearly, such a granule is not needed in $\mathsf{Body}\langle\varphi\rangle$. Hence (c) holds.

(c) ⇒ (b): Assume that (c) holds. For any $g_1' \in \mathsf{Body}\langle\varphi\rangle$, it is clear that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \neg\mathsf{PrReduct}\langle\varphi : \{g_1'\}\rangle$, since there is no way that $(\mathsf{Head}\langle\varphi\rangle \sqsubseteq_{\mathfrak{S}} \mathsf{Body}\langle\varphi\rangle \setminus \{g_1'\})$ can hold, owing to the disjointness of distinct granules of G_1. Hence φ is resolved minimal.

(b) ⇒ (a): Assume that φ is resolved minimal. Then for any $g_1' \in \mathsf{Body}\langle\varphi\rangle$, $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models \neg(\mathsf{PrReduct}\langle\varphi : \{g_1'\}\rangle)$. Since distinct granules of G_1 are disjoint, this implies that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1', \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})$. On the other hand, let $g_1'' \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \setminus \mathsf{Body}\langle\varphi\rangle$. If $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \not\models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1'', \mathsf{Head}\langle\varphi\rangle\} = \bot_{\mathfrak{S}})$, then there must be a model σ of $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle$ for which $\sigma \in \mathsf{Models}_{\mathfrak{S}}\langle(\bigsqcap_{\mathfrak{S}}\{g_1'', \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})\rangle$ also. In that case, owing to the disjointness of distinct granules of G_1, it would necessarily be the case that $g_1'' \in \mathsf{Body}\langle\varphi\rangle$, a contradiction. Hence it must be the case that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1'', \mathsf{Head}\langle\varphi\rangle\} = \bot_{\mathfrak{S}})$, and so $\langle G_1, G_2\rangle$ is locally disjointness resolvable at $\mathsf{Head}\langle\varphi\rangle$, as required. □

The above result provides in particular a succinct characterization of the subsumption join order $\circledcirc$ in terms of subsumption join rules. Notice that, in contrast to the case for $\trianglelefteq$, resolved minimality must be asserted explicitly.

Theorem 4.9 (Characterization of subsumption join order). *Let $\langle G_1, G_2\rangle$ be a granularity pair. The following conditions are equivalent.*

(a) $G_1 \circledcirc_{\mathfrak{S}} G_2$.

(b) *For each $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$,*
$g_2 \sqsubseteq_{\mathfrak{S}} \overset{\text{rmin}}{\bigsqcup}_{\mathfrak{S}}\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} \neq \bot_{\mathfrak{S}})\}$,
and this is the only possibility for a resolved minimal rule φ with
$\mathsf{Head}\langle\varphi\rangle = g_2$ *and* $\mathsf{Body}\langle\varphi\rangle \subseteq \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$.

Furthermore, if either (a) *or* (b) *holds, then $\langle G_1, G_2\rangle$ is both fully biresolvable and fully equiresolvable.*

Proof. Follows directly from Proposition 4.8 using Definition 4.3(sj-ord). □

For the special case of equality join, the results of Proposition 4.8 may be refined as follows, establishing resolved minimality, local biresolvability and equiresolvability, as well as characterization of the body in terms of both subsumption and nondisjointness.

Proposition 4.10 (Resolved minimality for equality join). *Let φ be an equality-join rule of type $\langle G_1, G_2\rangle$ with the property that $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} \varphi$. The following properties then hold.*

(a) *φ is resolved minimal.*

(b) *$\langle G_1, G_2\rangle$ is locally biresolvable as well as locally equiresolvable at $\mathsf{Head}\langle\varphi\rangle$.*

(c) $\mathsf{Body}\langle\varphi\rangle = \{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_1 \sqsubseteq_{\mathfrak{S}} \mathsf{Head}\langle\varphi\rangle)\}$
$= \{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\bigsqcap_{\mathfrak{S}}\{g_1, \mathsf{Head}\langle\varphi\rangle\} \neq \bot_{\mathfrak{S}})\}$.

Proof. Part (a) follows immediately from Proposition 4.7, Proposition 3.5, and Proposition 4.8(b), whereupon the equality of the first and third expressions of (c) follows from Proposition 4.8(c). To complete the proof, it suffices to note that, by the very definition of disjoint-join equality rule (Summary 2.5(cjrule-iii)), $(g \sqsubseteq_{\mathfrak{S}} \mathsf{Head}\langle\varphi\rangle)$ for every $g \in \mathsf{Body}\langle\varphi\rangle$. Since granules of G_1 are pairwise disjoint, and since $\mathsf{Head}\langle\varphi\rangle = \bigsqcup_{\mathfrak{S}}\mathsf{Body}\langle\varphi\rangle$, is follows that no granule $g \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \setminus \mathsf{Body}\langle\varphi\rangle$ can have the property that $(g \sqsubseteq_{\mathfrak{S}} \mathsf{Head}\langle\varphi\rangle)$. Hence, the remaining equality of (c) holds, from which (b) then follows directly. □

A characterization of equality join order $\trianglelefteq$, similar to that of Theorem 4.9 but expanded to include subsumption, may now be established.

Theorem 4.11 (Characterization of equality-join order). *Let $\langle G_1, G_2\rangle$ be a granularity pair. The following conditions are equivalent.*

(a) $G_1 \trianglelefteq_{\mathfrak{S}} G_2$.

(b) *For each* $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$,

$$g_2 = \overset{\mathsf{rmin}}{\biguplus}_{\mathfrak{S}}\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_1 \sqsubseteq_{\mathfrak{S}} g_2)\}$$
$$= \overset{\mathsf{rmin}}{\biguplus}_{\mathfrak{S}}\{g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\textstyle\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} \neq \bot_{\mathfrak{S}})\},$$

and this is the only possibility for a minimal rule φ with
$\mathsf{Head}\langle\varphi\rangle = g_2$ *and* $\mathsf{Body}\langle\varphi\rangle \subseteq \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$.

Furthermore, if either (a) *or* (b) *holds, then $\langle G_1, G_2\rangle$ is both fully biresolvable and fully equiresolvable.*

Proof. Follows directly from Proposition 4.10 using Definition 4.3(ej-ord). □

Discussion 4.12 (Consequences of the characterizations). The main thrust of the results developed so far in this section is that even though there may be many granule structures which are models for the constraints associated with $G_1 \otimes_{\mathfrak{S}} G_2$ and $G_1 \trianglelefteq_{\mathfrak{S}} G_2$, all of these models agree on which granules of G_1 are and are not disjoint from granules of G_2. Furthermore, this disjointness information is sufficient to recover completely the join rules. This information is represented via the relation nondisjointness relation $\mathsf{NRel}_{\mathfrak{S}:\langle -,-\rangle}$, as introduced in Sec. 1. The corresponding relation $\mathsf{SRel}_{\mathfrak{S}:\langle -,-\rangle}$ for subsumption is similarly used, as its special properties will prove to be useful in the representation of rules associated with $\trianglelefteq_{\mathfrak{S}}$. The formalization of these ideas are found in Definition 4.13 and Theorem 4.14 below.

Definition 4.13 (The fundamental relations of a granularity pair). Define the *nondisjointness relation* for $\langle G_1, G_2\rangle$ as

$$\mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle} = \{\langle g_1, g_2\rangle \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \times \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (\textstyle\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} \neq \bot_{\mathfrak{S}})\}.$$

Similarly, define the *subsumption relation* for $\langle G_1, G_2\rangle$ as

$$\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle} = \{\langle g_1, g_2\rangle \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \times \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle \mid \mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_1 \sqsubseteq_{\mathfrak{S}} g_2)\}.$$

Note that if $\langle G_1, G_2\rangle$ is fully equiresolvable (Definition 4.5), in particular if $G_1 \trianglelefteq_{\mathfrak{S}} G_2$ (Theorem 4.11), then $\mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle} = \mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$.

The main theorem for implicit representation is the following.

Theorem 4.14 (Representation of bigranular join rules using fundamental relations).

(a) *If $G_1 \otimes_{\mathfrak{S}} G_2$ holds, then for every $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$ and every $S \subseteq_f \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$,*

$$\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_2 \sqsubseteq_{\mathfrak{S}} \textstyle\bigsqcup_{\mathfrak{S}} S) \textit{ iff } \{g_1 \mid \langle g_1, g_2\rangle \in \mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}\} \subseteq S.$$

In particular,

$$\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_2 \sqsubseteq_{\mathfrak{S}} \overset{\mathsf{rmin}}{\biguplus}_{\mathfrak{S}} S) \textit{ iff } S = \{g_1 \mid \langle g_1, g_2\rangle \in \mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}\}.$$

(b) *If $G_1 \trianglelefteq_{\mathfrak{S}} G_2$ holds, then for every $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$ and every $S \subseteq_f \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$, $\mathsf{Constr}^{\pm}\langle\mathfrak{S}\rangle \models_{\mathfrak{S}} (g_2 = \bigsqcup_{\mathfrak{S}} S)$ iff*

$$S = \{g_1 \mid \langle g_1, g_2\rangle \in \mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}\} = \{g_1 \mid \langle g_1, g_2\rangle \in \mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}\}.$$

Proof. The proof follows immediately from Theorem 4.9 and Theorem 4.11. □

Discussion 4.15 (Equality-join order is transitive). It is easy to see that the equality-join order relation is transitive. More precisely, if $G_1 \trianglelefteq_{\mathfrak{S}} G_2$ and $G_2 \trianglelefteq_{\mathfrak{S}} G_3$ both hold, then so too does $G_1 \trianglelefteq_{\mathfrak{S}} G_3$. This follows immediately from the first equality of Theorem 4.11(b) and the fact that the subsumption relation $\sqsubseteq_{\mathfrak{S}}$ is transitive. To illustrate the utility of this observation via example, referring to the hierarchy to the left in Fig. 2, since both $\mathsf{Province} \trianglelefteq_{\mathfrak{C}} \mathsf{Region}$ and $\mathsf{County} \trianglelefteq_{\mathfrak{C}} \mathsf{Province}$, it is also the case that $\mathsf{County} \trianglelefteq_{\mathfrak{C}} \mathsf{Region}$, and, furthermore,

$$\mathsf{SRel}_{\mathfrak{C}:\langle \mathsf{County},\mathsf{Region}\rangle} = \mathsf{SRel}_{\mathfrak{C}:\langle \mathsf{County},\mathsf{Province}\rangle} \circ \mathsf{SRel}_{\mathfrak{C}:\langle \mathsf{Province},\mathsf{Region}\rangle},$$

with $\circ$ denoting relational composition. Thus, it is not necessary to represent all pair of the form $G_i \trianglelefteq_{\mathfrak{S}} G_j$, but rather only a base set, from which the others may be obtained via transitivity. In both diagrams of Fig. 2, the edges labelled with $\trianglelefteq$ identify such base sets.

This transitivity property is not shared by the subsumption-join order relation $\otimes_{\mathfrak{S}}$, as is easily verified by example.

Discussion 4.16 (Implementation of bigranular constraints via implicit representation). A PostgreSQL-based system, providing multigranular features, is under development at the University of Concepción. Called MGDB, it is based upon the theory of [8], employing further the ideas elaborated in this paper. MGDB supports neither detailed spatial models (based upon regions in $\mathbb{R}^2$) nor the detailed spatial operations described in [4]. Rather, it is a relational extension which supports multigranular attributes. A main feature is support for basic spatial relationships, such as nondisjointness, subsumption, and join, without the need for an elaborate $\mathbb{R}^2$ model. A second feature is that spatial and temporal attributes are both recaptured using the same underlying formalism.

Currently, MGDB is implemented via additional relations on top of an ordinary relational schema. Thus, each multigranular attribute $\mathfrak{S}$ is represented as an ordinary attribute, together with additional relations which recapture its special properties. In particular, for each such attribute and each granularity pair $\langle G_1, G_2\rangle$, the relations $\mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ and $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ are stored, either fundamentally or as views (see below for more detail), to the extent that the associated information is known. In addition, there is a special ternary relation $\mathsf{GrPrProp}_{\mathfrak{S}}$, with a tuple of this relation of the form $\langle G_1, G_2, c\rangle$, with c a code which identifies the relationship between the granularities G_1 and G_2. The code may represent combinations of $G_1 \leq_{\mathfrak{S}} G_2$, $G_1 \trianglelefteq_{\mathfrak{S}} G_2$, and $G_1 \otimes_{\mathfrak{S}} G_2$, as well as other relationships not covered in this paper. Given a granule $g_2 \in \mathsf{Granules}\langle \mathfrak{S}|G_2\rangle$, and a request to determine which granules of G_1 are related to it via a join rule which is a consequence of a bigranular property, it is only necessary to look in $\mathsf{GrPrProp}_{\mathfrak{S}}$ to determine the type of join rule (e.g., equality or subsumption), and then to determine the body via a lookup, in $\mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$, which granules of G_1 form the body of that rule. Since the rules are recovered via retrieval of the appropriate tuples in these relations, and not directly as formulas, the representation is termed *implicit*.

For economy, some of the relations of the form $\mathsf{DRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ and $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ are implemented as views. For example, if either of $G_1 \leq_{\mathfrak{S}} G_2$ or $G_1 \trianglelefteq_{\mathfrak{S}} G_2$ holds, then $\mathsf{DRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ and $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle}$ are the same relation, so only one need be stored explicitly. Likewise, $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_3\rangle} = \mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_2\rangle} \circ \mathsf{SRel}_{\mathfrak{S}:\langle G_2,G_3\rangle}$ if either of $G_1 \leq_{\mathfrak{S}} G_2 \leq_{\mathfrak{S}} G_3$ or $G_1 \trianglelefteq_{\mathfrak{S}} G_2 \trianglelefteq_{\mathfrak{S}} G_3$ holds, so $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_3\rangle}$ may then be represented as a view defined by relational join. This means that relationships such as equality join, as sketched in Discussion 4.15, require virtually no additional storage for representation. While a tuple of the form $\langle G_1, G_3, c\rangle$ must be present in $\mathsf{GrPrProp}_{\mathfrak{S}}$, no additional space is required to represent $\mathsf{SRel}_{\mathfrak{S}:\langle G_1,G_3\rangle}$ or $\mathsf{NRel}_{\mathfrak{S}:\langle G_1,G_3\rangle}$.

A substantial superset of the hierarchies shown in Fig. 2, including electoral as well as administrative subdivisions of Chile in the spatial case, forms the core of the test database. All such data are obtained from publicly available sources. This spatial hierarchy is very rich in granularity pairs related by $\trianglelefteq_{\mathfrak{C}}$ and $\circledcirc_{\mathfrak{C}}$. Time intervals, as illustrated in the rightmost hierarchy of Fig. 2, form part of the test database as well. The system will be discussed in more detail in a future paper.

Discussion 4.17 (Relationship to other work). An extensive literature comparison for the general multigranular framework used in this paper may be found in [8, Sec. 6]. Only literature relevant to the topics of this paper which are not developed in [8] are noted here. A fairly extensive presentation of granular relationships may be found in [1], including in particular the equality join relation $\trianglelefteq$, there called *groups into*, as well as the combination of ordinary granularity order $\leq$ and equality join $\trianglelefteq$, there called *partitions*. It does not cover the subsumption join relation $\circledcirc$. Although [1] is specifically about the time domain, many of the concepts presented there apply equally well to spatial and other domains. This is reinforced not only by the work of this paper, but also by papers such as [2] and [10], which apply the concepts of [1] to the spatial domain. In addition, [12] provides a development of the equality-join operator $\trianglelefteq$ for the spatial domain, there denoted $\models$. Reference [5] provides further insights into the multigranular framework within the context of time granularity.

5 Conclusions and Further Directions

A method for representing bigranular join rules implicitly in a multigranular relational DBMS has been developed. As such rules occur frequently in practice, the technique promises to prove central to an implementation. Indeed, they have already been used in an early implementation of the system MGDB.

There are two main avenues for future work. First, the main reason that the techniques of this paper were developed is that direct implementation of join rules proved too inefficient in practice. While most rules are bigranular, there are often some which are not. One topic of future work is to find a way to integrate the methods of this paper with representation of non-bigranular rules, in a way which preserves the efficacy of the implementation. A second and very major topic is to extend MGDB with its own query language and interface. Currently, MGDB is a testbed for ideas, but to be useful as a stand-alone system, it must be augmented to have its own query language and interface, so that the implementation of the multigranular features is transparent to the user.

Acknowledgment: The work of M. Andrea Rodríguez, as well as three visits of Stephen J. Hegner to Concepción, during which many of the ideas reported here were developed, were funded in part by Fondecyt-Conicyt grant number 1170497.

References

1. C. Bettini, C. E. Dyreson, W. S. Evans, R. T. Snodgrass, and X. S. Wang. A glossary of time granularity concepts. In O. Etzion, S. Jajodia, and S. Sripada, editors, *Temporal Databases, Research and Practice*, volume 1399 of *Lecture Notes in Computer Science*, pages 406–413. Springer, 1998.
2. E. Camossi, M. Bertolotto, and E. Bertino. A multigranular object-oriented framework supporting spatio-temporal granularity conversions. *International Journal of Geographical Information Science*, 20(5):511–534, 2006.
3. B. A. Davey and H. A. Priestly. *Introduction to Lattices and Order.* Cambridge University Press, second edition, 2002.
4. M. J. Egenhofer. Deriving the composition of binary topological relations. *J. Vis. Lang. Comput.*, 5(2):133–149, 1994.
5. J. Euzenat and A. Montanari. Time granularity. In M. Fisher, D. M. Gabbay, and L. Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence*, volume 1, pages 59–118. Elsevier, 2005.
6. R. Fagin. Horn clauses and database dependencies. *J. Assoc. Comp. Mach.*, 29(4):952–985, 1982.
7. S. J. Hegner and M. A. Rodríguez. Integration integrity for multigranular data. In J. Pokorny, M. Ivanovic, B. Thalheim, and P. Saloun, editors, *Advances in Databases and Information Systems, 20th East European Conference, ADBIS 2016, Prague, Czech Republic, August 29-31, 2016, Proceedings*, volume 9809 of *Lecture Notes in Computer Science*, pages 226–242. Springer, 2016.
8. S. J. Hegner and M. A. Rodríguez. A model for multigranular data and its integrity. *Informatica, Lith. Acad. Sci.*, 28:45–78, 2017. Open access at `https://www.mii.lt/informatica/htm/INFO1132.htm`.
9. M. Kifer, A. Bernstein, and P. M. Lewis. *Database Systems: An Application-Oriented Approach.* Addison-Wesley, second edition, 2006.
10. M. A. Mach and M. L. Owoc. Knowledge granularity and representation of knowledge: Towards knowledge grid. In Z. Shi, S. Vadera, A. Aamodt, and D. B. Leake, editors, *Intelligent Information Processing V - 6th IFIP TC 12 International Conference, IIP 2010, Manchester, UK, October 13-16, 2010. Proceedings*, volume 340 of *IFIP Advances in Information and Communication Technology*, pages 251–258. Springer, 2010.
11. J. D. Monk. *Mathematical Logic.* Springer-Verlag, 1976.
12. S. Wang and D. Liu. Spatio-temporal database with multi-granularities. In *Advances in Web-Age Information Management WAIM*, volume 3129 of *Lecture Notes in Computer Science*, pages 137–146. Springer, 2004.