April 2023

An analysis of positionalism's roles in use

C. Maria KEET [1]
University of Cape Town, South Africa
ORCiD ID: C. Maria Keet https://orcid.org/0000-0002-8281-0853

Abstract. Roles as argument places in a positionalist account of relations are pervasive in conceptual data modelling and linguistics, known also as components of a relationship and as semantic, thematic, or deep roles as parts of a verb or verb class, respectively. They are also planned to be used in Abstract Wikipedia that seeks to combine them. There is, however, no insight in systematic or ontologically sound usage of such roles, in contradistinction to the ample attention given to aligning classes to nouns and relationships to verbs. Roles, as identifiable argument places, may benefit from similar efforts toward an ontology of roles. We aim to take a first step in that direction in a two-pronged approach. First, we conducted an analysis of a set of 101 conceptual data models on their use of roles. Second, we analysed VerbNet, an authoritative linguistic knowledge base on thematic roles. The results show promise for improvements of naming roles in conceptual data models. VerbNet's roles are challenging to align to an ontology due to its mixing of the ontological and linguistic layers and flexibility of natural language. The insights obtained also indicate ample avenues for further research.

Keywords. Roles, Positionalism, Ontology-driven conceptual data modelling, VerbNet

1. Introduction

The term 'role' and what it refers to mean different things in different specialist communities. It may refer to the social roles people play [1,2] (e.g., the role of student), roles in the context of an ontology of relations [3,4,5] as an identifiable element that forms part of the fundamental furniture of the universe, the linguistic viewpoint with semantic or thematic roles like in VerbNet (VN) [6], roles as part of fact types in Object-Role Modeling [7] in the context of conceptual data modelling, and roles in Description Logics where they are n-ary predicates where $n \geq 2$ [8]. The ontology of relations may admit roles [4,9], also called positions or argument places, as part of a relation, i.e., *positionalism*, which some dismiss for antipositionalism to avoid roles [3,5]; see Fig. 1 for an example. Either way, such roles of relationships are very useful for conceptual data modelling and linguistics and natural language processing. The former uses roles mainly to attach constraints to them to enhance data integrity in databases and software applications, and the latter uses it for both the theoretical understanding of verbs and for parsing text. This paper considers those roles of relationships or verbs only. Consider, e.g., the verb knead in the VN verb class knead-26.5 that may link to a food processing ontology. VN's sam-

[1]Corresponding Author: C. Maria Keet, Department of Computer Science, University of Cape Town, South Africa. mkeet@cs.uct.ac.za

Figure 1. Left: Object-Role Modeling diagram about animals (Q729 in Wikidata) being endemic to a region (Q82794), where "... is endemic to ... / ... is home to ... " is a reading label in both directions, originate could be the name of the relationship or Wikidata's P183 binary property (not shown in the diagram), and [inhabitant] and [location] are the roles as part of that relationship (respectively possibly Q22947 and Q17334923 in Wikidata). Right: typing of the relationship and mandatory participation of Animal, in $\mathcal{DLR}$ notation. The verb 'originate' is a member of the verb class establish-55.5-1 in VerbNet that has thematic roles *agent* and *theme*.

ple sentences include "I kneaded the dough into a loaf" and "I kneaded the dough", with suitable referents for the terms in VN: there is the role of *agent* that does the kneading that is played by (or: the slot is filled by) "I", there is a *material* role played by an amount of matter, "the dough", and there may be a concrete object (here: the "a loaf") coming out of the process referred to by that verb, which plays the *product* role[2]. In this setting, from the perspective of ontological analysis, we should be able to state something about 1) the relation that the verb refers to, 2) the roles that participate in the relation, 3) the role fillers, i.e., the kind of things that can play that role, and 4) how the former three interact with each other. The latter can be addressed by the $\mathcal{DLR}$ description logic [10] or the positionalist $\mathcal{DC}_p$ [11], so that a formalisation of knead-26.5 is as follows:

$\mathsf{knead} \sqsubseteq \mathsf{agent} : (\mathsf{Animate} \sqcup \mathsf{Machine}) \sqcap \mathsf{material} : \top \sqcap \mathsf{product} : \top$

for typing the relation and, for the context of kneading specifically:

$(\mathsf{Animate} \sqcup \mathsf{Machine}) \sqsubseteq \exists[\mathsf{agent}]\mathsf{knead}$ and $\top \sqsubseteq \exists[\mathsf{material}]\mathsf{knead}$

to state that the agent and material roles are mandatory, requiring some object to play the role. This does not provide an ontology of roles and it glosses over the fact that the Animacy is a semantic (in linguistics) and therewith grammatical feature of a natural language and Machine refers to physical objects in reality, i.e., that the restriction on what can play the agent role is an ontologically incoherent union. This begs the question what the role fillers may be restricted to, as well as what to do with a domain entity like Machine that is not present in a foundational ontology that otherwise might guide making the role fillers more precise. For the roles, questions include whether they are the right ones for the relation that knead refers to and, more generally, whether relations or their verb classes can be defined by their roles and relata.

To focus answering such questions, we will use as motivational use case an ontology-driven information system for the prospective "abstract representation" language for the "constructors" of Abstract Wikipedia [12,13], which are *de facto* small conceptual data model fragments that are to be verbalised into sentences that are to be put together to automatically create Wikipedia articles from the structured data stored in Wikidata as input. Its developers agree roles are essential for the "abstract representation", but the what and how are unclear[3]. This use case entails the need for a praxis-informed analysis to help gauging user needs for improving modelling practice with ontological principles. We therefore take a bottom-up approach as first step and focus on two main directions to answer:

[2] `https://uvi.colorado.edu/verbnet/knead-26.5`

[3] `https://meta.wikimedia.org/wiki/Abstract_Wikipedia/Wikidata_Abstract_Representation`, and the 'Semantic roles & slots' section specifically (d.d. 29-11-2022).

1. How are roles used in conceptual data models? Are they named and if so, how, and do they map usefully into semantic roles as specified in linguistic resources or ontologies? Can this *modus operandi* be copied over to Abstract Wikipedia's abstract representation?
2. To what extent do those verb classes with their roles and fillers in the authoritative linguistic resource VN adhere to ontological principles? Can that be improved upon further, using basic modelling guidance from ontology development and without the need for major theoretical 'overhead' for an end user writing Abstract Wikipedia's constructors?

The analysis of roles in 101 conceptual data models showed that about half of the roles are named, with the most roles named in UML class diagrams compared to EER and ORM. When named, they mostly are of the type of 'deep', or subject domain-specific, roles (also called ontological roles). This may possibly be improved upon with hints from VN's roles. The VN roles, however, required improvement in the role hierarchy and greater specificity in the ontological categories for role fillers, for which the DOLCE foundational ontology [2] was used. These insights, in turn, may inform the development of an ontology of positions.

The remainder of this paper is structured as follows. We first discuss related work in Section 2. The conceptual model analysis is described in Section 3 and VN in Section 4. We discuss in Section 5 and conclude in Section 6.

2. Related work

Roles, or argument positions as part of a relation(ship), are present in all major conceptual data modelling languages [14] but not in most logics, with a few exceptions [11,15], notably the $\mathcal{DLR}$ family of description logic languages [10] with its four extensions where $\mathcal{DLR}_{\mathsf{ifd}}$ fits best with EER it was proposed for, but also with UML class diagrams [16] and a fragment of ORM [17], and the $\mathcal{DC}_p$ core profile for all three [11]. They are the only logics that were designed explicitly for providing a logic-based reconstruction of conceptual data models. While it shows that such a logic can be defined, it does not provide insight into roles and, to the best of our knowledge, no ontological analysis of such roles in extant conceptual data models has been done, nor have the linguistic resources for roles been used to assist in conceptual data modelling. Philosophical inquiry into roles mainly focuses on the argument whether positionalism is justified or not, notably in [3,4,18,5,9], which is a foregone conclusion in praxis. When in favour of roles, they zoom in on aspects that do not help their ontologically sound usage. For instance, while Orilia's "onto-thematic roles" are noteworthy, he tries to impose an ordering [9]—precisely that what conceptual modellers and linguists want to avoid with roles—and though Gilmore's idea for a mereological view on roles may assist toward an ontological foundation to $\mathcal{DLR}$ and $\mathcal{DC}_p$, its treatise is narrowed down to propositions only [18].

The linguistics angle to roles has received ample attention, both regarding distinguishing between semantic, thematic, and deep roles, and which ones there may be. *Semantic macroroles* consist of Actor and Undergoer (sometimes named Agent and Patient, respectively), where the former is assumed usable with all verbs and the latter only with transitive verbs. *Semantic roles* have around 10 types of role specified in a set, refining Undergoer into, say, and Experiencer and a Recipient, which blurs into the set of

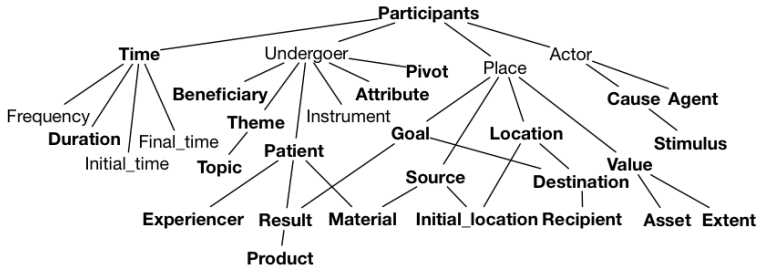

Figure 2. VerbNet's hierarchy of roles (Source: VerbNet guidelines).

thematic roles. A popular effort for thematic roles is VN [6], where they "refer to the semantic relationship between a predicate and its arguments". A key characteristic of VN's roles is that their names are reminiscent of a foundational ontology of sorts; see Fig. 2. The list of roles in VN varies across versions and sources: the general thematic roles[4] lists 30 roles, the hierarchy of [19] has 37, the VN guideline lists 36 and its hierarchy (Fig. 2) includes a subset (n=34). The VN guideline does not have, among others, Trajectory and Co-Patient. If one were to declare disjointness between siblings, inconsistencies will be obtained due to the multiple inheritance in the hierarchy. Multiple inheritance typically indicates see-sawing between two desiderata for how to structure information and possibly an under-representation of the semantics. A recent logic-based approach for VN [20] did not solve this, since the formalisation concerns an event calculus to describe what is happening during the event referred to by the verb. Last, there are *deep roles*, also called *subject domain roles* or *predicate-specific semantic roles*; e.g., FrameNet (FN) [21] has both thematic and deep roles specific to a verb. For instance, borrowing has as "core" roles Borrower, Lender, and Theme and losing has as core roles Owner and Possession. As with VN, they also may have optional roles to cater for other parts in a sentence; e.g., Means for how something was lost and Place for where it was lost. Like VN's verb classes, it also permits predicate-specific verbs; e.g., 'lose' and 'misplace' for losing.

Several efforts have been devoted to link linguistic resources at the system (schema) and at the content level, including PredicateMatrix [22,23], PreMOn [24] and Framester [25]. PredicateMatrix [22,23] integrates WordNet, FN, VN, and PropBank by building on top of SemLink that consisted of mapping pairs of ⟨lexical entry, semantic class⟩ between VN and PropBank and VN and FN and their roles. Framester was developed in parallel, integrating FN, WordNet, and VN and adding links to BabelNet, DBpedia, Yago, and DOLCE-Zero. The Framester schema has semanticRole[5], and FN is linked to DOLCE-Zero, but not its roles and, hence, not the VN roles indirectly either. In an adjacent effort, Wikidata lists 15 thematic roles[6]. PreMOn, focuses on the ontology for modelling semantic classes and semantic roles by extending the *lemon* model, and it also has the RDF dataset with alignments for PropBank, NomBank, VN, and FN. The ontology is about the kind of things in the resources, such as that it has Restriction and that SemanticRole $\sqsubseteq \exists$thematic role.ThematicRole, rather than aligning the roles or restrictions to an ontology. We could not find articles on semantic or thematic roles from linguistics and either of the established foundational ontologies. The nearest are align-

[4] `http://verbs.colorado.edu/verb-index/vn/reference.php`

[5] `https://github.com/alammehwish/framester/blob/master/schema/framester.owl`

[6] `https://www.wikidata.org/wiki/Special:WhatLinksHere/Q613930`; last counted d.d. 12-1-'23.

ments to WordNet, both for DOLCE [26] and SUMO [27], and VN's latest alignment to WordNet is available from its website. All the mappings concern the verbs, not the roles or selectional restrictions explicitly.

3. Usage of roles in Conceptual Data Models

The aim of the experimental assessment is to ascertain which roles are used by conceptual modellers, of which type, and how. The conceptual data modelling languages (CDMLs), by design, do have roles as components of relationship. This is graphically shown most clearly in ORM diagram with a separate element in addition to optional naming, as shown in Fig. 1. Examples in textbook and language standards show them to be mostly 'deep', or subject domain specific, roles, such as the [inhabitant] in Fig. 1, rather than semantic or thematic roles, except, perhaps, for UML Class Diagram's whole and part roles of the aggregation association (graphically, not named) and, by the same reasoning, the parent and child roles for class subsumption (also only graphically). It is expected that at least a few thematic roles emerge from the data, and insight is expected to be obtained from the naming of deep roles as well as the praxis of how they are used.

Design. We reuse the data set of [28] that contains 101 conceptual data models across three CDML families, being UML Class Diagrams, ER and EER, and ORM and ORM2, because it already contains an analysis of elements present in those models. The following steps were carried out.

1. For each diagram, count the relationships, roles, and which of them are named.
2. Data collection steps for the vocabulary collection and analysis, by CDML, for UML class diagrams, EER, and ORM/ORM2, respectively: record the names of [associations ends/relationship component names/user-declared role names], the [class/entity type/object type] associated with it, and the [association/relationship/fact type] name (if present).
3. Data analysis: count aggregates for roles and relationships and their names and compare across the modelling language families; examine whether there may be any naming patterns between roles, relationships, classes; extract meaningful deep and thematic roles.

Results. The quantitative aggregate data is presented in Table 1 (see also supplementary material sheets 'RolesCalc' and 'relRoleClass'). Naming of association ends, i.e., roles, in UML class diagrams is mandatory, yet only about half are named, and of those 345 that are, at least 136 were named by default by the software and never changed (the GenMyModel models labels them with the participating class name). Discarding those as cases of user-defined naming reduces it to only 28.7% and therewith brings the total down to 13.4% of active role naming. In some diagrams, the layout and name given make it ambiguous which class is participating in the role or whether they were intended as association names (model `REA-UML`) or are named with default naming/numbering of fact types and roles (e.g., R1 and Ft8, in total 32 ORM roles) and are therewith not interesting regarding semantics. Also, there are 6 ORM fact types with the is of/has combination for entity types that relate to a value type, which is also the default option that is typically used as reading label. In the EER diagrams, if roles are named, then they are typically those of recursive relations, intended for disambiguation. For these, as well as other properly named roles, they are generally deep roles, such as upperValue

Table 1. Aggregate data of role and relationship naming in the three main CDML families; Rel. = relationships; Pct. = percentage.

CDML family	$\|Rel.\|$	$\|Rel.\|$ **named**	**Pct. named Rel.**	$\|Role\|$	$\|Role\|$ **named**	**Pct. named Role**	**Both named**
UML	363	79	21.8	728	345	47.4	15
ORM	521	42	8.1	1125	107	9.5	0
EER	307	229	74.6	630	17	2.7	7
Total	*1191*	*350*	*29.4*	*2483*	*469*	*18.9*	*22*

for a ValueSpecification, or have role names that commonly refer to names of relations and are verbs or verb-like, such as contains and worksFor, or are nominalised verbs or roles proper, such as buyer and seller by Client, Physician playing the specialist role in the hasSpeciality relationship, and Detective playing the interrogator role in a conducts relationship. Regarding the recurring roles, there are:

- roles of commonly generic relationships: 6 participant roles, 8 member roles, and there are 100 aggregation associations in the UML diagrams and thus 100 whole roles and at least 100 part roles.
- 6 parent (parent_of etc.) relationships or roles or classes (e.g., recursively on IIoMSCalendar in `uml_calendars_default`, Parent *sensu* person), indicating that parent/child is generic and may be considered thematic rather than deep.
- 4 work[s for/in] roles with role players being persons or an employee. The role worker may be considered a deep role that is specific to humans as employees and their roles in the organisation within that universe of discourse.
- 4 manages that are either a relationship or a role name, and has as participant Manager or Employee (or other type of job) across models. The manager role may also be considered a deep role, for the same reason as the worker role.
- 3 Based-On relations (in the same model) with roles Original and New. One may rather use original and derived, which can apply to many things, like movie remakes, a movie based on a book, derivatives of software code. They may be a candidate for a thematic role, or at least 'highly reusable deep' roles.

These low numbers of commonality, except for the roles of common relations like participation and membership, should be seen in the light that there were 1191 relationships and 2483 roles in the diagrams, and that the aggregates presented combine variant spelling options of the names. There are only 22 relationships where both it and its roles were named, which turned out to be highly specific to the universe of discourse of the conceptual model, such as to and from roles for a call, and detective and investigation roles in the conductedBy relationship.

Observe also the differences between the three language families in Table 1, which exhibit effects of both practices and affordances of the language and of tooling. UML does not have an association naming requirement and consequently few named associations. EER is relationship oriented and the diamond shape invites labelling it and thus has a high percentage of named relationships. Only one diagram with diamonds did not have names for the relationships (`erd1`) and the remaining unnamed ones are mainly due to isolated cases in two models, i.e., mainly two sloppy modellers that did not name any relationship in crows feet notation-based models (`asp-net-membership-er-diagram` and `dnd1996120204`) rather than lapses pervasive throughout all models. ORM tooling

focuses on reading labels and has therefore very few explicit user-defined naming of either fact types or roles. Moreover, popular editors auto-generate unique role and fact type names in the background that are not shown by default in the diagrams whereas reading labels are made compulsory.

Discussion and conclusions. In sum, while roles are present in the CDMLs, and thus an unequivocal commitment to the ontological commitment of positionalism, the practice of actually naming them is limited. Even for mandatory association end naming to obtain only 28.7% of the roles to have been named by a modeller is low and a 2.7% just for disambiguation in EER is negligible. When the roles are named actively by the modeller, they tend to be deep roles rather than thematic. Besides roles for common relations, such as participant, member, part, and whole, for which the taxonomy of part-whole relations [29] may assist in specifying the roles more precisely, the few emerging ones that may possibly be thematic are parent (/child), original and derived, and several deep roles typical for all organisations' information systems, such as manager and worker, whose naming and intent aligns with the notion of social roles in [1] and thus also with such entities in other ontologies, like BFO's role [30].

The data cannot answer *why* comparatively few roles are named by the modeller; e.g., it may be that role naming does not matter in software and database design or that is it too hard to do. For the latter, one may try to embed common roles as autocomplete options in the CDMLs to improve the naming practices and clarify the semantics.

4. Thematic roles from language and linguistics

For examining roles of verbs, the analogue to usage in CDMs is not a set of annotations of text, because the annotations are based on already modelled roles; it is those pre-defined roles for annotation that must be assessed. We do this with VN. Of them, we seek to:

- A: Determine whether there are verb classes (candidate relationships) that have the same combination of roles and participation mode of their role fillers; if none, then that combination makes them definable by their roles.
- B: Examine the thematic roles and their fillers against a foundational ontology, and analyse and classify them accordingly, including any possible updates to the VN role hierarchy.
- C: Compare the naming practices of conceptual data models with their (mostly) deep roles to VN's thematic roles. It is expected that there will be little overlap between them, but they may have subsumption alignments.

VN annotates each verb class with thematic roles that may have "selectional restrictions", i.e., role fillers, such as Animate and Pointy, or their negation, frames with permissible sentence structure and the use of one or more roles and further semantic functions, and members of the verb category (the verbs). To answer Task A, we first examined existing files on which to run the reasoner. They were either instances due to SKOS usage (PreMon), very large Turtle files with all the verbs as RDF data (Framester), or a subset (FRED). Therefore, we encoded the knowledge in OWL to determine equivalences and the hierarchy, as follows. 1) Take an increasing number of verb classes (until equivalence has been detected) and encode their role(s) in an OWL file as object properties and selectional restriction(s) as existentially quantified axiom when all frames use that role or universally quantified axiom when not all frames use that role. 2) Classify the

ontology and record modifications in the taxonomy, including number of equivalences and subsumption within a verb class and across verb classes. Task B will be addressed by aligning it to a foundational ontology that is informed by, among others, linguistics, being DOLCE [31,2], so as to increase the chances of obtaining equivalence or subsumption mappings. We use the latest v3.3 of VN's thematic roles and their descriptions in the guidelines (see fn. 4) of the downloaded version in XML files. The five sources on VN roles have different lists of roles and documentation thereof (see supplementary material, sheet "RolesinVNsources"), leaving 30 roles for assessment with 36 role fillers ("selectional restrictions").

4.1. Results

We first describe the results on verbs and their classes, and then the alignment to a foundational ontology.

4.1.1. Task A: Do the constituent roles define verb classes?

The first five VN verb classes, being numbers 9-13, were encoded in OWL. This resulted in a test ontology of 767 classes, 17 object properties, 1673 axioms, overall remaining within $\mathcal{ALC}$ expressiveness (OWL DL). Of the 767 classes, there are 17 Restriction (i.e., role filler) classes, 39 Thematic roles, and 48 Verb classes; the remainder of the classes are their respective 'members' as subclasses; e.g., Equip-13.4.2 has as VN members (OWL subclasses) charge, invest, and ply. Verb classes were made defined classes; e.g.,

$$\mathsf{Equip\text{-}13.4.2} \equiv \exists \mathsf{agent}.(\mathsf{Animate} \sqcup \mathsf{Organization}) \sqcap \exists \mathsf{recipient}.(\mathsf{Animate} \sqcup \mathsf{Organization}) \sqcap \exists \mathsf{theme}.\top$$

Two versions were created, since the respective XML file of a verb class does not state whether the roles are mandatory: `VN33exist.owl` has all the verb class properties coded as existentially quantified and `VN33.owl` has those properties optional (only universally quantified) if the provided English syntax has a sample pattern that does not have that particular role.

Classifying the ontology, there are only minor differences in the deductions that do not affect the conclusions; therefore, we only report on the `VN33.owl` deductions. There are three sets of inferred equivalences among 9 verb classes, being Deprive-10.6.2 and Cheat-10.6.1 with the same definition, Put-9.1, Put_spatial-9.2, and Funnel-9.3, and Give-13.1, Fulfilling-13.4.1, and Contribute-13.2 and Equip-13.4.2. Put differently, under the assumption that the represented knowledge is exhaustive, then they are ontologically the same if the roles and relata were to suffice to define the meaning of a verb class denoting a relation. But, e.g., depriving and cheating are not synonyms; hence, either the inferred equivalence must be due to underspecification in VN or it might be that a verb class cannot be defined by its roles, role fillers, and other constraints. For this case, Deprive-10.6.2 has an additional $\forall \mathsf{goal}.\top$, which, given the underspecification elsewhere (on roles as object properties), does not make a difference logically. If, say, $\forall \mathsf{goal}.\mathsf{Concrete}$, then $\mathsf{Deprive\text{-}10.6.2} \sqsubseteq \mathsf{Cheat\text{-}10.6.1}$. Thus, resolving Task A: currently, not all verb classes in VN can be uniquely characterised by their role specifications.

In addition, there are 36 inferred subsumptions on the main verb classes. They appear both within a verb class and across verb classes, i.e., a different hierarchy is inferred for the verb classes than declared in VN. For instance, it deduced $\mathsf{Fire10.10} \sqsubseteq \mathsf{Hire13.5.3}$, which is plausible once one realises the same roles and role

fillers must be involved. Since the set of relation instances that are fired must be a subset of those that are hired, it can serve as useful constraint in an ontology or conceptual model, but, intensionally, firing is not a type of hiring. Fire10.10 has a property ∃source.Organization that Hire-13.5.3 does not, but one easily could argue that it should be added to Hire-13.5.3's definition, and therewith leading to equivalence. The two verbs have the opposite meaning, however, which is still elusive in the representation. Considering their respective parent classes, Removing10 and ChangeOfPossession13, firing's semantics may be argued to not be one of removing, but also a change of possession, in that in the firing act, the organisation does not possess the employee anymore. Examples of the converse also exist; e.g., Equip13.4.2 ⊑ Resign10.11 is deduced and also here merely changing the range of goal (to, say, Abstract or Concrete) already solves the undesirable deduction. Encoding all verb classes likely may result in further analysis and semantic refinement of the current specifications in VN. As such, the approach of encoding the knowledge of the verb classes in an ontology and examining the deductions shows to be a useful tool for further assessment and VN refinement.

4.1.2. Task B: VN's thematic role hierarchy assessment with ontological principles

The assessment faces three main challenges: 1) terminology regarding the roles and role fillers, 2) sibling desiderata, and 3) multiple inheritance. Key problems observed, which we will illustrate in the next three paragraphs, are the following. Many VN thematic role descriptions indicate it is neither like one would assume it to be based on a dictionary definition nor like any of the entities with the same or similar names in any of the foundational ontologies. The names and descriptions are incoherent when read with those meanings in mind. The subsumption alignments in the taxonomy thus also must be read in conjunction with the thematic role descriptions. A confounding aspect is the 'duality' of the VN terms as roles: either as if they all should align directly to DOLCE's Social Object, which would not get us one step further in the ontological analysis, or as indications of the ontological nature of the role players. Further, the name of the role may not be ontologically a role, but instead aiming to indicate the category or domain entity of its filler that will play that role in the relation. After substantiating these observations, we shall move on to a conservative alignments of the role fillers to DOLCE and a revised hierarchy of VN roles.

Role versus role filler. Unlike the clear distinction in the knead example in the Introduction where Machine is the role filler of the agent role, consider the Material thematic role in VN. Generally, material is that what an object is made of, but in VN, it is both an Undergoer (see Table 2) and a Place and, ultimately, a Participant, which is a role, and so therefore Material is a role, not a physical entity. In addition, the VN description states that Material can be concrete or abstract, but the vast majority of verbs where it is used with their sample sentences refer to concrete things, like build-26.1-1 with "Martha carved a toy out of a piece of wood." where the 'piece of wood' is the player in the Material role. Material is also used for things that are cultivated, like table manners and a puppy's habits[7], which, while less concrete than wood, does not make them abstract, not even in DOLCE. An exception may be the 'action verbs' playing the Material role in "Formulate your sentences using action verbs ...". They are all entities that *fill* or *play* the

[7]Examples of these material roles are taken from the sample sentences at `https://verbs.colorado.edu/html_groupings/cultivate-v.html`.

Table 2. Descriptions of roles in VerbNet (VN), for Material and its super-roles. (Source: VerbNet guidelines).

VN Role	VN Description
Material	*Patient* that exists at the starting point of action (inheritance from *Source*), which is transformed through the event into a new entity; concrete or abstract.
Patient	*Undergoer* in an event that experiences a change of state, location or condition, that is causally involved or directly affected by other participants, and exists independently of the event.
Undergoer	*Participant* in a state or event that is not an instigator of the event or state.
Participant	*Entity* involved in a state or event.
Entity	– (undefined)

role, which is different from the nature of the role *itself*, and thus not that the role itself could be concrete or abstract.

Features and desiderata of sibling roles in the VN role hierarchy. Another line of inquiry for the alignment process and position in a role hierarchy is to examine the focal entity to its siblings. Material's sibling Co-patient is for two patients that "participate equally in the event" and sibling Experiencer "is aware of the event undergone", neither of which are said to have something "transformed" like Material's description does. Co-patient and Experiencer are distinguished by different ways of participating. They thus may not be disjoint, because something that plays the Material role may also be a Co-patient, in VN. Conversely, Theme and Pivot are siblings in the VN hierarchy in Fig. 2, but should not be so based on their descriptions: "does not have control over the way the event occurs, is not structurally changed by the event", and "Theme that participates in an event with another theme unequally. Pivot is much more central to the event", respectively; rather, Pivot $\sqsubseteq$ Theme may hold. Juxtaposing Theme would be something that *is* structurally changed by the event, which is the Patient only, not the Instrument with which the action is performed nor the Attribute where the undergoer is a property of an entity (the 'does have control'-part of Theme is covered by the other main branch of Actor).

Multiple inheritance in the role hierarchy. The consideration of the hierarchy brings us to the second challenge: the multiple inheritance cases in Fig. 2. Consider Result, a "Goal that comes into existence through the event" and its use indicates an outcome that may be a physical object that is created or a state that is the outcome of an action, such as the broken glass resulting from breaking it. So that may be physical objects or states in DOLCE. It is, however, a direct sub-role of both Patient and Goal, but the former is an undergoer and the latter a place, which presumably would be disjoint. If it were to be a Goal, then it contradicts its sub-role Product, which is filled by concrete objects (e.g., toys, acorns, and sandwiches in the build-26.1, grow-26.2, and preparing-26.3 sample sentences) that is an unlikely sub-role of Place, or Product needs to be a subrole elsewhere in the hierarchy. Place is described as a "Participant that represents the state in which an entity exists.", hence, clearly excluding the physical objects of Product. If it were a state, it then should not subsume different physical locations either, for a state exists *at* a location, not that a location is a state. Place is not used in any verb class, so it is non-disruptive to change the description to more accurately reflect its meaning, being a placeholder for various location entities. These location entities include entities such as (concrete physical) Location, but also the more broadly construed Source (starting point of action) and Destination, which can be argued to be roles played by a quality, rather than the locations (physical or abstract qualities) themselves. Besides these and other challenging cases (omitted due to space limitations), there are also unobjectionable

Table 3. Conservatively, DOLCE categories of role fillers ('restrictions') of the VerbNet's thematic roles (see text for details).

DOLCE category as role filler	**Verbnet thematic roles**
Particular	Participant
Endurant or Perdurant	Goal
Endurant	Beneficiary, Co-Patient, Co-Theme, Experiencer, Instrument, Material, Patient, Pivot, Recipient, Theme, Topic, Undergoer, Actor, Agent, Cause, Co-Agent, Stimulus
Physical object or State	Result
Physical object	Product
Social object (inheres in Physical Quality) or Quality	Destination, Source
Social object or Quality or Region	Place
Quality	Attribute, Extent, Trajectory
Abstract quality	Asset, Frequency
Temporal quality	Final_time, Initial_Time, Time
Physical quality	Initial_Location, Location
Region	Value

roles; the revised hierarchy is shown in Fig. 3.

From the reverse perspective, i.e., that of trying to align VN into ontologies, it can be argued that DOLCE, and, in fact, also all those not motivated at least in part by linguistics and cognitive science, is not specific enough for the VN roles, since there is only a social role. Even setting that aside, there are alignment challenges for the role fillers. Assessing VN's Restrictions for verb classes, one also has to conclude that they bear no resemblance to ontological categories either. For instance, Co-patient can be filled by something that is Animate or Abstract in the amalgamate-22.2-2 verb class, with as example 'company B' in "The merger associated company A with company B.", which is neither animate nor abstract. Also here, as with knead-26.5, there is a mixing of linguistic restrictions and semantic restrictions regarding the kind of entities that can play that role. For roles as part of relations, it is the semantics of the entities that play those roles we are interested in, not the linguistic constraints on the words for the role fillers in a natural language sentence. They are partially used, as with aforementioned Machine, but also fillers that are not endurants; e.g., Solid is a filler for Instrument in cooking-45.3, with 'the oven' supposedly the solid in "The oven baked the potatoes.": the oven is a solid object, but it is a property of the physical object only, i.e., an indirect feature and thus the direct filler is DOLCE's physical object, rather, and likewise for the Pointy filler (a 'needle' in the sample sentence). Neither of the two would be included in DOLCE or other foundational ontologies, however, since they are not deemed top-level entities.

The revised role fillers, specifically those that refer to the kind of entity only, i.e., without any linguistic terminology on grammatical features of the respective linguistic realisations in a natural language, and with caution, are listed in Table 3. It may still be a sub-optimal mapping, especially when noting the long list of role fillers that are aligned to DOLCE's Endurant as the extreme case and the unusual union for Place. This first characterisation of the role fillers with respect to the subject domain semantics may serve as a baseline for further research.

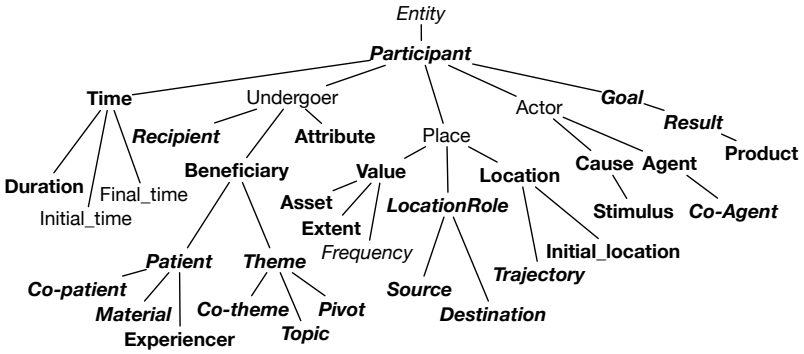

Figure 3. Redesigned taxonomy of VerbNet roles, where the ones in bold are used by VerbNet in their verb class specifications and in italics those that changed cf. Fig. 2.

4.2. Task C: Enhanced VerbNet roles with conceptual data modelling roles

Given the improved hierarchy and the initial alignment of the role fillers to categories in DOLCE, let us take that together with the roles in conceptual data models to address Task C. First, the only overlap in name is participant. It also suggests a similar meaning, with VN's description of "Entity involved in a state or event". A conceptual model's participant is typically more specific than the generic 'entity', being at least an endurant, if not more precisely the union of DOLCE's physical object and social object, which is the case in foundational ontologies as well. Thus, it would practically amount to a sub-role of VN's participant role. Roles that indicate physical and social objects for role filler may make it more precise; notably, Recipient, Patient, or Agent may be good candidates for refinement, and excluding time, place or goal.

The other two main roles in conceptual data models, membership and part-whole relations, do not have a direct 1:1 mapping, because, in English at least, neither role is a constituent of a verb. Verbs approaching membership, such as join, connect, and unite (of mix-22.1-2-1 and cooperate-73.1-1) either have Patient and Co-patient or Agent and Theme, and likewise for the act of signing up, like a student who enrols for (i.e., becomes a member of) a course, that is also a different relationship from parthood or membership. Neither entity that plays the part, the whole, or the member or the collective is changed or instigate anything, i.e., they are in a state of being; thus, they can be sub-roles of Theme.

The second set of roles from conceptual data modelling that seemed thematic, being the parent and child roles, appear not so. Without extensive examples of their use, it is difficult to make it more precise than sub-roles of Participant and not declaring them disjoint from any other main branch in the hierarchy. The third set of roles, being those from the enterprise-oriented conceptual data models with employees and managers, are in the Actor and Undergoer branch of the revised VN role hierarchy, requiring a case-by-case analysis as to which of their sub-role applies. This is also holds for the 'based-on' relationship with the original and derived roles.

Alignments of conceptual modelling roles are not limited to aforementioned VN roles. For instance, `uml_calendars_default.png` has a repetitions role with as player DaySequenceRepetition, which aligns to VN's (hitherto unused) Frequency role and its intervals relates to the Duration role. In sum, there are no straightforward equivalence

or subsumption alignments of the named roles in the conceptual data models and of VN, but some of them can be matched.

5. Discussion

The naming practices and use of roles in conceptual data models and the popular linguistic resource VN was shown to be overwhelmingly non-ontological. Improving the practice of naming roles in conceptual data models with those from VN as-is does not offer a hoped-for ontological rigour. The revised hierarchy of roles may assist better, as will the alignment of the 'restrictions' (role fillers) to DOLCE, if the entity types in the conceptual data models are categorised into such categories, alike in OntoUML [32].

The coarse-grained alignment invites for additional roles to at least structure the hierarchy better, rather than lumping things together. One could argue that the domain-level roles of the conceptual models may have a better alignment with FN roles than with VN, such as with its `[Employee]` and `[Employer]` roles in its `Being_employed` frame, but it comes at a loss of top-level roles and a hierarchy of roles, since they are not positioned in a hierarchy. The FN frame for `Participation` is scattered with 5 core roles, including event, institution, and three versions of participant and thus would need further analysis as well, and it has another 7 non-core roles. The `Membership` may look more promising initially, but also here there's ontological imprecision, both regarding its fillers and the hierarchy, such as the distinction between set and collective entity and it is asserted to be inheriting from `Be_subset_of`, but a member is not the same as a singleton subset. Put differently: also this resource cannot be used off-the-shelf for enhancing role naming in conceptual data models or for an ontology of positionalism.

The recognition of roles as argument places or positions as proposed in the philosophy literature [3,4,18,5,9] also does not solve these issues, since they stop at the existence of roles rather than which ones and how to use them. Orilia's recognition of what he terms "o-roles" [9] as distinct from linguistics' thematic roles is, however, a useful one that also surfaced from the VN data: *ontological fillers*, i.e., entities in the reality, must be separated from *linguistic fillers* as criteria on the categories of words that fill the positions in a natural language sentence.

These insights do not yet amount a sound and comprehensive ontology of roles (as relation[ship] components). Compared to the starting position with just a few logics with roles, we showed that some structuring of roles in a shallow hierarchy without multiple inheritance is feasible, that 'verb classes' as proxies for predicates or relations cannot be uniquely defined by their roles and role fillers based on currently declared knowledge in VN, and that the role fillers need to be cleared up between the entities that fill them and the features of the words in the sentences that fill them. The latter suggest that maybe two ontologies, or one ontology and one language model could be developed: one for roles and relations and one for linguistic roles, linguistic information on the role fillers, and their verb classes. An argument in favour of the latter is that the separation is conceptually cleaner and therewith then also should be amenable to computation and computational verification to ensure quality, which is beneficial for usage scenarios such as the Natural Language Generation as part of Abstract Wikipedia. For example, VN's knead-26.5 from the Introduction would then have at least two encodings: first, the

$$\mathsf{knead} \sqsubseteq \mathsf{agent} : (\mathsf{PhysObj} \sqcup \mathsf{Machine}) \sqcap \mathsf{material} : \mathsf{AmtMatter} \sqcap \mathsf{product} : \mathsf{PhysObj}$$

for the ontological layer to represent (part of) the semantics of kneading, and, second, at least one for the natural language aspects, which may be either with the required grammatical features of the role fillers for valid natural language sentences, such as a, say,

$knead_{vc} \mapsto$ *agent:(Animate | Inanimate) & material:Inanimate & product:Inanimate*

or concretely for a particular natural language and Part-Of-Speech categories, alike a

$knead_{vc\_pos} \mapsto$ *agent:(CountN | NP) & material:MassN & product:(CountN | NP)*

where the left-hand-side acts as a pivot. Likewise, for the snippet of Wikidata information in Fig. 1 on animals being endemic to a region, we could keep the formalisation as-is (albeit with P and Q items and their labels shown), but need to add an assertion about the matching verb from VN, being 'originate' as a member of the verb class establish-55.5-1; e.g., and using VN's role fillers:

originate $\mapsto$ *agent:(Animate | Organization) & theme:*$\top$

For this to work reliably, it also will need to be asserted which ontological role maps to which linguistic role, like the [inhabitant] to *agent* and [location] to *theme* in this example. The design of such a framework and evaluation thereof is left for future work.

6. Conclusions

A bottom-up investigation into the ontology of roles showed that while they are used widely, there are many modelling issues compared to classes and relations. Conceptual data models have less than half of the roles named, and mainly of the type of 'deep' or domain-level roles. The linguistic roles in VerbNet were partially formalised, the role hierarchy improved, and the role fillers disambiguated and aligned to DOLCE, which may contribute to their use for, and alignment to, conceptual data modelling. Future work includes further ontological analysis on roles, notably the hierarchy and role fillers, and its application in Abstract wikipedia and NLP more broadly.

Acknowledgments The author is grateful for the discussions with Kutz Arrieta, which helped clarify notions of roles in linguistics and the narrative of the paper.

Supplementary Material Statement The VerbNet data used for the test ontologies, the test ontologies, and the and the data analysis of the conceptual data models and VerbNet are available at `https://doi.org/10.5281/zenodo.7861468`.

References

[1] Masolo C, Vieu L, Bottazzi E, Catenacci C, Ferrario R, Gangemi A, et al. Social Roles and their Descriptions. In: Proc. of KR'04; 2004. Whistler, Canada, June 2004.
[2] Borgo S, Ferrario R, Gangemi A, Guarino N, Masolo C, Porello D, et al. DOLCE: A descriptive ontology for linguistic and cognitive engineering. Applied Ontology. 2022;17(1):45-69.
[3] Fine K. Neutral Relations. The Philosophical Review. 2000;109(1):1-33.
[4] Gilmore C. Slots in Universals. Oxford Studies in Metaphysics. 2013;8:187–233.
[5] Leo J. Modeling relations. Journal of Philosophical Logic. 2008;37:353-85.
[6] Palmer M, Bonial C, Hwang JD. VerbNet: Capturing English verb behavior, meaning and usage. In: Chipman SEF, editor. The Oxford Handbook of Cognitive Science. OUP; 2017. p. 315-36.
[7] Halpin T, Morgan T. Information modeling and relational databases. 2nd ed. Morgan Kaufmann; 2008.
[8] Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF, editors. The Description Logics Handbook – Theory and Applications. 2nd ed. Cambridge University Press; 2008.

[9] Orilia F. Relational Order and Onto-Thematic Roles. Metaphysica. 2011;12:1-18.

[10] Calvanese D, De Giacomo G, Lenzerini M. On the decidability of query containment under constraints. In: Proc. of PODS'98; 1998. p. 149-58.

[11] Fillottrani P, Keet CM. Evidence-based lean conceptual data modelling languages. Journal of Computer Science and Technology. 2021 Oct;21(2):e10.

[12] Vrandecic D. Architecture for a multilingual Wikipedia. Arxiv. 2020;abs/2004.04733.

[13] Vrandecic D. Building a multilingual Wikipedia. Commun ACM. 2021;64(4):38-41.

[14] Keet CM, Fillottrani PR. An ontology-driven unifying metamodel of UML Class Diagrams, EER, and ORM2. Data & Knowledge Engineering. 2015;98:30-53.

[15] Fillottrani PR, Keet CM. An analysis of commitments in ontology language design. In: Proc. of FOIS'20. vol. 330 of FAIA. IOS Press; 2020. p. 46-60.

[16] Berardi D, Calvanese D, De Giacomo G. Reasoning on UML class diagrams. Artificial Intelligence. 2005;168(1-2):70-118.

[17] Keet CM. Mapping the Object-Role Modeling language ORM2 into Description Logic language $\mathscr{DLR}_{ifd}$. Free University of Bozen-Bolzano, Italy; 2009. arXiv:cs.LO/0702089v2.

[18] Gilmore C. Parts of propositions. In: Mereology and Location; 2014. p. 156-208.

[19] Hwang JD. Identification and representation of caused motion constructions [PhD Thesis]. University of Colorado, Boulder, Colorado, USA; 2014.

[20] Brown SW, Bonn J, Kazeminejad G, Zaenen A, Pustejovsky J, Palmer M. Semantic Representations for NLP Using VerbNet and the Generative Lexicon. Frontiers in Artificial Intelligence. 2022;5.

[21] Baker CF, Fillmore CJ, Lowe JB. The Berkeley FrameNet Project. In: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1. Montreal, Quebec, Canada: ACL; 1998. p. 86-90.

[22] Lopez De Lacalle M, Laparra E, Rigau G. Predicate Matrix: extending SemLink through WordNet mappings. In: Proc. of LREC'14. Reykjavik, Iceland: ELRA; 2014. .

[23] Lopez de Lacalle M, Laparra E, Aldabe I, Rigau G. A Multilingual Predicate Matrix. In: Proc. of LREC'16. Paris, France: ELRA; 2016. .

[24] Corcoglioniti F, Rospocher M, Aprosio AP, Tonelli S. PreMOn: a Lemon Extension for Exposing Predicate Models as Linked Data. In: Proc. of LREC'16. Portorož, Slovenia: ELRA; 2016. p. 877-84.

[25] Gangemi A, Alam M, Asprino L, et al. Framester: A Wide Coverage Linguistic Linked Data Hub. In: Proc. of EKAW'16. Springer; 2016. p. 239-54.

[26] Gangemi A, Guarino N, Masolo C, Oltramari A. Sweetening WORDNET with DOLCE. AI Magazine. 2003;24(3):13.

[27] Niles I, Pease A. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Proc. of IKE'03. CSREA Press; 2003. p. 411-6.

[28] Keet CM, Fillottrani PR. An analysis and characterisation of publicly available conceptual models. In: Proc. of ER'15. vol. 9381 of LNCS. Springer; 2015. p. 585-93. 19-22 Oct, Stockholm, Sweden.

[29] Keet CM, Artale A. Representing and Reasoning over a Taxonomy of Part-Whole Relations. Applied Ontology. 2008;3(1-2):91-110.

[30] Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology. USA: The MIT Press; 2015.

[31] Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. Ontology Library; 2003. Http://wonderweb.semanticweb.org. WonderWeb Deliverable D18 (ver. 1.0, 31-12-2003).

[32] Guizzardi G, et al. Endurant Types in Ontology-Driven Conceptual Modeling: Towards OntoUML 2.0. In: Proc. of ER 2018. vol. 11157 of LNCS. Springer; 2018. p. 136-50.