Towards Corpus-Based Population and Visualization of ISO 24617-8 Ontology

Dariusz Czerski, Maciej Ogrodniczuk
Institute of Computer Science, Polish Academy of Sciences
dariusz.czerski@ipipan.waw.pl, maciej.ogrodniczuk@ipipan.waw.pl

Abstract

This paper presents an extension of the ISO 24617-8 ontology for discourse relations through the integration of corpus-based examples and the development of a dedicated Ontology Viewer. The goal is to bridge the gap between formal ontological representations and practical corpus-based linguistic analysis, making discourse annotation frameworks more accessible to researchers. The proposed approach introduces a method for populating the ISO ontology with instances derived from three corpora (in Polish and English) compliant with the ISO 24617-8 standard. These instances formally connect discourse relations, argument roles, and explicit connectives within a unified semantic model. The Ontology Viewer enables intuitive browsing, filtering, and full-text searching of examples by language, relation type, and connective, offering both a relation-oriented and connective-oriented perspective. The experiment demonstrates the feasibility and effectiveness of this corpus-driven instantiation method and its visualization. The system provides a foundation for future integration of multilingual discourse corpora and contributes to the development of interoperable language resources for the Semantic Web and Natural Language Processing applications.

Keywords: discourse relations, ISO 24617-8, discourse ontology, ontology viewer

1. Introduction

While existing discourse formalisms provide conceptual schemata for discourse representation, their formalization into Web Ontology Language (OWL) ontologies is crucial for the Semantic Web, knowledge graph construction, and interoperable Natural Language Processing. A notable proposal of linking these two worlds is the set of Ontologies of Linguistic Annotation (OLiA; Chiarcos, 2014)[1], containing a Discourse Extensions[2] framework to cover discourse-related phenomena. The primary goal of the discourse part is to enhance the semantic interoperability of discourse annotations across major corpora and formalisms.

Although ontologies such as OLiA offer a powerful formalism for representing linguistic concepts and their interrelations, they can be difficult for linguists to use directly. Most linguists are not trained in ontology engineering or description logics, and the abstract syntax and formal constraints of OWL often obscure the linguistic content they aim to model. Specialized ontology editors such as Protégé[3] (Musen, 2015) provide useful visualization methods, but they remain too general-purpose, optimized for ontology builders rather than users who don't need to be aware of complex class hierarchies or logical restrictions. To make ontology-based linguistic resources more accessible and practically useful, there is a strong need for a dedicated interface abstracting away from the formal representation and instead presenting ontology content in user-friendly terms, supporting corpus-based exploration, and allowing intuitive linking between linguistic annotations, corpus examples and ontology concepts.

Following this path of thought, this article intends to propose two contributions: (1) the corpus-based method of ontology population limited to one discourse representation formalism – ISO 24617-8 (ISO, 2016), also with respect to discourse marker inventory induction, tested on two small ISO-compatible mini-corpora and one larger discourse corpus, and (2) presentation of the newly developed ontology viewer tailored to corpus-centered presentation of discourse relations as compared to previously proposed discourse marker-centered applications.

2. Related Work

Here we present three components that directly influenced our work: the ontology building framework concentrated on discourse, the ontology browser and a connective lexicon which can be also used as a connective-oriented discourse corpus browser.

The OLiA Discourse Extensions currently formalize nine annotation schemes covering corefer-

[1] https://acoli-repo.github.io/olia/
[2] https://acoli-repo.github.io/olia/discourse.html
[3] https://protege.stanford.edu/

Proceedings of the 22nd Joint ACL - ISO Workshop on Interoperable Semantic Annotation and Representation (ISA-22), pages 54–61
12 May 2026. ©ELRA Language Resources Association (ELRA), 2026

ence, information status, information structure, and discourse structure for multiple languages. Each scheme is represented as an OWL/DL Annotation Model linked to a unified Reference Model via declarative linking models. This modular approach enables interoperability between different discourse annotation frameworks, serving as a terminological bridge and foundation for future integration with additional discourse and pragmatics resources. The work on the ontology is ongoing; since November 2024, the Reference Model for the Discourse Extensions has been integrated with the OLiA Reference Model.

One of the ontologies integrated in OLiA models is the ISO 24617-8 ontology, limited to listing all 20 discourse relations and their argument roles. Instead of directly annotating discourse relations, the ontology emphasizes argument roles (e.g. REASON and RESULT), from which the associated discourse relations (here: CAUSE) can be inferred via entailment. These entailment relations are formalized through `rdfs:subClassOf` links between argument roles and discourse relation classes, these in turn being subclasses of SYMMETRICDREL and ASYMMETRICDREL and grandchildren of DREL classes. For asymmetric discourse relations, argument roles are traditionally named ARG1 and ARG2 (subclasses of ASYMMETRICARGUMENTROLE).

The most widely used platform for viewing and editing OWL ontologies is PROTÉGÉ (Musen, 2015), an open-source framework developed at Stanford University and also available as a Web application[4] (Tudorache et al., 2013). WEBPROTÉGÉ offers intuitive access to classes, properties, and instances through an entity-centered interface. The browsing capabilities of the tool are offered by the Entity Description Browser which displays complete logical definitions and annotations of selected entities, Users can easily search and filter entities within projects, explore imported ontologies, and inspect both class hierarchies and instance data.

CONNECTIVE-LEX.INFO[5] (Stede et al., 2019) is a multilingual platform primarily designed to document and describe discourse connectives – words and expressions that signal semantic or pragmatic relations between discourse units (e.g. *because*, *although* or *however*). It provides fine-grained descriptions of connectives across multiple languages, linking them to standardized inventories of discourse relations. While concentrating on discourse connectives, the platform integrates information from various discourse-annotated sources, offering a view of corpus-based examples.

While general-purpose corpus query tools such as PML-Tree Query[6] (Štěpánek and Pajas, 2010) or ANNIS[7] (Krause and Zeldes, 2014) provide powerful multi-layer searching, they often require complex query languages (e.g., AQL). Our Ontology Viewer differs by being schema-native: it is built specifically to reflect the ISO 24617-8 hierarchy. Unlike Connective-Lex.info, which is connective-centric, our tool provides a balanced dual perspective: relation-oriented and connective-oriented, specifically tailored for validating the ISO semantic model.

3. Ontology Extension

OLiA defines ISO relations and their arguments[8] in a `discourse.ISO.owl` file as a series of classes following the naming from ISO 24617-8 standard. Figure 10 presents such definition for the PURPOSE relation. This asymmetric discourse relation takes two arguments: GOAL and ENABLEMENT which ISO standard describes as:

> Arg2 (Goal) is the goal or purpose of the situation described by Arg1 (Enablement).

Our goal was to extend this ontology with representation of corpus examples (from various corpora, in various languages) and all necessary constructs linking typed relations, arguments and connectives into a common formal description. To achieve that, the following design principles were assumed:

1. base definition of ISO relations are preserved in the original `discourse.ISO.owl` file
2. project extensions to the ontology are stored in `discourse.ISO.Ext.owl` file
3. corpus-based instance definitions are stored in a series `discourse.ISO.<corpus>.owl` files, e.g. `discourse.ISO.PDC.owl`.

3.1. Connective Property

The only project extension to the ISO ontology, independent of individual corpora, is the association of connectives (defined in the top-level ontology file `olia.owl`) to discourse relations. Contrary to

[4] http://webprotege.stanford.edu/
[5] http://connective-lex.info

[6] https://ufal.mff.cuni.cz/pmltq
[7] https://corpus-tools.org/annis/
[8] This unobvious decision is motivated by its better compatibility with other annotation models which reflects "that most argument roles are recognized as independent discourse relations in other schemes" – see comment in the top `owl:Ontology` element in `discourse.ISO.owl` and (Bunt and Prasad, 2016). The authors of OLiA are of course well aware that "In terms of ISO DR-Core, however, these are not in a hierarchical structure, but only in an entailment relation" (see another comment at DREL class definition).

relation arguments, which need to be assigned to a specific relation, connectives can be assigned to any relation so we use DREL class from `discourse.ISO.ow` file as the property domain:

```
<owl:ObjectProperty rdf:about=
  "http://purl.org/olia/discourse/
   discourse.ISO.Ext.owl
   #hasConnective">
   <rdfs:domain rdf:resource=
     "http://purl.org/olia/discourse/
      discourse.ISO.owl#DRel"/>
   <rdfs:range rdf:resource=
     "http://purl.org/olia/
      olia.owl#ExplicitConnective"/>
</owl:ObjectProperty>
```

3.2. Object Properties

To formally assign typed arguments to the corpus instance, we define properties of relation classes which would point to respective arguments (here: HASGOAL and HASENABLEMENT for the PURPOSE relation/class):

```
<owl:ObjectProperty rdf:about=
  "http://purl.org/olia/discourse/
   discourse.ISO.Ext.owl#hasGoal">
   <rdfs:domain rdf:resource=
     "http://purl.org/olia/
      discourse/discourse.ISO.owl
      #Purpose"/>
   <rdfs:range rdf:resource=
     "http://purl.org/olia/
      discourse/discourse.ISO.owl
      #Goal"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about=
  "http://purl.org/olia/discourse/
   discourse.ISO.Ext.owl
   #hasEnablement">
    <rdfs:domain rdf:resource=
      "http://purl.org/olia/
       discourse/discourse.ISO.owl
       #Purpose"/>
    <rdfs:range rdf:resource=
      "http://purl.org/olia/
       discourse/discourse.ISO.owl
       #Enablement"/>
</owl:ObjectProperty>
```

3.3. Instance Definitions

Finally, we construct corpus examples (instances of respective OLiA classes) using the original and newly defined building blocks. One example is an `owl:NamedIndividual` of a given type, carrying the text of the example split into relation arguments and (optionally) a discourse marker (connective).

An instance of a given relation carries its complete textual representation as the `label` and `skos:notation` value is used to represent the identifier of the sample. Then, the description uses previously defined object properties to in-line assign arguments to the relation. `skos:order` is used to preserve the linear sequence of text segments in the original discourse.

3.4. Design Rationale and Alternatives

The choice of `owl:NamedIndividual` with inline object properties was driven by the need for direct compatibility with existing Semantic Web reasoners. While RDF-star was also considered, we opted for a standard OWL 2 approach to ensure broader interoperability with legacy tools like Protégé. In the era of Large Language Models, where custom annotation tools can be generated rapidly, our approach prioritizes a standardized ontological schema. This model ensures that discourse data remains machine-actionable, queryable via SPARQL, and integrated into the global Linked Data cloud.

4. Ontology Population

Eventually the ontology is intended to be used to represent corpus examples from all future corpora converted to ISO representation in an ongoing project. Currently we decided to test its applicability to this task by representing three test corpora for two languages, based on the annotation guidelines (Ogrodniczuk, 2021) for the Polish Discourse Corpus (Ogrodniczuk, 2024), the ISO 24617-8 standard for English and the complete PDC (Ogrodniczuk et al., 2024). The samples from the two first sources were extracted from respective documents (the technical report and the ISO standard description) and stored as two mini-corpora (see their statistics in Table 4), with identifiers of examples automatically constructed. Polish examples were numbered according to their order within each relation description (and they are taking the form of `relation name-number`, e.g. `Cause-1`) while the ISO standard used letters for consecutive examples and they were also preserved (e.g. `Purpose-a`). Additionally, the complete PDC was encoded.

Dialogue act-related relations (FUNCTIONAL DEPENDENCE and FEEDBACK DEPENDENCE) were not represented in the source document because these type of relations were omitted both from PDC annotation based on the assumption that their complexity exceeds the current capabilities of the annotation project. This decision goes in line with both the composition of ISO guidelines where dialogue acts are described by a separate part of the standard

Relation	PDC guidelines	ISO 24617-8	PDC (full)
Conjunction	3	2	8 286
Cause	1	7	1 753
Contrast	3	2	1 511
Asynchrony	1	4	1 057
Disjunction	3	3	815
Condition	1	2	807
Concession	1	2	712
Synchrony	3	3	528
Purpose	1	2	517
Exemplification	1	3	282
Substitution	1	3	227
Similarity	3	2	139
Manner	1	2	102
Restatement	3	2	78
Exception	1	2	32
Elaboration	1	3	12
Expansion	1	5	8
Negative Condition	1	3	5
Functional dependence	0	2	0
Feedback dependence	0	2	0
Total	**30**	**56**	**17 088**

Table 1: Number of examples in the ontology representing discourse relations of individual ISO types from the converted corpora

(ISO, 2020) and the decisions made by OLiA creators, deliberately excluding dialogue and speech act models from the ontologies.

5. Ontology Viewer

To facilitate exploration of discourse relation examples and annotations compatible with the ISO 24617-8 standard we have developed an interactive, Web-based Ontology Viewer powered by the ontologies described in the previous section. It integrates instances from available ISO 24617-8-compatible multilingual corpora, allowing users to browse and search through discourse relations, connectives and text samples across individual corpora and languages.

As for the final design of the tool, we consulted both Web-based versions of Protégé and CONNECTIVE-LEX.INFO. While WEBPROTÉGÉ claims to be highly customizable with respect to user interface, its editing interface prevails over browsing and search capabilities which makes it less adaptive to our needs as the interface of CONNECTIVE-LEX.INFO from which we decided to borrow the basic concepts of filter-oriented views and examples-centered search result lists

Figure 1 presents the main interface of the viewer. Each corpus entry in the left-hand panel shows the number of available annotated examples and the primary language of annotation. The top panel offers powerful filtering and search capabilities. Users can filter examples by language, connective, relation type, or relation symmetry. They can also perform full-text searches through all example sentences. This facilitates both relation-oriented analyses (e.g., inspecting how a relation such as PURPOSE or CONTRAST is realized in a specific language) and comparison across individual use of connectives within relations.

In the main panel individual discourse-relation instances are displayed with their internal structure: relation arguments (here: ENABLEMENT and GOAL), the connective linking them (*so that*), and the relation type (PURPOSE). Each example links to its unique ontology identifier, ensuring traceability to the underlying OWL definition and corpus source.

6. Limitations and Replicability

Currently, the viewer focuses exclusively on the discourse layer. It does not yet visualize lower linguistic layers such as morphology or syntax (based on available CoNLL-U annotation). The current population is limited to three corpora, representing a pilot phase of the project.

To support open science, the Ontology Viewer source code is available at: `https://github.com/ipipan/ontology-viewer`. A live demo of the Ontology Viewer is accessible at: `https://ontology-viewer.ipipan.waw.pl/`

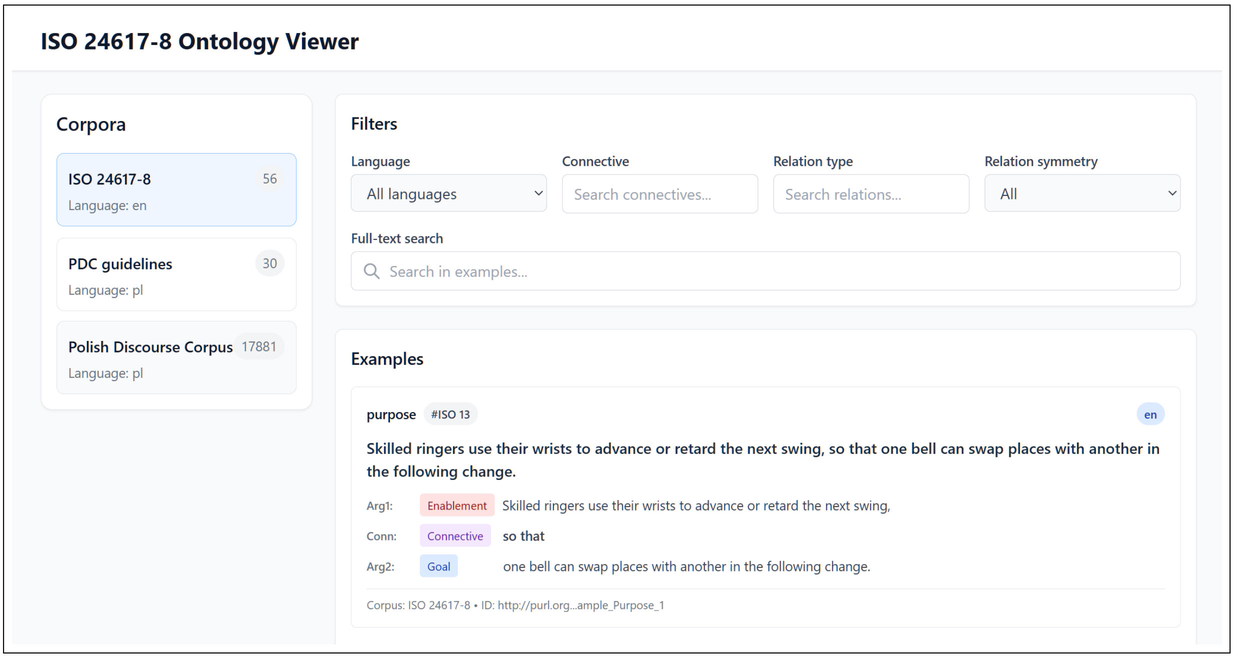

Figure 1: Interface of the ontology viewer

7. Conclusions and Future Work

We believe that the proposed population of ISO ontology with corpus examples and implementation of the ontology viewer can act as a bridge between ontology data and corpus-based linguistic evidence, providing both a formal description and an intuitive environment for linguists and annotators to explore, validate, and compare ISO 24617-8 discourse-relation annotations.

Both the representation of the instances of discourse relations from individual corpora in the ISO ontology and the flexible interface of the viewer were developed as a starting point for integration of a set of 10–15 discourse corpora in several formalisms and languages to be converted to the ISO 24617-8 standard in the ongoing Universal Discourse project[9] (see Acknowledgements).

The experiment showed that both the proposed method of providing corpus-based instantiation of ISO ontology classes and its presentation in a dedicated viewed are feasible and effective. We believe that the interface of the ontology viewer constitutes another convenient method of browsing through the corpus samples, filter them and offer and easy-to-use search across languages, relation types, connectives and textual content.

The viewer also offers another method of presenting the discourse corpora through the lens of discourse connectives, similar to CONNECTIVE-LEX.INFO and other corpus-based methods (see e.g. Das et al., 2018; Silvano et al., 2022) augmenting the inventory of discourse markers with reference material.

Even though it was not intended to be used for other discourse representation formalisms, the ontology viewer can be easily adapted to be used for displaying ‘flat’ relation models such as PDTB.

[9] http://udisc.org/

8. Acknowledgements

This research was funded in whole by the National Science Centre, Poland, grant 2023/50/A/HS2/00559 (”Universal Discourse: a multilingual model of discourse relations”).

9. Bibliographical References

Harry Bunt and Rashmi Prasad. 2016. ISO DR-Core (ISO 24617-8): Core Concepts for the Annotation of Discourse Relations. In *Proceedings of the 10th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 45–54.

Christian Chiarcos. 2014. Towards interoperable discourse annotation. Discourse features in the Ontologies of Linguistic Annotation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pages 4569–4577, Reykjavik, Iceland. European Language Resources Association (ELRA).

Debopam Das, Tatjana Scheffler, Peter Bourgonje, and Manfred Stede. 2018. Constructing a Lexicon of English Discourse Connectives. In *Pro-*

ceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 360–365, Melbourne, Australia. Association for Computational Linguistics.

ISO. 2016. ISO 24617-8:2016: Language resource management – Semantic annotation framework (SemAF) – Part 8: Semantic relations in discourse, core annotation schema (DR-core).

ISO. 2020. ISO 24617-2:2020: Language resource management – Semantic annotation framework (SemAF) – Part 2: Dialogue acts.

Thomas Krause and Amir Zeldes. 2014. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31(1):118–139.

Mark A. Musen. 2015. The Protégé Project: A Look Back and a Look Forward. *AI Matters*, 1(4):4–12.

Maciej Ogrodniczuk. 2021. Założenia dla korpusu metatekstowego (EN: Discourse corpus guidelines). Technical report, CLARIN-PL. Institute of Computer Science, Polish Academy of Sciences.

Maciej Ogrodniczuk, Aleksandra Tomaszewska, Daniel Ziembicki, Sebastian Żurowski, Ryszard Tuora, and Aleksandra Zwierzchowska. 2024. Polish Discourse Corpus (PDC): Corpus design, ISO-compliant annotation, data highlights, and parser development. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 12829–12835, Torino, Italy. ELRA and ICCL.

Purificação Silvano, Mariana Damova, Giedrė Valūnaitė Oleškevičienė, Chaya Liebeskind, Christian Chiarcos, Dimitar Trajanov, Ciprian-Octavian Truică, Elena-Simona Apostol, and Anna Baczkowska. 2022. ISO-based Annotated Multilingual Parallel Corpus for Discourse Markers. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, pages 2739–2749, Marseille, France. European Language Resources Association.

Manfred Stede, Tatjana Scheffler, and Amália Mendes. 2019. Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives. *Discours*, pages 1–38.

Jan Štěpánek and Petr Pajas. 2010. Querying Diverse Treebanks in a Uniform Way. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Tania Tudorache, Csongor Nyulas, Natalya F. Noy, and Mark A. Musen. 2013. WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. *Semantic Web*, (1):89–99.

10. Language Resource References

Ogrodniczuk, Maciej. 2024. *Polish Discourse Corpus*. Institute of Compute Science, Polish Academy of Sciences. PID https://zil.ipipan.waw.pl/PolishDiscourseCorpus.

Appendix

A. Definition of PURPOSE discourse relation class in OLiA

```
<owl:Class rdf:about="http://purl.org/olia/discourse/
                      discourse.ISO.owl#Purpose">
    <rdfs:subClassOf rdf:resource="http://purl.org/olia/discourse/
                                   discourse.ISO.owl#AsymmetricDRel"/>
    <rdfs:comment>Arg2 is the goal or purpose of Arg1
                  Purpose: Goal, Enablement</rdfs:comment>
    <rdfs:label>Purpose</rdfs:label>
</owl:Class>

<owl:Class rdf:about="http://purl.org/olia/discourse/
                      discourse.ISO.owl#Goal">
    <rdfs:subClassOf rdf:resource="http://purl.org/olia/discourse/
                                   discourse.ISO.owl#Arg1"/>
    <rdfs:subClassOf rdf:resource="http://purl.org/olia/discourse/
                                   discourse.ISO.owl#Purpose"/>
    <rdfs:label>Goal</rdfs:label>
</owl:Class>

<owl:Class rdf:about="http://purl.org/olia/discourse/
                      discourse.ISO.owl#Enablement">
    <rdfs:subClassOf rdf:resource="http://purl.org/olia/discourse/
                                   discourse.ISO.owl#Arg2"/>
    <rdfs:subClassOf rdf:resource="http://purl.org/olia/discourse/
                                   discourse.ISO.owl#Purpose"/>
    <rdfs:label>Enablement</rdfs:label>
</owl:Class>
```

B. Example instances of PURPOSE relation from the test English discourse corpus

```
<owl:NamedIndividual rdf:about="http://purl.org/olia/discourse/
  discourse.ISO.EN.owl#Example_Purpose_1">
  <rdf:type rdf:resource="http://purl.org/olia/discourse/
                          discourse.ISO.owl#Purpose"/>
  <rdfs:label xml:lang="en">Skilled ringers use their wrists to advance
    or retard the next swing, so that one bell can swap places with
    another in the following change.</rdfs:label>

  <skos:notation>ISO 13</skos:notation>

  <hasEnablement>
    <owl:NamedIndividual rdf:about="http://purl.org/olia/discourse/
      discourse.ISO.EN.owl#Example_Enablement_1">
      <rdf:type rdf:resource="http://purl.org/olia/discourse/
        discourse.ISO.owl#Enablement"/>
      <rdfs:label>Skilled ringers use their wrists to advance
        or retard the next swing,</rdfs:label>
      <skos:order rdf:datatype="http://www.w3.org/2001/
        XMLSchema#integer">1</skos:order>
    </owl:NamedIndividual>
  </hasEnablement>
```

```xml
  <hasConnective>
    <owl:NamedIndividual rdf:about="http://purl.org/olia/discourse/
      discourse.ISO.EN.owl#Connective_Purpose_1">
      <rdf:type rdf:resource="http://purl.org/olia/olia.owl
        #ExplicitConnective"/>
      <rdfs:label>so that</rdfs:label>
      <skos:order rdf:datatype="http://www.w3.org/2001/
        XMLSchema#integer">2</skos:order>
    </owl:NamedIndividual>
  </hasConnective>

  <hasGoal>
    <owl:NamedIndividual rdf:about="http://purl.org/olia/discourse/
      ISO.EN.owl#Example_Goal_1">
    <rdf:type rdf:resource="http://purl.org/olia/discourse/
      discourse.ISO.owl#Goal"/>
    <rdfs:label>one bell can swap places with another in the following
      change.</rdfs:label>
    <skos:order rdf:datatype="http://www.w3.org/2001/
      XMLSchema#integer">3</skos:order>
    </owl:NamedIndividual>
  </hasGoal>
</owl:NamedIndividual>
```