The Talkidz Ontology: Supporting Clinical Studies through Semantic Representation of the Speech Pathology Domain

Lia Draetta[1], Nicola Barbaro[1], Francesco Petriglia[2], Andrea Meirone[2], Rossana damiano[1] and Cristina Gena[1]

[1]University of Turin, Italy, Department Of Computer Science
[2]Fondazione Paideia, Italy

Abstract

In recent years, Semantic Web technologies have gained increasing attention for their ability to enhance data integration, reusability, and interoperability across various domains, including healthcare. In line with the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles, ontologies provide structured frameworks for representing domain-specific knowledge across different resources and applications, facilitating clinical research, and improving decision-making processes, but their potential remains underexplored in the domain of speech therapy. Existing digital tools for speech therapy often lack interoperability and full alignment with FAIR principles. To bridge this gap, we propose the Talkidz Ontology, which is designed to support the development of applications that assist speech therapists in detecting and treating speech disorders. The Ontology was developed through a participatory approach involving speech therapists, ontology engineers, and clinical researchers. Its structure was validated through a survey of domain experts assessing the relevance of key classes and competency questions. Additionally, we demonstrate a real-world clinical application, showcasing the ontology's ability to facilitate epidemiological studies, and support evidence-based decision-making. By providing a standardized, expert-validated resource, the Talkidz Ontology aims to enhance the efficiency and effectiveness of speech therapy practices. It serves as a crucial step toward improving knowledge sharing and data-driven research in the field of language disorders.

Keywords

Domain Ontology, FAIR Principles, Speech Therapy, Healthcare

1. Introduction

In recent years, the Semantic Web and its technologies have gained increasing attention, particularly for their impact on various domains such as e-learning, social media, healthcare, and digital humanities [1]. In healthcare, semantic technologies play a crucial role in ensuring data uniformity, reusability, and shareability. Following the FAIR principles [2], they provide a shared and unambiguous semantics for the representation of data and metadata, bridging fragmented data sources. Additionally, they facilitate epidemiological and clinical research by uncovering patterns in medical reports [3, 4], supporting monitoring systems [5], and enabling ontology-based decision-making [6].

Despite their success in various medical fields, the application of semantic technologies in speech therapy domain remains underexplored [7, 8]. Developing an ontology-based tool for sharing therapy data and analyzing the prevalence of speech disorders in relation to demographic factors can provide valuable epidemiological insights and assist therapists in selecting optimal treatment approaches. This is particularly relevant given the significant prevalence of developmental language disorders, for instance, in Italy, 7.01% of children are diagnosed with Specific language

Proceedings of FOIS 2025 Satellite events co-located with the 15th International Conference on Formal Ontology in Information Systems (FOIS 2025), September 10-12, 2025, Catania, Italy

lia.draetta@unito.it (L. Draetta); nicola.barbaro@unito.it (N. Barbaro); francesco.petriglia@centropaideia.org (F. Petriglia); andrea.meirone@centropaideia.org (A. Meirone); rossana.damiano@unito.it (R. damiano); cristina.gena@unito.it (C. Gena)
0009-0004-6479-5882 (L. Draetta); 0009-0003-8467-2946 (N. Barbaro); 0000-0001-9866-2843 (R. damiano); 0000-0003-0049-6213 (C. Gena)

disorders [9]. In addition, childhood speech disorders can severely impact education and social development, leading to learning difficulties [10], behavioral and psychiatric issues [11], and challenges in emotional and social adaptation [12]. Early identification of at-risk children is essential for timely interventions that maximize developmental outcomes [13, 14].

In recent years, it has become increasingly evident that data and knowledge sharing are powerful tools for advancing research across all fields, contributing to more data-driven and empirically supported clinical studies and ontologies provide a structured and unambiguous framework for representing large volumes of domain-specific data. In speech therapy, they have been applied mainly for clinical data storage and visualization [8], [15], to support early detection of speech pathologies [16], [17], and as base for educational tools in assisting speech therapists [7]. However, few ontological resources have been rigorously evaluated by experts or fully comply with FAIR principles while ensuring alignment with others existing resources.

To address these gaps, we propose an ontology[1] designed to integrate and analyze data from diverse medical clinics. The Talkidz Ontology aims to support epidemiological studies, facilitate clinical research, and enhance diagnostic and therapeutic decision-making for speech therapists. To do so, the ontology represent speech disorders, therapies, and relationships between patients, therapists, and conditions while incorporating demographic information. Developed and validated with input from speech therapists and domain experts, the Ontology has been assessed through a survey to validate competency questions and ontology classes. In this paper we describe the ontology and we demonstrate a clinical application of our resource, showcasing its potential for integrating heterogeneous data sources and assisting speech therapists in obtaining a comprehensive and standardized view to support evidence-based decision-making.

The present work is organized as follows: in Section 2, we review the most relevant studies that have addressed speech disorders using ontologies, highlighting both their practical applications and the ontology engineering processes, and underscoring whether they comply with the FAIR principles. In Section 3, we present the multidisciplinary research initiative to which this ontology belongs. In Sections 4 and 5, we provide an in-depth description of our methodological approach and the validation process of the ontology's main classes and competency questions, followed by the description of the full ontology structure. Finally, we validate the presented work with a case study, and in Section 6, we discuss its limitations and outline future work.

2. Background and Motivations

The field of Speech Therapy has been extensively studied for many years by speech therapists and language experts. It encompasses the prevention, assessment, and treatment of communication impairments, as well as the scientific study of speech, spoken and written language, and related disorders. Speech Language Therapy plays a crucial role in addressing difficulties associated with impairments of sound production by developing effective treatment programs to improve spoken communication and prevent future literacy challenges [18]. In the following sections, we first review key works that leverage ontologies to develop applications supporting speech therapists, highlighting the main addressed tasks and objectives. In subsequent paragraphs, we focus the review on the ontology engineering process, emphasizing the goals and validation methods of existing resources while underscoring the differences and novelties of our resource compared to others.

2.1. Ontology-based Services in Speech Therapy Domain

In the healthcare domain in general, several attempts have been made to represent the field by leveraging semantic representation, aiming to establish a standard for the classification of diseases and pathologies. For example, the Human Disease Ontology (DO) [19] is a resource

[1] https://github.com/liadraetta/talkidz-ontology

that plays a central role in classifying rare, common, and complex human diseases, as well as linking them to genetic information. Although it is a comprehensive and standardized resource for representing diseases in a general way, it lacks specialization within specific fields.

In recent years semantic web technologies have been increasingly applied to the field of speech pathology for various purposes, including data sharing, support for Speech Therapy and speech therapists in the diagnostic process. Notably, the use of ontologies in guiding Speech Therapy design has been steadily increasing and is expected to serve as a valuable support for developing efficient, solution-based therapies. As highlighted by Usip and colleagues [20], well-founded ontologies play a crucial role in accelerating scientific research; the researchers highlighted that their structured nature ensures effective discovery, automation, integration, and reuse across diverse application platforms, providing a comprehensive framework for representing relevant concepts and their relationships.

Exploiting the potential of ontologies, Robles-Bykbaev and colleagues [21] developed a system to provide several services related with information querying, reports generation, inference of intervention strategies, with the aim of helping speech therapist in the detection process. To build and validate the ontology they used clinical information about 152 real patients cases. In another study Robles-Bykbaev et al. [22] developed Speech Therapy knowledge-based model tools to support speech and language pathologists, doctors, students, patients and their relatives. Martín-Ruiz and colleagues [16] developed an ontology-based smart solution to support the detection of language disorders among young children during pediatricians routine visits. They built the ontology with the help of pediatricians and speech therapists creating a representation that helps pediatrician in asking the right clinical questions according to the age of the child at the time of evaluation. The competency questions (CQs) were built engaging four experts in neuropediatrics, neonatology and language disorders and the tool finally was validated against 21 cases obtained from a center oriented to the detection and treatment of language disorders.

In the same line, the Martín Ruiz and colleagues [17] leveraged an ontology to develop and evaluate a web-based Clinical Decision Support System. This web-based system relies on an ontology and can be used by pediatricians with the aim of detecting communication disorders in children. The ontology contains information related with acquisition and development of language in children as well as a rule set to generate alerts when language and communication disorders are detected. Ontologies were also exploited with the aim of building decision support system for planning therapy sessions. To do this, Robles-Bykbaev and colleagues [23] built an ontology-based expert system with the aim of suggesting activities that can be part of the therapy plan. The resource was developed with classes and relations that describe the speech-language therapy elements, aiming at answering questions such as "Determine with which speech-language categories must be carried out the therapy for a given patient."

Ontology-based applications have also been used to develop tools to assist future speech therapists during their training. In [7] researchers developed an ontology to model patients' anamnesis and speech-language milestones (exercises, skills), representing clinical profiles that include personal data, medical records, and speech-language records. The goal was to create a system capable of automatically generating tests and educational exercises to enhance skill development for designing effective therapy plans.

2.2. Speech Disorders Ontologies

As discussed in the previous section, various ontological resources have been developed in recent years aiming at representing the speech therapy domain. However, some of these resources fail to fully meet the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR). In other cases, a comprehensive validation process of the ontology structure and competency questions is lacking.

Robles-Bykbaev and colleagues developed [8] an ontological resource to represent the domain of speech-language therapy with the aim of assisting speech therapists in the detection process.

The resource is built in collaboration with specialists but is not accessible and, therefore, not reusable. Additionally, no alignments with other resources are provided, nor are case studies presented for the resource's evaluation. The authors state that the resource was validated by collecting 130 real cases of children with speech disorders; however, no further details on the evaluation results are given, and no competency questions are provided. On the same line García et al. [15] aiming at developing an ontology as a tool to support therapists in detection and suggest potential treatment, collected a corpus by crawling web data using specific keywords and their semantic relations (synonyms, hyponyms, and hyperonyms), ultimately obtaining 1,097 documents. The authors then constructed a taxonomy based on the collected documents, focusing on the representation of speech disorder, therapy strategies, individuals, and signs and symptoms. From this taxonomy, they engineered the ontology by identifying key classes using both top-down and bottom-up strategies [24]. However, the resulting resource is not publicly available, and no evaluation of the ontology or application case studies have been reported. Chuchuca-Mendez and colleagues [7] propose an ontology designed to model key aspects of speech therapy, with the goal of automatically generating tests for future therapists. The ontology includes elements such as patient anamnesis, profiles, and speech-language clinical history. The researchers further populated the ontology by collecting 383 samples of rehabilitation activities used in therapy. However, no competency questions, evaluation process, or access to the resource are provided.

Table 1
The table shows the reviewed works and their compliance with FAIR principle and the presence, or not, of competency questions (CQs), an evaluation strategy and a case study.

Resource	FAIR	CQs	Evaluation	Case study
Vladimir Robles-Bykbaev et al. [8]	✗	✗	✗	✗
García et al. [15]	✗	✓	✗	✗
Chuchuca-Ménde et al. [7]	✗	✗	✗	✓
Robles-Bykbaev et al.[21]	✗	✗	✓	✓
Robles-Bykbaev et al. [22]	✗	✗	✓	✗
Martín-Ruiz et al. [16]	✗	✓	✓	✓
Martín Ruiz et al. [17]	✗	✗	✓	✓
Usip et al. [20]	✗	✗	✓	✗

The review of existing works clearly highlights persistent gaps in the representation of speech disorders using ontologies. Firstly, many studies fail to adopt standardized classifications, such as the International Classification of Diseases (ICD-11)[2]. More concerning, however, is that none of the analyzed resources are openly shared or freely accessible. It appears that these ontologies are primarily designed to meet specific project goals rather than to adhere to FAIR principles, which would ensure their reusability by other researchers and institutions. Table 1 summarizes the key characteristics of the reviewed works. From this brief analysis, it becomes clear that several gaps remain to be addressed.

3. The Talkidz Project

The ontology presented in this work is part of a larger project that aims to automatically transcribe children's pathological language. The recent advancements in Natural Language Processing (NLP) and Machine Learning (ML) have significantly impacted various sectors, including the clinical domain. NLP techniques have been applied to the study of both typical language development [25], [26] and atypical language patterns [27], [28]. These technological

[2]https://icd.who.int/en

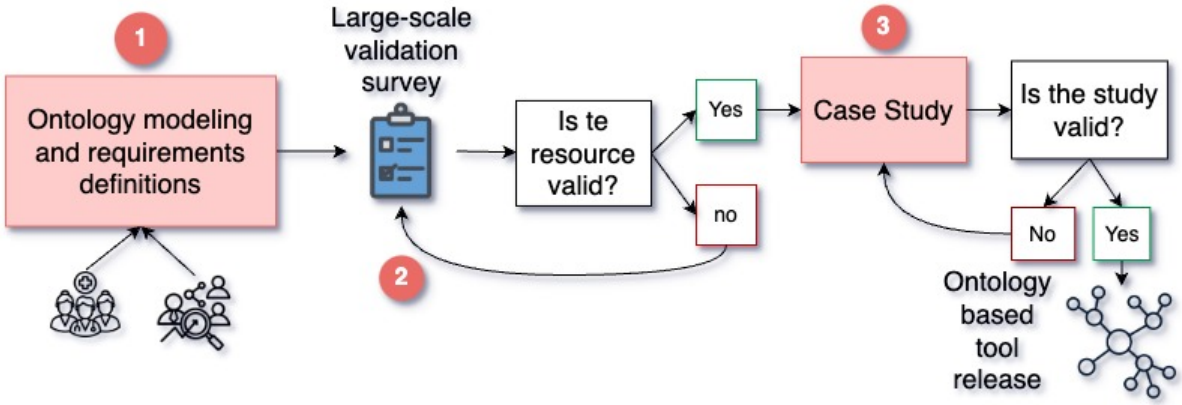

Figure 1: Explanation of the community-based approach we adopt. Our three-fold validation pipeline consists of: (1) building the ontology within a multidisciplinary team, (2) validating it through a survey distributed to domain experts and professionals, and (3) presenting a case study for final validation if the majority of experts consider the resource valid.

advancements have also facilitated the development of NLP and ML based tools to support clinical practice [29]. Despite the promising results of these applications, there is still a lack of models specifically designed for the automatic analysis of pathological language in the Italian language. Currently, the evaluation of expressive language disorders is often conducted using paper tests, with no digital support available to help clinicians reduce scoring time and result analysis. In recent years, the use of software for language analysis has gained traction, as these tools can significantly decrease correction time, making the process faster, automated, and more accurate [30]. Starting from these premises, the Talkidz Project[3] aims to automate and accelerate the analysis of language assessment tests by providing an automatic transcription of pathological language in the International Phonetic Alphabet (IPA) using Deep Learning Techniques. Additionally, it preserves errors and automatically generates phonetic and phonological, lexical and morphosyntactical analyses of the processed language. In the context of the explained Project, the ontology presented in this work aims to support and enhance the tool offered to speech therapist, providing a software capable of processing and automatically analyzing patients' atypical speech. Additionally, it includes an ontology that integrates data from various medical studies, aiming to offer statistical insights and a comprehensive overview of speech and language disorders and their distribution.

4. The Talkidz Ontology: Design

The Talkidz Ontology was developed by a multidisciplinary team of experts across various fields, including speech therapists, occupational therapists, ontology engineers, machine learning experts, and human-machine interaction experts. This diverse expertise facilitated a comprehensive development process, ensuring that the ontology addressed multiple perspectives, from real clinical needs to ontology alignment, as well as case studies and practical applications. In the following section, we detail the ontology's development and validation process, including the design methodology and the formulation of competency questions.

4.1. Design Methodology

When developing an ontology in the healthcare domain, a fundamental challenge is deciding whether to model the main classes "from scratch" or to adhere to international standards. The

[3]The project "Talkidz: Software for Language Analysis", funded by Fondazione CRT (Ordinary Grants Call 2023, RF = 106465 / 2023.1740)

Table 2
The table provide an overview of the competency question developed for the presented resource and the average scores they obtained in the validation survey and the standard deviation.

Competency Question	Avg. scores	Std dev.
Which therapeutic programs are most effective for a specific disorder?	3,61	± 0,66
What are the at-risk groups that could benefit from early treatment?	3,85	± 0,38
What are the differences in gender, age, socioeconomic status, and geographical origin that may influence the onset of language disorders?	3,14	± 0,86
What are the correlations between the onset of language disorders and certain clinical conditions?	3,42	± 0,99
What are the correlations between the onset of language disorders and linguistic exposure (e.g.,shared reading, bilingualism, preschool attendance)?	3,64	± 0,66
Are there correlations between language and motor, social, or emotional skills?	3,61	± 0,51
Are there relationships between language skills and cognitive development (e.g., IQ scores)?	3,14	± 0,71
What is the distribution of linguistic indices (distribution of phoneme classes, mean length of utterance, percentage of errors in a language sample) in the general population?	2,92	± 1,16
What are the timing and frequency of early language habilitation programs?	3,64	± 0,66

choice between standardized and customized classes remains an open issue in healthcare ontology development [31], as both approaches offer distinct advantages and disadvantages. Spoladore and colleagues [31] found that, in the development of disability-related ontologies, the majority of reviewed works (26 out of 43) adopted a "from-scratch" approach. This choice was primarily driven by the need to better address the empirical requirements of the modeled resources. Such ontologies rarely aim to holistically represent an entire domain, instead focusing on a limited set of classes, allowing them to narrow down a complex domains into a more well-defined scope. In addiction the authors highlighted that the reuse of existing ontologies is not hindered by a "from scratch" approach, indeed it is documented in more than 50% of the articles analyzed. Aware of this and of the benefits that participatory design can bring to the development of computational resources [32], we structured our design methodology into multiple steps (Figure 1). First, we discussed the ontology design with two domain experts from a local foundation that deals with children speech pathology and occupational therapy. Our methodology then consists in conducting a survey to validate the proposed classes taxonomy. Specifically, we presented a taxonomy of speech disorders based on ICD-11 and speech therapy classifications based on possible clinical needs. We then asked experts to assess its acceptability, and if they found it unsuitable, they were invited to explain why and suggest alternatives so as to expand and enrich our model. By building on a standardized framework while incorporating expert input, we tailored our design onto the needs of those who will use the resource.
The survey was then filled by a total of 21 speech therapist aged between 24 and 66 years old, 85,5% were female and 14,5% male. Regarding the validation of classes we asked experts to validate the categorization of language disease and therapies ("Regarding the categorization of language disorders proposed, do you think the presented classification reflects the one most commonly used in clinical practice?", "Regarding the categorization of treatment types, do you think the presented classification reflects the one most commonly used in clinical practice?"). As for the first question, 73.3% responded positively, while 26.7% answered negatively and suggested modifications. Regarding the second question, concerning the categorization of treatment, all respondents confirmed the acceptability of the proposed taxonomy for clinical uses. Given that the majority of experts we consulted validated it, we decided to retain both classifications as originally proposed, setting the goal of further analyzing the discrepancy emerged from the answers to the first question as future work.

4.2. Competency Questions

When developing an ontology from scratch, one of the first challenges is defining its requirements [33]. In this context, competency questions (CQs) serve to clarify the ontology's purpose, scope, and target users. Following an participatory approach, we engaged experts and key stakeholders to validate a preliminary list of potential uses and applications for the ontology. Trough the survey we also ask experts to evaluate each presented CQ on a scale from 0 to 4, assessing its relevance, utility, and potential applications. From the survey, we obtained homogeneous results: 8 out of 9 of the CQs we presented were scored higher than 3. In Table 2 2 we present all the competency questions with the obtained scores.

5. Ontology Description

As specified below, the main aim of this resource is twofold. On one hand, we aim to build a semantically enriched resource that addresses the real needs of stakeholders, and on the other, to aimed for compliance with FAIR principles. For this reason, when developing classes and relations, we incorporate both standard classifications (such as speech and language disorders from ICD-11) and a taxonomy more tailored to clinical needs. In the following sections, we will provide a detailed description of the resource's structure, review the main classes, present a validation strategy, and discuss a case study.

5.1. Core Model

The Talkidz Ontology model is structured to provide a clear and comprehensive resource that supports extensibility and practical application in clinical settings. The Talkidz Ontology includes six top-level disjoint classes, namely `Person`, `LinguisticArea`, `SpeechError`, `SpeechOrLanguageDisorder`, `SpeechTherapy`, and `Severity`.

The classification of persons within the top class `Person` (`Patient`, `Therapist`, and `AdultInCharge` with its further specializations), and the relations over these classes, including caring (inChargeOf) and treatment (isBeingTreatedBy) allow for indexing the responsibilities and interactions between every part involved in the speech therapy. This class can be aligned with the `Agent` class in foundational ontologies such as Dolce [34].

The top class `LinguisticArea`, modeled as an enumeration of individuals representing the various linguistic levels affected by disorders, allows mapping the range of speech language disorders (`SpeechOrLanguageDisorder`) onto the specific language areas they affect through the relation affects.

The class `SpeechOrLanguageDisorder`, representing the language disorders, is aligned with the Language Disorder class of the Disease Ontology [19], which provides a standardized reference framework for multiscale biomedical data integration and analysis across thousands of clinical, biomedical, and computational research studies.

The `Severity` top class (enumeration class), standardizes disorder classification according to qualitatively different degrees and aids in treatment prioritization. It is linked trough the relation hasSeverity to the class (`SpeechOrLanguage Disorder`), thus supporting queries aimed at understanding the incidence of specific speech disorders in a subset of (filtered and anonymized) patients (answering a need identified in a CQ, see Table 2).

`SpeechError` subclasses (a type of `syndrome` in Disease Ontology) capture distinct categories of articulation and cognitive speech issues, and through the property makesErrors (linking each speech disorder with a specific `Patient`), facilitates targeted searches for therapeutic purposes.

With the support of domain experts, we designed the class `SpeechTherapy` (a `Plan` in Dolce) and its subclasses trough a "from-scratch" approach, aiming at creating a taxonomy that could represent the real clinical needs. The pivotal class is linked to the `Therapist` through the property appliesTherapy, to the `LanguageDisorder` trough the property isTreatedWithTherapy

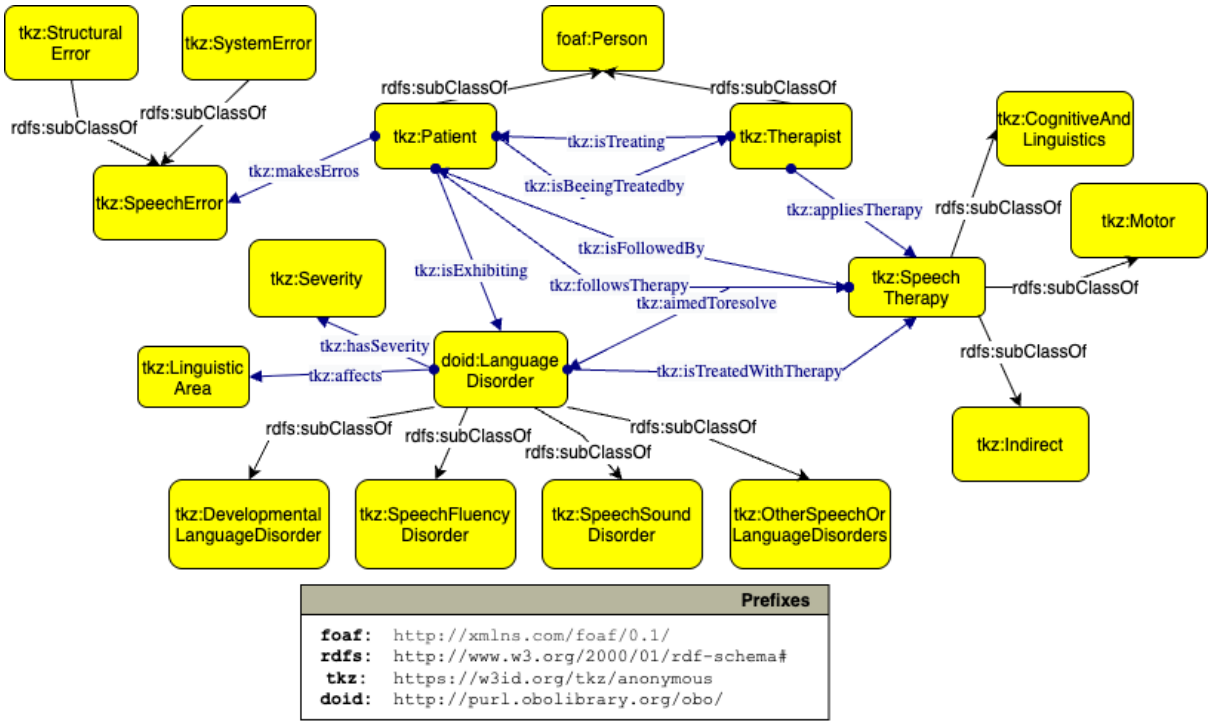

Figure 2: Graphic representation of ontology top classes and relations.

and to the `Patient` through the property isFollowedBy. We decided not to introduce specific Role classes. However, after discussing this issue with the therapists, we chose to handle separately the cases in which a single individual may be both a patient and a therapist.

5.2. Ontology Validation

To ensure that our ontology can effectively support research in child Speech Therapy, and due to the fact that synthetic datasets in the last years have been used in computer science domain with different applications and promising results [35], we created for testing purposes a synthetic dataset designed to mimic real-world clinical data. The generated data were modeled for validation purposes based on real data stored in the databases of a speech therapy clinic, with the aim of providing a data structure that is as realistic as possible while addressing real needs. This dataset consists of anonymized, artificially generated patient records, with distributions of speech disorders modeled after incidence rates reported in the literature. To enhance realism, speech therapists have validated these distributions ensuring they align with clinical expectations. By testing our SPARQL queries and hypotheses against this dataset, we assessed the ontology's ability to retrieve relevant aggregated information and evaluated different therapeutic approaches, all while maintaining full data privacy.

5.3. Use Case: Speech Therapists in Clinical Decision-Making

To demonstrate the applicability, we present a use case illustrating its role in assisting therapists with clinical recommendations. This scenario will showcase how the ontology integrates data to provide a structured view of speech disorders, therapies and patient progress. A speech therapist from a speech therapy clinic is evaluating a 5-year old child (`Patient`) who has been accepted due to concerns about speech development. The child's caregiver (`AdultInCharge`) is recorded, and evidences the child inability of pronouncing such and such words. The therapist aims to diagnose the condition, determine the severity, and design an intervention plan based on evidence-based practices. Figure 3 shows the modeling of the following case study through

ontology classes and properties. The structured process followed by the therapist is described as follows:

1. Assessment And Detection:
    - The patient undergoes the standardized language assessment test, identifying errors to record in any subclass of `Speech Error`.
    - The speech therapist then maps the error to the `LinguisticArea` affected by the disorder.
    - Based on test results, the patient is diagnosed wit a `Developmental LanguageDisorder`, a subclass of `SpeechOrLanguageDisorder`.
    - The `Severity` is classified for each disorder and mapped to the patient's description.
2. Therapy Selection:
    - The speech therapists suggest possible treatment options and links, through the `AimedToResolve` property, the speech disorder to an appropriate `SpeechTherapy` type (e.g.: the `CognitiveLinguisticTherapy` is chosen, which is usually aligned with treating `DevelopmentalLanguage` Disorder.
    - The system records that the `Therapist` appliesTherapy to the `Patient`.
3. Monitoring and Adaptation:
    - Over time, therapy sessions and patient progress are recorded, updating values like the therapy start and end date data properties (therapyStartDate and therapyEndDate) or the severity (hasSeverity).
    - If new occurrences of `SpeechError` emerge or the progress is slower than expected, alternative therapy strategies can be recommended based on the alternatives listed in the Ontology. However, the adaptation is not currently modeled in the ontology as it requires an apparatus for dealing with temporal aspects.

This use case highlights how our Ontology supports speech therapists during the treatment process, providing comprehensive information for clinical assessments and interventions.

6. Conclusion and Future Work

In this work, we presented a semantic resource designed to represent the domain of language disorders with the dual goal of assisting speech pathologists in the detection phase and providing large-scale epidemiological insights. We reviewed previous efforts in the semantic representation of language disorders and highlighted a critical gap: most existing resources lack compliance with FAIR principles and fail to undergo a comprehensive validation process. To address this shortcoming, we developed the presented Ontology as part of a larger project aimed at supporting speech therapists in clinical practice.

Our approach embraces a community-based methodology, engaging experts throughout different phases, from defining the ontology's taxonomy to designing the case study. The result is a semantically enriched resource that not only aligns with standardized frameworks but also responds to real-world needs, offering a valuable tool for specialists in the field. As a limitation we are aware that the pilot validation survey we proposed is not representative and we plan to further investigate our taxonomy and competency questions on a larger scale, involving more experts from different countries and considering the various approaches that could be taken into account during language therapy. We also acknowledge that a semantic resource of this kind cannot claim to be fully representative or universally applicable across all cases. For this reason, we emphasize the importance of a collaborative approach to minimize the risk of developing a resource that, while perfectly aligned with international standards and taxonomies, could ultimately fail to address real-world needs.

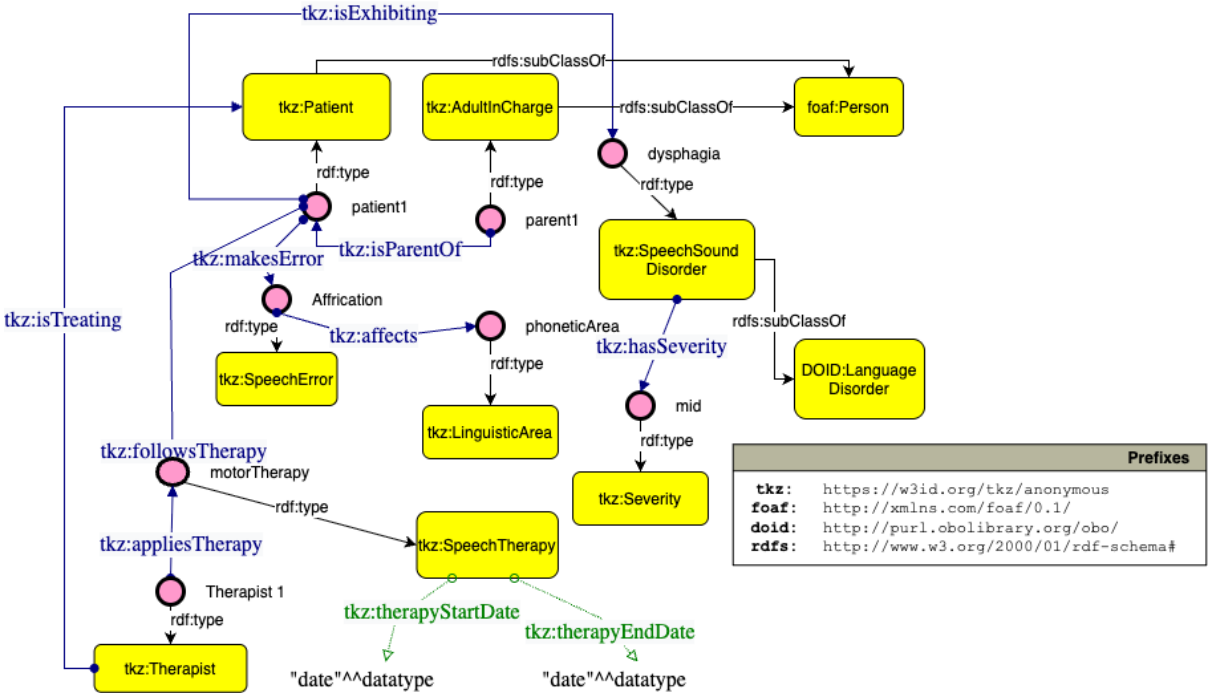

Figure 3: Example of the presented case study modeled with the Ontology classes

In this line, we strongly support multidisciplinary collaboration and believe that researchers from different fields must continue to work together to develop a realistic and practical resource for speech therapy. As future work, following a review phase, we plan to update the ontology by adapting its classes and relations to the needs that will emerge from field use. This alignment should necessarily involve the integration in the ontology of the capability to express temporal aspects, needed for the monitoring of therapies. Additionally, aligning our Ontology with other resources in the healthcare domain, such as those related to cognitive or motor disorders, would be an interesting direction, aiming to provide an even more comprehensive perspective.

7. Limitations

This contribution presents a preliminary phase of a broader project that aims to support speech therapists in daily work and in patient monitoring. However, this paper has some limitations that we plan to address in future work. First, regarding alignment with the FAIR principles, while the ontology is available online, compliance with these principles is not fully detailed. As future work, we intend to apply and document the FAIR principles more thoroughly. Secondly, we acknowledge that the current case study may be limited, as the syntactic dataset and SPARQL queries are not publicly available. We plan to conduct and describe in detail a second case study with real usage examples to strengthen our validation. As future work, we plan to enrich this contribution by presenting additional case studies that demonstrate the potential of the proposed resource.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] S. K. Narayanasamy, K. Srinivasan, Y.-C. Hu, S. K. Masilamani, K.-Y. Huang, A contemporary review on utilizing semantic web technologies in healthcare, virtual communities, and ontology-based information processing systems, Electronics 11 (2022) 453.

[2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[3] A. B. Haque, B. Arifuzzaman, S. A. N. Siddik, A. Kalam, T. S. Shahjahan, T. Saleena, M. Alam, M. R. Islam, F. Ahmmed, M. J. Hossain, Semantic web in healthcare: a systematic literature review of application, research gap, and future research avenues, International Journal of Clinical Practice 2022 (2022) 6807484.

[4] F. Okikiola, A. Ikotun, A. Adelokun, P. Ishola, A systematic review of health care ontology, Asian J Res Comput Sci 5 (2020) 15–28.

[5] H. B. Elhadj, F. Sallabi, A. Henaien, L. Chaari, K. Shuaib, M. Al Thawadi, Do-care: A dynamic ontology reasoning based healthcare monitoring system, Future Generation Computer Systems 118 (2021) 417–431.

[6] D. Spoladore, E. Pessot, Collaborative ontology engineering methodologies for the development of decision support systems: Case studies in the healthcare domain, Electronics 10 (2021) 1060.

[7] F. Chuchuca-Méndez, V. Robles-Bykbaev, P. Vanegas-Peralta, J. Lucero-Saldaña, M. López-Nores, J. Pazos-Arias, An educative environment based on ontologies and e-learning for training on design of speech-language therapy plans for children with disabilities and communication disorders, in: IEEE CACIDI 2016-IEEE Conference on Computer Sciences, IEEE, 2016, pp. 1–6.

[8] V. Robles-Bykbaev, J. J. Pazos-Arias, M. López-Nores, J. García-Duque, J. Ochoa-Zambrano, Modelling domain knowledge of speech and language therapy with an owl ontology and openehr archetypes, in: International Conference on Health Informatics, volume 2, SCITEPRESS, 2015, pp. 585–591.

[9] P. Rinaldi, A. Bello, I. Simonelli, M. C. Caselli, Is specific learning disorder predicted by developmental language disorder? evidence from a follow-up study on italian children, Brain Sciences 13 (2023) 701.

[10] G. Conti-Ramsden, M. C. St Clair, A. Pickles, K. Durkin, Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence (2012).

[11] G. Conti-Ramsden, P. L. Mok, A. Pickles, K. Durkin, Adolescents with a history of specific language impairment (sli): Strengths and difficulties in social, emotional and behavioral functioning, Research in developmental disabilities 34 (2013) 4161–4169.

[12] M. C. St Clair, A. Pickles, K. Durkin, G. Conti-Ramsden, A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (sli), Journal of communication disorders 44 (2011) 186–199.

[13] A. Sansavini, M. E. Favilla, M. T. Guasti, A. Marini, S. Millepiedi, M. V. Di Martino, S. Vecchi, N. Battajon, L. Bertolo, O. Capirci, et al., Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. a scoping review, Brain sciences 11 (2021) 654.

[14] T. J. Spaulding, E. Plante, K. A. Farinella, Eligibility criteria for language impairment, Language, Speech, and Hearing Services in Schools 37 (2006) 61–72.

[15] M. J. S. García, S. V. González, I. H. P. Torres, C. P. de Celis Herrero, A diagnostic tool for speech disorders based on nlp with ontological reasoning., Res. Comput. Sci. 147 (2018) 75–86.

[16] M. L. Martín-Ruiz, M. Á. V. Duboy, I. P. de la Cruz, Deployment and validation of a smart system for screening of language disorders in primary care, Sensors 13 (2013) 7522–7545.

[17] M. L. Martín Ruiz, M. A. Valero Duboy, C. Torcal Loriente, I. Pau de la Cruz, Evaluating a web-based clinical decision support system for language disorders screening in a nursery school, Journal of medical Internet research 16 (2014) e139.
[18] S. Crosbie, A. Holm, B. Dodd, Intervention for children with severe speech disorder: a comparison of two approaches, International Journal of Language & Communication Disorders 40 (2005) 467–491.
[19] L. M. Schriml, J. B. Munro, M. Schor, D. Olley, C. McCracken, V. Felix, J. A. Baron, R. Jackson, S. M. Bello, C. Bearer, et al., The human disease ontology 2022 update, Nucleic acids research 50 (2022) D1255–D1261.
[20] P. U. Usip, M. E. Ekpenyong, Towards ontology-driven application for multilingual speech language therapy, Human Language Technologies for Under-Resourced African Languages: Design, Challenges, and Prospects (2018) 85–101.
[21] V. E. Robles-Bykbaev, W. Guamán-Murillo, D. Quisi-Peralta, M. López-Nores, J. J. Pazos-Arias, J. García-Duque, An ontology-based expert system to generate therapy plans for children with disabilities and communication disorders, in: 2016 ieee ecuador technical chapters meeting (etcm), IEEE, 2016, pp. 1–6.
[22] V. Robles-Bykbaev, M. López-Nores, J. Pazos-Arias, D. Quisi-Peralta, J. García-Duque, An ecosystem of intelligent ict tools for speech-language therapy based on a formal knowledge model, in: MEDINFO 2015: eHealth-enabled Health, IOS Press, 2015, pp. 50–54.
[23] V. Robles-Bykbaev, M. Guamán-Heredia, Y. Robles-Bykbaev, J. Lojano-Redrován, F. Pesántez-Avilés, D. Quisi-Peralta, M. López-Nores, J. Pazos-Arias, Onto-speltra: A robotic assistant based on ontologies and agglomerative clustering to support speech-language therapy for children with disabilities, in: Advances in Computing: 12th Colombian Conference, CCC 2017, Cali, Colombia, September 19-22, 2017, Proceedings 12, Springer, 2017, pp. 343–357.
[24] M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, A. Gangemi, Introduction: Ontology engineering in a networked world, Springer, 2012.
[25] B. MacWhinney, Understanding spoken language through talkbank, Behavior research methods 51 (2019) 1919–1927.
[26] K. Sagae, A. Lavie, B. MacWhinney, Automatic measurement of syntactic development in child language, in: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), 2005, pp. 197–204.
[27] J. Boorse, M. Cola, S. Plate, L. Yankowitz, J. Pandey, R. T. Schultz, J. Parish-Morris, Linguistic markers of autism in girls: evidence of a "blended phenotype" during storytelling, Molecular autism 10 (2019) 1–12.
[28] I. Chojnicka, A. Wawer, Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction, PLoS One 15 (2020) e0229985.
[29] L.-Y. Guo, S. Eisenberg, N. B. Ratner, B. MacWhinney, Is putting sugar (sampling utterances of grammatical analysis revised) into language sample analysis a good thing? a response to pavelko and owens (2017), Language, Speech, and Hearing Services in Schools 49 (2018) 622–627.
[30] M. J. Pezold, C. M. Imgrund, H. L. Storkel, Using computer programs for language sample analysis, Language, Speech, and Hearing Services in Schools 51 (2020) 103–114.
[31] D. Spoladore, M. Sacco, A. Trombetta, A review of domain ontologies for disability representation, Expert Systems with Applications 228 (2023) 120467.
[32] F. Delgado, S. Yang, M. Madaio, Q. Yang, The participatory turn in ai design: Theoretical foundations and the current state of practice, in: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, 2023, pp. 1–23.
[33] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The neon methodology for ontology engineering, in: Ontology engineering in a networked world, Springer, 2011, pp. 9–34.
[34] S. Borgo, R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. M. Sanfilippo,

L. Vieu, Dolce: A descriptive ontology for linguistic and cognitive engineering, Applied ontology 17 (2022) 45–69.

[35] A. Bauer, S. Trapp, M. Stenger, R. Leppich, S. Kounev, M. Leznik, K. Chard, I. Foster, Comprehensive exploration of synthetic data generation: A survey, arXiv preprint arXiv:2401.02524 (2024).