

Status Quo der Entwicklungen von Ontologien Rhetorischer Figuren in Englisch, Deutsch und Serbisch

Ramona Kühn und Jelena Mitrović

Universität Passau
{ramona.kuehn,jelena.mitrovic}@uni-passau.de

1 Einleitung und Motivation

Rhetorische Figuren sind seit Jahrhunderten Gegenstand der Untersuchung in den Bereichen der Linguistik und der Rhetorik. Diese Figuren, wie zum Beispiel Metapher, Ironie, Alliteration und viele weitere, können als Abweichung vom normalen Sprachgebrauch verstanden werden (Fahnestock, 1999). Die daraus resultierende Wirkung ist, dass Texte oder Worte, die rhetorische Figuren enthalten, einprägsamer, überzeugender oder auch beeinflussender sein können. Wir möchten dabei Figuren unterscheiden, die sprachunabhängig funktionieren, wie zum Beispiel die Wiederholung eines Wortes, oder die Verwendung des gleichen Anfangslauts (z. B. bei der Figur Alliteration) und Figuren, die auf einem gemeinsamen kulturellen Verständnis und Hintergrundwissen basieren (z. B. Metaphern wie “jemandem das Herz brechen”) und oft nicht wörtlich, sondern nur im übertragenen Sinne verstanden werden können. Besonders bei Übersetzungen zwischen zwei Sprachen ist es immens wichtig, die Nuancen zu erkennen und richtig zu interpretieren.

Bei der Analyse dieser übertragenen Figuren können kultur- und sprachübergreifende Vergleiche angestellt werden, um Gemeinsamkeiten und Unterschiede in ihrer Verwendung und Wirkung über verschiedene kulturelle Kontexte hinweg zu untersuchen. Auch können neue Einblicke in menschliches Verhalten oder die soziale Dynamik innerhalb bestimmter Gemeinschaften gewonnen werden. Um dies zu ermöglichen, benötigt man große Datenmengen, die von Expert:innen annotiert werden. Diese manuelle Arbeit ist jedoch sehr zeit- und kostenintensiv. Darüber hinaus sind versierte Fachkräfte auf diesem Forschungsgebiet vergleichsweise selten anzutreffen. Daher ist es wichtig, dass rhetorische Figuren maschinell annotiert werden können. Eine maschinenlesbare Darstellung und Beschreibung rhetorischer Figuren ist daher notwendig. Häufig werden dafür formale, domänenspezifische Ontologien verwendet.

In diesem Artikel gehen wir der Frage nach, wie der aktuelle Status Quo der Ontologie-Entwicklung rhetorischer Figuren ist. Unser Fokus liegt dabei vor allem auf englisch-, serbisch- und deutschsprachigen Modellen. Wir geben zuerst einen Überblick über die Entwicklung und Kategorisierung von rhetorischen Figuren, um die Notwendigkeit für Ontologien hervorzuheben. Anschließend präsentieren wir aktuelle Ontologien in Serbisch, Deutsch und Englisch,

sowie eine mehrsprachige Ontologie, die eine Vereinigung der vorigen Ontologien darstellt. Am Ende des Artikels geben wir einen Ausblick, wie die Zukunft bezüglich automatisierter Erkennung rhetorischer Figuren im Kontext großer Sprachmodelle (“Large Language Models”) aussehen könnte.

2 Einblicke in die Beschreibungen Rhetorischer Figuren und ihrer Kategorisierung

Bereits Aristoteles, Sokrates und Plato beschäftigten sich mit den Grundlagen der Rhetorik und ihrer Überzeugungsfähigkeit in Argumentationen. Die erste konkrete Beschreibung rhetorischer Figuren findet sich im Buch IV der Schrift “Rhetorica ad Herennium”, dessen unbekannter Autor das Werk ca. 80 v. Chr. verfasst hat. Die Figuren werden dort in die zwei Gruppen “Figuren der Diktion” (zum Beispiel Epanaphora, Antithesis, Klimax, Metapher) und “Gedankenfiguren” (zum Beispiel Vergleich, Personifikation, Untertreibung, oder lebendige Beschreibung) eingeteilt (Caplan und Winterbottom, 2016). Man kann jedoch in dieser Einteilung kein einheitliches Schema erkennen. Fahnestock (1999) kritisiert, dass zusätzlich indirekt eine dritte Gruppe mit Ausnahmen eingeführt wird, sodass diese Einteilung in Gruppen inkonsistent und sogar verwirrend erscheint.

Etwa zur gleichen Zeit entstand Ciceros Werk “De inventione”, welches sich zum einen mit Grundbegriffen der Rhetorik und zum anderen mit Argumentationstechniken befasst. Es ähnelt in vielen Bereichen der Schrift Rhetorica ad Herennium. Quintilian wagt in seiner Schrift “Institutio Oratoria” (Quintilian und Butler, 1922) einen neuen Versuch, rhetorische Figuren zu kategorisieren. Er verwendet dabei die Kategorien aus dem Werk Rhetorica ad Herennium, führt aber eine Unterscheidung zwischen “Figuren” und “Tropen” ein. Bereits in dieser Schrift fällt auf, dass es verschiedene Kategorisierungen und Meinungen gibt, die sich nur schwer in ein einheitliches Format bringen lassen. Dieses Dilemma zieht sich durch die Jahrhunderte hinweg, von Richard Whatelys “Elements of Rhetoric” im Jahre 1828 über Kenneth Burkes rhetorische Theorien (1945-1966), bis Lausberg (1960) mit seinem “Handbuch der literarischen Rhetorik” einen neuen Versuch wagt. Das Buch ist eine enorme Sammlung von Erkenntnissen der vorigen Jahrhunderte und versucht, eine Systematik zu etablieren. Mit Hilfe mehrsprachiger Beispiele sollen den Leser:innen rhetorische Figuren und ihre Eigenheiten verdeutlicht werden. Jedoch merkt man in diesem Werk, dass es schwierig ist, allgemeingültige Standards in einem Gebiet zu etablieren, das es seit nahezu 2000 Jahren nicht geschafft hat, Kategorien ohne Widersprüche zu formulieren. Möchte man dies modellieren, wäre es notwendig, die verschiedenen Datenquellen und Definitionen zu integrieren. Außerdem müsste eine Modellierung gerade abstrakt genug sein, um auch konkurrierende Kategorien abbilden zu können, aber dennoch spezifisch genug, um die Eigenheiten der verschiedenen Figuren reflektieren zu können.

3 Ontologien zur Standardisierung und ihre Relevanz für die Digital Humanities

Wie soll man es nun schaffen, aus diesem Wirrwarr der Kategorien einen einheitlichen Standard zu entwickeln, der zugleich maschinenlesbar ist? Warum sollte dies überhaupt nötig sein? Rhetorische Figuren spielen auch in unserer heutigen Zeit eine große Rolle, sei es bei politischen Reden, in der Werbung oder in sozialen Medien. Denn auch heute haben sie die gleiche Wirkung wie vor tausenden Jahren: Sie können die Leserschaft überzeugen, Argumente plausibler erscheinen lassen, oder eine Botschaft einprägsamer machen. Da eine manuelle Analyse zu zeit- und kostenintensiv ist, wird eine computergestützte Erkennung immer wichtiger. Dafür sind allerdings eine einheitliche Benennung, ein gemeinsames Verständnis, sowie ein Set an Regeln erforderlich.

Ontologien sind hierfür das geeignete Mittel, da sie eine formale, einheitliche Struktur darstellen, die sowohl von Menschen als auch Maschinen lesbar und verarbeitbar ist. Sie können Daten aus heterogenen Quellen integrieren (Rehbein, 2017) und bestehende Wissensbestände vereinen (Mladenović et al., 2014). Sie dienen der Wissensrepräsentation und -organisation. Eine semantische Verarbeitung wird dadurch ermöglicht (Kaiya und Saeki, 2006). Zudem helfen sie, ein gemeinsames Verständnis der vorhandenen domänenspezifischen Informationen zu vermitteln (Noy et al., 2001). Zusammenhänge können durch die graphische Darstellung der Relationen in Form eines Netzwerkes noch leichter erkannt werden.

Konsistenz, Validität und das Einhalten logischer Regeln lassen sich mit Ontologien ebenfalls überprüfen. Ontologien haben bereits in vielen Bereichen bewiesen, dass sie unterschiedliche Definitionen und Beschreibungen vereinheitlichen und Sachverhalte konsistent modellieren können, zum Beispiel in der Medizin (Hu, 2006) oder im Katastrophenschutz (Daher et al., 2022). Die wohl bekannteste Ontologie im Bereich der Digital Humanities ist das CIDOC Conceptual Reference Model (Bruseker et al., 2017), welches Objekte im Bereich des kulturellen Erbes beschreibt. Ein anderes Projekt in diesem Gebiet ist beispielsweise "GOLEM" (Pianzola et al., 2023), welches mit Hilfe von Ontologien kulturelle Merkmale der Online-Fiktion in verschiedenen Sprachen miteinander vergleicht. Langmead et al. (2016) untersuchen die Relevanz von Ontologien in den Digital Humanities mit einem Fokus auf historische Netzwerke der Frühen Neuzeit.

Im Bereich rhetorischer Figuren sind Ontologien wichtige Werkzeuge für Linguist:innen und Informatiker:innen, die ein tieferes Verständnis dafür entwickeln wollen, wie Sprache in bestimmten Kontexten oder Genres verwendet wird. Rhetorische Figuren sind komplexe sprachliche Phänomene, die in verschiedenen Texten und Sprachen auftreten. Die Ontologien vereinen dabei die Vielzahl an heterogenen Datenquellen, die es im Bereich der rhetorischen Figuren gibt. Eine Ontologie rhetorischer Figuren ermöglicht die systematische Erfassung und Strukturierung des Wissens über diese Figuren. Sie hilft dabei, die verschiedenen Figuren, ihre Definitionen, Eigenschaften und Beziehungen zueinander in einer formalen und einheitlichen Weise darzustellen. Ontologien, in denen ver-

wandte rhetorische Figuren nach ihren Konstruktionsregeln oder Eigenschaften gruppiert sind, dienen dabei unter anderem als Nachschlagewerk und Sammlung des Wissens über rhetorische Figuren. Darüber hinaus können sie in der Erkennung rhetorischer Figuren in Texten unterstützen und dabei helfen, Annotationsrichtlinien zu erstellen (Reiter, 2020). Diese Annotationen können dann für das Training von großen Sprachmodellen verwendet werden, um eine automatische Identifizierung in Texten zu ermöglichen.

Die Bedeutung von Ontologien für rhetorische Figuren in den Digital Humanities liegt in ihrer Fähigkeit, semantische Strukturen und Beziehungen zwischen den Figuren zu erfassen und zu organisieren. Zudem können solche Ontologien dazu beitragen, das Textverständnis zu steigern, da die darin erkannten Figuren direkt mit ihrer Wirkung verknüpft werden können, die sie auf die Zielgruppe haben. Zudem können kultur- und sprachübergreifende Analysen durchgeführt werden, da bestimmte Figuren in manchen Sprachen eventuell häufiger verwendet werden als in anderen. Hinzu kommt, dass Definitionen mancher Figuren oder sogar Figuren selbst sprachabhängig existieren und dadurch Gemeinsamkeiten und Unterschiede in den Sprachen und über verschiedene kulturelle Kontexte hinweg untersucht werden können.

4 Status Quo der Entwicklungen von Ontologien Rhetorischer Figuren und ihre Anwendungen

Im Folgenden stellen wir verschiedene Ontologien rhetorischer Figuren und ihre Entwicklung vor. Dabei gehen wir nur auf Ontologien ein, die mehrere verschiedene rhetorische Figuren abbilden und sich nicht auf eine einzelne Figur fokussieren. Den Beginn der Entwicklung solcher Ontologien markiert das sogenannte RhetFig Projekt (Kelly et al., 2010), dessen Ziel es war, ein theoretisches Konzept für eine mögliche Modellierung von Ontologien rhetorischer Figuren zu erstellen. Dabei werden rhetorische Figuren in drei Gruppen kategorisiert, nämlich Tropen, Schema und Chroma. Zudem werden linguistische Domänen (z. B. morphologisch, phonologisch) eingeführt, die jeder Figur zugeordnet werden. Dieses Konzept diente als Basis für die serbische RetFig (ohne “h”) Ontology (Mladenović und Mitrović, 2013). In dieser wird die Einteilung in linguistische Domänen und Gruppen fortgeführt. 98 serbische Figuren wurden dabei formal in der W3C Web Ontology Language (OWL) modelliert und beschrieben. Wang et al. (2021) entwickelten dann die Ploke Ontologie, die Figuren der perfekten lexikalen Wiederholung modelliert. Diese Ontologie basiert ebenfalls auf der Grundstruktur des RhetFig Projekts. In dieser Ontologie werden zusätzlich neurokognitive Eigenschaften eingeführt, die von einer Wiederholungsfigur ausgelöst werden (z. B. “Attention Effect”). Eine deutsche Ontologie für rhetorische Figuren stellt die GRhOOT Ontologie (Kühn et al., 2022) dar, die wiederum auf der serbischen RetFig Ontologie basiert, aber insgesamt 110 verschiedene Figuren modelliert. Dafür wurde für jede Figur einzeln untersucht, ob sie im Deutschen ebenfalls existiert und ob die Definitionen gleich sind. Dabei wurde festgestellt, dass 12 serbische Figuren so nicht im Deutschen existieren (z. B. die serbische

rhetorische Figur “Wunsch” hat kein Pendant im Deutschen). Circa 60 Figuren konnten gleich modelliert werden. Allerdings gab es auch Fälle wie die Figur “Syllepsis”, deren deutsche Definition besagt, dass ein Worтеlement weggelassen wird, während bei der serbischen Figur “Silepsa” ein Worтеlement hinzugefügt wird. Inspiriert durch diese sprachlichen Unterschiede wurde die “Multilingual Ontology” (Wang et al., 2022) modelliert, die versucht, Ploke Figuren – also Figuren mit perfekter lexikaler Wiederholung – der serbischen, deutschen und englischen Ontologie zu vereinen. Wie wir bereits anfangs festgestellt haben, sollten diese Figuren sprachlich unabhängig sein. Jedoch wurden dabei ebenfalls Unterschiede festgestellt. Beispielsweise lautet die Definition der Figur Epizeuxis im Deutschen, dass ein Wort oder Wortgruppe mindestens drei Mal wiederholt wird. Die englische Definition spezifiziert jedoch nicht näher, wie oft ein Wort wiederholt werden muss, wodurch bereits eine zweimalige Wiederholung auf die Figur Epizeuxis hindeuten kann.

Diese Beispiele zeigen deutlich, dass existierende Ontologien rhetorischer Figuren nicht einfach in andere Sprachen übersetzt werden können. Für jede einzelne Figur muss verglichen werden, ob sie so in der Zielsprache existiert und die Definition gleich ist. Zudem müssen die Figuren gefunden werden, die es in der Sprache der existierenden Ontologie nicht gibt, allerdings in der Zielsprache als rhetorische Figur angesehen wird. Genau so wurde verfahren, um eine englische Ontologie zu modellieren, die mehrere verschiedene Figuren abbildet. Mit dieser sogenannten ESTHER Ontologie (Kühn et al., 2023) wurde das englische Äquivalent zur deutschen und serbischen Ontologie geschaffen. Sie ist eine Weiterentwicklung der deutschen GRhOOT Ontologie und beinhaltet zusätzlich Wissen aus weiteren englischen Quellen. Auch wurden hierarchische Abhängigkeiten und Co-Lokationen zwischen verschiedenen Figuren modelliert, zum Beispiel, dass die Figur Symploke stets auch eine Anapher und Epipher ist, oder dass eine Klimax aus der Kombination von Gradatio und Inkrementum gebildet wird (O’Reilly et al., 2018). Eine zeitliche Übersicht der hier genannten Ontologien findet sich in Abbildung 1.

Doch wie sollen diese Ontologien nun dazu beitragen, große Datenmengen zu analysieren und rhetorische Figuren darin automatisch zu identifizieren? Kühn und Mitrovic (2023) zeigen dafür einige Anwendungsbeispiele auf. Sie nennen als größtes Problem für die automatisierte Erkennung das Fehlen annotierter Datensätze. Die Ontologien sollen dabei unterstützen, diese Hürde zu überwinden, um dann mit annotierten Daten große Sprachmodelle für diese spezielle Aufgabe zu trainieren. Konkret können die Ontologien dazu genutzt werden, Personen in der manuellen Annotation zu schulen, um die wenigen Expert:innen darin zu unterstützen. Da die Ontologien in OWL geschrieben sind, können die Daten mit der Anfragesprache SPARQL angefragt werden. Zudem beschreiben Kühn und Mitrovic (2023) das theoretische Konzept einer grafische Benutzeroberfläche, die ein niedrigschwelliges Angebot zur Annotation rhetorischer Figuren für viele Nutzer:innen darstellt. Dabei handelt es sich um eine Website mit Auswahlmöglichkeiten, die die Nutzer:innen aktiv dabei unterstützt, Figuren in einem Text bestimmen zu können. Die Anfragen im Hintergrund basieren da-

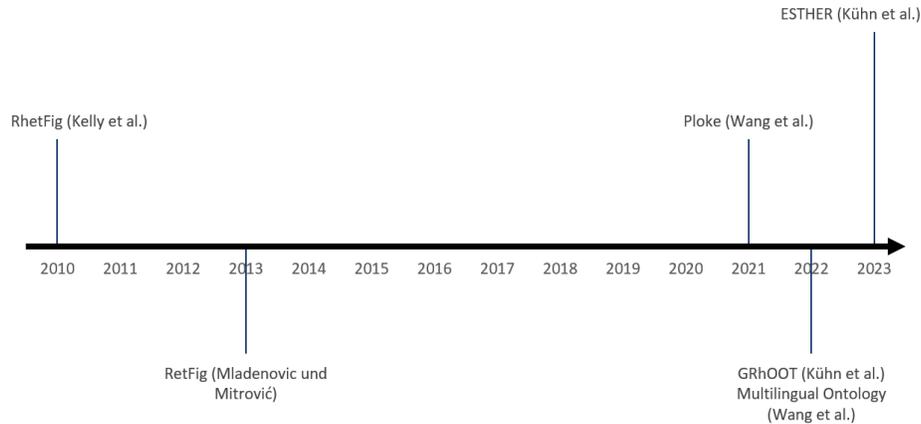


Abb. 1. Zeitliche Übersicht der Ontologie-Entwicklungen.

bei auf SPARQL-Anfragen an die jeweilige Ontologie. Diese Oberfläche befindet sich allerdings noch in der Entwicklung. Da die serbische RetFig, die deutsche GRhOOT und die englische ESTHER Ontologie auf der gleichen Architektur basieren, kann dieses Tool jedoch dann sprachübergreifend eingesetzt werden. Zusätzlich können Gamification Elemente das Angebot noch attraktiver für eine größere Nutzerschaft machen.

5 Zusammenfassung und Ausblick

Wir haben gezeigt, wie aktuelle Entwicklungen bezüglich maschinenlesbarer und formaler Ontologien im Bereich rhetorischer Figuren voranschreiten. Dadurch, dass die Ontologien alle aufeinander aufbauen aber dennoch unterschiedliche Sprachen betrachten, ist eine strukturelle Kompatibilität gegeben.

Die in den Ontologien spezifizierten Konstruktionsregeln rhetorischer Figuren können Sprachmodelle trainiert werden. Diese können dabei auch dazu dienen, Sätze oder ganze Texte mit gewünschten rhetorischen Figuren zu generieren. Somit könnte der Mangel an Daten von gewissen Figuren in diesem Bereich überwunden werden. Außerdem muss erforscht werden, inwiefern Sprachmodelle bereits jetzt rhetorische Figuren in Texten oder Sätzen erkennen und korrekt annotieren können. Sie würden damit auch eine kostengünstige Alternative zur manuellen Annotation schaffen. Bis dies jedoch möglich ist, brauchen wir unsere Ontologien und manuelle Bemühungen, denn die Sprachmodelle sind nur so gut wie die verfügbaren Trainingsdaten. Dazu gehören ethische Überlegungen wie die Vermeidung von Bias und die Gewährleistung der Transparenz bei der Generierung von Texten. Außerdem können Sprachmodelle nicht ohne menschliche Überprüfung als alleinige Quelle der Wahrheit betrachtet werden, da sie

auf vorherigen Textdaten trainiert werden und daher deren Vorurteile und Einschränkungen reflektieren können.

Wenn wir jedoch in der Lage sind, rhetorische Figuren automatisiert zu erkennen und zu identifizieren, bekommen wir einen besseren Einblick in die Art und Weise, wie wir sowohl in schriftlicher als auch in mündlicher Form effektiv kommunizieren - etwas, das zweifellos im Laufe der Zeit relevant bleiben wird, unabhängig von den technologischen Fortschritten, die in den kommenden Jahren gemacht werden.

Acknowledgment

SPONSORED BY THE



The project on which this report is based was funded by the German Federal Ministry of Education and Research (BMBWF) under the funding code 01—S20049. The author is responsible for the content of this publication.

Literaturverzeichnis

- G. Bruseker, N. Carboni, und A. Guillem. Cultural heritage data management: The role of formal ontology and cidoc crm. *Heritage and archaeology in the digital age: acquisition, curation, and dissemination of spatial cultural heritage data*, Seiten 93–131, 2017.
- H. Caplan und M. Winterbottom. Rhetorica ad herennium. In *Oxford Research Encyclopedia of Classics*. 2016.
- J. B. Daher, T. Huygue, P. Stolf, und N. J. Hernandez. An ontology and a reasoning approach for evacuation in flood disaster response. In *17th International Conference on Knowledge Management: "Knowledge, Uncertainty and Risks: From individual to global scale" (ICKM 2022)*, Seiten 117–131. UNT Digital Library, 2022.
- J. Fahnestock. *Rhetorical figures in science*. Oxford University Press, USA, 1999.
- X. Hu. Natural language processing and ontology-enhanced biomedical literature mining for systems biology. In *Computational systems biology*, Seiten 39–56. Elsevier, 2006.
- H. Kaiya und M. Saeki. Using domain ontology as domain knowledge for requirements elicitation. In *14th IEEE International Requirements Engineering Conference (RE'06)*, Seiten 189–198. IEEE, 2006.
- A. R. Kelly, N. A. Abbott, R. A. Harris, C. DiMarco, und D. R. Cheriton. Toward an ontology of rhetorical figures. In *Proceedings of the 28th ACM International Conference on Design of Communication*, Seiten 123–130, 2010.
- R. Kühn und J. Mitrović. Multilingual domain ontologies of rhetorical figures and their applications. *UniDive 1st General Meeting, Paris*, 2023.
- R. Kühn, J. Mitrović, und M. Granitzer. Esther: Ontology of rhetorical figures in english. *Proceedings of the Join Ontology Workshops. To be published.*, 2023.
- A. Langmead, J. M. Otis, C. N. Warren, S. B. Weingart, und L. D. Zilinski. Towards interoperable network ontologies for the digital humanities. *International Journal of Humanities and Arts Computing*, 10(1):22–35, 2016.
- H. Lausberg. *Handbuch der literarischen Rhetorik*, volume 2. Hueber München, 1960.
- M. Mladenović und J. Mitrović. Ontology of rhetorical figures for serbian. In *Text, Speech, and Dialogue: 16th International Conference, TSD 2013, Pilsen, Czech Republic, September 1-5, 2013. Proceedings 16*, Seiten 386–393. Springer, 2013.
- M. Mladenović, J. Mitrović, und C. Krstev. Developing and maintaining a wordnet: Procedures and tools. In *Proceedings of the Seventh Global Wordnet Conference*, Seiten 55–62, 2014.
- N. F. Noy, D. L. McGuinness, et al. *Ontology development 101: A guide to creating your first ontology*, 2001.
- C. O'Reilly, Y. Wang, K. Tu, S. Bott, P. Pacheco, T. W. Black, und R. A. Harris. Arguments in gradatio, incrementum and climax; a climax ontology. In *Pro-*

- ceedings of the 18th workshop on Computational Models of Natural Argument. Academic Press, 2018.*
- F. Pianzola, X. Yang, N. Visser, M. van der Ree, und A. van Cranenburgh. Constructing the golem: Graphs and ontologies for literary evolution models. In *Digital Humanities 2023. Collaborations as Opportunity (DH2023)*. Graz, Austria, 2023.
- Quintilian und H. E. Butler. *The Institutio Oratoria*, volume 8. Heinemann, 1922.
- M. Rehbein. Ontologien. *Digital Humanities: Eine Einführung*, Seiten 162–176, 2017.
- N. Reiter. Anleitung zur Erstellung von Annotationsrichtlinien. *Nils Reiter/Axel Pichler/Jonas Kuhn (Hg.), Reflektierte Algorithmische Textanalyse. Interdisziplinäre (s) Arbeiten in der CRETA-Werkstatt, Berlin/Boston*, Seiten 193–201, 2020.
- Y. Wang, R. A. Harris, und D. M. Berry. An ontology for plope: Rhetorical figures of lexical repetitions. In *JOWO*, 2021.
- Y. Wang, R. Kühn, R. A. Harris, J. Mitrovic, und M. Granitzer. Towards a unified multilingual ontology for rhetorical figures. In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. Valletta, Malta: SCITEPRESS-Science and Technology Publications*, Seiten 117–127, 2022.