|  |
| --- |
| DeNpKongru – Ein Erkenner zur Bestimmung der nominalen Kongruenz in deutschen Nominalphrasen aus Lernerkorpora |
| **Christopher Michael Chandler** |
| Straße  Postleitzahl, Stadt  [christohper.chandler@rub.de](mailto:christohper.chandler@rub.de) |

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Eu mi bibendum neque egestas congue quisque egestas diam in. Ante in nibh mauris cursus mattis. Risus commodo viverra maecenas accumsan lacus vel facilisis. Sagittis id consectetur purus ut. Convallis tellus id interdum velit laoreet id. Eget felis eget nunc lobortis mattis. Tempus iaculis urna id volutpat lacus laoreet non. Imperdiet massa tincidunt nunc pulvinar sapien et. Sed ullamcorper morbi tincidunt ornare massa eget egestas. Vitae turpis massa sed elementum. Turpis egestas sed tempus urna et pharetra pharetra massa. Venenatis tellus in metus vulputate eu scelerisque. Venenatis urna cursus eget nunc scelerisque viverra. Sed ullamcorper morbi tincidunt ornare massa eget egestas. Vitae turpis massa sed elementum. Turpis egestas sed tempus urna et pharetra pharetra massa. Venenatis tellus in metus vulputate eu scelerisque. Venenatis urna cursus eget nunc scelerisque viverra.

1. Einleitung

Kongruenz als eine grammatikalische Notwendigkeit im Sinne von Satzkorrektheit ist in vielen Sprachen vertreten. Das gilt beispielsweise bei europäischen Sprachen (Englisch, Deutsch, Englisch, Isländisch , Polnisch, Russisch, Spanisch), semitischen Sprachen (Arabisch, Hebräisch), uto-akztekischen Sprachen (Hopi) und auch bei isolierten Sprachen (Baskisch) (Corbett, 2003; Hofherr, 2019). Man kann also daraus schließen, dass die Kongruenz an sich ein sprachübergreifendes Phänomen ist.

Kongruenz im Allgemeinen bezeichnet das regelhafte linguistische Phänomen, dass bestimmte linguistische Elemente, wie z.B. Verben, Nomen, etc. mit anderen linguistischen Elementen, wie z.B. Pronomen, Adjektiven, etc. innerhalb einer linguistischen Äußerung miteinander übereinstimmen müssen (Corbett, 2003; Hofherr, 2019). Um das näher definieren zu können, kann man also in diesem Fall dann von morphologischer Kongruenz sprechen.

Da Kongruenz ein so stark präsentes Phänomen ist, ist es von Bedeutung festzulegen, wie man sie am besten beschreiben und untersuchen kann, denn Kongruenz bereitet nicht nur Nicht-Muttersprachlern große Schwierigkeiten, sondern auch Sprachsystemen, wo z.B. die Semantik-, oder Syntaxerkennung oft von Kongruenz abhängig ist (Corbett, 2003; Spinner & Juffs, 2008).

Kongruenz kann verschiedene Ausprägungen annehmen und wie diese Ausprägungen ausfallen hängt direkt von der jeweiligen Sprache ab. Deswegen lässt sich nicht verallgemeinert sagen, wie genau die Kongruenz auszusehen hat und welche Formen es geben soll, dennoch gibt es ein paar Kongruenzarten, die hier aufgeführt werden, um diese zu veranschaulichen.

Um die Kongruenz und deren Merkmale klarer zu untersuchen wird als Objektsprache das Deutsche herangezogen, da diese Sprache auch im Laufe des Projekts als Objektsprache diente.

Bei beispielsweise Subjekt-Verb-Kongruenz müssen Subjekt und Verb bezüglich Numerus und Kasus miteinander übereinstimmen. Genus spielt hierbei keine Rolle (Hofherr, 2019).

|  |  |
| --- | --- |
| Kongruenz | Nicht-Kongruenz |
| Ich gehe | Ich gehst\* |
| 1 SG.NOM gehen.pres1 SG | 2 SG.NOM  gehen.pres1 SG |

1. Beispiel der Subjet-Verb-Kongruenz

*Anmerkung*. In Anlehnung an *Agreement Morphology*, von P. C. Hofherr, 2019, S. 2. Copyright durch Oxford Research Encyclopedia of Linguistics.

Diese Form von Kongruenz, die in Tabelle 1 veranschaulicht wurde, ist in einigen Sprachen (Französisch, Spanisch, Russisch) vorhanden (Hofherr, 2019). Somit dürfte sie auch den meisten Sprechern der erwähnten Sprachen vertraut vorkommen.

Eine andere Variante der Kongruenz ist die Kongruenz bei Nomen, wo Numerus, Genus und Kasus in einer linguistischen Phrase alle eine Rolle spielen. Diese wird infolge als nominale Kongruenz bezeichnet. Bei der nominalen Kongruenz spielt der Kopf der Nominalphrase eine entscheidende Rolle, denn alle betroffenen linguistischen Elemente müssen mit dem Kopf kongruieren (Hofherr, 2019).

|  |  |
| --- | --- |
| Kongruenz | Nicht-Kongruenz |
| Der kleine Hund | Die kleine Hund\* |
| 1 SG.NOM bestimmt.mask.sing Artikel + mask.sing.nomen + mask.sing.nominativ adjketiv | 1 SG.NOM bestimmt.fem.sing Artikel + fem.sing.nomen + fem.sing.nominativ adjketiv |

1. Beispiel der nominalen Kongruenz

*Anmerkung*. In Anlehnung an *Agreement Morphology*, von P. C. Hofherr, 2019, S. 2. Copyright durch Oxford Research Encyclopedia of Linguistics.

Wenn ein Element in einer Nominalphrase fehlt oder nicht richtig eingesetzt wurde, ist die gesamte Äußerung falsch. Im Vergleich zu anderen Sprachen wie z.B. Englisch, Italienisch, Französisch, weil so ein Kasussystem fehlt, gilt die deutsche Sprache als bei Nicht-Muttersprachlern bzw. Lernenden als Schwierig (Spinner & Juffs, 2008).

Um die genauen Schwierigkeiten in Bezug auf nominale Kongruenz im Deutschen bestimmen zu können, werden Texte von Anglophonen und Frankophonen aus dem Lernerkorpus Merlin (Wisniewski et al., 2018) extrahiert. Gewählt wurden diese beiden Gruppen unter der Annahme, dass Frankophonen am Anfang eher weniger Schwierigkeiten bei nominaler Kongruenz als Anglophone haben, da Kongruenz, auch wenn weniger komplex, im Französischen stärker präsent ist als im Englischen. Es wird jedoch erwartet, dass diese Differenz mit steigendem Niveau abnimmt.

Die Ergebnisse sollen Auskunft darüber liefern, inwieweit man Strategien oder Ansätze entwickeln kann, um Lernenden effektivere Lernmethoden bezüglich des Erwerbs der nominalen Kongruenz im Deutschen anbieten zu können.

1. Verwandte Literatur

Die Kernthematik bezieht sich hier auf morphologische Kongruenz, insbesondere Kongruenz bei Nominalphrasen bzw. nominale Kongruenz. In folgenden Kapiteln wird ein regelbasiertes Vorgehen aufgestellt, jedoch gibt es schon LSTM-basierte Ansätze (Ravfogel et al., 2018), um Kongruenz bei Phrasen, wo nominale Kongruenz und (Subjekt)-Verb-Kongruenz eine Rolle spielen. Das RNN-Model benutzt lokale Heuristiken, die aus einer baskischen Phrase extrahiert werden können, um die Kongruenz zu bestimmen. Nichts destotrotz sind fundierte grammatikalische Kenntnisse der Sprache unerlässlich, wenn die Kongruenz bestimmt werden sollte (Ravfogel et al., 2018).

Bei Sprachen, wo die morphologischen Merkmale wesentlich weniger Komplex ausweisen, wie z.B. Englisch wurde auch Kongruenzbestimmung an einem LSTM-Ansatz erprobt. LSTMs weisen zwar eine sehr hohe Accuracy bzw. Genauigkeit auf, dennoch gilt dies nur bei Äußerungen einer gewissen Länge, wo strukturelle und sequentielle Informationen nicht miteinander harmonieren (Linzen et al., 2016).

1. Nominale Kongruenz

Die Kongruenzart, die im Rahmen dieses Projekts untersucht werden soll, ist die nominale Kongruenz. Das Ziel ist es, automatisch bestimmen zu können, ob eine Nominalphrase korrekt bzw. kongruent oder inkorrekt bzw. nicht-kongruent.

Eine Nominalphrase gilt als korrekt, wenn alle seine Elemente mit dem Kopf der Nominalphrase kongruieren bzw. übereinstimmen. Gleichfalls gilt eine Kongruenz nicht, wenn mindestens ein Element nicht mit dem Kopf kongruiert.

Termini

Um die Kongruenz genauer untersuchen und beschreiben zu können, müssen bestimmte Termini festgelegt werden:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Domäne | |  |  |  |  |
|  | Ziel |  | Kontroller |  |  |  |
|  | Das blaue |  | Haus |  |  | Bedingung |
|  |  |  |  |  |  |  |
|  | Merkmal:  Numerus, Genus, Kasus  Wert: Singular, Neutrum, Nominativ | | |  |  |  |

1. Umriss der Kongruenz-Termini

*In Anlehnung* an *Agreements: Terms and Boundaries,* von C. Greville, S. 2., 2003. Copyright durch SMG conference papers.

Der Kontroller bestimmt die Kongruenz und das Ziel muss mit den Eigenschaften des Kontrollers übereinstimmen. Die Domäne ist die syntaktische Umgebung, worin die Nominalphrase sich befindet. Merkmale sind die Eigenschaften der Kongruenz und es gibt Bedingungen außerhalb der Domäne, die Auswirkungen auf die Kongruenz haben (Corbett, 2003; Hofherr, 2019).

Intention des Sprechers

Obwohl Zielhypothesen der einzeln Texte gemacht wurde, in dem eine korrigierte Version festgelegt wurde (Boyd et al., 2014), wird die Intention des Sprechers bei der Bestimmung der Nominalphrasenarten nicht berücksichtigt. Der Hauptgrund dafür ist, dass diese sich nicht so leicht automatisch bestimmen lässt und es könnte mehrere Möglichkeit als richtig gelten:

|  |  |
| --- | --- |
| Satzarten | Sätze |
| Lernersatz | Ich sehe die Hund\* |
| Zielhypothese 1 | Ich sehe den Hund |
| Zielhypothese 2 | Ich sehe die Hunde |

1. Zielhypothesen

Wenn man beispielsweise, wie in Tabelle 4, annimmt, dass die Nominalphrase Ich sehe die Hund\* falsch ist, hat man zwei Möglichkeiten, den Satz zu korrigieren. Ohne Weiteres ist die Intention des Lerners unbekannt und somit wird sie hier nicht beachtet. Der Satz wird als solcher behandelt. Es ist daher möglich, die Annahme zu vertreten, dass man den Satz als solchen behandeln soll.

1. Vorgehensweise

In diesem Projekt ging es hauptsächlich darum, die Kongruenz der einzelnen Nominalphrasen zu bestimmen, die aus bestimmten Lernerkorpustexten extrahiert wurden. Nominalphrasen bestimmen zu können setzt voraus, dass der Eingangstext schon entsprechend tokenisiert und nach Wortklasse getaggt wurde. Diese Aufgabe wurde im Vorfeld von dem ParZu-Tagger (Sennrich et al., 2009)

übernommen, sodass die Texte im Rahmen des

Projekts nicht geparst werden müssen.

Folgende Wortklassen nach dem STTS-Schema (Schiller et al., 1999) sind für die Nominalphrasen relevant: NOUN, PREP, ART und ADJA. Ausgeschlossen wurden alle anderen Wortklassen.

Nach dem erfolgreichen Parsen wurden die daraus stammenden Conll-Dateien nach geeigneten Nominalphrasen untersucht. Geeignete bedeutet an dieser Stelle, dass mindestens ein Nomen in einer Äußerung unbestimmter Länge vorkommen muss.

Die Nominalphrasen wurden extrahiert und in einer gesonderten Datei gespeichert. Ein Nominalphrase-Eintrag hat dann beispielsweise die folgende Form:

|  |  |
| --- | --- |
| Nominalphrase | meine Fahrkarte |
| Morphologische Information | meine ART Fem|Acc|Sg,  Fahrkarte N Fem|Acc|Sg |
| Satz | Ich gebe Dir meine Fahrkarte . |

1. Nominalphraseneintrag

Nominalphrasen können verschiedene Formen annehmen und können somit unterschiedlichen Kategorien zugeordnet. Die Nominalphrasen wurden erst in unterschiedliche Kategorien eingeteilt. Innerhalb dieser Kategorien wurden einfache Heuristiken angewandt, um die Kongruenz bestimmen zu können. Die Kategorien kann man Tabelle 6 entnehmen.

|  |  |
| --- | --- |
| Arten | Beispiel |
| Einfach | Stadt |
| ART | Das Leben /  Welche Freizeitaktivitäten, |
| PREP | Mit Kindern  Für neuen Aufgaben\*,  In der Woche |
| Rechtschreibfehler | Reche Pfliche\*,  Wiviel\* |
| Eigennamen | Katharina,  Maria Meier |
| Redewendungen | Liebe Julia,  Mit freundlichen Grüßen |

1. Kongruenzarten

Nachdem einer Nominalphrase einer Kategorie zugeordnet wurden, wurde sie innerhalb dieses Kontexts auf Kongruenz untersucht. Dazu wurden Kongruenz-codes vergeben, sodass es festgelegt werden kann, ob Kongruenz vorhanden ist und wenn ja und welche Art.

|  |  |
| --- | --- |
| Code | Bedeutung |
| 0 | EINFACH |
| 1 | ART |
| 2 | PREP |
| 3 | Eigennamen |
| 4 | Redewendung |

1. Kongruenz-Codes

Wenn die Kongruenz innerhalb einer Kategorien nicht vorhanden ist, wird die Nominalphrase mit einem Code versehen, der angibt, dass die Kongruenz nicht zutrifft.

|  |  |
| --- | --- |
| Code | Bedeutung |
| 10 | EINFACH |
| 11 | ART |
| 12 | PREP |
| 9 | Unbekannt |

1. Nicht-Kongruenz-Codes
2. Datensatz

Der Merlin-Korpus wurde als Datensatz für das Projekt verwendet. Es ein fehlerannotiertes schriftliches Lernerkorpus für Deutsch, Italienisch und Tschechisch (Wisniewski et al., 2018). Lerner unterschiedliche Niveaus nach dem Gemeinsamer europäischer Referenzrahmen für Sprachen bzw. CEFR haben Texte nach einer vorgegebener Aufgabe z.B. Alltagsszenarien oder Arbeitsszenarien verfasst (Wisniewski et al., 2018). Diese wurden dann anschließend bewertet und ggf. korrigiert.

Die Texte, die in italienischer und tschechische Sprache verfasst wurden, werden in diesem Projekt nicht berücksichtigt. Deswegen werden nur die deutschen Texte analysiert, die von Lernern geschrieben wurden.

|  |  |  |
| --- | --- | --- |
| Fair CEFR Level | | Deutsch |
| Anfänger | A1 | 57 |
| A2 | 199 |
| A2+ | 107 |
| Selbständig | B1 | 219 |
| B1+ | 115 |
| B2 | 219 |
| B2+ | 73 |
| Fortgeschritten | C1 | 42 |
| C2 | 4 |
| Textanzahl | | 1,035 |
| Wortanzahl | | 125,927 |

1. Merlin-Korpus-Textanzahl

Da nur 4 Texte auf C2 Niveau verfasst wurden, werden diese Texte auch nicht berücksichtigt. Es bleiben also Texte von A1 bis C1 zu analysieren.

Aufbau der einzelnen Einträge

Um die Texte nachher auswerten und sortieren zu können, wurde jeder Text mit Metadaten versehen.

Zusätzlichen zum Lernertext wurden für einige Texte in diesem Datensatz eine korrigierte Version bzw. eine Zielhypothese angelegt (Boyd et al., 2014), jedoch wurde diese bei der Analyse im Rahmen dieses Projekts nicht berücksichtigt.

Projektdatensatz

Zu den Muttersprachen der Teilnehmer gehören die folgenden Sprachen: Arabisch, Chinesisch, Tschechisch, Englisch, Französisch, Ungarisch, Italienisch, Polnisch, Portugiesisch, Russisch, Spanisch und Türkisch. Einige Sprecher haben entweder ihre Muttersprache nicht angegeben oder sie wurde nicht erfasst.

Relevant für dieses Projekt sind die Texte, die von Anglophonen und Frankophonen verfasst wurden.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A1 | A2 | B1 | B2 | C1 | Gesamt |
| English | 15 | 1 | 11 | 9 | 10 | 46 |
| French | 5 | 1 | 2 | 19 | 12 | 39 |
| Gesamt | 20 | 2 | 13 | 28 | 22 | 85 |

Damit die Datensätze einigermaßen gleich aufgeteilt sind, wird nur ein Bruchteil der Daten benutzt

Tabelle

1. Evaluation

Da es sich hier um ein regelbasiertes Model und nicht um ein statistisches Model handelt, musste das Programm an sich nicht trainiert werden.

In jedem Bereich waren nicht ge

Jedoch wurden die Textdaten so aufgeteilt, dass 50 % als Trainingsdaten dienten, um die Algorithmen zu verbessern und die restlichen 50% als tatsächliche Testdaten.

DeNpKongru-Performanz

Die Performanz des Programms wurde anhand Sci-kit learn durchgeführt. Es wurden 6 Dateien als Referenz-Dateien verwendet, woraus die Gold-Dateien manuell erstellt wurden.

Trainingsphase

Während der Trainingsphrase wurden die folgenden Textsorten behandelt.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A1 | A2 | B1 | B2 | C1 | Gesamt |
| English | 2 | 0 | 1 | 4 | 5 | 12 |
| French | 2 | 0 | 1 | 4 | 5 | 12 |
| Gesamt | 4 | 0 | 1 | 8 | 10 | 24 |

Testphase

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A1 | A2 | B1 | B2 | C1 | Gesamt |
| English | 2 | 1 | 1 | 4 | 5 | 13 |
| French | 2 | 1 | 1 | 4 | 5 | 13 |
| Gesamt | 4 | 2 | 2 | 8 | 10 | 26 |

1. Diskussion

Im Rahmen des Projekts wurden Frankophonen und Anglophonen wurden mit einander verglichen. Es hat sich herausgestellt, dass

Im Allgemeinen lässt sich schwer sagen, ob so eine Methode geeignet ist, um Differenzen zwei Sprechergruppen festzulegen. Zum einen ist der Korpus dafür zu klein und zum anderen waren die Nominalphrasen, die daraus entstanden sind, aus einigen Gründen zum Teil nicht aussagekräftig genug

1. Fazit

Solche Experimente durchzuführen ist auf jeden Fall hilfreich, da

Danksagung

Dieses Projekt wurde im Rahmen des computerlinguistischen Kurses

*Korpuslinguistische Analysen der Nominalflexion im Deutschen (050041-SoSe23)*an der Ruhr-Universität Bochum in Deutschland entwickelt. Ideen und Code von den Mitkommilitonen und der Kursleiterin Frau Dr. Ronja Laarman-Quante wurden hier verwendet und waren bei der Planung und Ausführung des Projekts ausschlaggebend.

Literaturverzeichnis

Boyd, A., Hana, J., Nicolas, L., Meurers, D., Wisniewski, K., Abel, A., Schöne, K., Štindlová, B., & Vettori, C. (2014). The MERLIN corpus : Learner language and the CEFR. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, 1281‑1288. http://www.lrec-conf.org/proceedings/lrec2014/pdf/606\_Paper.pdf

Corbett, G. (2003). Agreement : Terms and boundaries. *SMG conference papers*.

Długosz, K. (2022). Processing gender agreement in an additional language : The more languages the better? *Second Language Research*, 02676583221113333. https://doi.org/10.1177/02676583221113333

Gonen, H., Kementchedjhieva, Y., & Goldberg, Y. (2019). How Does Grammatical Gender Affect Noun Representations in Gender-Marking Languages? *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 463‑471. https://doi.org/10.18653/v1/K19-1043

Hofherr, P. C. (2019). *Agreement Morphology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.500

Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Transactions of the Association for Computational Linguistics*, *4*, 521‑535. https://doi.org/10.1162/tacl\_a\_00115

Ravfogel, S., Goldberg, Y., & Tyers, F. (2018). Can LSTM Learn to Capture Agreement ? The Case of Basque. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 98‑107. https://doi.org/10.18653/v1/W18-5412

Sá-Leite, A. R., Luna, K., Fraga, I., & Comesaña, M. (2020). The gender congruency effect across languages in bilinguals : A meta-analysis. *Psychonomic Bulletin & Review*, *27*(4), 677‑693. https://doi.org/10.3758/s13423-019-01702-w

Schiller, A., Teufel, S., Stöckert, C., & Thielen, C. (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Universität Tübingen. http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf

Sennrich, R., Schneider, G., Volk, M., & Warin, M. (2009). *A New Hybrid Dependency Parser for German*. https://doi.org/10.5167/uzh-25506

Spinner, P., & Juffs, A. (2008). L2 grammatical gender in a complex morphological system : The case of German. *Iral-international Review of Applied Linguistics in Language Teaching - IRAL-INT REV APPL LINGUIST*, *46*, 315‑348. https://doi.org/10.1515/IRAL.2008.014

Wisniewski, K., Abel, A., Vodičková, K., Plassmann, S., Meurers, D., Woldt, C., Schöne, K., Blaschitz, V., Lyding, V., Nicolas, L., Vettori, C., Pečený, P., Hana, J., Čurdová, V., Štindlová, B., Klein, G., Lauppe, L., Boyd, A., Bykh, S., & Krivanek, J. (2018). *MERLIN Corpus | Resources for research and practice related to foreign language learning* (1.1) [jeu de données]. https://www.merlin-platform.eu/

(Boyd et al., 2014; Corbett, 2003; Długosz, 2022; Gonen et al., 2019; Hofherr, 2019; Ravfogel et al., 2018; Sá-Leite et al., 2020; Spinner & Juffs, 2008; Wisniewski et al., 2018)

Eigenständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Ich erkläre weiterhin, dass ich alles gedanklich, inhaltlich oder wörtlich von anderen (z.B. aus Büchern, Zeitschriften, Zeitungen, Lexika, Internet usw.) Übernommene als solches kenntlich gemacht, d.h. die jeweilige Herkunft im Text oder in den Anmerkungen belegt habe. Dies gilt ggf. auch für Tabellen, Skizzen, Zeichnungen, bildliche Darstellungen usw.

Ich nehme zur Kenntnis, dass die nachgewiesene Unterlassung der Herkunftsangabe als versuchte Täuschung bzw. als Plagiat („geistiger Diebstahl“) gewertet wird. Unkenntnis der in der Wissenschaft gebräuchlichen Regeln gilt nicht als Entlastung.

Ich anerkenne hiermit, dass bei Vorliegen eines Plagiats die Arbeit nicht als selbständige Leistung gewertet wird mit der Folge, dass mein Anspruch auf einen Leistungsnachweis (auch Teilnahmenachweis) in der Lehrveranstaltung und die Möglichkeit einer Nachbesserung oder Wiederholung der Arbeit entfällt.

Christopher Michael Chandler

Ort, Datum

Anhang

|  |  |
| --- | --- |
|  | Definition |
| Author id | Identifikationsnummer des Texts |
| Test language | Die Sprache, worin der Text verfasst wurde |
| CEFR level of test | Sprachniveau der Aufgabe |
| Task | Die Aufgabe, nach der der Text verfasst wurde |
| Mother tongue | Muttersprache |
| Age | Alter |
| Gender | Geschlecht |
| Overall CEFR rating | Gesamtbewertung des Textes |
| Grammatical accuracy | Grammatikalische Richtigkeit |
| Orthography | Rechtschreibung in Bezug auf Tippfehler, Zeichensetzung, Groß- und Kleinschreibung, Wortgrenzen, etc. |
| Vocabulary range | Lexikalische wissen, lexikalische Fehler bzw. falsche Wortwahl |
| Vocabulary Control | Inwieweit man mit dem eigenen Wortschatz geben kann |
| Coherence/Cohesion | Konnektoren und Aufbau des Textes |
| Sociolinguistic  appropriateness | Anreden, Anfragen innerhalb eines gegebenes Kontexte |

|  |  |
| --- | --- |
|  | Definition |
| Author id | 1023\_0108888 |
| Test language | German |
| CEFR level of test | B2 |
| Task | Au pair writes letter of complaint to Agency - ID: 1023\_100303 |
| Mother tongue | French |
| Age | 27 |
| Gender | Female |
| Overall CEFR rating | B2 |
| Grammatical accuracy | B2 |
| Orthography | B2 |
| Vocabulary range | B2 |
| Vocabulary Control | B1 |
| Coherence/Cohesion | B2 |
| Sociolinguistic  appropriateness | B2 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | general\_mother\_tongue | general\_cefr | txt\_len\_in\_char |
| 1071\_0024802 | English | A1 | 405 |
| 1071\_0024814 | English | A1 | 266 |
| 1071\_0024816 | French | A1 | 209 |
| 1071\_0024691 | French | A1 | 703 |
| Tabelle | Training | Anfänger | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART | EINFACH | PREP |
| 1071\_0024802 | 0 | 1 | 0 |
| 1071\_0024814 | 0 | 1 | 0 |
| 1071\_0024816 | 0 | 1 | 0 |
| 1071\_0024691 | 0 | 1 | 0 |
| Tabelle | Training | Anfänger | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| general\_author\_id | EIGENNAMEN | REDEWENDUNGEN |  |
| 1071\_0024802 | 3 | 0 |  |
| 1071\_0024814 | 3 | 0 |  |
| 1071\_0024816 | 1 | 0 |  |
| 1071\_0024691 | 7 | 0 |  |
| Tabelle | Training | Anfänger | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART\_NICHT | PREP\_NICHT | EINFACH\_NICHT |
| 1071\_0024802 | 0 | 0 | 0 |
| 1071\_0024814 | 0 | 1 | 3 |
| 1071\_0024816 | 0 | 1 | 5 |
| 1071\_0024691 | 0 | 0 | 2 |
| Tabelle | Training | Anfänger | 4 |
|  |  |  |  |
| general\_author\_id | GESAMT\_UNBEKANNT | GESAMT\_WAHR | GESAMT\_FALSCH |
| 1071\_0024802 | 13 | 4 | 0 |
| 1071\_0024814 | 3 | 4 | 4 |
| 1071\_0024816 | 1 | 2 | 6 |
| 1071\_0024691 | 12 | 8 | 2 |
| Tabelle | Training | Anfänger | 5 |
|  |  |  |  |
| AVG\_WAHR\_MEAN | AVG\_WAHR\_MODE |  |  |
| 4,5 | 4 |  |  |
| Tabelle | Training | Anfänger | 6 |
|  |  |  |  |
| AVG\_FALSCH\_MEAN | AVG\_FALSCH\_MODE |  |  |
| 3 | 0 |  |  |
| Tabelle | Training | Anfänger | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_UNBEKANNT\_MEAN | AVG\_UNBEKANNT\_MODE |  |  |
| 7,25 | 13 |  |  |
| Tabelle | Training | Anfänger | 8 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | general\_mother\_tongue | general\_cefr | txt\_len\_in\_char |
| 1061\_0120283 | English | B1 | 260 |
| 1061\_0120343 | French | B1 | 581 |
| 1023\_0101851 | English | B2 | 1165 |
| 1023\_0109401 | English | B2 | 1006 |
| 1023\_0109951 | English | B2 | 1504 |
| 1023\_0101841 | English | B2 | 1440 |
| 1023\_0108888 | French | B2 | 1392 |
| 1023\_0101895 | French | B2 | 1245 |
| 1023\_0109880 | French | B2 | 1172 |
| 1023\_0108752 | French | B2 | 1612 |
| Tabelle | Training | Mittelstuffe | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART | EINFACH | PREP |
| 1061\_0120283 | 0 | 0 | 0 |
| 1061\_0120343 | 0 | 1 | 0 |
| 1023\_0101851 | 0 | 5 | 0 |
| 1023\_0109401 | 0 | 1 | 0 |
| 1023\_0109951 | 0 | 6 | 0 |
| 1023\_0101841 | 0 | 7 | 0 |
| 1023\_0108888 | 0 | 9 | 0 |
| 1023\_0101895 | 0 | 2 | 0 |
| 1023\_0109880 | 0 | 1 | 0 |
| 1023\_0108752 | 0 | 3 | 0 |
| Tabelle | Training | Mittelstuffe | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | EIGENNAMEN | REDEWENDUNGEN |  |
| 1061\_0120283 | 4 | 0 |  |
| 1061\_0120343 | 2 | 0 |  |
| 1023\_0101851 | 8 | 1 |  |
| 1023\_0109401 | 4 | 1 |  |
| 1023\_0109951 | 7 | 1 |  |
| 1023\_0101841 | 6 | 1 |  |
| 1023\_0108888 | 6 | 0 |  |
| 1023\_0101895 | 5 | 1 |  |
| 1023\_0109880 | 7 | 1 |  |
| 1023\_0108752 | 2 | 1 |  |
| Tabelle | Training | Mittelstuffe | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART\_NICHT | PREP\_NICHT | EINFACH\_NICHT |
| 1061\_0120283 | 0 | 2 | 1 |
| 1061\_0120343 | 0 | 1 | 4 |
| 1023\_0101851 | 0 | 4 | 10 |
| 1023\_0109401 | 0 | 3 | 14 |
| 1023\_0109951 | 0 | 3 | 13 |
| 1023\_0101841 | 0 | 6 | 21 |
| 1023\_0108888 | 0 | 4 | 16 |
| 1023\_0101895 | 0 | 10 | 11 |
| 1023\_0109880 | 0 | 3 | 12 |
| 1023\_0108752 | 0 | 4 | 11 |
| Tabelle | Training | Mittelstuffe | 4 |
|  |  |  |  |
| general\_author\_id | GESAMT\_UNBEKANNT | GESAMT\_WAHR | GESAMT\_FALSCH |
| 1061\_0120283 | 3 | 4 | 3 |
| 1061\_0120343 | 10 | 3 | 5 |
| 1023\_0101851 | 16 | 14 | 14 |
| 1023\_0109401 | 16 | 6 | 17 |
| 1023\_0109951 | 18 | 14 | 16 |
| 1023\_0101841 | 20 | 14 | 27 |
| 1023\_0108888 | 17 | 15 | 20 |
| 1023\_0101895 | 17 | 8 | 21 |
| 1023\_0109880 | 17 | 9 | 15 |
| 1023\_0108752 | 41 | 6 | 15 |
| Tabelle | Training | Mittelstuffe | 5 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_WAHR\_MEAN | AVG\_WAHR\_MODE |  |  |
| 9,3 | 14 |  |  |
| Tabelle | Training | Mittelstuffe | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_FALSCH\_MEAN | AVG\_FALSCH\_MODE |  |  |
| 15,3 | 15 |  |  |
| Tabelle | Training | Mittelstuffe | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_UNBEKANNT\_MEAN | AVG\_UNBEKANNT\_MODE |  |  |
| 17,5 | 17 |  |  |
| Tabelle | Training | Mittelstuffe | 8 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | general\_mother\_tongue | general\_cefr | txt\_len\_in\_char |
| 1031\_0003133 | English | C1 | 1320 |
| 1031\_0003132 | English | C1 | 1380 |
| 1031\_0003155 | English | C1 | 1234 |
| 1031\_0003357 | English | C1 | 1503 |
| 1031\_0003356 | English | C1 | 1239 |
| 1031\_0003170 | French | C1 | 1334 |
| 1031\_0002199 | French | C1 | 1404 |
| 1031\_0003136 | French | C1 | 1682 |
| 1031\_0003023 | French | C1 | 2004 |
| 1031\_0001998 | French | C1 | 1122 |
| Tabelle | Training | Fortgeschritten | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART | EINFACH | PREP |
| 1031\_0003133 | 0 | 3 | 0 |
| 1031\_0003132 | 0 | 9 | 0 |
| 1031\_0003155 | 0 | 11 | 0 |
| 1031\_0003357 | 0 | 3 | 0 |
| 1031\_0003356 | 0 | 6 | 2 |
| 1031\_0003170 | 0 | 1 | 2 |
| 1031\_0002199 | 0 | 1 | 0 |
| 1031\_0003136 | 0 | 3 | 0 |
| 1031\_0003023 | 0 | 5 | 2 |
| 1031\_0001998 | 0 | 3 | 0 |
| Tabelle | Training | Fortgeschritten | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | EIGENNAMEN | REDEWENDUNGEN | |
| 1031\_0003133 | 6 | 0 |  |
| 1031\_0003132 | 3 | 0 |  |
| 1031\_0003155 | 3 | 0 |  |
| 1031\_0003357 | 3 | 0 |  |
| 1031\_0003356 | 8 | 0 |  |
| 1031\_0003170 | 5 | 0 |  |
| 1031\_0002199 | 8 | 0 |  |
| 1031\_0003136 | 8 | 0 |  |
| 1031\_0003023 | 7 | 0 |  |
| 1031\_0001998 | 3 | 0 |  |
| Tabelle | Training | Fortgeschritten | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | ART\_NICHT | PREP\_NICHT | EINFACH\_NICHT |
|  |  |  |  |
| 1031\_0003133 | 0 | 0 | 8 |
| 1031\_0003132 | 0 | 3 | 5 |
| 1031\_0003155 | 0 | 0 | 10 |
| 1031\_0003357 | 0 | 5 | 14 |
| 1031\_0003356 | 0 | 6 | 13 |
| 1031\_0003170 | 0 | 3 | 2 |
| 1031\_0002199 | 0 | 4 | 6 |
| 1031\_0003136 | 0 | 3 | 12 |
| 1031\_0003023 | 0 | 2 | 7 |
| 1031\_0001998 | 0 | 1 | 3 |
| Tabelle | Training | Fortgeschritten | 4 |

|  |  |  |  |
| --- | --- | --- | --- |
| general\_author\_id | GESAMT\_UNBEKANNT | GESAMT\_WAHR | GESAMT\_FALSCH |
| 1031\_0003133 | 20 | 9 | 8 |
| 1031\_0003132 | 27 | 12 | 8 |
| 1031\_0003155 | 27 | 14 | 10 |
| 1031\_0003357 | 35 | 6 | 19 |
| 1031\_0003356 | 20 | 16 | 19 |
| 1031\_0003170 | 20 | 8 | 5 |
| 1031\_0002199 | 29 | 9 | 10 |
| 1031\_0003136 | 25 | 11 | 15 |
| 1031\_0003023 | 37 | 14 | 9 |
| 1031\_0001998 | 19 | 6 | 4 |
| Tabelle | Training | Fortgeschritten | 5 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_WAHR\_MEAN | AVG\_WAHR\_MODE | |  |
| 10,5 | 9 |  |  |
| Tabelle | Training | Fortgeschritten | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_FALSCH\_MEAN | AVG\_FALSCH\_MODE | |  |
| 10,7 | 8 |  |  |
| Tabelle | Training | Fortgeschritten | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
| AVG\_UNBEKANNT\_MEAN | AVG\_UNBEKANNT\_MODE | |  |
| 25,9 | 20 |  |  |
| Tabelle | Training | Fortgeschritten | 8 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | P\_VALUE | ALTERNATIVE | DF |
| A1 | 0,79516724 | two.sided | 1 |
| Tabelle | Training | statistical\_data | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | ESTIMATE | CONF\_INT\_LOWER | CONF\_INT\_UPPER |
| A1 | 4 | -39,118614 | 37,1186142 |
| A1 | 5 | -39,118614 | 37,1186142 |
| Tabelle | Training | statistical\_data | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | STATISTIC | METHOD |  |
| A1 | -0,3333333 | Welch Two Sample t-test |  |
| Tabelle | Training | statistical\_data | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | P\_VALUE | ALTERNATIVE | DF |
| A2 | UNK | UNK | UNK |
| Tabelle | Training | statistical\_data | 4 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | ESTIMATE | CONF\_INT\_LOWER | CONF\_INT\_UPPER |
| A2 | UNK | UNK | UNK |
| Tabelle | Training | statistical\_data | 5 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | STATISTIC | METHOD |  |
| A2 | UNK | UNK |  |
| Tabelle | Training | statistical\_data | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | P\_VALUE | ALTERNATIVE | DF |
| B1 | UNK | UNK | UNK |
| Tabelle | Training | statistical\_data | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | ESTIMATE | CONF\_INT\_LOWER | CONF\_INT\_UPPER |
| B1 | UNK | UNK | UNK |
| Tabelle | Training | statistical\_data | 8 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | STATISTIC | METHOD |  |
| B1 | UNK | UNK |  |
| Tabelle | Training | statistical\_data | 9 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | P\_VALUE | ALTERNATIVE | DF |
| B2 | 0,40380036 | two.sided | 5,99376299 |
| Tabelle | Training | statistical\_data | 10 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | ESTIMATE | CONF\_INT\_LOWER | CONF\_INT\_UPPER |
| B2 | 12 | -4,3136329 | 9,31363291 |
| B2 | 9,5 | -4,3136329 | 9,31363291 |
| Tabelle | Training | statistical\_data | 11 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | STATISTIC | METHOD |  |
| B2 | 0,89802651 | Welch Two Sample t-test |  |
| Tabelle | Training | statistical\_data | 12 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | P\_VALUE | ALTERNATIVE | DF |
| C1 | 0,44647517 | two.sided | 7,49721834 |
| Tabelle | Training | statistical\_data | 13 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | ESTIMATE | CONF\_INT\_LOWER | CONF\_INT\_UPPER |
| C1 | 11,4 | -3,4276218 | 7,02762177 |
| C1 | 9,6 | -3,4276218 | 7,02762177 |
| Tabelle | Training | statistical\_data | 14 |

|  |  |  |  |
| --- | --- | --- | --- |
| GRUPPEN\_NAMEN | STATISTIC | METHOD |  |
| C1 | 0,80337932 | Welch Two Sample t-test |  |
| Tabelle | Training | statistical\_data | 15 |

|  |  |  |  |
| --- | --- | --- | --- |
| A1\_EN\_AVG\_WAHR | A1\_EN\_AVG\_FALSCH | A1\_FR\_AVG\_UNBEKANNT |  |
| 4 | 2 | 6,5 |  |
| Tabelle | training | averages | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| A2\_EN\_AVG\_WAHR | A2\_EN\_AVG\_FALSCH | A2\_FR\_AVG\_UNBEKANNT |  |
| UNK | UNK | UNK |  |
| Tabelle | training | averages | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| B1\_EN\_AVG\_WAHR | B1\_EN\_AVG\_FALSCH | B1\_FR\_AVG\_UNBEKANNT |  |
| 4 | 3 | 10 |  |
| Tabelle | training | averages | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| B2\_EN\_AVG\_WAHR | B2\_EN\_AVG\_FALSCH | B2\_FR\_AVG\_UNBEKANNT |  |
| 12 | 18,5 | 23 |  |
| Tabelle | training | averages | 4 |

|  |  |  |  |
| --- | --- | --- | --- |
| C1\_EN\_AVG\_WAHR | C1\_EN\_AVG\_FALSCH | C1\_FR\_AVG\_UNBEKANNT |  |
| 11,4 | 12,8 | 26 |  |
| Tabelle | training | averages | 5 |