

Pressemitteilung

FernUniversität in Hagen

Carina Grewe

26.09.2024

<http://idw-online.de/de/news840271>

Forschungsprojekte
Informationstechnik, Sprache / Literatur
überregional



Mit KI dem Spracherwerb auf der Spur

Zum „Tag des Sprachenlernens“ am 26. September: Computerlinguist Dr. Josef Ruppenhofer vom Forschungszentrum CATALPA der FernUniversität erforscht, wie KI die Qualität von Sprachtests verbessern kann.

Für Studierende aus dem Ausland steht vor Studienbeginn eine Sprachprüfung an. Zum Studium wird nur zugelassen, wer nach dem „Gemeinsamen europäischen Referenzrahmen für Sprachen“ (GER) mindestens Niveau C1, also „fachkundige Sprachkenntnisse“ vorweisen kann. Doch wie gut lässt sich mit den aktuellen Verfahren wirklich der Lernstand der Getesteten abbilden? Mit großen Datensätzen will Computerlinguist Dr. Josef Ruppenhofer vom Forschungszentrum CATALPA der FernUniversität in Hagen eine Grundlage dafür schaffen, dass Testverfahren künftig noch aussagekräftiger werden.

„Bisher gingen Teile der Forschungscommunity davon aus, dass der Spracherwerb in Stufen erfolgt, und zwar unabhängig vom Alter oder der Vorbildung von Lernenden“, erklärt Ruppenhofer. „Zu diesem Konzept gibt es auch schon viele Untersuchungen – aber immer nur mit eher kleinen Gruppen von Lernenden.“ Fassbar werden die Spracherwerbsstufen beispielsweise durch die Wortstellung von Verben in einem Satz. Zuerst werden einfache Sätze gelernt, etwa „ich suche eine neue Wohnung“. Stufe für Stufe werden die Sätze komplexer, bis schließlich auch solche Nebensätze korrekt gebildet werden können, bei denen das Verb am Ende stehen muss („weil ich eine neue Wohnung suche“). Die Zwischenstufen enthalten unter anderem typische Fehler im Satzbau.

Forschende aus Hagen und Leipzig arbeiten zusammen

Gängige Sprachtests basieren auf diesem Konzept der Erwerbsstufen. Ob das Konzept aber tatsächlich für alle Lernenden einheitlich passt, ist in der neueren linguistischen Forschung umstritten. Hier will das vom BMBF geförderte Forschungsprojekt DAKODA, kurz für „Datenkompetenzen in DaF/DaZ“, mehr Klarheit schaffen. Ruppenhofer arbeitet in dem Projekt gemeinsam mit dem Hagener Computerlinguisten Prof. Dr. Torsten Zesch und einem Team der Universität Leipzig rund um Prof. Dr. Katrin Wisniewski zusammen. Die Forschenden wollen mit Künstlicher Intelligenz eine differenziertere Analyse ermöglichen und dadurch genauere Aussagen über die Qualität des Erwerbsstufen-Ansatzes treffen können.

Unterschiedliche Textformen als Basis

Dafür ist viel Vorarbeit gefragt: „Wenn bisher die Sprache von Lernenden erhoben wurde, hatte man meist keine großen Fallzahlen“, erklärt Ruppenhofer. Damit sich mit Hilfe von KI Aussagen über die Sprachkenntnisse der Lernenden treffen lassen, muss er daher zunächst zahlreiche Datensätze mit erfassten Texten – sogenannte Textkorpora – zusammenführen.

Das ist aber nicht ganz einfach, denn die Textkorpora unterscheiden sich sowohl inhaltlich als auch in der Struktur. „Mal handelt es sich um Briefe, die die Lernenden schreiben sollten, mal um Sachtexte. Es gab auch Ton-Aufnahmen von Gesprächen mit Gastarbeitern aus den 1980er-Jahren, die später verschriftlicht wurden“, erzählt er. Auch scheinbar einfache Daten, wie das Alter der Lernenden, wurden nicht einheitlich erfasst. „Mal haben wir nur das Geburtsjahr, mal das Alter in Jahren, mal war das Alter in Jahren und Monaten notiert.“

Diese Textkorpora fügt Ruppenhofer nun zusammen – um sie selbst zu untersuchen, aber auch, um sie weitgehend auch für andere Forschende verfügbar zu machen. „Nicht immer ist das möglich“, erklärt er. „Bei Erhebungen in den 1980er Jahren hat man natürlich noch nicht daran gedacht, dass Daten im Internet verfügbar sein könnten. Entsprechend gibt es dafür keine ausreichende Einverständniserklärung der Beteiligten.“ Diese Teile können daher nicht weitergegeben werden, sondern werden ausschließlich innerhalb des Projekts analysiert.

Internationales Interesse an Datenschatz

Trotzdem bleibt noch ein umfangreiches Textkorpus, das veröffentlicht und nach unterschiedlichen Kriterien durchsucht werden kann. So gibt es dann auch großes Interesse an Ruppenhofers Datenschatz aus der internationalen Forschungscommunity. In dem interdisziplinären Projekt werden zahlreiche Workshops für Nachwuchsforschende angeboten, die mit den Textkorpora arbeiten wollen. Josef Ruppenhofer und Torsten Zesch nehmen dabei die Rolle der Datenexperten ein. „Wir haben Workshop-Teilnehmende beispielsweise aus China oder aus anderen europäischen Ländern“, erzählt Ruppenhofer. „Insgesamt ist das ein wirklich spannender Austausch.“

Aktuell präsentiert Ruppenhofer einen Teil seiner Forschungsergebnisse bei der „Learner Corpus Research Conference“ in Tartu in Estland. Das Datum des Konferenzbeginns passt zum Thema: Sie startet am 26. September - also am Internationalen Tag des Sprachenlernens.

Redaktion: Christina Lüdeke

wissenschaftliche Ansprechpartner:

Dr. Josef Ruppenhofer

josef.ruppenhofer@fernuni-hagen.de

<https://www.fernuni-hagen.de/english/research/clusters/catalpa/about-catalpa/members/josef.ruppenhofer.shtml>

URL zur Pressemitteilung:

<https://www.fernuni-hagen.de/forschung/schwerpunkte/catalpa/forschung/projekte/DAKODA.shtml>



Dr. Josef Ruppenhofer
Hardy Welsch
FernUniversität in Hagen/Hardy Welsch