

## Medienmitteilung

### Spracherkennung für die 4. Landessprache: RTR bringt dem System Rätoromanisch bei

Chur, 17.05.2021 – **«Speech to text» ist für Sprachen wie Deutsch, Italienisch oder Französisch nichts Neues – für das Rätoromanische hingegen schon. Jetzt soll der Computer lernen, auch Rätoromanisch zu verstehen und zu schreiben. Das rätoromanische Medienhaus Radiotevisiun Svizra Rumantscha (RTR) hat gemeinsam mit der Walliser Firma *recapp* ein Projekt für die schriftliche Wiedergabe von gesprochenem Rätoromanisch lanciert.**

«Speech to text» ist eine automatische und teilweise auch unmittelbare Umsetzung von gesprochener Sprache in einen schriftlichen Text. Dabei spricht man von der direkten Variante, d.h. der Computer schreibt unmittelbar was die eine Person spricht, und von der indirekten, bei welcher der Computer zuerst das Gesprochene analysiert und es erst danach in eine schriftliche Form bringt.

Um dem Computer überhaupt Rätoromanisch beibringen zu können, hat RTR zunächst Daten beschafft, mit denen das System gefüttert werden konnte. Dazu nutzte das Medienhaus die Wörterbücher der Lia Rumantscha (LR) und öffentlich zugängliche Texte wie etwa Zeitungsartikel. Die Abteilung Sprache der LR unterstützte das Projekt bei sprachspezifischen Fragen. RTR liefert für jedes Idiom und für Rumantsch Grischun jeweils rund 30 Stunden Tonmaterial mit den dazugehörigen Transkriptionen. Diese insgesamt 180 Stunden wurden von RTR, in Zusammenarbeit mit Studierenden der Pädagogischen Hochschule Graubünden, in kleine Segmente portioniert, damit der Text exakt den entsprechenden Audiosequenzen zugeordnet werden konnte. Mit diesem Material bringen die Spezialistinnen und Spezialisten von *recapp* nun dem Computer Rätoromanisch bei. Der Lernprozess des Systems ist dabei ähnlich wie bei einer Schülerin oder einem Schüler: Zuerst muss der Computer das Alphabet kennenlernen, danach die Wörter, kurze Sätze und später kann er dann alles verstehen.

#### Rumantsch Grischun, Sursilvan und Vallader bereits in Gebrauch

Bis heute hat der Computer bereits Rumantsch Grischun, Sursilvan und Vallader gelernt. Es wird jedoch deutlich, dass die Lokalkolorite von Tujetsch und Medel über das Lugnez bis hinunter in die Foppa dem Computer die Arbeit nicht leicht machen. Auch für das Vallader (Unterengadin) und die Mundart des Münstertals, das Jauer, war es schwierig, das System zu schulen. Diesen Sommer folgt das Puter (Oberengadin) und im Jahr 2022 wird RTR das Projekt mit dem Sutsilvan zum Abschluss bringen. Damit ist das Vorhaben aber nicht einfach zu Ende. Das Rätoromanische – so wie andere Sprachen auch – verändert sich stetig und sein Wortschatz wird grösser.

«Speech to text» ist die technologische Grundlage für die Digitalisierung der rätoromanischen Sprache. So wird es in Zukunft möglich sein, Gemeindeversammlungen oder Debatten im Grossen Rat zu transkribieren. RTR hat überdies bereits das Projekt «Text to text» in Angriff genommen. Hier geht es um die maschinelle Übersetzung von rätoromanischen Texten ins Deutsche. Ohne diese Technologien ist das Rätoromanische in der digitalen Welt nicht präsent und kann für Anwendungen wie «voice recognition» oder Google Translate und DeepL nicht verwendet werden.

«Speech to text» kommt in den RTR-Redaktionen zur Anwendung, um die redaktionelle Arbeit zu erleichtern. Durch dessen Integration im Produktionssystem unterstützt es die Journalistinnen und Journalisten bei der Arbeit. Sie können Texte transkribieren lassen, um sie dann für Radio- und Videobeiträge aufzubereiten oder sie der Fundaziun Medias Rumantschas (FMR) zur Verfügung zu stellen. Zudem finden sich im Archiv von RTR zahlreiche alte Sendungen ohne Manuskripte. Dies stellt eine enorme Wortfülle dar, die dank «Speech to text» in eine schriftliche Form gebracht werden kann.

**Mehr Informationen:** RTR | Tamara Deflorin | Leiterin Kommunikation und Marketing | 058 136 20 23 | tamara.deflorin@rtr.ch