

**Mittwoch, 27.02., 09–11 Uhr – Fortgeschrittenenkurs
Fortgeschrittene Datenvisualisierung**

Erläuterung: Wir importieren und visualisieren beschreibende Feature-Vektoren und Sprachraumcluster zum Mittelrheinischen Sprachatlas.

Aufgabe: Diese Daten verfügen über Marker, die in Verbindung mit wenig beachteten Einstellungen neue Visualisierungsmöglichkeiten bieten. Wir vergleichen die Änderung von Features zwischen zwei Generationen an Hand von Balkendiagrammen und visualisieren die Güte eines Clusterings mittels der Symbolgröße. Zum Schluss vergleichen wir ein stark dimensionreduziertes Datenset mit möglichen Schibboleths.

Übungseinheit – Gruppierung und Skalierung

Schritt 1	Importieren der CSV-Dateien in REDE und Speichern als Ebene.	<p>Wie gehabt!</p> <p>Laden Sie die CSV-Dateien</p> <ol style="list-style-type: none"> 1. „2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv“ (Enthält beschreibende Featurevektoren zu den beiden Generationen des MRhSA.) 2. „2019_SprachGIS Workshop_Handout_6-WARD5ALLNORMASIL.csv“ (Enthält ein 5er-Clustering nach Ward und die entsprechenden Silhouettenkoeffizienten zu den Orten.) 3. „2019_SprachGIS Workshop_Handout_6-ALT-PCA1.csv“ (Enthält eine beschreibende Größe, ermittelt mittels Principal-Component-Analysis.) <p>von https://www.regionalsprache.de/workshopmaterialien.aspx herunter und importieren Sie diese in das SprachGIS.</p> <p>Die Datei „2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv“ hat sehr viele Spalten, so dass man nicht wirklich viel in der Vorschau erkennen kann, aber das soll uns erstmal nicht stören.</p>
-----------	--	---

Schritt 2	<p>Datei: 2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv</p> <p>Vorverarbeiten der Daten. Abwahl der nicht benötigten Spalten und Auswahl der Gruppierungsspalte (gen).</p>	<p>Aktivieren Sie zunächst die Ebene zu „2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv“. Blenden Sie die anderen Ebenen so lange aus.</p> <p>Öffnen Sie das Visualisierungswerkzeug und wählen sie den Balkendiagrammen-Reiter aus.</p> <p>Öffnen Sie das „Datenformat zuweisen“-Menü. Setzen Sie alle Felder auf „ignorieren“ bis auf „gen“ und die mit „p=“ beginnenden. Für eine bessere Übersicht können Sie die p=Felder auch ignorieren bis auf die Eigenschaft, die Sie vergleichen wollen. (Spoiler: Alveolar, LoweredClose, Nil und Velar sind interessant). Für „gen“ wählen Sie „Gruppieren nach“ aus. Klicken Sie auf „Datenformat aktualisieren“ und schließen Sie das Menü wieder.</p>
Schritt 3	<p>Visualisieren Sie die linguistischen Daten mittels Balkendiagrammen.</p>	<p>Sie sehen zwei Ordner zu den beiden Generationen „jung“ und „alt“. Wählen Sie aus jeder Generation eine zugehörige Eigenschaft aus und ändern Sie die Farben für eine bessere Unterscheidung. Klicken Sie auf „Visualisieren“.</p> <p>Passen Sie die Größe der Balken an. (20 ist ein guter Wert.)</p> <p>Sie sehen die 0-1-normierte Feature-Ausprägung der älteren und der jüngeren Generation des MRhSA im Vergleich.</p>
Schritt 4	<p>Datei: 2019_SprachGIS Workshop_Handout_6-WARD5ALLNORMASIL.csv</p> <p>Gewichten von nominalen Werten an Hand einer Skalierung.</p>	<p>Aktivieren Sie nun die Ebene zu „2019_SprachGIS Workshop_Handout_6-WARD5ALLNORMASIL.csv“ und wählen Sie im Visualisierungswerkzeug den Symbole-Reiter aus. Blenden Sie nicht benötigte Ebenen solange aus.</p> <p>Öffnen Sie wieder „Datenformat zuweisen“. Ignorieren Sie alle Felder bis auf „cls“ und „silhouette“. Ändern Sie „cls“ zu „Nominale Werte“ und „silhouette“ zu „Skalieren an“.</p> <p>Aktualisieren Sie das Datenformat.</p>
Schritt 5	<p>Skalierung aktivieren und visualisieren.</p>	<p>Aktivieren Sie in den erweiterten Optionen „Skalieren“ und visualisieren Sie die Daten.</p> <p>Die Symbole sind nun an Hand des Wertes in „silhouette“ zwischen 0.75 und 1.25 skaliert. Größere Symbole implizieren eine höhere Konfidenz zur Clusterzugehörigkeit, kleine Symbole eine niedrige.</p>

Schritt 6	<p>Datei: 2019_SprachGIS Workshop_Handout_6-ALT-PCA1.csv</p> <p>Visualisierung der dimensionenreduzierten Daten.</p>	<p>Aktivieren Sie die Ebene zu „2019_SprachGIS Workshop_Handout_6-ALT-PCA1.csv“ und visualisieren Sie das relevante Datenfeld (0) als Choroplethenkarte. Nicht benötigte Felder können wieder über das „Datenformat zuweisen“-Menü ignoriert werden.</p>
Schritt 7	<p>Vergleich mit „2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv“.</p>	<p>Wechseln Sie zur „2019_SprachGIS Workshop_Handout_6-FILTERED_MINMAX.csv“-Ebene und wählen Sie aus der älteren Generation ein Feature aus, das Sie Ebenfalls als Choropleth visualisieren.</p> <p>Vergleichen Sie dies mit Ebene „2019_SprachGIS Workshop_Handout_6-ALT-PCA1.csv“.</p> <p>Sie können die Features vergleichen, indem Sie zum Beispiel die Ebenen ein- und ausblenden oder verschiedene Größen für die Kreise wählen.</p> <p>Suchen Sie nach Features, die einen ähnlichen Farbverlauf haben (ggf. die Farbskala umkehren).</p> <p>(Spoiler: Bilabial und Velar sind interessant.)</p> <p>Features, die einem ähnlichen Verlauf folgen, haben wahrscheinlich einen größeren Einfluss auf die Form der Daten.</p>