

Übungen zur Textanalyse mit TXM

1. Korpus laden

Bitte wählen Sie eines der folgenden Beispielkorpora aus, die Sie bei Zenodo herunterladen können:

- roman19.txm (<https://zenodo.org/record/44547>)
- nohispam.txm (<https://zenodo.org/record/47218>)

Laden Sie das Korpus in TXM und verwenden Sie es für die folgenden Übungen.

2. Wortformen / Lemmata / POS

Lexicon

- Welches sind die 5 häufigsten und welches die 5 seltensten Wortformen („word“) im Korpus?
- Welches sind die 3 häufigsten und welches die 3 seltensten Wortarten (POS) im Korpus?

Index

- Wie viele Tokens mit der Wortart „Verb“ gibt es im Korpus?
- Wie häufig kommt die Wortform „Straße“ („calle“/„rue“) vor?
- Erstellen Sie einen Index für bzw. suchen Sie das Lemma „Straße“ („calle“, „rue“).
 - Wie viele Tokens gibt es?
 - Wie viele und welche verschiedenen Wortformen kommen vor?
- Suchen Sie nach Wortformen, die „calle“ bzw. „rue“ enthalten. Lassen Sie neben der Wortform in der Ergebnisliste auch das Lemma anzeigen. Finden Sie hier weitere Lemmata, die Sie mit „Straße“ assoziieren würden? Welche?
- In welchen Texten kommt das Lemma „callejero“ bzw. „ruelle“ vor?

3. Konkordanzen und Kookkurrenzen

- Erstellen Sie eine Konkordanz für die Wortform „mexicana“ bzw. „parisienne“. Sehen Sie sich einen interessanten Treffer im Kontext an.

- Erstellen Sie für die Wörter „campo“ bzw. „campagne“ und „ciudad“ bzw. „ville“ jeweils eine Kookkurrenz. Welches andere Substantiv (und welches Adjektiv), das zusammen mit diesen Wörtern vorkommt, hat jeweils den höchsten Score?

4. Erweiterte Suchen

- Suchen Sie nach dem Lemma „arbeiten“ („trabajar“ bzw. „travailler“) gefolgt von Adverbien.
- Suchen Sie nach Verben, die im Abstand von 0 bis 3 Wörtern vor dem Lemma „Arbeit“ („trabajo“ bzw. „travail“) stehen.
- Suchen Sie nach Adverbien, vor denen entweder das Lemma „leben“ („vivir“, „vivre“) oder das Lemma „sterben“ („morir“, „mourir“) vorkommt.
- Suchen Sie nach Wortformen, die 4 (für Spanisch) bzw. 5 (für Französisch) aufeinander folgende Vokale enthalten.

5. Vergleichende Abfragen

Progression

- Erstellen Sie eine Progression für die Lemmata „Freude“ („alegría“ bzw. „joie“) und „Angst“ („miedo“ bzw. „peur“) und sehen Sie sich mit Hilfe des Verlaufs an, wie stark die Wörter im Korpus vertreten sind und in welchen Texten sie besonders häufig vorkommen.
- Erstellen Sie einen Verlauf für Interjektionen.

Partitionen und Spezifika

- Partitionieren Sie das Korpus nach Untergattungen.
 - Untersuchen Sie, in welchen Untergattungen die Wörter „Stadt“ und „Land“ besonders häufig vorkommen.
 - Untersuchen Sie, in welchen Untergattungen Interjektionen besonders häufig vorkommen.
- Partitionieren Sie das Korpus nach Autoren.
 - Suchen Sie sich zwei Autoren aus und untersuchen Sie, welche Lemmata für diese Autoren spezifisch sind.
 - Untersuchen Sie, welche Wortarten für die einzelnen Autoren spezifisch sind.

Lösungen

Die Lösungen schließen nur die bei den Übungen benötigten Suchbefehle und die ggf. vorzunehmende Änderung von Anzeige-Einstellungen ein. Der Aufruf der TXM-Funktionen „Lexicon“, „Index“, usw. wird vorausgesetzt.

2. Wortformen / Lemmata / POS

Index

- Wie viele Tokens mit der Wortart „Verb“ gibt es im Korpus?
 - NOHISPAM: [espos="V.*"]
 - ROMAN19: [frpos="VER.*"]
- Wie häufig kommt die Wortform „Straße“ („calle“/„rue“) vor?
 - NOHISPAM: [word="calle"]
 - ROMAN19: [word="rue"]
- Erstellen Sie einen Index für bzw. suchen Sie das Lemma „Straße“ („calle“, „rue“).
 - NOHISPAM: [eslemma="calle"]
 - ROMAN19: [frlemma="rue"]
- Suchen Sie nach Wortformen, die „calle“ bzw. „rue“ enthalten.
 - NOHISPAM: [word=".*calle.*"]
 - ROMAN19: [word=".*rue.*"]
- In welchen Texten kommt das Lemma „callejero“ bzw. „ruelle“ vor?
 - NOHISPAM: [eslemma="callejero"] mit Einstellung „Properties: eslemma“
 - ROMAN19: [frlemma="ruelle"] mit Einstellung „Properties: frlemma“

3. Konkordanzen und Kookkurrenzen

- Erstellen Sie eine Konkordanz für die Wortform „mexicana“ bzw. „parisienne“.
 - NOHISPAM: [word="mexicana"]
 - ROMAN19: [word="parisienne"]
- Erstellen Sie für die Wörter „campo“ bzw. „campagne“ und „ciudad“ bzw. „ville“ jeweils eine Kookkurrenz. Welches andere Substantiv (und welches Adjektiv), das zusammen mit diesen Wörtern vorkommt, hat jeweils den höchsten Score?
 - NOHISPAM: [eslemma="campo"] bzw. [eslemma="ciudad"] mit Einstellung „Cooccurents properties: word; espos“
 - ROMAN19: [frlemma="campagne"] bzw. [frlemma="ville"] mit Einstellung „Cooccurents properties: word; frpos“

4. Erweiterte Suchen

- Suchen Sie nach dem Lemma „arbeiten“ („trabajar“ bzw. „travailler“) gefolgt von Adverbien.
 - NOHISPAM: [eslemma="trabajar"][espos="ADV"]
 - ROMAN19: [frlemma="travailler"][frpos="ADV"]
- Suchen Sie nach Verben, die im Abstand von 0 bis 3 Wörtern vor dem Lemma „Arbeit“ („trabajo“ bzw. „travail“) stehen.
 - NOHISPAM: [espos="V.*"][]{0,3}[eslemma="trabajo"]
 - ROMAN19: [frpos="VER.*"][]{0,3}[frlemma="travail"]
- Suchen Sie nach Adverbien, vor denen entweder das Lemma „leben“ („vivir“, „vivre“) oder das Lemma „sterben“ („morir“, „mourir“) vorkommt.
 - NOHISPAM: [eslemma="vivir|morir"][espos="ADV"]
 - ROMAN19: [frlemma="vivre|mourir"][frpos="ADV"]
- Suchen Sie nach Wortformen, die 4 (für Spanisch) bzw. 5 (für Französisch) aufeinander folgende Vokale enthalten.
 - NOHISPAM: [word=".*[aeiou]{4}.*"]
 - ROMAN19: [word=".*[aeiou]{5}.*"]

5. Vergleichende Abfragen

Progression

- Erstellen Sie eine Progression für die Lemmata „Freude“ („alegría“ bzw. „joie“) und „Angst“ („miedo“ bzw. „peur“) ...
 - NOHISPAM: [eslemma="alegría"] bzw. [eslemma="miedo"]
 - ROMAN19: [frlemma="joie"] bzw. [frlemma="peur"]
- Erstellen Sie einen Verlauf für Interjektionen.
 - NOHISPAM: [espos="ITJN"]
 - ROMAN19: [frpos="INT"]

Partitionen und Spezifika

- Untersuchen Sie, in welchen Untergattungen die Wörter „Stadt“ und „Land“ besonders häufig vorkommen.
 - NOHISPAM: [eslemma="ciudad|campo"] mit Einstellung „Properties: eslemma“
 - ROMAN19: [frlemma="ville|campagne"] mit Einstellung „Properties: frlemma“
- Untersuchen Sie, in welchen Untergattungen Interjektionen besonders häufig vorkommen.
 - NOHISPAM: [espos="ITJN"] mit Einstellung „Properties: espos“
 - ROMAN19: [frpos="INT"] mit Einstellung „Properties: frpos“