**Umgang mit mehrfachen Lextypes an einem Ort**

Stand 8.7.23

Problemstellung

Die mit deutlichem Abstand häufigsten Fälle sind eindeutige Übereinstimmung der Lextypes: Anzahl Lextypes in Maurer und DWA=1 Lextyp ist identisch oder In Maurer ODER DWA ist ein einzelner Lextyp der beim anderen Datensatz vorkommt (egal wie viele weitere Lextypes vorkommen).

Die Levensthein Methode für diese Fälle ist eine „Best Match“ Matrix: jeder Komma separierte Phontyp Maurer gegen jeden Komma separierten Phontyp DWA. Da in einem Datensatz (zB Maurer) nur 1 Lextyp vorhanden ist, muss die höchste Übereinstimmung in der Matrix die Paarung der Phontyps sein, welche zu dem Lextyp gehören.

In wenigen Fällen (siehe Tabelle 1) gibt es Orte, an welchen beide Datensätze mehr als 1 Lextyp aufweisen („MultiLex“).

|  |  |
| --- | --- |
| Item | N places with MuliLex |
| Ameise | 7 |
| Begräbnis | 26 |
| Ziege | 6 |
| Deichsel | 4 |
| Gurke | 33 |
| Pflaume | 2 |
| Hebamme | 10 |
| Hagebutte | 9 |

Bei zwei identischen Lextypes in beiden Datensatz wo die Phontypes ebenfalls in beiden Datensätzen identisch sind (MultiLex Double Indentical) kann für beide Lexklasse für die Ähnlichkeit eine 1 (100%) vergeben werden.

Wenn ein Lextyp übereinstimmt, unabhängig von der Anzah der Lextypes in beiden Datensätzen (MultiLex Eindeutig), kann die „Best Match“ Methode eingesetzt werden.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *dwa\_phontype* | *mau\_phontype* | *dwa\_lextype* | *mau\_lextype* | *Lexclass Übereinstimmung* | *Levensthein Methode* |
| *arSkrätzer, butte* | *arSkrätzer* | *Arschkrätzer, Butte* | *Arschkrätzer* | *Eindeutig* | *Best Match* |
| *arSkratzer, butte* | *arSkrätzer, butt* | *Arschkrätzer, Butte* | *Arschkrätzer, Butte* | *MultiLex Double unequal* | *By Position* |
| *butt* | *butt, butte, hambüttel* | *Butte* | *Butte, Hambüttel* | *Eindeutig* | *Best Match* |
| *buttenmus, hagapfela* | *botte, butt, butte* | *Buttenmus, Hagäpfele* | *Butte* | *Keine* | *Keine* |
| *butte, hagebutte* | *butte, hagebutte* | *Butte, Hagebutte* | *Butte, Hagebutte* | *MultiLex Double Identical* | *Beide Klassen = 1* |
| *hagebuette, hiefe, hife* | *hife, reiselebeis* | *Hagebutte, Hiefe* | *Hiefe, Reiselebeis* | *MultiLex Eindeutig* | *Best Match* |

Bei zwei identischen Lextypes in beiden Datensatz wo die Phontypes NICHT identisch sind (MultiLex Double unequal) könnte der zum Lextype gehörende Phontyp über die Position ermittelt und dann vergleichen werden. Dies setzt aber vorraus, dass die Reihenfolge Lextype und Phontyp im Datensatz immer gleich ist.

Beispiele von „MultiLex“ Vorkommen bei Hagebutte

Fazit