

312 semantically associated pairs of German nouns with equal word length

Olaf Dimigen, Reinhold Kliegl, & Werner Sommer *
Email: olaf.dimigen@hu-berlin.de

May 9, 2012

This document describes the word materials used in:

Dimigen, O., Kliegl, R., Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *Neuroimage*. DOI:10.1016/j.neuroimage.2012.04.006

Materials are provided as a .csv text file (comma separated values) which contains the set of 312 German noun pairs together with full word statistics and rating data. The organization of the .csv file is described further below. The materials may be freely used for research purposes, provided that the original publication is cited.

1 Construction of word pairs

Preview-targets pairs The basis for the construction of target nouns and their preview nouns was a set of 312 pairs of semantically associated German nouns (e.g. “Frau” – “Dame”, woman – lady; or “Wald” – “Baum”, forest – tree). Pairs were selected from a larger set of pairs after a semantic rating (see below). The nouns of each pair had the same length (between 4-7 letters, $M = 5.4$) and were either synonyms, associatively related, conceptually related (e.g., part-whole relation), or both associatively and conceptually related. Additionally, a few strongly associated antonyms (e.g., “Ebbe” – “Flut”, ebb tide - high tide) were included. Mean word frequency was 31.6 ($SD = 54.0$) for previews and 19.2 ($SD = 40.1$) for targets. From these 312 related pairs, 312 unrelated pairs were created by exchanging the previews of two pairs with the same word length, yielding two new pairs without noteworthy semantic associations (woman – tree, forest – lady).

Semantic rating Pairs were rated for association strength by fifteen native German-speaking university students (13 women, $M = 27.4$ years, range: 20-43 years). They used a five-point rating scale (from 1 to 5) on a keyboard to indicate how related both nouns are. In each trial, the two nouns of a pair appeared next to each other on the midline of the screen in Courier monospaced font. All related and unrelated pairings were presented in random order and participants were asked to give their first, spontaneous impressions. There was not time limit for the response (mean reaction time was 2.2 seconds). With a mean score of 4.50 ($SD: 0.34$), related pairs were rated as significantly more related than their recombinations, $t(311) = 129.1$, $p < .0001$, which received a mean score of

*Benthe Kornrumpf, Susann Meyberg, Ulrich Reinacher, and Katharina Schäfer helped to construct the word pairs. Kay-Michael Würzner computed the collocation norms.

1.52 (SD: 0.26). In addition, all preview-target units fulfilled the following criteria: (1) both related pairs scored > 3.5 , (2) both unrelated pairs scored < 2.5 , and (3) the two related differed from the two unrelated pairs by > 1.5 points.

Mutual Information As an additional measure of semantic association, we computed collocation norms for all pairs from the DWDS corpus. The resulting measure, Mutual Information describes the likelihood that two words co-occur within the same sentence of the text corpus. As expected, Mutual Information (computed logarithmized to base 10) was higher for related than for unrelated pairs, $M = 5.12$ vs. 4.71 , $t(311) = 10.5$, $p < .0001$.

2 Matching of pairs on non-semantic dimensions

Since related and unrelated word pairs comprised of the same words, lexical properties (such as word frequency) were matched for both types of pairs. Nevertheless, it is still possible that semantically related words are more similar to each other than unrelated words on a visual, orthographic, or phonological level. To rule out priming based on non-semantic properties, preview and target never started with the same letter or phoneme. In addition, related pairs were matched to unrelated pairs on 15 non-semantic measures of word similarity (summarized in Table 1 of Dimigen et al., 2012). For this matching, a simple resampling method was used. On each of 100,000 iterations, our algorithm randomly sampled 312 preview-target units (two related pairs plus their recombinations) from a total pool of 317 units that were considered suitable after the semantic rating. For each random sample of 312 units, it then computed a weighted sum of several measures of non-semantic word similarity, separately for related and the unrelated pairs. The best-matching solution was kept and is presented here.

Visual similarity Visual similarity between targets and previews was measured by comparing letters located at corresponding positions in the two strings. As a first measure, we counted the letters that shared the same stroke direction. Stroke direction matches if both letters have an ascender (e.g., ‘d’ and ‘f’), both have a descender (‘g’ and ‘j’), or both have neither ascender nor descender (‘a’ and ‘c’). As a second measure, we used empirical data on letter confusability. Confusability is the likelihood of confusing a letter with another letter after brief visual presentation. For upper-case letters, we used a confusability matrix by Bouma (1971; his Table 3). For lower-case letters, we used the matrix from Mueller and Weidemann (2008, their Table 3), after transforming the data to a format similar to that used by Bouma.

Orthographic similarity Orthographic similarity was determined location-specific and location-invariant (i.e., irrespective of letter position). As a location-specific measure, we counted instances where the same letter occurred at the same position in both strings. This was summarized as the Hamming Distance (Hamming, 1950), which is the edit distance (Levenshtein Distance) for two strings of equal length. As a location-invariant measure, we counted the letter bigrams and letter trigrams in the preview that also occurred somewhere in the target. Orthographic similarity based on bigrams was also operationalized as the Dice Coefficient (Dice, 1945) based on letter bigrams. The Dice coefficient was computed following Lambert et al. (2002).

Phonological similarity Phonological similarity was defined as the edit distance (Levenshtein distance) based on phonemes (instead of letters). For this, all nouns were translated into the SAMPA phonetic alphabet. SAMPA codes were retrieved from the

BOMP dictionary (Bonn machine-readable pronunciation dictionary; Portele, Krämer, and Stock, 1995). Some of the nouns (39 nouns) were not contained in the BOMP dictionary; for these cases, SAMPA codes were extracted from the dlexDB database using an automatic algorithm.

3 List of pairs

Target nouns are followed by their related and unrelated preview. Please note that related previews from subsequent pairs were exchanged to create unrelated previews. Numbers in parentheses denote the mean association rating between targets and the respective preview.

Four letters (62)

1. Wind - Luft (4.73) / Blut (1.27)
2. Mord - Blut (4.53) / Luft (1.13)
3. Geld - Euro (4.87) / Säge (1.33)
4. Holz - Säge (4.47) / Euro (1.27)
5. Ober - Wirt (4.53) / Flut (1.40)
6. Ebbe - Flut (4.73) / Wirt (1.20)
7. Saum - Naht (4.53) / Brot (1.13)
8. Mehl - Brot (4.47) / Naht (1.27)
9. Tüte - Sack (4.67) / Ziel (1.20)
10. Plan - Ziel (4.47) / Sack (1.40)
11. Hall - Echo (4.73) / Grün (1.33)
12. Blau - Grün (4.47) / Echo (1.33)
13. Wall - Burg (4.40) / Bett (1.33)
14. Sofa - Bett (4.53) / Burg (1.13)
15. Berg - Höhe (4.60) / Zahn (1.20)
16. Arzt - Zahn (4.53) / Höhe (1.53)
17. Rose - Dorn (4.47) / Halt (1.53)
18. Stop - Halt (4.80) / Dorn (1.40)
19. Tuch - Hals (4.27) / Reis (1.67)
20. Korn - Reis (4.73) / Hals (1.13)
21. Kahn - Ufer (3.80) / Ehre (1.07)
22. Ruhm - Ehre (4.73) / Ufer (1.27)
23. Rohr - Ofen (4.27) / Meer (1.47)
24. Sand - Meer (4.73) / Ofen (1.40)
25. Idol - Star (4.47) / Bier (1.47)
26. Krug - Bier (4.47) / Star (1.33)
27. Neid - Hass (4.60) / Tanz (1.67)
28. Ball - Tanz (4.60) / Hass (1.40)
29. Dame - Frau (4.87) / Wald (1.80)
30. Baum - Wald (4.67) / Frau (1.60)
31. Sekt - Wein (4.60) / Kanu (1.13)
32. Boot - Kanu (4.73) / Wein (2.13)
33. Herd - Topf (4.47) / Gang (1.53)
34. Flur - Gang (4.67) / Topf (1.53)
35. Saal - Raum (4.47) / Mann (1.60)
36. Herr - Mann (4.80) / Raum (1.60)
37. Nerv - Hirn (4.67) / Welt (1.80)
38. Erde - Welt (4.60) / Hirn (1.60)
39. Kern - Nuss (4.33) / Lärm (1.27)
40. Hupe - Lärm (3.87) / Nuss (1.13)
41. Auto - Taxi (4.67) / Ruhe (1.87)
42. Muße - Ruhe (4.33) / Taxi (1.40)
43. Sieb - Loch (4.07) / Kind (1.53)
44. Baby - Kind (4.53) / Loch (1.40)
45. Brut - Nest (4.40) / Glas (1.60)
46. Vase - Glas (4.33) / Nest (1.53)
47. Horn - Jagd (4.27) / Wand (1.73)
48. Putz - Wand (4.20) / Jagd (1.27)
49. Helm - Kopf (4.13) / Text (1.07)

50. Satz - Text (4.20) / Kopf (1.80)
51. Korb - Obst (4.07) / Nase (1.27)
52. Auge - Nase (4.20) / Obst (1.60)
53. Land - Feld (4.07) / Bein (1.60)
54. Knie - Bein (4.53) / Feld (1.60)
55. Eile - Hast (4.73) / Dank (1.93)
56. Lohn - Dank (4.13) / Hast (1.73)
57. Hand - Ring (4.27) / Park (2.07)
58. Bank - Park (4.60) / Ring (1.73)
59. Urne - Sarg (4.47) / Buch (1.67)
60. Wort - Buch (4.20) / Sarg (2.00)
61. Wahl - Jury (4.00) / Lust (2.07)
62. Gier - Lust (4.40) / Jury (1.53)

Five letters (118)

63. Narbe - Wunde (4.73) / Bohne (1.07)
64. Erbse - Bohne (4.60) / Wunde (1.13)
65. Niere - Leber (4.60) / Moral (1.20)
66. Ethik - Moral (4.67) / Leber (1.20)
67. Depot - Lager (4.87) / Suppe (1.33)
68. Kelle - Suppe (4.60) / Lager (1.33)
69. Kakao - Milch (4.60) / Anzug (1.40)
70. Jacke - Anzug (4.60) / Milch (1.07)
71. Probe - Übung (4.40) / Tulpe (1.13)
72. Nelke - Tulpe (4.47) / Übung (1.00)
73. Kutte - Mönch (4.53) / Feuer (1.33)
74. Brand - Feuer (4.67) / Mönch (1.20)
75. Bluse - Ärmel (4.47) / Organ (1.20)
76. Lunge - Organ (4.67) / Ärmel (1.33)
77. Decke - Stuck (4.47) / Krieg (1.27)
78. Bombe - Krieg (4.53) / Stuck (1.13)
79. Möbel - Regal (4.60) / Küste (1.60)
80. Bucht - Küste (4.67) / Regal (1.07)
81. Karre - Wagen (4.60) / Braut (1.40)
82. Kleid - Braut (4.67) / Wagen (1.27)
83. Beton - Stahl (4.20) / Nebel (1.27)
84. Dunst - Nebel (4.87) / Stahl (1.20)
85. Hecke - Busch (4.53) / Lippe (1.20)
86. Stift - Lippe (4.33) / Busch (1.13)
87. Kamin - Asche (4.40) / Blüte (1.13)
88. Honig - Blüte (4.47) / Asche (1.20)
89. Kerze - Docht (4.67) / Salat (1.27)
90. Blatt - Salat (4.60) / Docht (1.47)
91. Druck - Zwang (4.27) / Tanne (1.27)
92. Buche - Tanne (4.67) / Zwang (1.13)
93. Stiel - Blume (4.40) / Mütze (1.07)
94. Schal - Mütze (4.47) / Blume (1.33)
95. Lotto - Glück (4.53) / Fluss (1.07)
96. Teich - Fluss (4.67) / Glück (1.67)
97. Orkan - Sturm (4.73) / Junge (1.73)

98. Knabe - Junge (4.87) / Sturm (1.40)
 99. Haken - Angel (4.53) / Sport (1.40)
 100. Halle - Sport (4.60) / Angel (1.40)
 101. Möhre - Gurke (4.47) / Sucht (1.47)
 102. Droge - Sucht (4.73) / Gurke (1.40)
 103. Hitze - Wärme (4.80) / Folge (1.93)
 104. Serie - Folge (4.93) / Wärme (1.47)
 105. Tisch - Stuhl (4.53) / Ärger (1.33)
 106. Frust - Ärger (4.73) / Stuhl (1.60)
 107. Bitte - Gebet (4.13) / Maler (1.27)
 108. Farbe - Maler (4.67) / Gebet (1.27)
 109. Palme - Insel (4.40) / Sahne (1.47)
 110. Torte - Sahne (4.60) / Insel (1.27)
 111. Luxus - Yacht (4.47) / Nadel (1.33)
 112. Stich - Nadel (4.20) / Yacht (1.13)
 113. Stoff - Faser (4.67) / Hotel (1.47)
 114. Suite - Hotel (4.47) / Faser (1.53)
 115. Acker - Bauer (4.53) / Bogen (1.67)
 116. Pfeil - Bogen (4.80) / Bauer (1.53)
 117. Wolke - Regen (4.67) / Birne (1.07)
 118. Apfel - Birne (4.60) / Regen (2.13)
 119. Boden - Grund (4.60) / Griff (1.67)
 120. Knauf - Griff (4.33) / Grund (1.20)
 121. Punkt - Komma (4.67) / Eisen (1.47)
 122. Nagel - Eisen (4.40) / Komma (1.60)
 123. Milde - Gnade (4.53) / Linie (1.53)
 124. Zeile - Linie (4.27) / Gnade (1.27)
 125. Stolz - Würde (4.67) / Leder (1.53)
 126. Schuh - Leder (4.33) / Würde (1.53)
 127. Glanz - Chrom (4.20) / Stamm (1.87)
 128. Rinde - Stamm (4.73) / Chrom (1.13)
 129. Feger - Besen (4.60) / Kiosk (1.33)
 130. Stand - Kiosk (4.20) / Besen (1.53)
 131. Grill - Wurst (4.53) / Nacht (2.20)
 132. Stern - Nacht (4.60) / Wurst (1.07)
 133. Angst - Sorge (4.87) / Party (1.80)
 134. Feier - Party (4.93) / Sorge (2.13)
 135. Sitte - Regel (4.07) / Taste (1.13)
 136. Knopf - Taste (4.60) / Regel (1.73)
 137. Krone - Prinz (4.33) / Geige (1.33)
 138. Cello - Geige (4.60) / Prinz (1.80)
 139. Knast - Zelle (4.53) / Wiese (1.20)
 140. Rasen - Wiese (4.27) / Zelle (1.87)
 141. Ferse - Hacke (4.00) / Dauer (1.20)
 142. Länge - Dauer (4.60) / Hacke (1.67)
 143. Perle - Kugel (3.87) / Eiche (1.20)
 144. Ahorn - Eiche (4.60) / Kugel (1.53)
 145. Kniff - Trick (4.47) / Bauch (2.00)
 146. Magen - Bauch (4.60) / Trick (1.40)
 147. Pirat - Säbel (4.40) / Staub (1.47)
 148. Dreck - Staub (4.47) / Säbel (1.73)
 149. Atlas - Karte (4.67) / Licht (1.47)
 150. Sonne - Licht (4.60) / Karte (2.20)
 151. Spiel - Pokal (4.40) / Kabel (1.67)
 152. Draht - Kabel (4.40) / Pokal (1.53)
 153. Wanne - Eimer (4.00) / Brief (1.07)
 154. Paket - Brief (4.33) / Eimer (1.80)
 155. Marke - Porto (3.80) / Alter (1.73)
 156. Greis - Alter (4.60) / Porto (1.20)
 157. Qualm - Rauch (5.00) / Agent (2.13)
 158. Spion - Agent (4.47) / Rauch (1.87)
 159. Elend - Armut (4.87) / Radio (1.93)
 160. Musik - Radio (4.60) / Armut (2.13)
 161. Bühne - Szene (4.53) / Zange (1.73)
 162. Feile - Zange (3.80) / Szene (1.20)
 163. Trieb - Zweig (3.93) / Grube (1.13)

164. Senke - Grube (3.93) / Zweig (1.33)
 165. Weite - Größe (4.33) / Haube (1.33)
 166. Kappe - Haube (4.00) / Größe (1.67)
 167. Thron - König (4.53) / Eifer (2.00)
 168. Fleiß - Eifer (4.60) / König (2.07)
 169. Faden - Leine (4.13) / Reise (1.47)
 170. Fahrt - Reise (4.67) / Leine (2.33)
 171. Harfe - Engel (3.93) / Lücke (1.80)
 172. Spalt - Lücke (4.33) / Engel (1.53)
 173. Hülle - Folie (4.07) / Seife (1.73)
 174. Blase - Seife (4.40) / Folie (1.87)
 175. Anker - Hafen (4.27) / Strom (1.80)
 176. Blitz - Strom (4.47) / Hafen (2.07)
 177. Truhe - Kiste (4.73) / Kanne (2.00)
 178. Tasse - Kanne (4.40) / Kiste (2.33)
 179. Koran - Bibel (4.73) / Stein (1.93)
 180. Mauer - Stein (4.33) / Bibel (2.40)

Six letters (80)

181. Gemüse - Tomate (4.73) / Metall (1.13)
 182. Silber - Metall (4.33) / Tomate (1.07)
 183. Kachel - Fliese (4.60) / Melone (1.07)
 184. Ananas - Melone (4.47) / Fliese (1.20)
 185. Käufer - Handel (4.47) / Schnee (1.33)
 186. Flocke - Schnee (4.87) / Handel (1.27)
 187. Diesel - Benzin (4.73) / Eskimo (1.07)
 188. Arktis - Eskimo (4.47) / Benzin (1.40)
 189. Tropen - Urwald (4.60) / Rummel (1.40)
 190. Kirmes - Rummel (4.87) / Urwald (1.40)
 191. Hopfen - Gerste (4.67) / Format (1.13)
 192. Rahmen - Format (4.13) / Gerste (1.07)
 193. Bonbon - Zucker (4.47) / Messer (1.27)
 194. Klinge - Messer (4.60) / Zucker (1.20)
 195. Gläser - Brille (4.67) / Steuer (1.13)
 196. Abgabe - Steuer (4.40) / Brille (1.33)
 197. Killer - Mörder (4.80) / Bäcker (1.60)
 198. Kuchen - Bäcker (4.53) / Mörder (1.13)
 199. Gesang - Stimme (4.67) / Kupfer (1.13)
 200. Bronze - Kupfer (4.53) / Stimme (1.53)
 201. Kittel - Doktor (4.40) / Anfang (1.20)
 202. Beginn - Anfang (4.73) / Doktor (1.40)
 203. Ferien - Urlaub (4.73) / Kanone (1.33)
 204. Pulver - Kanone (4.47) / Urlaub (1.40)
 205. Hefter - Ordner (4.67) / Soldat (1.20)
 206. Gewehr - Soldat (4.60) / Ordner (1.67)
 207. Sattel - Reiter (4.60) / Daumen (1.27)
 208. Finger - Daumen (4.60) / Reiter (1.53)
 209. Weiher - Tümpel (4.33) / Filter (1.20)
 210. Kaffee - Filter (4.67) / Tümpel (1.47)
 211. Felsen - Klippe (4.67) / Becher (1.13)
 212. Schale - Becher (4.00) / Klippe (1.27)
 213. Herbst - Sommer (4.33) / Profit (1.20)
 214. Gewinn - Profit (4.80) / Sommer (1.73)
 215. Pommies - Imbiss (4.73) / Anhang (1.13)
 216. Zusatz - Anhang (4.47) / Imbiss (1.87)
 217. Husten - Grippe (4.40) / Leiter (1.20)
 218. Treppe - Leiter (4.00) / Grippe (1.13)
 219. Becken - Wasser (4.33) / Puzzle (1.40)
 220. Rätsel - Puzzle (4.47) / Wasser (1.33)
 221. Infekt - Fieber (4.33) / Schiff (1.67)
 222. Kutter - Schiff (4.60) / Fieber (1.27)
 223. Stelle - Posten (4.13) / Tumult (1.53)
 224. Unruhe - Tumult (4.73) / Posten (1.33)
 225. Pappel - Fichte (4.47) / Rücken (1.13)
 226. Wirbel - Rücken (3.87) / Fichte (1.27)
 227. Tusche - Wimper (4.53) / Kosten (1.53)

228. Spesen - Kosten (4.07) / Wimper (1.20)
 229. Kiefer - Gebiss (3.67) / Furcht (1.33)
 230. Bammel - Furcht (4.87) / Gebiss (1.40)
 231. Schema - Modell (4.60) / Risiko (2.00)
 232. Wagnis - Risiko (4.73) / Modell (1.53)
 233. Planet - Rakete (4.00) / Banane (1.27)
 234. Frucht - Banane (4.20) / Rakete (1.20)
 235. Chemie - Physik (4.67) / Teufel (1.87)
 236. Hörner - Teufel (4.27) / Physik (1.40)
 237. Jurist - Anwalt (4.87) / Statue (1.93)
 238. Museum - Statue (4.13) / Anwalt (1.53)
 239. Götter - Tempel (4.47) / Umfang (1.80)
 240. Ausmaß - Umfang (4.60) / Tempel (1.80)
 241. Beutel - Tasche (4.47) / Spitze (1.80)
 242. Gipfel - Spitze (4.07) / Tasche (1.27)
 243. Wurzel - Knospe (4.20) / Ritual (2.33)
 244. Brauch - Ritual (4.93) / Knospe (1.40)
 245. Ausweg - Lösung (4.60) / Träger (1.40)
 246. Stütze - Träger (4.07) / Lösung (1.93)
 247. Kummer - Trauer (5.00) / Marine (1.93)
 248. Flotte - Marine (4.40) / Trauer (2.13)
 249. Muskel - Athlet (4.47) / Wecker (1.40)
 250. Morgen - Wecker (4.47) / Athlet (2.27)
 251. Blende - Kamera (4.47) / Gewalt (1.80)
 252. Waffen - Gewalt (4.60) / Kamera (2.13)
 253. Nation - Flagge (4.33) / Folter (2.13)
 254. Tortur - Folter (4.20) / Flagge (1.33)
 255. Galgen - Henker (4.67) / Fehler (2.40)
 256. Irrtum - Fehler (4.87) / Henker (2.07)
 257. Talent - Stärke (3.93) / Gebiet (2.20)
 258. Fläche - Gebiet (4.53) / Stärke (1.67)
 259. Gruppe - Partei (4.27) / Straße (2.27)
 260. Häuser - Straße (4.20) / Partei (2.33)
- Seven letters (52)*
261. Meldung - Bericht (4.87) / Getränk (1.47)
 262. Flasche - Getränk (4.67) / Bericht (1.20)
 263. Mission - Auftrag (4.80) / Alkohol (1.40)
 264. Schnaps - Alkohol (4.47) / Auftrag (1.13)
 265. Anstieg - Zunahme (4.67) / Polizei (1.53)
 266. Gendarm - Polizei (4.73) / Zunahme (1.20)
 267. Fransen - Teppich (4.40) / Anzeige (1.20)
 268. Inserat - Anzeige (4.80) / Teppich (1.40)
 269. Prüfung - Klausur (4.93) / Werbung (1.93)
 270. Reklame - Werbung (5.00) / Klausur (1.40)
 271. Applaus - Beifall (4.93) / Seemann (1.40)
 272. Matrose - Seemann (4.80) / Beifall (1.73)
 273. Spritze - Impfung (4.40) / Aufruhr (1.27)
 274. Krawall - Aufruhr (4.80) / Impfung (1.40)
 275. Skelett - Knochen (4.67) / Abstand (1.60)
276. Distanz - Abstand (4.87) / Knochen (1.47)
 277. Apparat - Telefon (4.67) / Schwert (1.67)
 278. Rüstung - Schwert (4.47) / Telefon (1.07)
 279. Zuhause - Wohnung (4.73) / Absicht (1.60)
 280. Vorsatz - Absicht (4.73) / Wohnung (1.53)
 281. Ballade - Gedicht (4.47) / Pudding (1.07)
 282. Dessert - Pudding (4.67) / Gedicht (1.73)
 283. Bandage - Verband (4.73) / Ankunft (1.20)
 284. Empfang - Ankunft (4.33) / Verband (1.60)
 285. Ausrede - Vorwand (4.47) / Vorhang (1.47)
 286. Gardine - Vorhang (4.87) / Vorwand (1.87)
 287. Premier - Kanzler (4.13) / Brunnen (1.13)
 288. Fontäne - Brunnen (4.13) / Kanzler (1.27)
 289. Edition - Ausgabe (4.60) / Gedanke (1.60)
 290. Einfall - Gedanke (4.47) / Ausgabe (1.60)
 291. Anklage - Vorwurf (4.60) / Merkmal (1.73)
 292. Symptom - Merkmal (4.47) / Vorwurf (1.60)
 293. Blamage - Skandal (4.13) / Eingang (1.40)
 294. Öffnung - Eingang (4.60) / Skandal (1.73)
 295. Schloss - Fahrrad (4.20) / Magazin (2.20)
 296. Journal - Magazin (4.87) / Fahrrad (1.40)
 297. Schwere - Gewicht (4.40) / Trainer (1.53)
 298. Spieler - Trainer (4.53) / Gewicht (2.07)
 299. Zeitung - Artikel (4.67) / Kleider (1.93)
 300. Schrank - Kleider (4.67) / Artikel (2.07)
 301. Sektion - Bereich (4.40) / Richter (1.80)
 302. Prozess - Richter (4.47) / Bereich (1.87)
 303. Fassung - Version (4.73) / Meister (2.07)
 304. Experte - Meister (4.93) / Version (2.47)
 305. Vorsitz - Leitung (4.40) / Theater (2.00)
 306. Kulisse - Theater (4.53) / Leitung (1.87)
 307. Dynamik - Energie (4.53) / Pflanze (2.40)
 308. Gärtner - Pflanze (4.53) / Energie (1.87)
 309. Gemälde - Malerei (4.67) / Fenster (2.40)
 310. Scheibe - Fenster (4.73) / Malerei (2.27)
 311. Gerücht - Legende (3.67) / Papiere (1.87)
 312. Ausweis - Papiere (4.67) / Legende (2.13)
- Animal names (60)*
- Adler, Amsel, Barsch, Biene, Bison, Eber, Elster, Eule, Falter, Fasan, Fliege, Floh, Frosch, Fuchs, Gepard, Grille, Hahn, Hammel, Hecht, Hering, Hummer, Hyäne, Igel, Iltis, Kakadu, Kalb, Kauz, Kobra, Krake, Kröte, Lama, Lamm, Laus, Luchs, Mammut, Marder, Meise, Ochse, Pfau, Pudel, Puma, Python, Ratte, Reiher, Robbe, Rochen, Schabe, Schaf, Schwan, Spinne, Taube, Tiger, Wanze, Wels, Wespe, Wiesel, Wisent, Wolf, Wurm, Zander

4 Data columns in .csv file

1. **prime** prime/preview word
2. **target** target word
3. **related** 0: unrelated/non-associated, 1: related/associated
4. **id_pair** unique ID for each related pair plus its recombination
5. **id_unit** unique ID for each preview-target unit
6. **wordLength** word length (in letters)
7. **rating_mean** mean score in association rating (scale from 1 to 5, N=12)
8. **rating_median** median score in association rating
9. **rating_std** standard deviation of score in asso. rating
10. **rtime_mean** mean reaction time in rating (there was no time pressure)
11. **rtime_median** median reaction time in rating
12. **rtime_std** standard deviation of reaction time in rating
13. **pairQuality** difference in rating score between related and unrelated pair
14. **unitQuality** difference in rating score between both related and both unrelated pairs of preview-target unit
15. **mutualInfo** MutualInformation, log-10 scaling, computed from DWDS corpus
16. **primeSAMPA** phonetic SAMPA code for prime
17. **targetSAMPA** phonetic SAMPA code for target
18. **prime_syl_dlex** prime syllables (automatically coded by dlexDB.de)
19. **target_syl_dlex** target syllables (automatically coded by dlexDB.de)
20. **prime_nsyl** number of syllables in prime (manually coded)
21. **target_nsyl** number of syllables in target (manually coded)
22. **prime_gender** grammatical gender prime (1:masculine,2:feminine,3:neuter)
23. **target_gender** grammatical gender target
24. **matchinStrokeDir** number of corresp. letters in prime and target whos stroke direction matches (e.g.: second letter has ascender in both words)
25. **confusability_let1** visual confusability for letter 1 of prime and target
26. **confusability_let23** visual confusability (mean for letters 2 and 3)
27. **confusability_mean** visual confusability (mean of all letters)
28. **same_let1** is letter 1 identical in prime and target?
29. **same_let2** is letter 2 identical in prime and target?
30. **same_let_final** is last letter identical in prime and target?

31. **hammingDist** Orthographic Hamming Distance (Edit Distance)
32. **n_sharedBigr** Letter bigrams (e.g. "ab") shared by prime and target
33. **n_sharedTrig** Letter trigrams (e.g. "abc") shared by prime and target
34. **diceCoeff_Bigr** Bigram-based Dice Coefficient (after Lambert et al., 2002)
35. **same_phon1** is phoneme 1 identical in prime and target?
36. **same_phon2** is phoneme 2 identical in prime and target?
37. **same_phon_final** is last phoneme identical in prime and target?
38. **phonLevenDist** Levenshtein distance (Edit Distance) based on SAMPA codes
39. **prime_typeFreq** type frequency of prime (dlexdb.de)
40. **prime_lemmaFreq** lemma frequency of prime (dlexdb.de)
41. **target_typeFreq** type frequency of target (dlexdb.de)
42. **target_lemmaFreq** type frequency of target (dlexdb.de)

5 References

- Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, 11, 459-474.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26, 297-302.
- Hamming, R. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29, 147-160.
- Lambert, B. L., Donderi, D., & Senders, J. W. (2002). Similarity of drug names: Comparison of objective and subjective measures. *Psychology and Marketing*, 19, 641-661.
- Mueller, S. T. & Weidemann, C. T. (2008). Alphabetic letter perceivability, similarity, and bias. *Acta Psychologica*, 139, 19-37.
- Portele, T., Krämer, J., & Stock, D. (1995). Symbolverarbeitung im Sprachsynthesystem Hadifix. Proc. 6. Konferenz Elektronische Sprachsignalverarbeitung 1995, Wolfenbüttel, 97-104

Deutsche Zusammenfassung

Dieses Dokument beschreibt eine Sammlung von 312 Wortpaaren, die jeweils aus zwei semantisch assoziierten oder relatierten deutschen Nomen bestehen. Beide Nomen eines Paares haben jeweils die gleiche Wortlänge, die zwischen vier und sieben Buchstaben beträgt. Aus den 312 assoziierten Wortpaaren (z.B. Wald - Baum; Frau - Dame) wurden mittels Rekombination 312 nicht-assozierte Wortpaare gebildet (z.B. Wald - Dame; Frau - Baum). Assoziierte und nicht-assozierte Wortpaare wurden hinsichtlich ihrer visuellen, orthographischen und phonologischen Ähnlichkeit angeglichen. Für die Wortpaare liegen Wortstatistiken und Normierungsdaten vor.