

Final project

2023-06-06

SaGA (Lücking et al., 2010; Kucherenko, 2021). 25

:

1. Lücking, A., Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010). *The Bielefeld Speech and Gesture Alignment Corpus (SaGA)*. In M. Kipp, J. - P. Martin, P. Paggio, & D. Heylen (Eds.), LREC 2010 Workshop: Multimodal Corpora—Advances in Capturing, Coding and Analyzing Multimodality (pp. 92-98).
2. Kucherenko, T., Nagy, R., Neff, M., Kjellström, H., & Henter, G. E. (2021). *Multimodal analysis of the predictability of hand-gesture properties*. arXiv preprint arXiv:2108.05762.

1) (.eaf), 2) (.npz), 3) 3 -
(.bvh). ELAN
Python.
(),
preprocessing_data.py.
(),
(, ,
,
Blühndorn (1994), 2 3. ,
Zeichendeiktika Lokaldeiktika Blühndorn',
,

:

1. Blühndorn, H. (1994). *Deixis und Deiktika in der deutschen Gegenwartssprache*. Deutsche Sprache, 21(1), 44-62.
2. *Präpositionaladverb*.
3. *Deiktisches Adverb*.

...

```
data <- read.csv("https://raw.githubusercontent.com/anna-a-m/andan-project/main/data.csv")
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2     3.4.2      v tibble    3.2.1
## v lubridate   1.9.2      v tidyr     1.3.0
## v purrr       1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag() masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
str(data)
```

```
## 'data.frame':    90776 obs. of  11 variables:
## $ Form          : chr  "ja" "also" "ich" "bin" ...
## $ Lemma         : chr  "ja" "also" "ich" "sein" ...
## $ Pos           : chr  "PTKANT" "ADV" "PPER" "VAFIN" ...
## $ Phrase        : chr  "" "" "" "" ...
## $ Hand          : chr  "" "" "" "" ...
## $ Phase         : chr  "" "" "" "" ...
## $ Practice      : chr  "" "" "" "" ...
## $ Semantic      : chr  "" "" "" "" ...
## $ HandShapeShape: chr  "" "" "" "" ...
## $ PalmDirection : chr  "" "" "" "" ...
## $ IsDeictic     : chr  "non-deictic" "non-deictic" "non-deictic" "non-deictic" ...
```

```
          90,776          11          ,          7          .          -
```

```
data %>%
  count(IsDeictic, sort = TRUE, name = "count") %>%
  mutate(part = round(count/nrow(data), digits=2))
```

```
##      IsDeictic count part
## 1 non-deictic 82465 0.91
## 2      deictic  8311 0.09
```

```
          ,          ,          .          10-          .
```

```
data[, 5:length(data)-1] %>%
  summarise_each(list(n_distinct)) %>%
  pivot_longer(cols=Phrase:PalmDirection, names_to="Variable", values_to="Distinct_count") %>%
  arrange(Distinct_count)
```

```
## Warning: `summarise_each()` was deprecated in dplyr 0.7.0.
```

```
## i Please use `across()` instead.
```

```
## i The deprecated feature was likely used in the dplyr package.
```

```
## Please report the issue at <https://github.com/tidyverse/dplyr/issues>.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## # A tibble: 7 x 2
##   Variable      Distinct_count
##   <chr>          <int>
## 1 Hand              3
## 2 Phase              6
## 3 Semantic          13
## 4 Phrase            14
## 5 Practice          25
## 6 PalmDirection     56
## 7 HandShapeShape    141
```

```
data %>%
  filter(Phase != "") %>%
  distinct(Phase, IsDeictic) %>%
  arrange(IsDeictic)
```

```
##      Phase  IsDeictic
## 1     prep    deictic
## 2    stroke    deictic
## 3 post.hold    deictic
## 4      retr    deictic
## 5 pre.hold    deictic
## 6     prep non-deictic
## 7    stroke non-deictic
## 8 post.hold non-deictic
## 9      retr non-deictic
## 10 pre.hold non-deictic
```

```
data %>%
  filter(Phrase != "") %>%
  distinct(Phrase, IsDeictic) %>%
  arrange(IsDeictic)
```

```
##      Phrase  IsDeictic
## 1     iconic    deictic
## 2      move    deictic
## 3     deictic    deictic
## 4 beat-iconic    deictic
## 5   discourse    deictic
## 6 deictic-iconic    deictic
## 7 beat-deictic    deictic
## 8      beat    deictic
## 9 beat-deictic-iconic    deictic
## 10 beat-discourse    deictic
## 11     iconic non-deictic
## 12 beat-iconic non-deictic
## 13   discourse non-deictic
## 14      move non-deictic
## 15      beat non-deictic
## 16     deictic non-deictic
```

```
## 17      beat-deictic non-deictic
## 18      deictic-iconic non-deictic
## 19 beat-deictic-iconic non-deictic
## 20      beat-discourse non-deictic
## 21      deictic-discourse non-deictic
## 22      discourse-iconic non-deictic
## 23      discourse-indexing non-deictic
```

```
data %>%
  select(Lemma, Phrase, IsDeictic) %>%
  filter(IsDeictic == 'non-deictic') %>%
  filter(stringr::str_detect(Phrase, 'deictic')) %>%
  distinct(Lemma)
```

```
##      Lemma
## 1      aeh
## 2      von
## 3      aus
## 4      und
## 5      Innenhof
## 6      stehen
## 7      jeweils
## 8      zwei
## 9      Seite
## 10     an
## 11     jed
## 12     sein
## 13     ein
## 14     treppe
## 15     Hufeisen
## 16     dies
## 17     Treppe
## 18     Straße
## 19     quasi
## 20     dann
## 21     Rathaus
## 22     ab
## 23     hinterm
## 24     man
## 25     zu
## 26     solch
## 27     Kirche
## 28     über
## 29     der
## 30     Kirchentür
## 31     eckig
## 32     gut
## 33     so
## 34     nebeneinander
## 35     durch
## 36     zwischen
## 37     vorbei
```

## 38	biegen
## 39	in
## 40	direkt
## 41	abgehen
## 42	entlang
## 43	auf
## 44	groß
## 45	wenn
## 46	weiter
## 47	immer
## 48	gucken
## 49	sehen
## 50	nochmal
## 51	also
## 52	Park
## 53	ja
## 54	ich
## 55	bis
## 56	können
## 57	einmal
## 58	See
## 59	rum
## 60	herum
## 61	drei
## 62	viertel
## 63	Rund
## 64	um
## 65	Ding
## 66	Ausgang
## 67	Haus
## 68	recht
## 69	link
## 70	Uhr
## 71	dran
## 72	ging
## 73	geradeaus
## 74	aber
## 75	glauben
## 76	haben
## 77	irgend
## 78	rot
## 79	Fall
## 80	Waldallee
## 81	wir
## 82	halt
## 83	genau
## 84	reinkommen
## 85	rumfahren
## 86	Eisstand
## 87	im
## 88	Punkt
## 89	beim
## 90	erst
## 91	Startpunkt

## 92	nach
## 93	raus
## 94	vom
## 95	Kirchplatz
## 96	Hecke
## 97	wieder
## 98	hoch
## 99	Gebäude
## 100	am
## 101	Brunn
## 102	Kapelle
## 103	bißchen
## 104	Ausweg
## 105	drunter
## 106	egal
## 107	ob
## 108	oder
## 109	einfach
## 110	nur
## 111	mit
## 112	beid
## 113	Kante
## 114	vorne
## 115	auch
## 116	zum
## 117	ander
## 118	Fenster
## 119	dreieckig
## 120	vorher
## 121	Fluß
## 122	Brücke
## 123	nah
## 124	grün
## 125	neben
## 126	Tür
## 127	weil
## 128	zuerst
## 129	Weg
## 130	schlängeln
## 131	sich
## 132	schräg
## 133	machen
## 134	leicht
## 135	Bogen
## 136	skulptur
## 137	Mitte
## 138	ziemlich
## 139	auffällig
## 140	eigentlich
## 141	dass
## 142	drauf
## 143	u
## 144	Rand
## 145	Kuppel

##	146	Kreuz
##	147	sondern
##	148	Turm
##	149	Platz
##	150	es
##	151	doch
##	152	du
##	153	trotzdem
##	154	dick
##	155	Baum
##	156	daneben
##	157	müssen
##	158	abbiegen
##	159	dein
##	160	halten
##	161	sofort
##	162	schnell
##	163	schon
##	164	zweit
##	165	Eingang
##	166	wissen
##	167	nicht
##	168	noch
##	169	geben
##	170	weit
##	171	Laterne
##	172	fangen
##	173	unmittelbar
##	174	fahren
##	175	los
##	176	halb
##	177	Kreisverkehr
##	178	sozusagen
##	179	Meter
##	180	Röhre
##	181	blau
##	182	fallen
##	183	nämlich
##	184	kein
##	185	gleich
##	186	farbe
##	187	wie
##	188	Kurve
##	189	stark
##	190	Haltepunkt
##	191	Anfang
##	192	scharf
##	193	Haken
##	194	Bus
##	195	unterschiedlich
##	196	ganz
##	197	mal
##	198	manchmal
##	199	Eisverkäufer

##	200	werden
##	201	Wagen
##	202	würd
##	203	empfehlen
##	204	parkähnlich
##	205	Landschaft
##	206	Bank
##	207	Teich
##	208	Front
##	209	jetzt
##	210	bleiben
##	211	Mülleimer
##	212	Richtung
##	213	bei
##	214	Bauklotz
##	215	sagen
##	216	alleine
##	217	säule
##	218	lang
##	219	falsch
##	220	aufer
##	221	wichtig
##	222	Moschee
##	223	rund
##	224	Kirchenkreuz
##	225	mehr
##	226	entweder
##	227	nehmen
##	228	anstatt
##	229	Eiskugel
##	230	Kugel
##	231	erkennen
##	232	vorm
##	233	durchgehen
##	234	Uni
##	235	ähnlich
##	236	symmetrisch
##	237	paar
##	238	überall
##	239	klein
##	240	Kirchturm
##	241	irgendwann
##	242	rein
##	243	laufen
##	244	Kreuzung
##	245	relativ
##	246	Eingangstor
##	247	Wasser
##	248	starten
##	249	außerhalb
##	250	Weggabelung
##	251	heißen
##	252	praktisch
##	253	Gabelung

254 davon
 ## 255 einfassen
 ## 256 rasen
 ## 257 erstmal
 ## 258 jedenfalls
 ## 259 Seitentrakt
 ## 260 mein
 ## 261 beide
 ## 262 letzt
 ## 263 wägelchen
 ## 264 Eistüte
 ## 265 Abbiegung
 ## 266 irgendwie
 ## 267 Überführung
 ## 268 nun
 ## 269 Spitzdach
 ## 270 beginnen
 ## 271 führen
 ## 272 kreisförmig
 ## 273 ungefähr
 ## 274 vier
 ## 275 sechs
 ## 276 wo
 ## 277 acht
 ## 278 aufbauen
 ## 279 Wendeltreppe
 ## 280 innen
 ## 281 Spitze
 ## 282 finden
 ## 283 Stahlbogen
 ## 284 umranden
 ## 285 Lampe
 ## 286 bitte
 ## 287 drübergehen
 ## 288 Rechtskurve
 ## 289 hindurch
 ## 290 Linkskurve
 ## 291 Häuserreihe
 ## 292 stellen
 ## 293 spitze
 ## 294 hindurchgehen
 ## 295 tun
 ## 296 Eismann
 ## 297 Ei
 ## 298 beziehungsweise
 ## 299 Stadt
 ## 300 umzäunen
 ## 301 Mauer
 ## 302 treffen
 ## 303 Prinzip
 ## 304 eigen
 ## 305 Ende
 ## 306 Breitseite
 ## 307 als

## 308	Indiz
## 309	eben
## 310	wollen
## 311	wiederum
## 312	abzweigen
## 313	bevor
## 314	hinführen
## 315	sie
## 316	parallel
## 317	gegenüber
## 318	sobald
## 319	Hall
## 320	Stück
## 321	Mal
## 322	mhm
## 323	darauf
## 324	zulaufen
## 325	zurückgehen
## 326	wirklich
## 327	offensichtlich
## 328	bekommen
## 329	Kapellenturm
## 330	münden
## 331	dadurch
## 332	Teil
## 333	rauslaufen
## 334	dritt
## 335	irgendwo
## 336	Umweg
## 337	sollen
## 338	Stelle
## 339	Eiswaffel
## 340	Öffnung
## 341	zurück
## 342	Tor
## 343	verwirren
## 344	immerhin
## 345	liegen
## 346	lieb
## 347	Wohngebiet
## 348	geschlängelt
## 349	einzig
## 350	welch
## 351	Band
## 352	Ring
## 353	dreistöckig
## 354	Stock
## 355	Vorderseite
## 356	ober
## 357	na
## 358	Litfaßsäule
## 359	Bodenhöhe
## 360	komplett
## 361	normal

## 362	Höhe
## 363	Angelpunkt
## 364	unter
## 365	rausgucken
## 366	Wand
## 367	weg
## 368	zählen
## 369	reinführen
## 370	dazu
## 371	zeigen
## 372	dieselbe
## 373	verbunden
## 374	Wanderweg
## 375	Sandweg
## 376	theoretisch
## 377	natürlich
## 378	Viertelkreis
## 379	Eiswagen
## 380	gleichmäßig
## 381	gerade
## 382	stehend
## 383	nachdem
## 384	Hügel
## 385	vielleicht
## 386	kennen
## 387	Schal
## 388	Kreis
## 389	Statue
## 390	nichts
## 391	zur
## 392	Auswahl
## 393	befinden
## 394	wovon
## 395	wobei
## 396	oft
## 397	rausfahren
## 398	Fuß
## 399	übereinander
## 400	ach
## 401	Flügel
## 402	neu
## 403	teilen
## 404	Uhrzeigersinn
## 405	Aufgang
## 406	losgehen
## 407	dauern
## 408	Weile
## 409	danach
## 410	viel
## 411	Grün
## 412	zugehen
## 413	weitergehen
## 414	reingehen
## 415	Eisverkaufsstand

416 Gras
 ## 417 offen
 ## 418 drehen
 ## 419 obendrauf
 ## 420 Ausgangspunkt
 ## 421 Podest
 ## 422 Außengebäude
 ## 423 miteinander
 ## 424 Balkon
 ## 425 Parkanlage
 ## 426 durchs
 ## 427 Rechtsbogen
 ## 428 ansonsten
 ## 429 eher
 ## 430 alle
 ## 431 zwanzig
 ## 432 fünfzig
 ## 433 Sehenswürdigkeit
 ## 434 zwar
 ## 435 gewiß
 ## 436 Art
 ## 437 Halbkreis
 ## 438 Beispiel
 ## 439 gegenüberliegend
 ## 440 Ausfahrt
 ## 441 vorstellen
 ## 442 langgefahren
 ## 443 Fläche
 ## 444 dahinten
 ## 445 zehn
 ## 446 breit
 ## 447 Bach
 ## 448 breiten
 ## 449 fünf
 ## 450 begrenzt
 ## 451 Frontbau
 ## 452 dazwischen
 ## 453 ohne
 ## 454 Einkerbung
 ## 455 Zentimeter
 ## 456 dreißig
 ## 457 zukommen
 ## 458 zufahren
 ## 459 Dach
 ## 460 Ecke
 ## 461 Würfel
 ## 462 bilden
 ## 463 inklusive
 ## 464 all
 ## 465 derselbe
 ## 466 Größe
 ## 467 denken
 ## 468 genauso
 ## 469 losfahren

470 Mittelpunkt
 ## 471 Beton
 ## 472 selbst
 ## 473 mögen
 ## 474 ändern
 ## 475 bestimmen
 ## 476 Auge
 ## 477 baulich
 ## 478 sowas
 ## 479 daran
 ## 480 Torbogen
 ## 481 tief
 ## 482 außerdem
 ## 483 Notausgang
 ## 484 nein
 ## 485 Kreiselverkehr
 ## 486 Horizont
 ## 487 Parkbank
 ## 488 Moment
 ## 489 Gebäudeteil
 ## 490 Fensterreihe
 ## 491 sowohl
 ## 492 Straßenverlauf
 ## 493 achten
 ## 494 Vergleich
 ## 495 kurz
 ## 496 Zeit
 ## 497 Pforte
 ## 498 Biegung
 ## 499 Stand
 ## 500 sicher
 ## 501 sonst
 ## 502 wer
 ## 503 Schluß
 ## 504 bestehen
 ## 505 Türmchen

• (,)

dies, jen, dann,

• ,

hinterm, über

• ,

Kirche, Park, See

• ,

stehen, biegen

•
quasi, aeh

,

```

non_deictics <- data %>%
  filter(IsDeictic == 'non-deictic') %>%
  filter(!stringr::str_detect(Phrase, 'deictic'))

non_deictics <- non_deictics[sample(1:nrow(non_deictics)), ]
non_deictics <- non_deictics %>%
  drop_na()
non_deictics <- non_deictics[1:8311,]

deictics <- data %>%
  filter(IsDeictic == 'deictic')

final_df <- bind_rows(deictics, non_deictics)

final_df %>%
  count(IsDeictic, name="count")

##      IsDeictic count
## 1      deictic  8311
## 2 non-deictic  8311

```

```

library(vcd)

```

```

## Loading required package: grid

mosaic(IsDeictic ~ Phrase, data=final_df,
  main = "Phrase feature distribution",
  direction=c("h","v"), spacing=spacing_increase(),
  gp = gpar(fill=c('lightblue', 'lightgreen'))))

```

The treemap visualizes the distribution of phrases across two categories: 'IsDeictic' (blue) and 'non-deictic' (green). The y-axis lists the phrases: 'move', 'iconic', 'discussion', 'statistical', 'iconic-deictic', 'deictic', 'beat', and 'thematic'. The x-axis represents frequency, with labels from 0 to 10. The 'IsDeictic' section (blue) contains phrases like 'move', 'iconic', 'discussion', 'statistical', and 'iconic-deictic'. The 'non-deictic' section (green) contains phrases like 'beat' and 'thematic'. The size of each rectangle corresponds to the frequency of the phrase in that category.

```
final_df <- final_df %>%
  mutate(across(where(is.character), as.factor))
fit1 <- glm(Phrase ~ IsDeictic, final_df, family = binomial)
summary(fit1)
```

15

```
## Call:
## glm(formula = Phrase ~ IsDeictic, family = binomial, data = final_df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.15658    0.03597  59.947 < 2e-16 ***
## IsDeicticnon-deictic -0.38018    0.04760  -7.986 1.39e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 12476  on 16621  degrees of freedom
## Residual deviance: 12411  on 16620  degrees of freedom
## AIC: 12415
##
## Number of Fisher Scoring iterations: 4
```

```
1.      -      .
2.      -      .
```