

# *Okay*, so what are all the *cool kids* talking about?

## Analyzing spoken corpora for anglicisms in German

Jaime Hunt

The University of Newcastle, Australia

X International Conference on Corpus Linguistics

Cáceres, Spain, May 9-11, 2018

# Questions

1. What are the most frequently occurring anglicisms in spoken German?
2. How are they used?
3. What are the collocations?

## How?

- Create / compile corpus of spoken German
- Identify top 10 most frequent anglicisms
- Analyse for collocations – using t-test

# Why spoken German?

- General focus on written language, especially press
- Highly edited
- Style often innovative (especially *der Spiegel*)
- Bloomfield (1933: 21) writing is “merely a way of recording language by means of visible marks”
- Speech is primary manifestation of language (Lyons 1981)
- Spoken corpora better representative of everyday language

# Super Spoken Corpus

- *Hempels' Sofa*
  - Bavarian Archive for Speech Signals
  - Types: 11,754, tokens: 167,097
- *Deutsch Heute* 'German Today'
  - Institute for German Language
  - Types: 35,752, tokens: 945,284
- *Forschungs- und Lehrkorpus für gesprochenes Deutsch* (FOLK) 'Research and teaching corpus for spoken German'
  - Databank for Spoken German,
  - Types: 56,880, tokens: 1,672,863
- Combined: 80,280 types, 2,785,244 tokens

# Anglicisms

Görlach (1994: 224): an anglicism is “a word or idiom that is recognizably English in its form (spelling, pronunciation, morphology, or at least one of the three)”

## Duden

Busse (1993)

## Etymological Dictionary

Körner (2004)

## Print media

Langer (1996), Plümer  
(2000), Onysko (2007)

## Spoken

Hunt (forth.)

3.5% / 2%\*

\*excluding partial loans

3%

0.6 – 1%

4.5% (types)<sup>+</sup>

0.8% (tokens)

<sup>+</sup>manually identified

# Frequencies – top 10

Type	Tokens (ppm)	Type	Tokens (ppm)
<b>Ignored</b>		Computer	475 (1323)
		Internet	396 (1103)
<i>okay</i>	<i>5974 (16,639)</i>	Test	180 (501)
<i>cool</i>	<i>491 (1367)</i>	PC	137 (3820)
<i>Sport</i>	<i>964 (2685)</i>	Handy	135 (376)
<i>Tunnel</i>	<i>288 (802)</i>	Mail	128 (356)
<i>Sprachcomputer</i>	<i>151 (420)</i>	Job	117(325)
		Stress	114 (318)
		Team	111 (309)
		Party	106 (295)

# Semantic fields

- Technology:
  - Computer
  - Internet
  - PC
  - Handy
  - Mail 'E-mail'
- Other:
  - Job
  - Stress
  - Team
  - Party
  - Test

# *t*-scores

- Indicator of meaningful associations
- T-score  $> 2$ , min 3-5 per 100 million statistically significant (Hunston 2002, Stubbs 1995)
- T-score more suitable for smaller corpora than MI
- Window span of 4L and 4R of node



# Collocations for *Computer*

Frequency	T-score	Word
91	15.08006	habe 'have'
84	13.77759	ich 'I'
160	13.60841	und 'and'
5	11.64426	am 'at the'
9	8.1658	den 'the' (acc.)
7	7.71345	dem 'the' (dat.)
22	7.45961	mit 'with'
44	7.05995	gespielt 'played'
43	6.94678	dann 'then'
37	6.46829	gearbeitet 'worked'

# Collocations for *Computer*

Frequency	T-score	Word
91	15.08006	habe 'have'
84	13.77759	ich 'I'
160	13.60841	und 'and'
5	11.64426	am 'at the'
9	8.1658	den 'the' (acc.)
7	7.71345	dem 'the' (dat.)
22	7.45961	mit 'with'
44	7.05995	gespielt 'played'
43	6.94678	dann 'then'
37	6.46829	gearbeitet 'worked'

# Collocations for *Computer*

Frequency	T-score	Word
236	15.08006	habe 'have'
50	7.05995	gespielt 'played'
42	6.46829	gearbeitet 'worked'
41	6.39873	gesessen 'sat'
31	5.32159	bin 'am'
27	5.1377	Stunde 'hour'
17	4.11384	gesetzt 'sat down'
17	4.09816	geschrieben 'wrote'
15	3.85555	Internet
14	3.69991	gegessen 'ate'

# *haben and sein*

- As lexical verbs:
  - Ich *habe* einen Computer ‘I have a computer’
  - Ich *bin* glücklich ‘I am happy’
- As auxiliary verbs in perfect constructions
  - Ich *habe* gespielt ‘I played, I have played’
  - Ich *bin* gesetzt ‘I sat, I have sat’

# Collocations for *PC*

Frequency	T-score	Word
52	7.03769	habe 'have'
16	3.99418	gearbeitet 'worked'
7	2.64172	erledigt 'completed'
6	2.44674	PC
6	2.43738	geschrieben 'wrote'
5	2.23567	Excel
5	2.23013	Post 'post, mail'
5	2.22159	Arbeit 'work'
5	2.2144	gelesen 'read'
5	2.19689	Stunde 'hour'

# Collocations for *Internet*

Frequency	T-score	Word
88	8.99552	habe 'have'
48	6.92716	gesurft 'surfed'
29	5.17297	bin 'am'
15	3.85555	Computer
14	3.72661	Internet
14	3.18904	haben 'have'
10	3.11799	gelesen 'read'
9	2.96465	Seite 'page/site'
12	2.84714	war 'was'
8	2.78575	Fragebogen 'questionnaire'

# Collocations for *Test*

Frequency	T-score	Word
248	15.74728	Test
23	4.45324	habe 'have'
17	4.11472	beide 'both'
15	3.75024	gut 'good'
13	3.33612	war 'was'
9	2.95414	sagt 'say, says'
9	2.9248	gemacht 'did/made'
7	2.64368	eingeladen 'invited'
7	2.54179	geht 'goes'
6	2.40725	mache 'do/make'

# Problems with *beide* & *Test*

so jetzt dürft Ihr beide mal Test sagen Test Te

so jetzt sagt mal beide Test bitte Martina Test

äh sagt mal beide Test bitte Test Test ol

ontest sagt mal beide Test bitte Test Test Te

sagt mal beide Test Test Test gut gef

testen sagt mal beide Test Test Test Test



# Collocations for *Handy*

Frequency	T-score	Word
27	4.74318	ist 'is'
22	4.68902	Handy 'mobile phone'
13	3.26379	habe 'have'
10	3.15754	Telefon 'telephone'
10	2.92072	hat 'has'
8	2.57083	war 'was'
7	2.37933	haben 'have'

# Collocations for *Mail*

Frequency	T-score	Word
23	4.55221	habe 'have'
11	3.30774	schreiben 'write'
11	3.30298	gelesen 'read'
6	2.44709	Mail 'email'
5	2.23409	geschickt 'sent'
5	2.22793	Internet
5	2.22367	geschrieben 'wrote'

# Collocations for *Job*

Frequency	T-score	Word
30	5.10479	ist 'ist'
24	4.76384	hat 'have'
10	2.82456	habe 'have'
6	2.44412	neuen 'new'
6	2.44025	bekommen 'got/received'
6	2.40478	wäre 'would be'
6	2.20009	haben 'have'

# Collocations for *Stress*

Frequency	T-score	Word
13	3.5674	hast 'have'
8	2.82678	Stress
9	2.80159	haben 'have'
7	2.41321	war 'was'
5	2.23579	pur 'pure'
5	2.22003	gab 'gave/(there)was'
5	2.19898	Schule

# Collocations for *Team*

Frequency	T-score	Word
22	4.27781	ist 'is'
12	3.46282	Team
8	2.62351	haben 'have'
5	2.23591	Germany
5	2.23576	Server
5	2.22785	gespielt 'played'

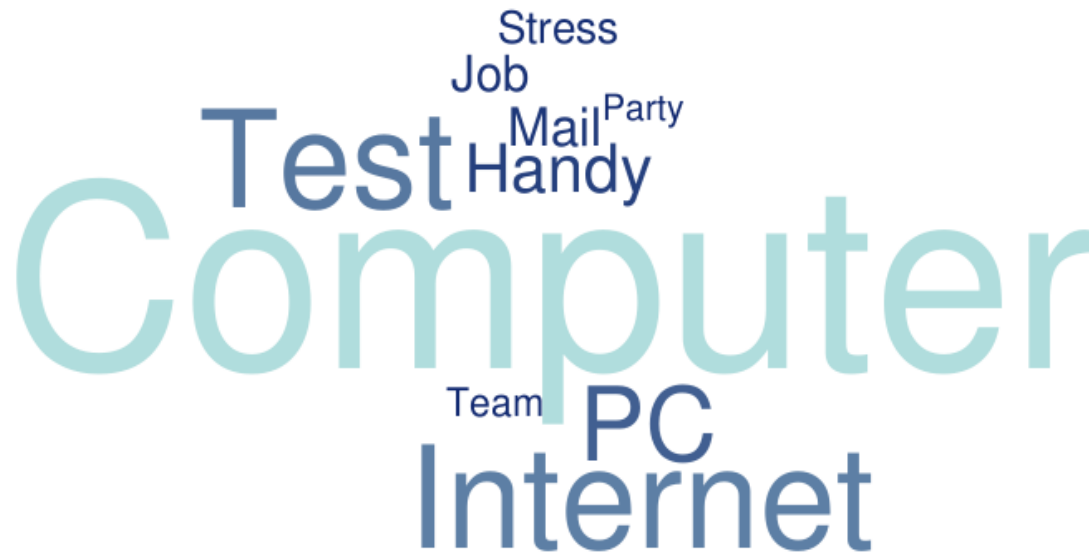
# Collocations for *Party*

Frequency	T-score	Word
19	431.280	machen 'do, make'
12	293.060	ist 'ist'
9	281.551	haben 'have'
6	244.784	Party
4	198.881	willst 'want'
4	186.002	gut 'gut'
5	180.337	habe 'have'
3	173.170	LAN 'local area network'
3	165.534	gemacht 'did, made'
3	163.854	geht 'goes'

# Topics in spoken German (Hunt, forth.)

Topic	Average
Advertising, Media	1.3%
Computing, Electronics, Technology	<b>29.3%</b>
Business, Finance	5.5%
Computer games	2.1%
<hr/>	
Sport	10.8%
Clothing, Fashion, & Cosmetics	1%
Leisure, Entertainment, Food	<b>24.5%</b>
Education, Teaching, Learning	4.2%
Other	20.8%

# So what are all the cool kids talking about? – top 10 anglicism tokens





# Collocate frequencies - types

haben	17	beide	1	Mail	1	sitzen	1
sein	11	bekommen	1	neuen	1	stress	1
machen	4	Computer	1	Party	1	surfen	1
schreiben	4	einladen	1	PC	1	Team	1
Internet	3	erledigen	1	Post	1	Telefon	1
lesen	3	essen	1	pur	1	Test	1
arbeiten	2	Excel	1	sagen	1	wollen	1
gehen	2	fragebogen	1	schicken	1		
gut	2	geben	1	schule	1		
spielen	2	Germany	1	Seite	1		
stunde	2	Handy	1	Server	1		
Arbeit	1	LAN	1	setzen	1		

# So what are all the cool kids talking about? – collocate frequencies types





# Conclusion

- *PCs* used for work-related activities (*worked, completed, wrote, work, read, Excel, post*), *Computers* less so (*played, ate, sat, sat down, wrote*)
- Verbs mostly in perfect aspect/past participle form
- German speakers use anglicisms when speaking about technology
- Integration of anglicisms – native collocate verbs

# References

Dąbrowska, Ewa. 2014. Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon* 9, no 3: 401–18.

Dodd, William. 2015. Under pressure? The anglicisms debate in contemporary germany as a barometer of german national identity today. *German Politics and Society* 33, no 114: 58-68.

Glahn, Richard. 2002. *Der Einfluß des Englischen auf gesprochene deutsche Gegenwartssprache: eine Analyse öffentlich gesprochener Sprache am Beispiel von "Fernsehdeutsch"* Frankfurt am Main: Peter Lang.

Gnatchuk, H. (2015). Anglicisms in the Austrian Newspaper *Kleine Zeitung*. *Glottometrics*, 31, 38-49.

Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Hunt, J. W. (forthcoming). Anglicisms in German: Tsunami or Trickle? In Amei Koll-Stobbe (ed.) *Informalization and Hybridization as a Consequence of Contact Processes*. Frankfurt am Main: Peter Lang

Knospe, Sebastian 2015. Pseudo-anglicisms in the language of the contemporary german press. In *False borrowings. English as a donor and recipient language*, eds Furiassi, C and Gottlieb, H, 99-122. Berlin: De Gruyter Mouton.

Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin: Walter de Gruyter.

Weisser, Martin. 2016. *Practical corpus linguistics: An introduction to corpus-based language analysis*. Hoboken, NJ Wiley-Blackwell.

Winter-Froemel, Esmé and Alexander Onysko. 2012. Proposing a pragmatic distinction for lexical anglicisms. In *The anglicization of european lexis*, eds Furiassi, C, Pulcini, V and González, FR, 43-64. Amsterdam Philadelphia: John Benjamins.

# Thank you

Comments and suggestions welcome 😊

Jaime Hunt  
The University of Newcastle, Australia  
[jaime.hunt@newcastle.edu.au](mailto:jaime.hunt@newcastle.edu.au)

X International Conference on Corpus Linguistics  
Cáceres, Spain, May 9-11, 2018