

# A Concept-Based Approach to Measuring the Success of Loanwords

Eline Zenner, Dirk Speelman, Dirk Geeraerts

## Borrowing and Borrowability

**likelihood of borrowing:** what is the chance that item A will be transferred from language X to language Y? Typically, linguists try to determine what features are common to items which are easily borrowed and use these to build clines of borrowability. It has, for instance, often been shown that nouns are more borrowable than any other POS.

**success of loanwords:** so far, less attention has been paid to what happens to a loanword once it is transferred. It is however an interesting question what determines to what extent the loanword will become a successful part of the receptor language-vocabulary. Also, several issues are to be noted in those studies dealing with the topic.

## Issues in Borrowability Research

unreliable measures of success

- success = raw frequency: problematic (Speelman *et al.* 2003)
- lack of set-external proof (Van Hout & Muysken 1994)

small datasets

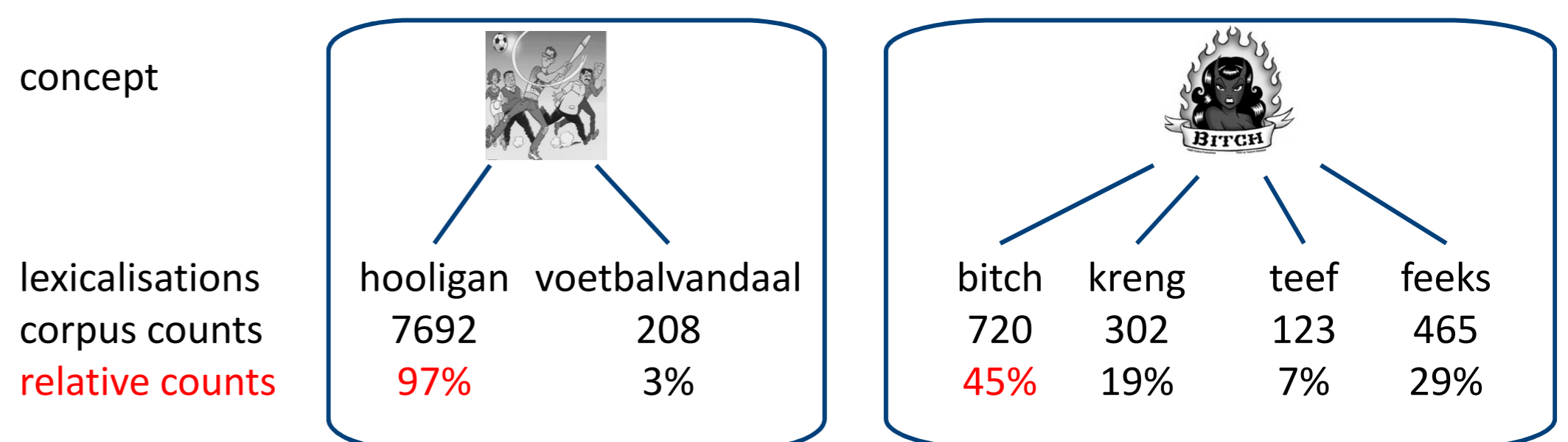
- manually counting number of tokens and processing results
- small corpora (see Geeraerts 2010)

limited attention for variation in success rates

- focus on a limited set of structural variables (e.g. POS)
- subjective categories (e.g. luxury vs. necessary anglicisms)
- lexical variation is largely ignored

## Concept-Based Approach

**onomasiological approach** to lexical variation (see e.g. Speelman *et al.* 2003) applied to loanwords: define the success of an anglicism as the relative preference for the anglicism vis-à-vis existing synonymous expressions



## English Person Reference Nouns in Dutch

selected loanwords

1. anglicisms: borrowed from English and recognizable as such
2. WordNet: automatic retrieval of all hyponyms of "person"
3. selection of over 125 anglicisms (*keeper, auditor* ...)

Dutch synonyms

1. coarse-grained (Edmonds & Hirst 2002)
2. combination of over 10 lexicographical sources
3. Word Space Models (Peirsman *et al.* 2007)

corpus counts

1. Dutch newspapers (> 1 billion words)
2. 4 subcorpora (2 regions, 2 registers)
3. semi-automatic token-counts

Example retrieval	Clean hit?
Miles Davis gebruikte bij het album <i>Bitches Brew</i> hetzelfde idee. (MD used the same idea for the album <i>Bitches Brew</i> )	proper noun
Elke voetbalsupporter werd aangesproken: "You wanna fight? You <i>hooligan</i> ?" (Every soccer fan was addressed: (...))	English context
BSE ontstaat doordat dieren veevoer eten waarin dierlijke <i>krenge</i> verwerkt zijn. (BSE occurs when animals eat feed containing carrion)	wrong meaning
Callas wordt in biografieën steeds opnieuw afgeschilderd als "een <i>feeks</i> " (In biographies, Calles is always portrayed as "a bitch")	clean hit
Ma, een uitgesproken <i>bitch</i> , doet het duidelijk beter dan pa. (Mum, a real bitch, clearly does better than dad)	clean hit

## Predictors and Results

**mixed effect linear regression**

response:

- 4 measurements per anglicism per subcorpus
- transformation from % to log(odds)

random effect:

- lemma

strength of the model

- R<sup>2</sup> for fixed effects only model **39%**
- reduction in StdDev ranefs **24%**

predictor/level	est.	StD	z/t	p-val	
(Intercept)	5.38	1.20	4.48	0.00	***
era introduction/borrowing:					
<1945 vs. 1945-1989	-0.46	0.25	-1.82	0.07	.
<1989 vs. >1989	-0.32	0.15	-2.17	0.03	*
age concept at intro anglicism:					
new concept vs. 1-80 yrs	-0.74	0.30	-2.50	0.01	*
<80 yrs vs. >80 yrs	-0.37	0.15	-2.52	0.01	*
conceptfrequency in RL (log)	-0.69	0.12	-5.56	0.00	***
lexical field: sports and leisure	-0.68	0.75	-0.90	0.37	
lexical field: moneymaking	-1.04	0.72	-1.45	0.15	
lexical field: social (neutral)	-0.42	0.71	-0.59	0.55	
lexical field: social (deviance)	-1.22	0.75	-1.63	0.10	.
<b>not significant</b>					
length word					
region					
register					

Not in final model

