

A multi-methodological study of *makan* 'eat': The challenge of divergent results

Gede Primahadi Wijaya Rajeg ^a; John Newman ^b; I Made Rajeg ^a

Universitas Udayana, Indonesia ^a

Monash University, Australia ^b

Evidence in linguistics

- Until mid 1990s, linguistic evidence from different methods is contested **against** each other (cf. Arppe & Järvikivi 2007:132; Kepser & Reis 2005)
 - Chiefly between **introspection** vs. **corpus-based** evidence
 - Judgement about (inf/sup)eriority of one type of method and data compared to the other (Arppe et al 2010:3)

- Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.
- Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zeschel. 2010. Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora* 5(1). 1–27.
- Kepser, Stephan & Marga Reis. 2005. Evidence in linguistics. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (Studies in Generative Grammar 85), 1–6. Berlin; New York: Mouton de Gruyter.

Renaissance in linguistic methods

- Since the early 2000s, usage-based/cognitive linguistics attempts to **combine** different methods leading to different types of data/evidence (Gilquin & Gries 2009; Gries, Hampe & Schönefeld 2010:59)
 - **Corpus-based AND experimental** methods
- The goal: **convergence** between, and **validation** of, different types of evidence

- Gilquin, Gaëtanelle & Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26. doi:[10.1515/CLLT.2009.001](https://doi.org/10.1515/CLLT.2009.001)
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2010. Converging evidence II: More on the association of verbs and constructions. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional Research*, 59–72. Stanford, CA: CSLI. (29 January, 2012).

Aiming for convergence

- Divjak, Dagmar & Stefan Th. Gries. 2008. Clusters in the mind?: **Converging evidence** from near synonymy in Russian. *The Mental Lexicon* 3(2). 188–213. doi:[10.1075/ml.3.2.03div](https://doi.org/10.1075/ml.3.2.03div).
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. **Converging evidence**: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4). 635–676.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2010. **Converging evidence II**: More on the association of verbs and constructions. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional Research*, 59–72. Stanford, CA: CSLI. (29 January, 2012).
- Wulff, Stefanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig & Chelsea J. Leblanc. 2009. The acquisition of Tense-Aspect: **Converging evidence** from corpora and telicity ratings. *The Modern Language Journal* 93(3). 354–369. doi:[10.1111/j.1540-4781.2009.00895.x](https://doi.org/10.1111/j.1540-4781.2009.00895.x).

Converging evidence

- Divjak & Gries (2008)
 - Investigate nine synonyms for TRY in Russian using corpus and experimental methods
 - Corpus-based finding of three distinct clusters within the synonym set
 - Converging evidence for the three clusters from sorting and gap-filling tasks

• Divjak, Dagmar & Stefan Th. Gries. 2008. Clusters in the mind?: Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3(2). 188–213. doi:[10.1075/ml.3.2.03div](https://doi.org/10.1075/ml.3.2.03div).

But evidence may diverge...

- Newman & Sorenson Duncan (2019)
 - Subject preference of the lemma ROAR
 - (i) sentence-elicitation task AND (ii) adult corpora from the *Corpus of Contemporary American English*
 - Elicitation data: LION as the most frequent subject
 - Corpus data: LION **never** occurs as the most preferred one in terms of **raw** frequency and statistically-based **association measures**, but CROWD, FIRE, ENGINE, and WIND do.
 - Divergence reflects different yet valid linguistic realities

Our project

Expanding the multi-methodological paradigm into Indonesian usage-based/cognitive linguistics

Our project

- Focuses on words encoding frequent, universal, and basic human experience, the so-called “basic verbs” (Newman & Rice 2006)
 - Hold a particular fascination for cognitive linguists
- INGESTION predicate: ***makan*** ‘eat’
 - Important source of metaphorical meaning extension
 - Complex semantic, lexical and morphosyntactic properties (see Newman 2009; Newman & Rice 2006)

- Newman, John & Sally Rice. 2006. Transitivity schemas of English EAT and DRINK in the BNC. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (Trends in Linguistics. Studies and Monographs 172), 225–260. Berlin: Mouton de Gruyter.
- Newman, John. 2009. A cross-linguistic overview of “eat” and “drink.” In John Newman (ed.), *The linguistics of eating and drinking* (Typological Studies in Language v. 84), 1–26. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Our aim

- To investigate the **usage patterns** of *makan* ‘eat’ based on data from different methods:
 - Usage patterns: n-grams or word-sequence containing *makan*
- Different methods:
 - Experimental: sentence elicitation task
 - Corpus-based: newspapers and Twitter

Our aim | *preliminary analysis*

- 2-gram patterns containing *makan* ‘eat’
 - n-grams are consecutive sequence of n-words
 - The 2-gram pattern is represented as [$word_1$ $word_2$]
- Focus on the top-10 2-grams where *makan* ‘eat’ fills the **second-word slot**.
 - For instance, [dia_{w1} ***makan*** $_{w2}$] ‘3SG eat’
 - Approximating subject retrieval
- **Would the contrasted data from different methods converge in their top-10 2-gram patterns?**

Methodology

Sentence-elicitation task based on *makan* 'eat'

- Administered via Google Form
- Task description only indicates each participant needs to **answer three questions**
- Task was split into three sections
- Four classes of English Department, Udayana University, Bali (200 students max)
 - **129 students** provided their responses
 - $129 * 3$ sentences = **387 sentences were produced**

Snippet of elicitation responses

```
# A tibble: 387 x 1
```

```
  SENT
```

```
  <chr>
```

```
1 saya mau makan nasi goreng
```

```
2 "kamu mau makan apa ? "
```

```
3 "siapa yang makan coklatku ? "
```

```
4 saya ingin makan masakan lokal
```

```
5 saya tidak bisa makan produk susu
```

```
6 "apa menu makan siang spesial hari ini ? "
```

```
7 "makan tuh batu ! "
```

```
8 awas kau dimakan sama dia
```

```
9 makan hati
```

```
10 saya makan malam dengan kakak saya
```

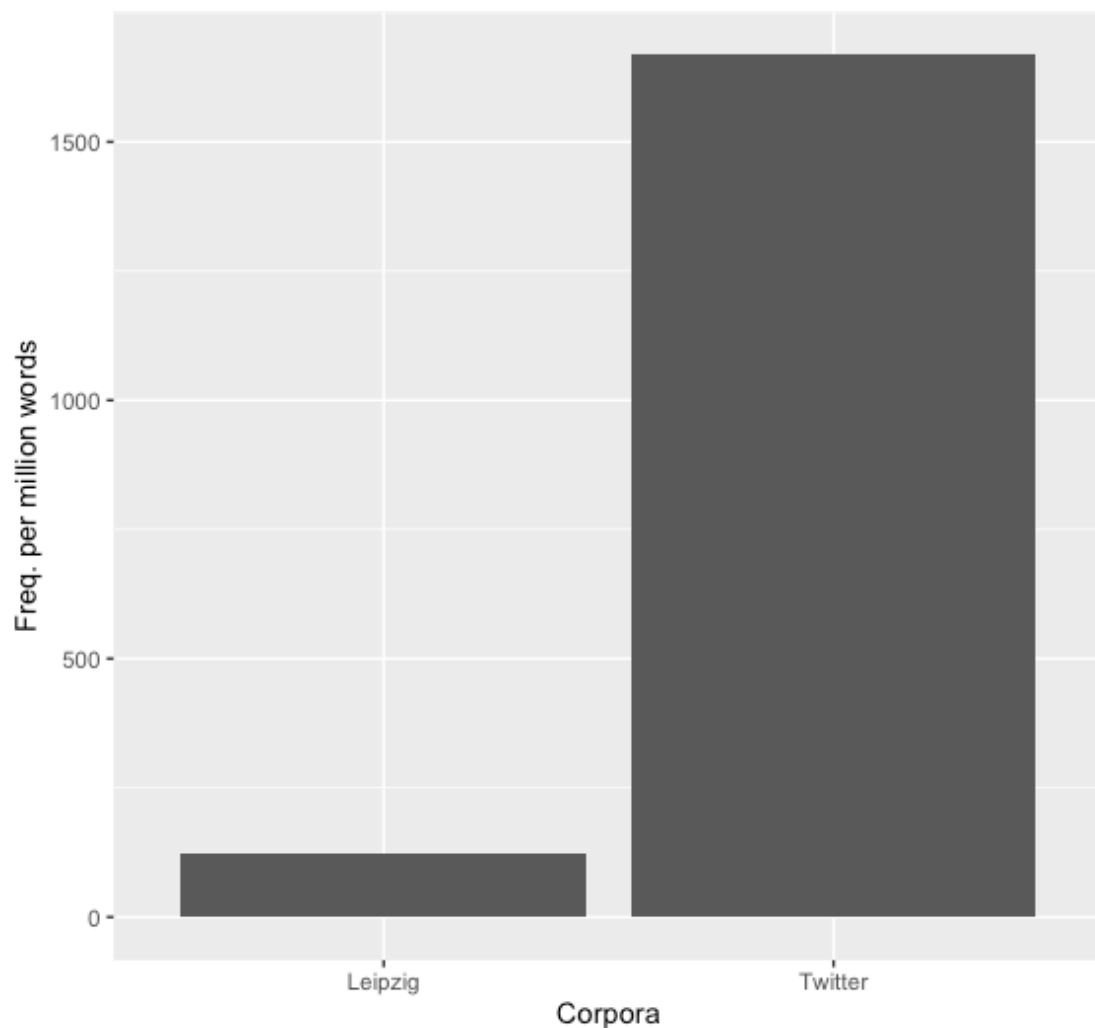
```
# ... with 377 more rows
```



Indonesian corpora

- Five newspapers files of the *Indonesian Leipzig Corpora* (**29,343,544** million word-tokens)
 - ind_news_2008_300K-sentences.txt
 - ind_news_2009_300K-sentences.txt
 - ind_news_2010_300K-sentences.txt
 - ind_news_2011_300K-sentences.txt
 - ind_news_2012_300K-sentences.txt
- Indonesian *Twitter* corpus (**9,764,055** million word-tokens)

Relative frequency of *makan* 'eat' in the corpus



makan is relatively much more frequent in the Twitter corpus, despite its much smaller total size than the studied Newspapers Leipzig Corpus

Results

Top-10 2-grams with *makan* ‘eat’ as the second word (Elicitation)

rank	ngrams	gloss	n
1	<i>saya_makan</i>	I eat	74
2	<i>suka_makan</i>	like to eat (sth.)	27
3	<i>sedang_makan</i>	PROG eat	16
4	<i>mau_makan</i>	want to eat	13
5	<i>sudah_makan</i>	PERF eat	12
6	<i>dia_makan</i>	3SG eats	11
7	<i>harus_makan</i>	must/have to eat	9
8	<i>ingin_makan</i>	want to eat	8
9	<i>tidak_makan</i>	NEG eat	8
10	<i>bisa_makan</i>	can eat	6
11	<i>itu_makan</i>	DET eat	6

Predominantly used in verbal context in a clause (e.g., as predicate head)

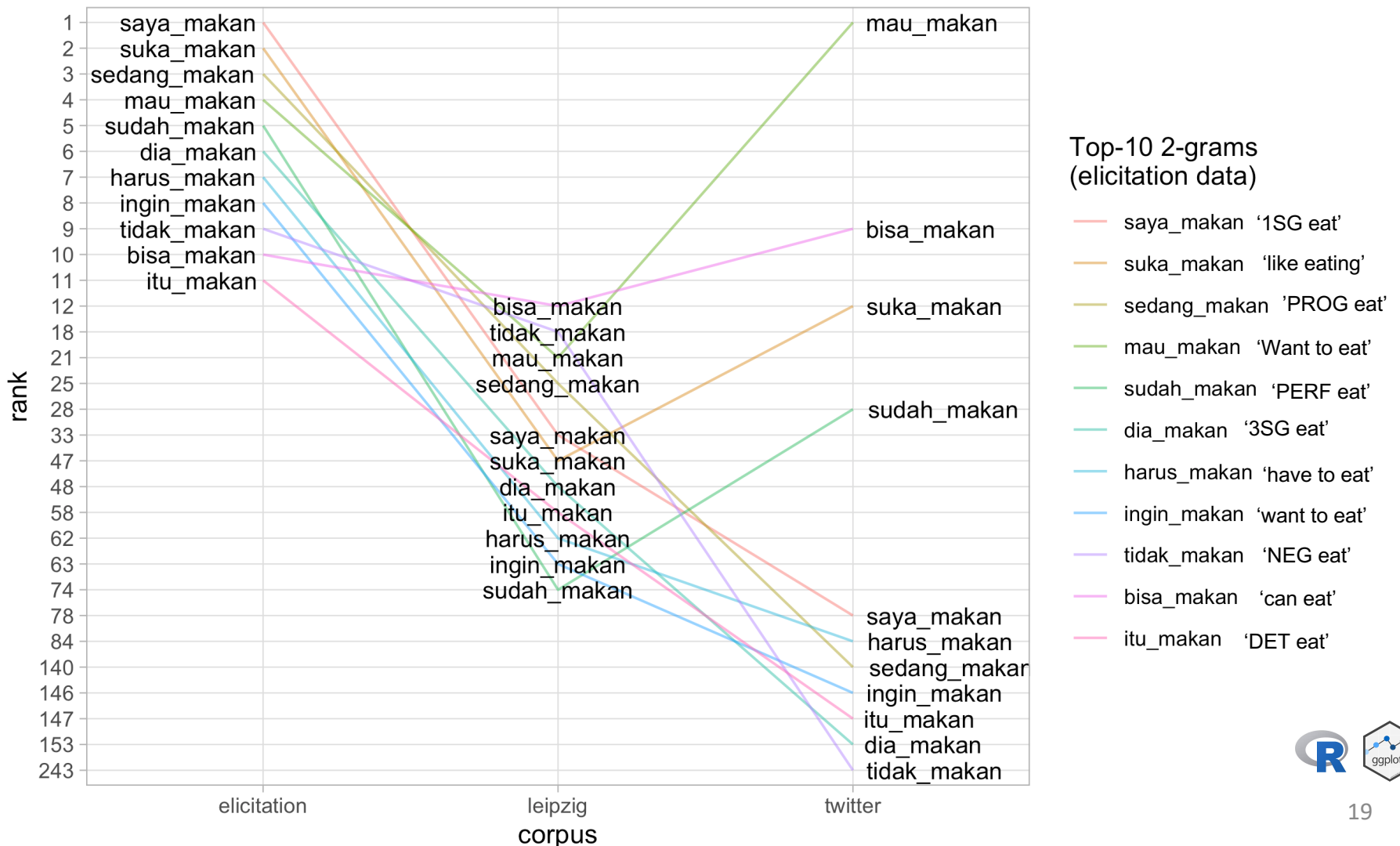
Top-10 2-grams with *makan* 'eat' as the second word (Elicitation)

rank	ngrams	gloss	n
1	<i>saya_makan</i>	I eat	74
2	<i>suka_makan</i>	like to eat (sth.)	27
3	<i>sedang_makan</i>	PROG eat	16
4	<i>mau_makan</i>	want to eat	13
5	<i>sudah_makan</i>	DEFE eat	12
6	<i>dia_makan</i>		
7	<i>harus_makan</i>		
8	<i>ingin_makan</i>		
9	<i>tidak_makan</i>		
10	<i>bisa_makan</i>		
11	<i>itu_makan</i>		

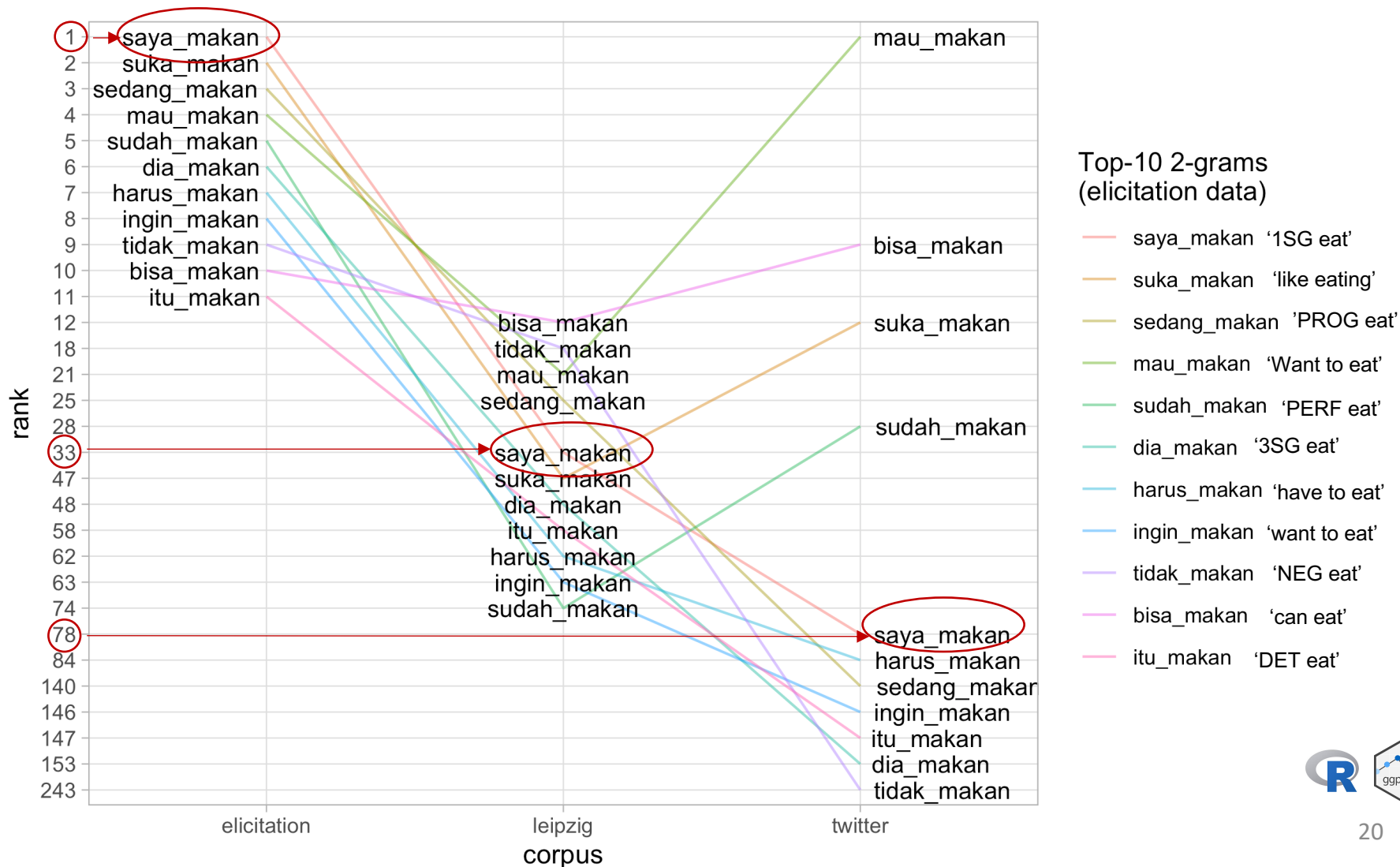
ayam_itu_makan 'that chicken eats'
gajah_itu_makan 'that elephant eats'
harimau_itu_makan 'that tiger eats'
***penelitian_itu_makan* 'that research eats/takes'**
perempuan_itu_makan 'that lady eats'
***proyek_itu_makan* 'that project eats/takes'**

Predominantly used in verbal context in a clause (e.g., as predicate head)

Top-10 2-grams in Elicitation and their distribution in other corpora



Top-10 2-grams in Elicitation and their distribution in other corpora

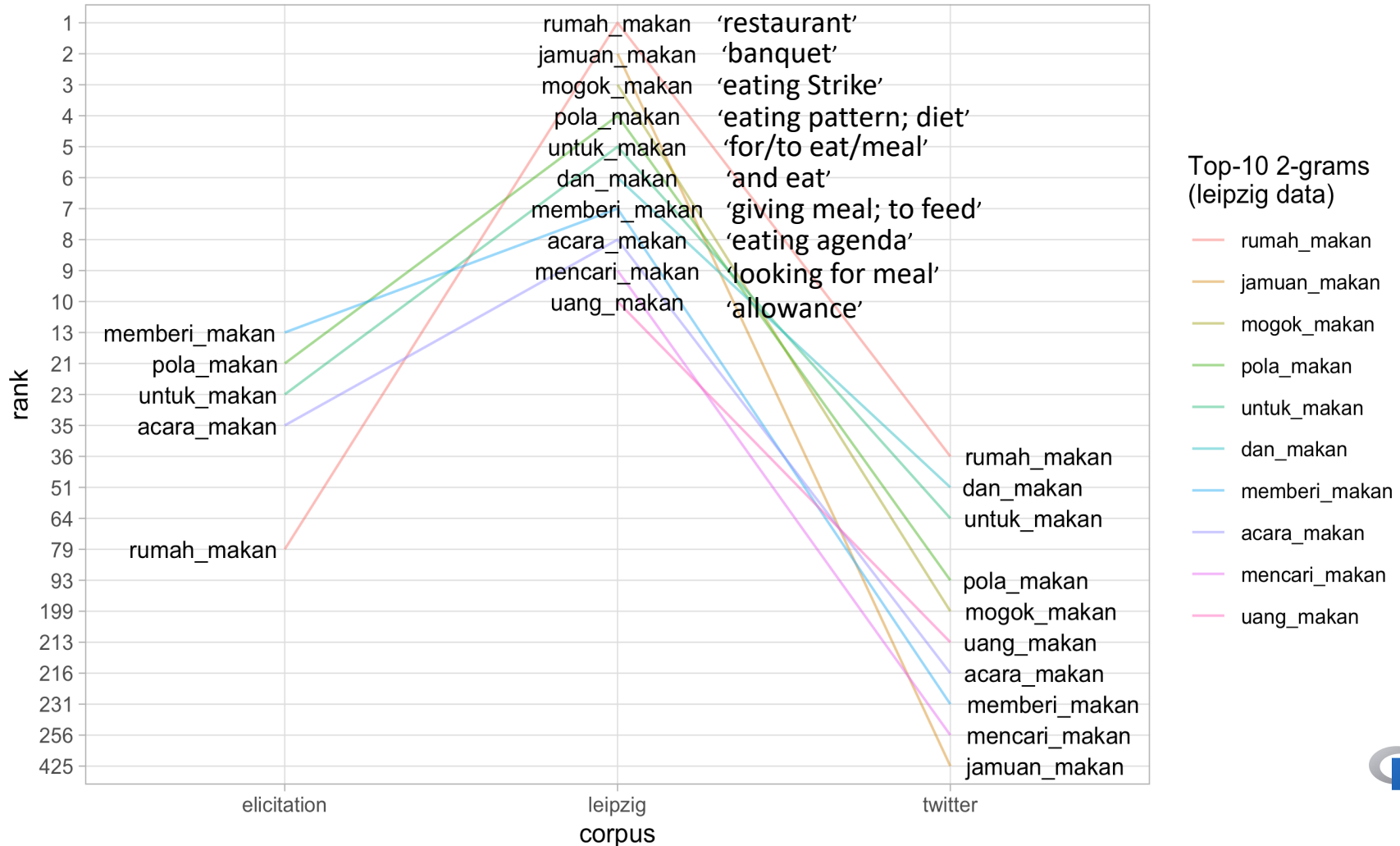


Top-10 2-grams with *makan* ‘eat’ as the second word (Leipzig)

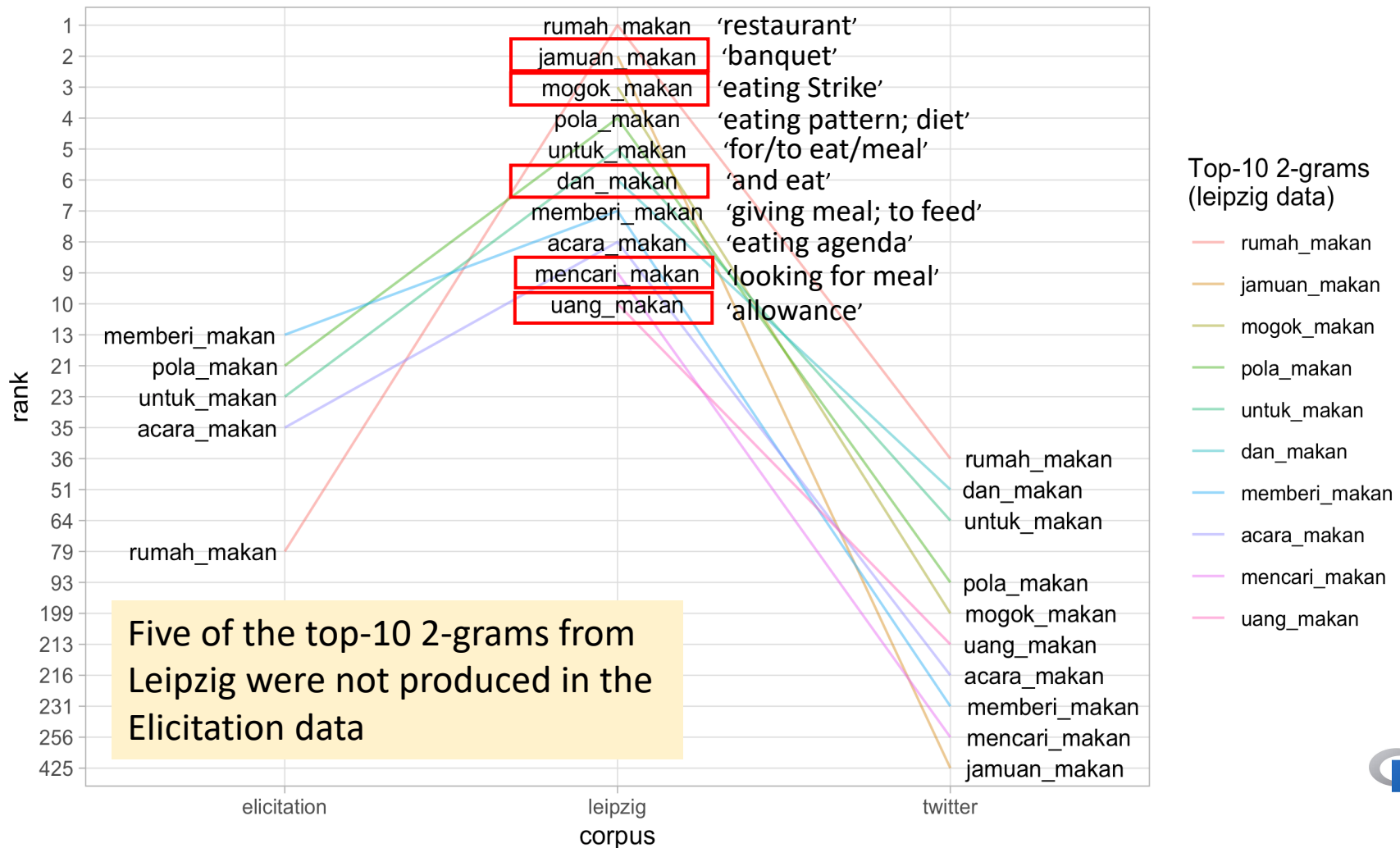
rank	ngrams	gloss	n
1	<i>rumah_makan</i>	house_eat; restaurant	284
2	<i>jamuan_makan</i>	service-to-guest_eat; banquet	215
3	<i>mogok_makan</i>	go-on-strike_eat; stop eating	166
4	<i>pola_makan</i>	pattern_eat; eating pattern/diet	156
5	<i>untuk_makan</i>	for/to_eat/meal	155
6	<i>dan_makan</i>	and_eat	91
7	<i>memberi_makan</i>	give_eat; to feed	88
8	<i>acara_makan</i>	agenda_eat; eating agenda	68
9	<i>mencari_makan</i>	look for_eat; look for meal	64
10	<i>uang_makan</i>	money_eat; allowance	56

Predominantly used in nominal contexts: (i) as verbal modifier for nominal compounds and (ii) as nominal direct object referring to ‘meal’ (e.g., 7 & 9)

Top-10 2-grams in Leipzig and their distribution in other corpora



Top-10 2-grams in Leipzig and their distribution in other corpora

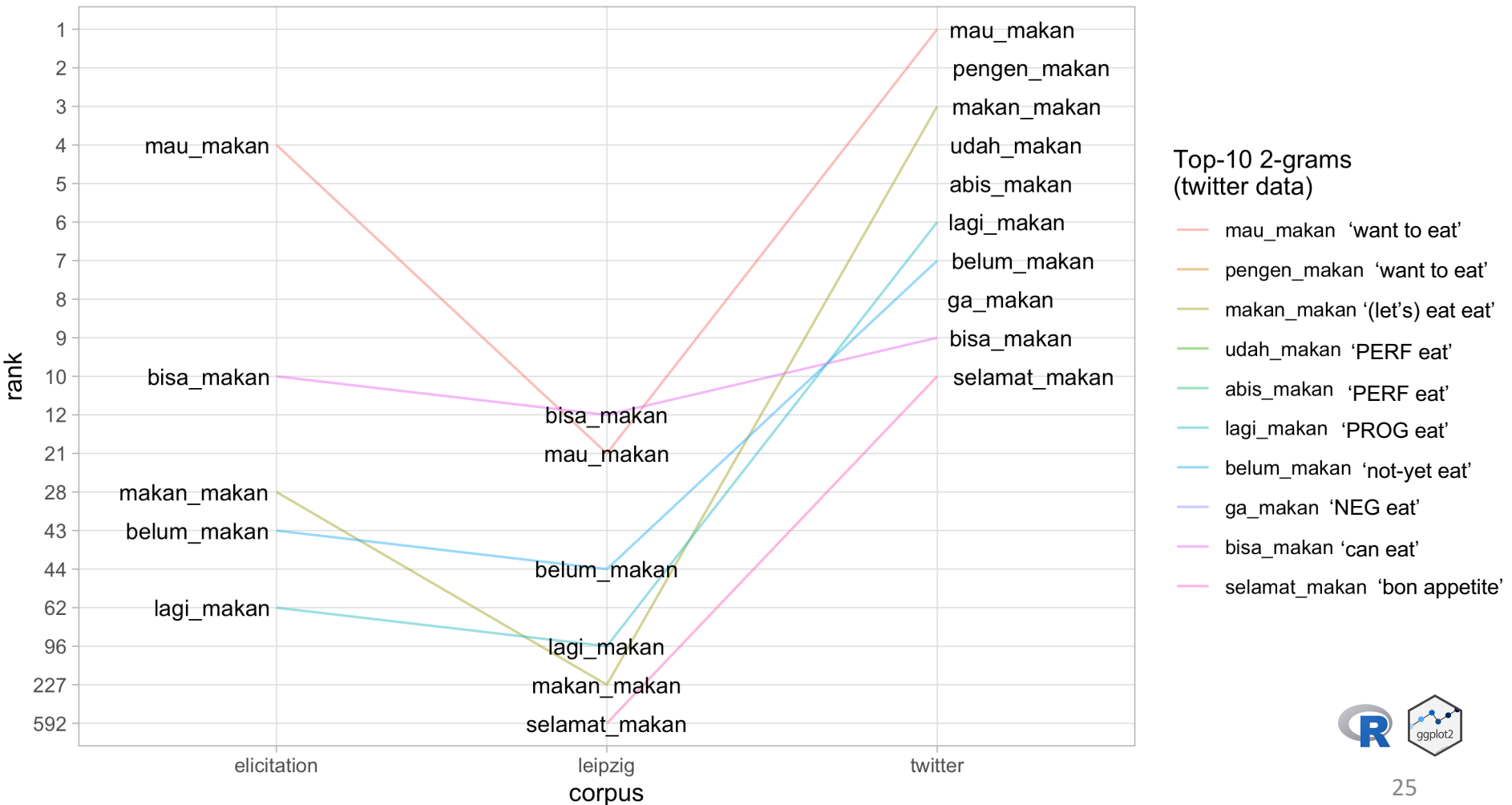


Top-10 2-grams with *makan* ‘eat’ as the second word (Twitter)

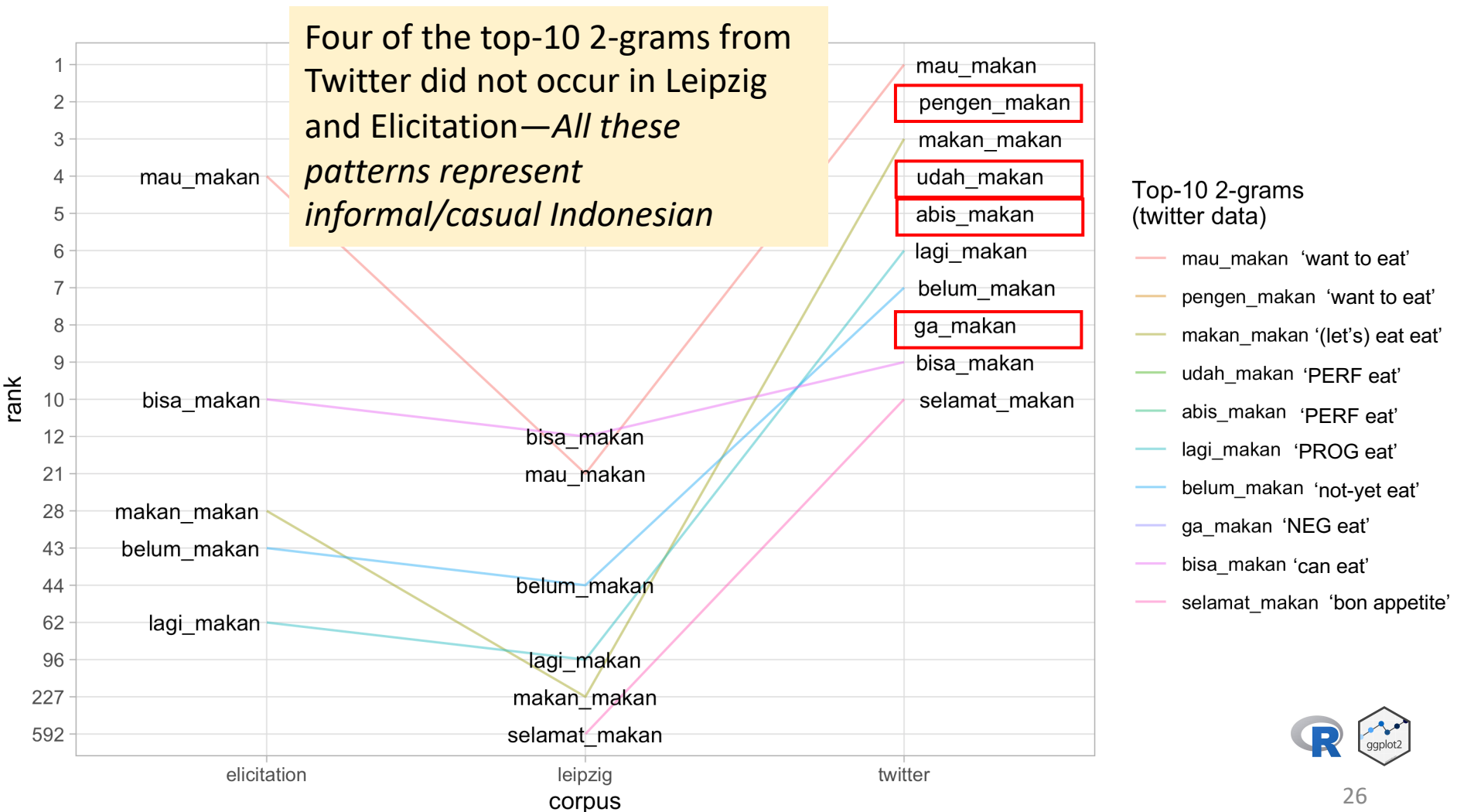
rank	ngrams	gloss	n
1	<i>mau_makan</i>	want to eat	485
2	<i>pengen_makan</i>	want to eat	382
3	<i>makan_makan</i>	eat eat	320
4	<i>udah_makan</i>	PERF eat	305
5	<i>abis_makan</i>	PERF eat; lit. finished eat	286
6	<i>lagi_makan</i>	PROG eat; lit. again eat	180
7	<i>belum_makan</i>	not-yet eat	160
8	<i>ga_makan</i>	NEG eat	151
9	<i>bisa_makan</i>	can eat	135
10	<i>selamat_makan</i>	happy eating; lit. save eating	125

Predominantly used in verbal context in a clause (e.g., as predicate head), as in the Elicitation data

Top-10 2-grams in Twitter and their distribution in other corpora



Top-10 2-grams in Twitter and their distribution in other corpora



Discussion

Experimental vs. Corpus results (I)

- Experimental

- Predominantly verbal usage of *makan*
- Directly related with bodily wants and desire (e.g., *saya makan* 'I eat'; *dia makan* '(s)he eats'; *suka makan* 'like to eat'; *mau makan* 'wants to eat')
- Approximate preferred syntactic subjects present

- Corpus (Leipzig)

- Predominantly used in nominal contexts as (i) modifier of (gastronomy-related) nominal compounds, and (ii) noun referring to 'meal'
- Patterns with potential syntactic subjects, e.g., *saya makan* '1SG eat' & *dia makan* '3SG eat', are not that prominent

Experimental vs. Corpus results (II)

- Experimental

- Predominantly verbal usage of *makan*
- Directly related with bodily wants and desire (e.g., *saya makan* 'I eat'; *dia makan* '(s)he eats'; *suka makan* 'like to eat'; *mau makan* 'wants to eat')
- Approximate preferred syntactic subjects present

- Corpus (Twitter)

- Similar flavour as in the Elicitation indicating bodily wants and desire
- Diglossic nature of Indonesian was revealed from several colloquial version of the 2-grams (e.g. *pengen makan* vs. *mau makan* 'want to eat'; *ga makan* vs. *tidak makan* 'NEG eat').

Convergence vs. Divergence (I)

- Strictly speaking, Elicitation vs. Leipzig evidence **diverge** wrt:
 - i. predominant verbal vs. nominal usage
 - ii. bodily wants vs. gastronomy-related terms
 - iii. **register** (Biber et al. 2002) of the top 2-grams
- Leipzig reflects *newspapers register* directed for wide audience, not directly interactive, conveying general information (proper nouns are more commons than pers. pronouns) (Biber et al. 2002)

Convergence vs. Divergence (II)

- Elicitation and Twitter data may **converge** wrt:
 - i. predominant verbal usage
 - ii. their *conversational register* or tone
- Conversational register represents interactive form of personal communication, abundant use of personal pronouns (esp. in Elicitation) (Biber et al. 2002:4-5)

Convergence vs. Divergence (III)

Dąbrowska (2014:411):

“patterns found in corpora need not necessarily reflect patterns in speakers’ minds”

What is frequent (usage pattern) in the corpus may not reflect the entrenchment and salience of the pattern in the speakers’ minds (e.g., when prompted to produce sentence using bodily-related word) (cf. Schmid 2010)

- Dąbrowska, Ewa. 2014. Words that go together: Measuring individual differences in native speakers’ knowledge of collocations. *The Mental Lexicon* 9(3). 401–418. doi:[10.1075/ml.9.3.02dab](https://doi.org/10.1075/ml.9.3.02dab).
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101–133. Berlin: Mouton de Gruyter.

Conclusion

- Sentence-elicitation task and corpus-based approaches are grounded in quite different realities:
 - ALL results should NOT necessarily converge
- Each data type has its own merit
- Converging AND diverging evidence enrich our understanding of alternative data and methods wrt a linguistic phenomenon (cf. Kepser & Reis 2005; Arppe & Järvi­kivi 2007, *inter alia*)

Thank you

Usage-based approach and Cognitive Linguistics (Dancygier 2017)

- “language study needs to be *usage-based*”
- “calling for an end to ‘armchair linguistics’”
- “thorough account of the facts of language, in its use and context, in its various instantiations, and in its connection to culture on the one hand and to the mind on the other”
- “language as such emerges from usage”

Usage-based approach and Cognitive Linguistics (Dancygier 2017)

- **“Actual usage is at the core of cognitive linguistic study”**
- “For many cognitive linguists, engaging with broadly construed and varied methods of data collection has also become important. There are no artificially drawn dividing lines – attested and responsibly gathered linguistic data are all subjects of study.”
- “The usage-based approach means not only that theoretical concepts represent the data well, but also that lines of investigation can be postulated on the basis of what usage suggests.”

Dancygier, Barbara. 2017. Introduction. In Barbara Dancygier (ed.), *The Cambridge handbook of Cognitive Linguistics* (Cambridge Handbooks in Language and Linguistics), 1–10. New York, NY: Cambridge University Press.