

NHH



FREQUENCY EFFECTS IN TRANSLATION: FURTHER REFINING THE NOTION OF GRAVITATIONAL PULL

SLE, Split, Croatia

18-21 September 2013





Today's talk: contextualizing

- 'Complementary methods and theory building' (email of 10.09)
- Patterns in translated text: possible explanations?



Today's talk

1. The gravitational pull hypothesis
 - a. Origins
 - b. Theoretically driven revisions
2. Methodological issues
3. The case
4. Preliminary analyses
 - a. factor one
 - b. factor two
 - c. factor three
5. Concluding remarks



1. The hypothesis

a. Origins

- Toury's laws (1980/1995)
- Baker's universals (1993, 1996)
- Evidence from SLA (e.g. Altenberg and Granger 2001; Ijaz 1986; Kellerman 1978, 1979; Liu and Shaw 2001)
- Cognitive Grammar (Langacker 1987, 2000, 2008)
- Distributed Feature Model of bilingual representation (de Groot 1992a, 1992b, 1993)



1. The hypothesis

The gravitational pull hypothesis (Halverson 2003)

Given a schematic network with a prototype or schema, the linguistic form linked to that prototype or schema will be overrepresented in translated, as opposed to non-translated text.

Posited as tentative explanation for simplification, generalization, normalization, sanitization, conventionalization, exaggeration of TL-specific features, leveling out

(Networks could represent e.g. polysemous lexemes or constructions, categories of synonymous lexemes or constructions)



1. The hypothesis

b. Theoretically driven revisions

- Need to allow for/incorporate cross-linguistic influence (cp. Jarvis & Pavlenko 2008). Bi- and multilinguals have a different 'multicompetence' (Cook 2003, Bassetti and Cook 2011). See translation as one of many bilingual modes of production.
- Need to incorporate insights from new models of bilingual representation and processing (e.g. Pavlenko 2009, Hartsuiker et al 2004, Hartsuiker 2013)



1. The hypothesis

b. Theoretically driven revisions

- Original presentation confounded three distinct (potential) sources of translational effects (all understood as frequency effects)
 - Prototype/schema in TL category ('pull proper')
 - Pattern of linkage within bilingual network (co-activation patterns and link weights)
 - Status of ST item in SL

(Cp. Halverson 2010)



2. Methodological issues

The status of corpus data: (see also references below)

‘Although corpus data do not reflect the characteristics of mental grammars directly, we do consider corpus data a legitimate source of data about mental grammars. Since the results of linguistic cognitive processes, e.g. corpus data, are not independent of, or unrelated to, the linguistic knowledge that is represented in the brain, we may assume with justification that characteristics observable in language usage reflect characteristics of the mental processes and structure yielding usage, even though we do not know the exact form of these mental representations.’ (Divjak and Arppe 2013: 229-230)



2. Methodological issues

Data types

Various corpora (monolingual, parallel, comparable) combined with psycholinguistic experimentation (Croft 1998; Sandra and Rice 1995; Sandra 1998; Tummers et al 2005; Heylen et al 2008, Gilquin 2008, Gilquin and Gries 2009)



3. The case

- Selection of case (*get*)
 - Independent semantic analyses (also crosslinguistic ones)
 - Rich polysemous network (cp basic verbs, Viberg 2002), with posited prototype
 - Sufficient isomorphism between semantic and syntactic distinctions



3. The case

Sense distinctions based on Gronemeyer (1999), Viberg (2002).
Sense 1 as prototype from Gronemeyer (1999); Johansson & Oksefjell (1996). Senses later confirmed by Berez and Gries (2008) in corpus analysis.

- get_1 – onset of possession
[1] Would you like me to go out and *get* some croissants?
- get_2 – stative possession
[2] *Have* you *got* any of those?
- get_3 - motion
[3] I don't want you to *get* there after dark.



3. The case

- get₄ – permission/obligation

You've got to take into account that I'm virtually single-handed here.

- get₅ – causation

[5] Despite his tuggings with the wrench he couldn't *get the screw to shift*.

- get₆ - inchoative

[6] Sit down and *get warm*.



3. The case

- get₇ - passive

[7] ... and you really do have to be a winklebrain to *get ejected* from there...

- get₈ - ingressive

[8] I'm going to *get moving*.

- get₉ – idioms



3. The case

Test no.	Theoretical issue	Proposition(s)	Design type	Data	Analysis/method
1	gravitational pull (TL network structure)	overrepresentation of schema/ prototype in transl.	cross-sectional	ENPC	distributions + stat. tests
2	gravitational pull (TL network structure + SL)	a)overrepresentation of schema/ prototype in transl. b) interaction with effect of SL network structure	cross-sectional comparative	a)BNC/TEC b) BNC/TEC + ENPC	distributions + stat. tests
3	gravitational pull of schemas	overrepresentation of schematic words in translated text	cross-sectional	BNC/TEC +??	word list comparisons + stat. tests
4	causal effect of network struct.	effect greater in translated text	cross-sectional	BNC/TEC	multivariate analysis
5	gravitational pull	order of production mirrors category structure	cross-sectional	Sentences	sentence generation task stat. analysis
6	gravitational pull	schematic or prototypical structures selected with less hesitation, more often edited out	case study	keystroke logs	pause units prior to structure, frequency of revision quant. analysis

(Halverson 2009)



4. Preliminary analyses: teasing apart factors

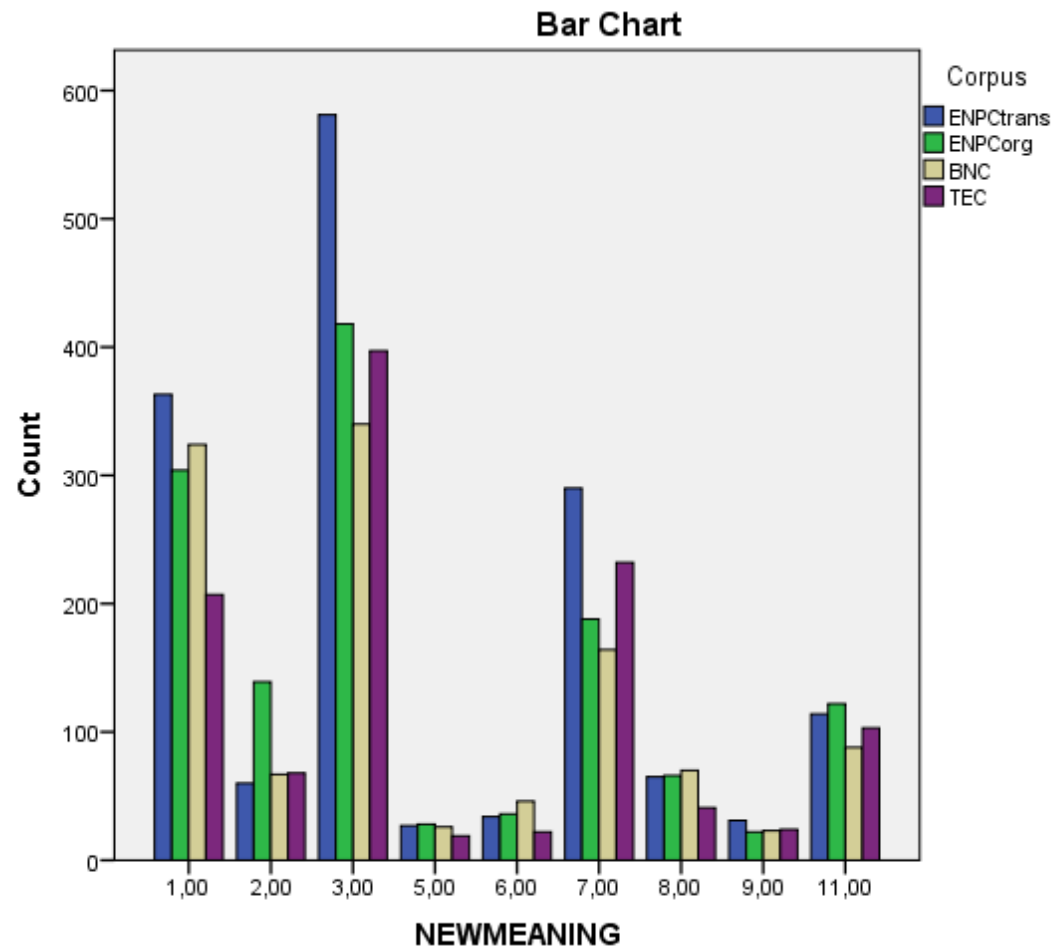
Factor one: gravitational pull proper (pull of prototype/most frequent members, ‘semasiological salience’ (Geeraerts 2000))

BUT: translation as analogous to bilingual priming (Hartsuiker et al 2004)

‘inverse preference effect’ in priming studies: ‘moderately infrequent structures [...] tend to be primed more strongly or reliably than more frequent structures [...]’ (Flett et al 2013:753).



4a. Preliminary analyses – factor one



$n=5149$, $X^2=137.54$, $df=24$, $p<.001$, Cramer's $V=.094$



4a. Preliminary analyses – factor one

- Checking TEC for ST effect

- Arabic, French, German, Spanish (n=774)

($\chi^2 = 56.54$, $df=30$, $p=.002$, Cramer's $V = .156$)



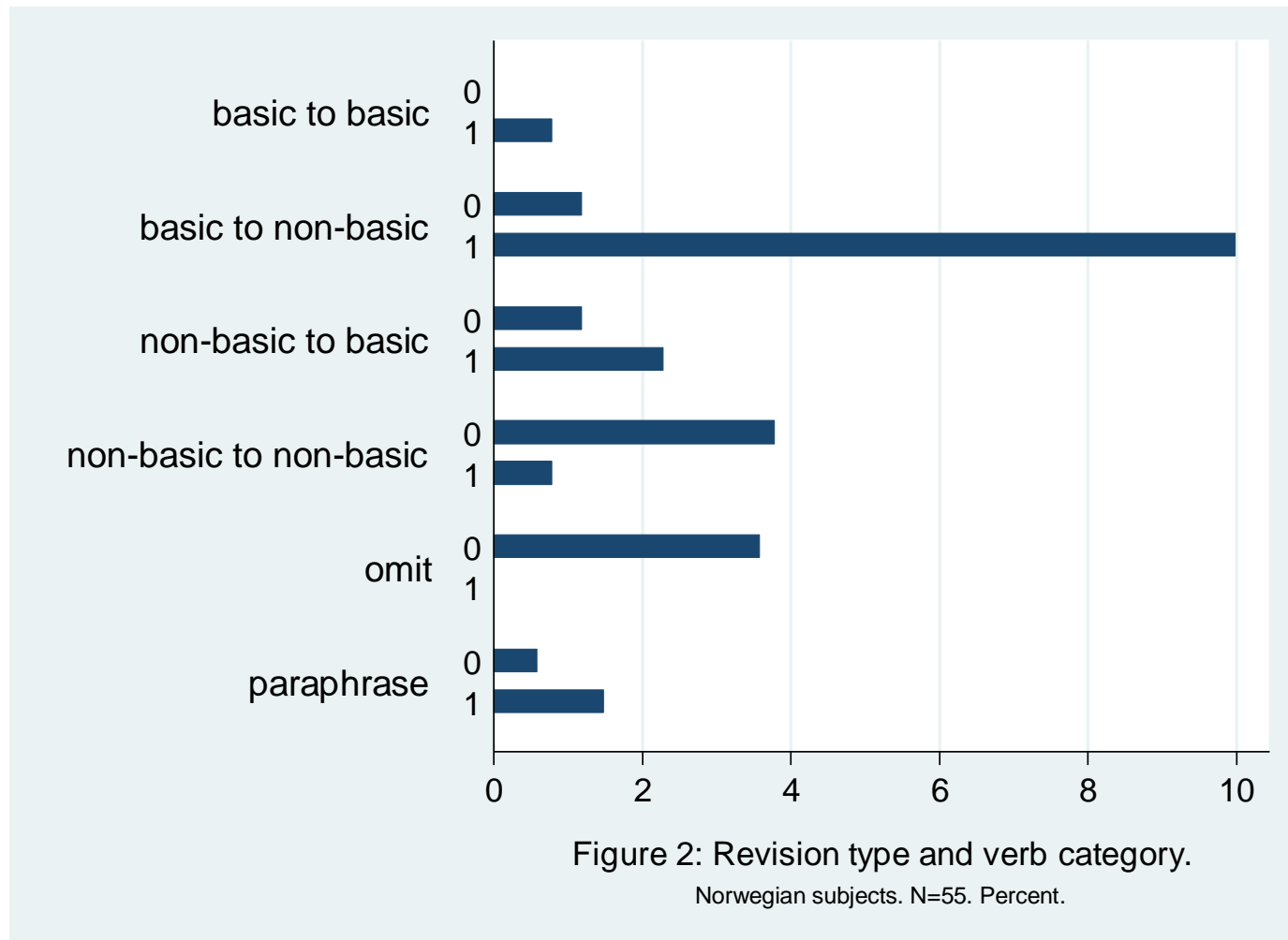
4a. Preliminary analyses: factor one

NOTE: This is corpus data, and it doesn't tell us anything direct about cognitive processes.

We need process (online) and performance data too.



4a. Preliminary analyses: factor one



(Halverson 2011)



4. Preliminary analyses

Factor one:

- highest frequency senses in Target category overrepresented (N-E translation)
- High frequency verbs revised more often
- Frequency relationship to range of translations requires further investigation

Factor two: linkage within network

Factor three: status of ST item in SL



4b. Preliminary analyses: factor two

Attraction-reliance measures (relative frequencies)
(Schmid 2010)

Attraction: ‘the degree to which a pattern attracts a particular noun’

Reliance: ‘the proportion of uses of nouns in the patterns vis-a-vis other usage-types of the same noun’
(Schmid 2010: 108)



4b. Preliminary analyses: factor two

For translational relationships (in translational corpus):

SC (source concentration) = frequency of item X as translation of Y compared to all occurrences of X in T

TC (target concentration) = frequency of item X as translation of Y compared to all occurrences of Y in S

4b. Preliminary analyses - linkage patterns for *få/bli* and *get* in ENPC



Preliminary figures:

	SC	TC
1 – onset (<i>få</i>)	43.25	25.20
2 – possession	n/a	n/a
3 – movement	n/a	n/a
4 – permission/ obligation (<i>få</i>)	35.29	5.02
5 – causation (<i>få</i>)	--	--
6 – inchoative (<i>bli</i>)	54.48	10.15
7 – passive (<i>bli</i>)/(<i>få</i>)	29.23/9.23	2.15/12.24

Data from Ebeling (2003)



4b. Preliminary analyses: factor two

When X is selected, chances are the starting point was Y. But from Y, only a few end up at X!

We need information about the other translation direction.

Still corpus data – not online. But, look at elicitation data:



4b. Preliminary analyses: factor two

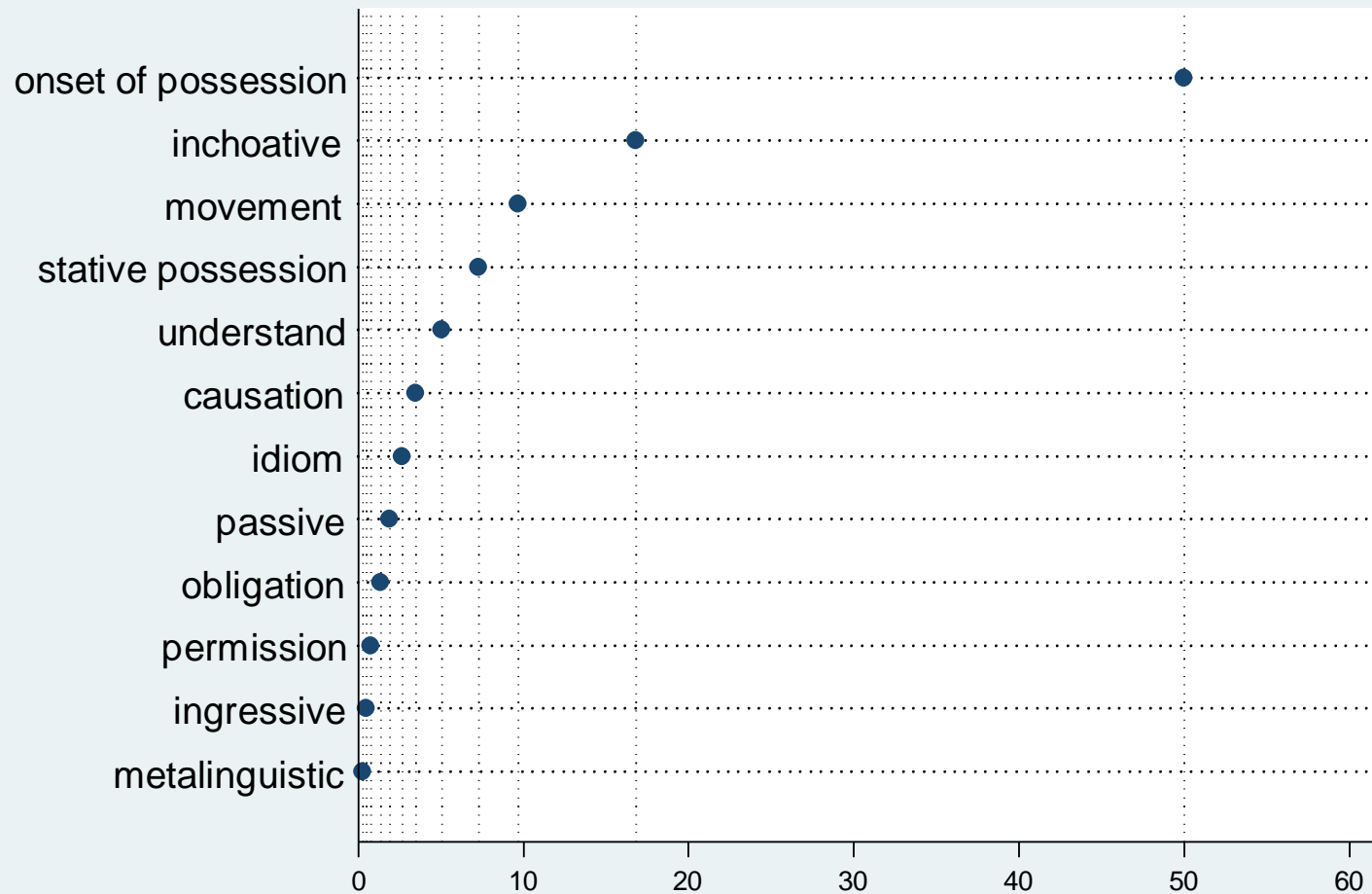


Figure 1: Sentence generation for 'get'.

Norwegian subjects. N=370. Percent.



4b. Preliminary analyses: factor two

Factor two:

- Source concentration much higher than target concentration for virtually all senses – is it easier to predict a source than a target?
- High source concentrations linked to overrepresentation in corpus (though not for passive)



4c. Preliminary analyses: factor three

Status of SL item in network (semasiological salience): *få* and *bli*

få - ENPC

	N	%
onset	623	40
poss/obl.	239	15
caus/ctrans	212	14
aspaux	157	10
mwv	244	16
passaux	49	3
Tditrans	18	1
TOTAL	1542	100

bli - ENPC

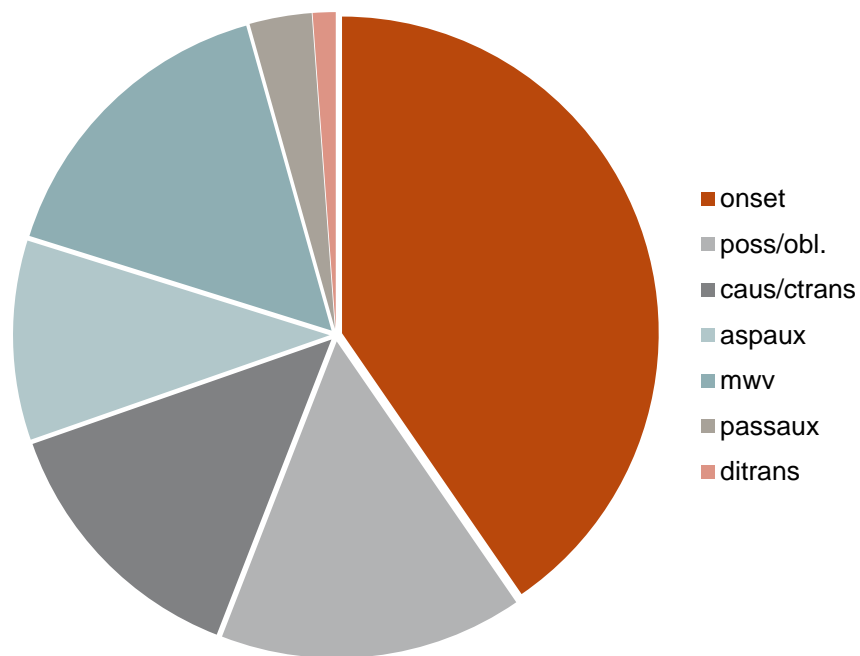
	N	%
cop+intr	1546	57
passaux	884	32
aspaux	177	6
mwv	119	4
TOTAL	2726	100



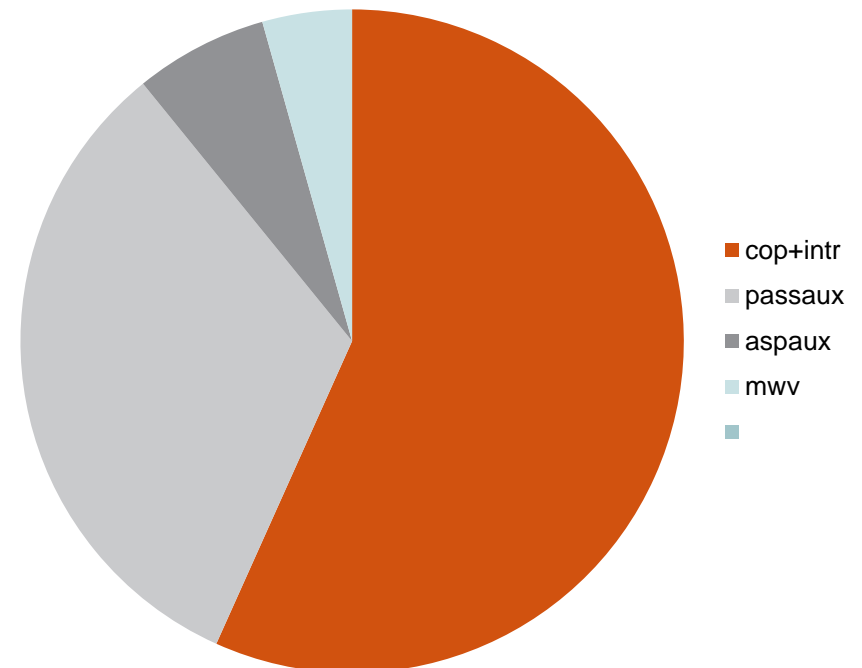
4c. Preliminary analyses: factor three

Status of SL item in its own network: *få* and *bli*

få - ENPC



bli - ENPC



Data from (Ebeling 2003)



4c. Preliminary analyses: factor three

Factor three:

Overrepresented senses in TT correspond to dominant/salient senses in frequent source lexemes



5. Concluding remarks

- For this case (*get*) in N-E translation:
 - Most frequent senses are overrepresented in translations.
 - These senses are also most frequently chosen in a sentence generation test.
 - There is a high concentration of sources (lexical items) for the most frequent senses (and some more infrequent senses).
 - For the dominant source lexemes in the above, the most frequent sense is the one that corresponds to the target sense here.



5. Concluding remarks

Preliminary interpretation: evidence of ‘semasiological salience’ (S and T) and network links playing into translational choices

Next step: multivariate analysis with semantic/syntactic behavioral profile and translational factor variables

Query: how to code these new variables? (factor two in particular, but also three)

Note: must test for interaction effects between these and other variables