



German Synthetic Compounds and the Architecture of the Grammar:

A Behavioral Analysis

Livio Gaeta
Dipartimento di Filologia Moderna "Salvatore Battaglia",
Università degli Studi di Napoli Federico II

Amir Zeldes
Korpuslinguistik und Morphologie,
Humboldt-Universität zu Berlin



Mediterranean Morphology Meeting 8, Cagliari, 14-17 September 2011

Synthetic Compounds and their Behavior

- **Synthetic compounds (SCs)**, German **Rektionskomposita** are compounds in which the modifier saturates an argument of the head (Roeper & Siegel 1978, Gaeta 2010), usually as a result of deverbal nominalization:

<i>X fährt ein Auto</i>	'X drives a car'
<i>Autofahrer</i>	'car driver'
? <i>Autofahrer eines Porsche</i>	'car driver of a Porsche'

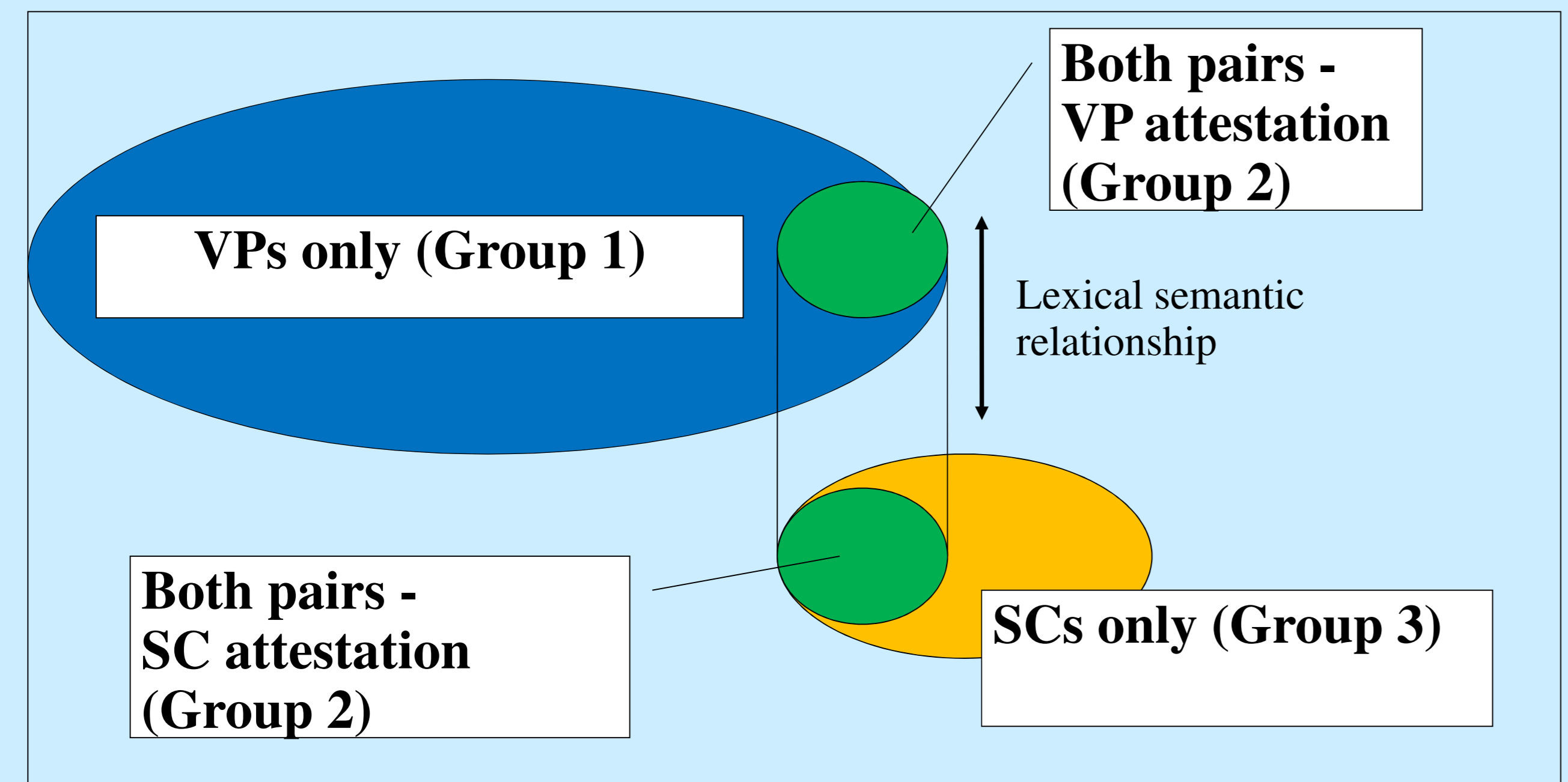
- Main questions:
 - Can SCs simply be derived from VPs (syntax below zero, see Spencer 2005) or are they an independent construction (Scalise & Guevara 2005)?
 - Can the selectional behavior of deverbal SCs in usage be predicted from that of corresponding VPs?
- Focus on German agent nominalizations in -er (see Meibauer et al. 2004)

$$[X_N \text{ fahr}_V\text{-er}_N]_{VP} \leftrightarrow [X_N [\text{fahr}_V\text{-er}_N]_N]_N$$

- Are the same objects attested? With similar frequency?
- Are there heads which prefer one pattern over the other?
- Does having many VP objects mean having many SCs?
- Is productivity as a VP head and as an SC head correlated?

Methodology

- Extract transitive VPs & SCs in -er from large corpus (deWaC, Baroni et al. 2009, ~1.7G tokens):
 - Use conservative patterns (verb final VPs with conjunction, subject, object compatible article not following a preposition)
 - Match verb as substring of compound
 - Correct for metathesis, Umlaut (*Träger*: *tragen*, *Sammler*: *sammeln*)
- Three groups of lexeme pairs are extracted:



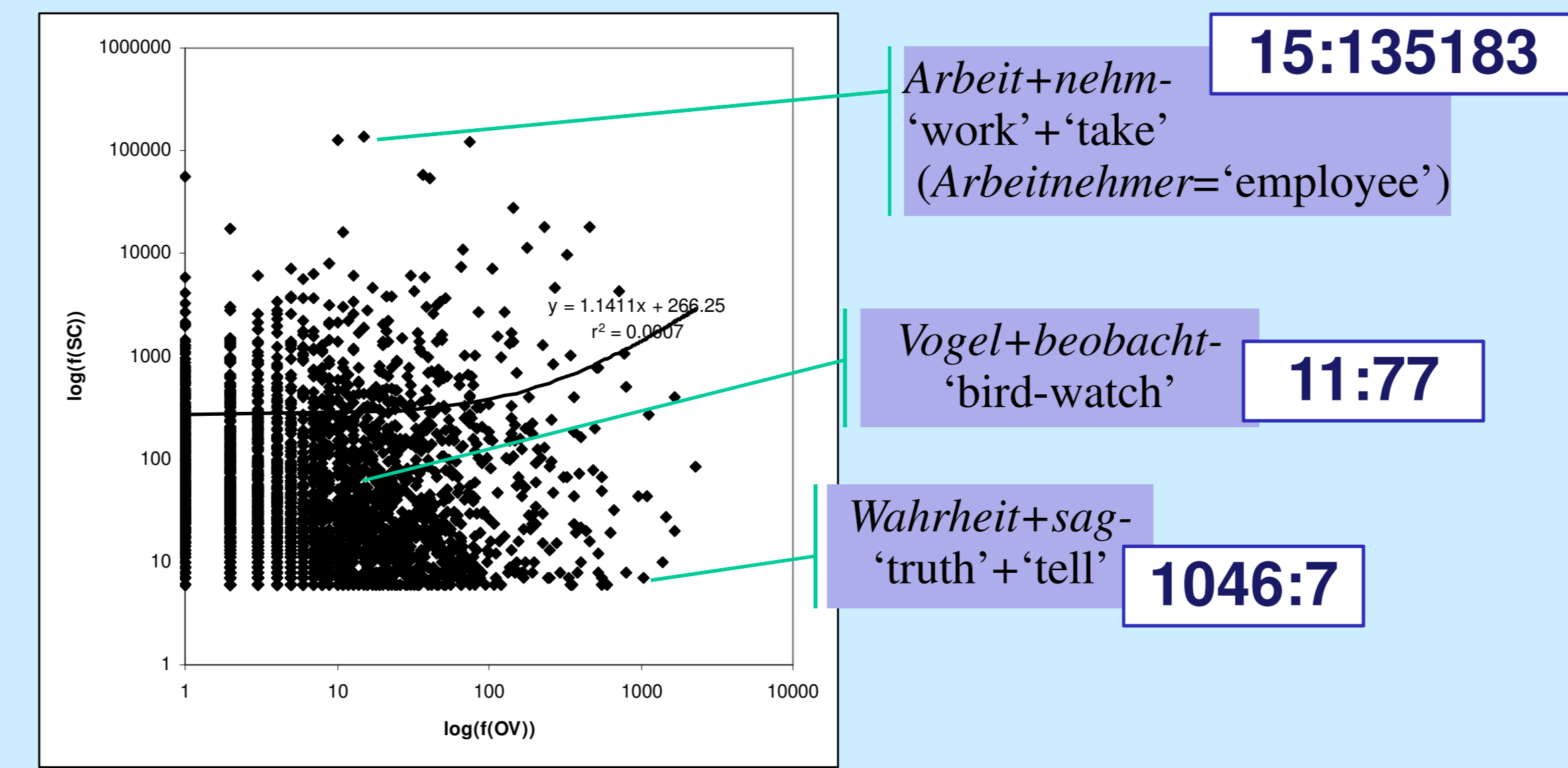
Group 1 – VP attestation only

- Many very frequent VPs have no corresponding SC
- Most cases can be divided into 3 groups:
 - Idiomatized phrase with preferred syntactic realization
 - Nominalization of head is avoided
 - Head nominalization has a different sense

VP	SC
<i>Gebrauch machen</i> make use 2134	? <i>Gebrauchmacher</i> use-maker 0
<i>Gedanken machen</i> give thought 1341	? <i>Gedankenmacher</i> thought-maker 0
<i>Kinder bekommen</i> get kids 806	? <i>Kinderbekommer</i> kid-getter 0
<i>Ziel erreichen</i> reach a goal 1544	? <i>Zielerreicher</i> goal-reacher 0
<i>Möglichkeit bieten</i> offer a possibility 592	? <i>Möglichkeitsbieter</i> possibility-bidder 0
<i>Rolle spielen</i> play a role 5088	<i>Rollenspieler</i> role-player 780

Group 2 – VP and SC attestation

- Here we find a gradient from syntactic to morphological preference (sorted by ratio SC/VP):
 - Highly lexicalized but transparent compounds
 - Balanced attestation, including collocated AND lexicalized cases
 - More compositional but highly collocated idioms
- However, there is **no significant correlation** between SC and VP attestation for each lexeme pair ($r^2=0.0007$, $p>0.05$)



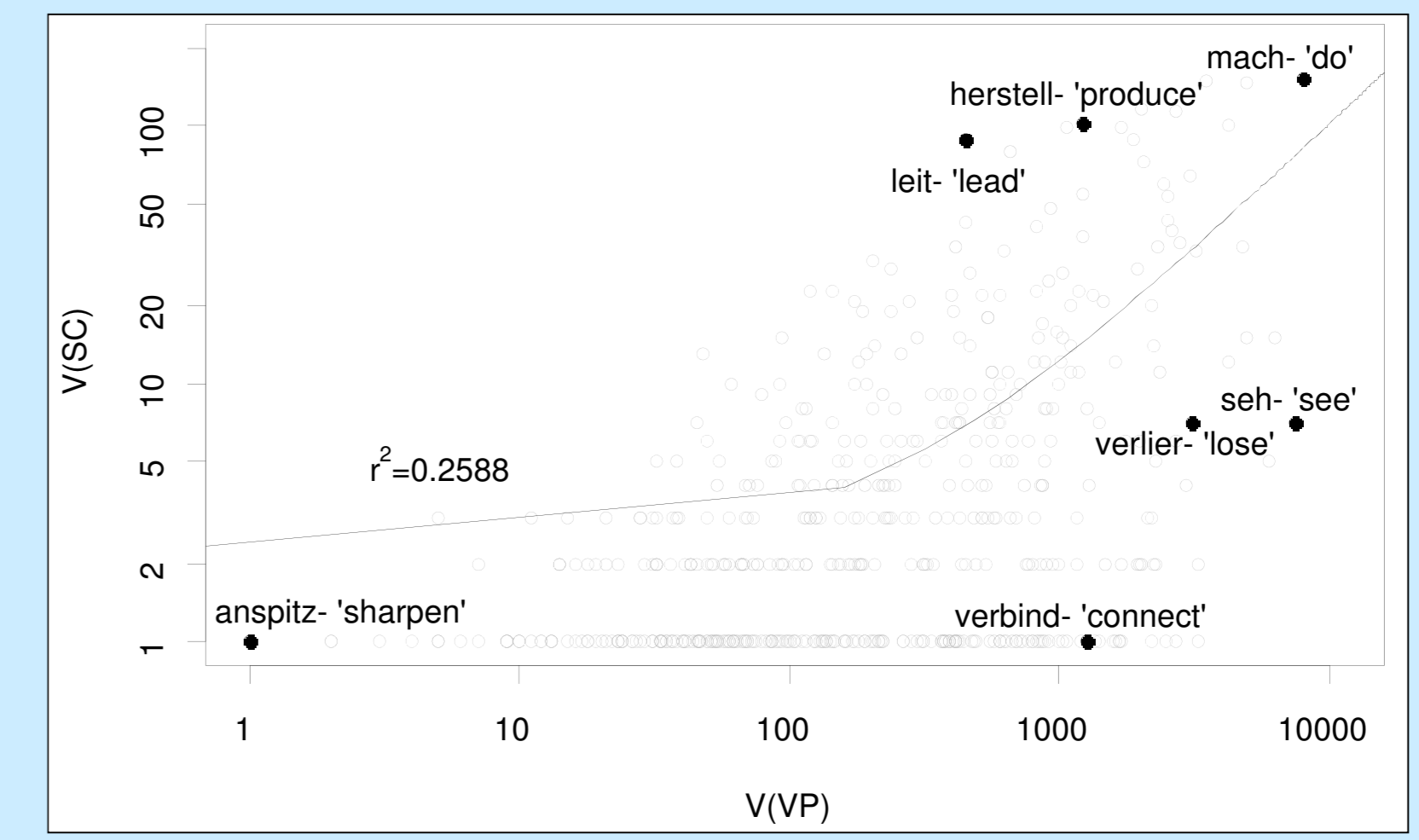
Group 3 – SC attestation only

- Lexicalizations (*Krankheitserreger* 'pathogen, lit. disease exciter')
- Suppletion (?*Unterrichter/Lehrer* 'teacher', ?*Haber/Besitzer* 'owner')
- Metonymy / ellipsis (*Erotikhersteller* 'erotics-manufacturer')
- Archaisms (*Staubsaugervertreter* 'vacuumcleaner sales rep')

SC	f(SC)	f(V)
<i>Versicherungsnehmer</i> insurance-taker, insuree	9355	958278
<i>Krankheitserreger</i> pathogen	5481	17018
<i>Musiklehrer</i> music teacher	1458	49788
<i>Arbeitsplatzbesitzer</i> work place owner	207	155563
<i>Reiseleiter</i> tour guide	2584	70686
<i>Pharmahersteller</i> pharma-producer	368	98433
<i>Staubsaugervertreter</i> vacuumcleaner sales rep	116	144465
<i>Automobilhersteller</i> automobile manufacturer	2923	98433

Productivity and Generation of Novel SCs

- The established lexeme types in Groups 1-3 may be **lexicalized**, and different lexicalizations for SC and VP may occur
- But if SCs are derived from VPs we **expect productive behavior to correlate** (non-lexicalized cases)
- Novel SCs **should be based on VPs**
 - We use Baayen's (2001) **morphological productivity** paradigm
 - Compare **type frequency (V)** and proportion of **hapax legomena (HL)**, forms with frequency=1 for each head lexeme



- **Correlation of type frequencies is fairly weak** – many stems are much more prolific in object selection either as SCs or as VP
- Similarly, many heads have **mainly VP-independent hapax SCs**:

SC head	hapax frequency	attested as VP	VP/SC
<i>Hersteller</i> manufacturer	1130	92	0.081416
<i>Leiter</i> head, leader, manager	1057	51	0.04825
<i>Besitzer</i> owner, possessor	802	178	0.221945
<i>Anbieter</i> provider, offerer	716	136	0.189944
<i>Vertreter</i> representative	664	71	0.106928
<i>Macher</i> maker, doer	629	240	0.381558
<i>Betreiber</i> operator	568	57	0.100352
<i>Lehrer</i> teacher	392	30	0.076531
<i>Sammler</i> collector	344	1	0.002907

Conclusion

- Lexical usage of SCs and VPs is **different and unpredictable**
- Constructional preferences, e.g. habitual/professional as SCs (*Leiter* 'leader', *Sammler* 'collector'), others as VPs (*sehen* 'see', *sagen* 'say')
- Often **little or no correlation** of vocabulary size, productive behavior
- **Frequent SC heads motivate novel SCs** in same pattern, not extant VPs with same lexemes (cf. Construction Morphology, Booij 2010)
- Well-behaved exceptions confirm **importance of lexical patterns**: lexicalizations, head blocking, metonymy and partial suppletion
- More work needed on exhaustive **classification** of all cases

References

Baayen, R. H. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.
 Baroni, M./Bernardini, S./Ferraresi, A./Zanchetta, E. 2009. The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *LRE* 43(3), 209-226.
 Booij, G. E. 2010. *Construction Morphology*. Oxford: Oxford University Press.
 Gaeta, L. 2010. Synthetic compounds. With special reference to German. In Scalise, S./Vogel, I. (eds.) *Cross-Disciplinary Issues in Compounding*. Amsterdam: Benjamins, 219-235.
 Roeper, T./Siegel, M. E. A. 1978. A lexical transformation for verbal compounds. *Linguistic Inquiry* 9(2), 199-260.
 Scalise, S./Guevara, E. 2005. The Lexicalist Approach to Word-formation and the Notion of the Lexicon. In Štekauer, P./Lieber, R. (eds.) *Handbook of Word-Formation*. Dordrecht: Springer, 147-187.
 Spencer, A. 2005. Word-formation and Syntax. In Štekauer, P./Lieber, R. (eds.), 73-97.