Intensifying Intensifiers: Variations in Expressivity

Roeland van Hout J. Nathanael Philipp Michael Richter

48. Annual Meeting of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS) 2026
University of Trier

25 February 2026

Expressivity and variation in intensification

- Intensifiers contribute little to nothing to the formal meaning of propositions, rather they are expressive in varying degrees.
- Intensifiers are an open, large, highly variable and innovative word class, especially in informal and social media discourses.

General focus: intensifiers in predicative sentences (following adjective)

Examples from our CMC corpora:

- *Das Video war mega toll*. ‘The video was mega awesome’.
- *Die ist sehr sehr gut gemacht* ... ‘It’s very, very well done ...’.
- *Versionen des T-shirts bitte: Tank top und V-Schnitt wäre super geil!* ‘Versions of the T-shirt please: tank top and V-neck would be awesome!’
- *Das Lied ist so geil – wieso könnt ihr so gut singen?* ‘This song is so awesome – how can you can sing so well?’

What are marked variants?

- Marked spelling variants of intensifiers are not random and not *noisy*. Rather they are devices to increase expressivity. Some examples, again from our CMC-data:
- **Lengthening:** Ich hasse lady gaga sie ist sooooooooooo hässlig.
  'I hate lady gaga she is sooooooooooo nasty'.
- **Repetition:** Ich habe dich sehr sehr sehr sehr sehr sehr sehr sehr lieb.
  'I love you very, very, very, very, very, very, very much.'.
- **Capitalisation:** Ich war schun mal forml 1 zu schauen und es ist MEGA geil !!
  'I've already watched Formula 1 and it's MEGA cool!!'.
- **Stacking:** ihr würdet euch direkt super krass gut verstehen...

Hypotheses: distributing information

- **Hypothesis 1:** Marked variants of intensifiers are less frequent and have a higher information value. A higher information value increases expressivity, also to announce a higher expressivity of the intensifier-adjective combination.

- **Hypothesis 2:** Marked variants intensify in the same way as less frequent intensifiers and resemble intensifiers with a high information value.

- **Information value:** Surprisal coefficients, based on expected occurrence.

Which intensifiers did we select to test our hypotheses?

We selected four intensifiers:

- Two older and frequent ones: *so*, *sehr*,
- Two more recent and less frequent *super*, *mega*,

The database

Corpus	Units	Tokens	Ø tokens/unit
YouTube 2014	71	181k	2553
YouTube 2015	78	249k	3198
YouTube 2016	86	232k	2694
YouTube 2017	93	255k	2741
YouTube 2018	104	203k	1957
Twitter 2022	8 563	4.20M	490
Twitter (other)	35 317	370k	10.5
Blogs	25 399	365k	14.4
total	**69 711**	**6.06M**	**86.9**

The Twitter (other) and Blogs come from the TwiBloCop [Scheffler and Seemann, 2024].

Two information-theoretic models

We test the hypotheses using two information-theoretic models and their coefficients:

- Model 1: Shannon information / IC / H (context-free, frequency-based)
- Model 2: TCM-surprisal (TCM) (context-dependent, topic-based)
- Expressivity is operationalised as the amount of information.

If marked intensifiers increase expressivity, they should carry higher IC and higher TCM values.

Two information theoretic models: more details

- Model 1: In philosophical and probability theory terms, Shannon information I is an expression of uncertainty. It is measured in bits. A bit is the set of yes/no-questions that can be used to reliably find an option in a probability space or decision tree. After tossing a coin, the result of which I do not know, I can ask, 'Is it "tails"?' If the answer is 'yes,' I know that it was indeed 'tails.' If the answer is 'no,' I know that the result was 'heads.' So I only need one question. The coin toss therefore has 1 bit of information. Shannon expressed this formally:

$$I(x) = -\log_2 P(x) \tag{1}$$

where x may be anything, and thus also a linguistic unit, like a letter or a word.

Two information theoretic models: more details

- Model 2:The concept of Surprisal was introduced by [Tribus, 1961] in an engineering context as a quantitative measure of unexpectedness,
- Surprisal is contextualised information (equivalent with Shannon information) and a common concept in psycholinguistics.

Two information theoretic models: more details

Model 2 [Hale, 2001] and [Levy, 2013] utilise surprisal S to model cognitive effort of sentence comprehension, operationalised by reading times:

$$S(w_i) = -\log_2 P(w_i \mid w_1, w_2, \ldots, w_{i-1}, C) \tag{2}$$

where C represents any context beyond the sentence.
When a linguistic message is highly informative and thus carries high surprisal, it requires high processing effort (*cognitive load*) on the part of the language processor.
Information distributed over sentences/utterances and their structures and over texts / discourse.

Two information-theoretic models: more details

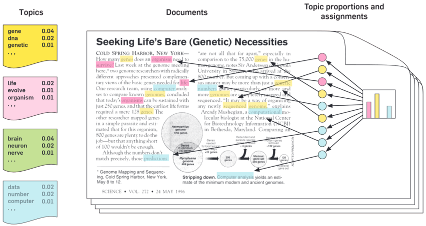
Model 2: TCM[1]

- TCM calculates surprisal for a word,
- TCM uses topics as context for surprisal,
- for topic detection, TCM employs the Latent Dirichlet Allocation technique (LDA),
- LDA models a topic as a probability-distribution of words (in a document), and a document on its turn is in LDA a probability-distribution of topics. The optimal distributions are achieved in an iterative process.

[1] `https://github.com/jnphilipp/tcm`

Two information theoretic models: more details

Model 2: The figure shows a document and its topics as probability distributions of words [Blei et al., 2003]:



- Given this word-topic-distribution: What could be the tendency of TCM-(surprisal) for a language processor if he/she had to process the word, say, 'Marzipan' later on?,
- since we assume that 'Marzipan' occurs anywhere in the document, it must have probabilities in the word-topic-distributions, but presumably very low ones, thus it is unexpectable.
- Will 'Marzipan' have a high or low surprisal value?

Two information theoretic models: more details

Model 2: Definition of TCM-surprisal

$$\overline{surprisal}(w_d) = -\phi \log_2 \frac{c_d(w_d)}{|d|} - \frac{1}{|T_d|} \sum_{i \in T_d} \log_2 \left(P(t_i|d) f(w_d, t_i) \right)$$

$$f(w_d, t_i) = \begin{cases} WT_{w_d, t_i} & \text{for } w \in WT \\ \overline{WT_{t_i}} & \text{for } w \notin WT \end{cases}$$

- $\phi \in \{0, 1\}$ controls the usage of the relative word-frequency
- $\frac{c_d(w_d)}{|d|}$ relative word w frequency $c_d(w_d)$ in a document d
- T_d topic dsitribution, indexed such that $P(t_i|d) \geq P(t_j|d)$ if and only if $i \leq j$, with a threshild $\mu \in (0, 1]$ and with k_μ be the largest integer such that $\sum_{i=1}^{k_\mu} P(t_i|d) \geq \mu$, for $k_\mu = 0,\ T_d = \{1\}$.
- $P(t_i|d)$ the probability of a topic t_i in a document d
- WT is the normalised word-topic matrix of the LDA

Four intensifiers, their variants and frequencies

variant	**intensifier**			
	super	**mega**	**so**	**sehr**
standard	191	320	3947	4985
capitalization	0	1	35	28
lengthening	2	18	589	19
stacking	4	11	13	37
repetition	0	0	0	125
standard	191	320	3947	4985
non-standard	6	30	637	209
standard	97%	92%	84%	93%
non-standard	3%	8%	16%	7%

Conclusions:

1. different frequencies of occurrence of intensifiers;
2. different frequencies of occurrence of non-standard variants;
3. different distributions of specific non-standard variants (see markings in red).

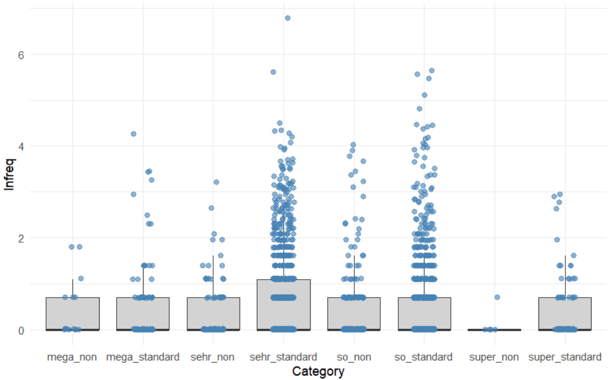
The five most frequent adjectives in combination with the four intensifiers

variant	**intensifier**				
	super	**mega**	**so**	**sehr**	total
gut	20	37	298	898	1253
schönn	0	21	339	284	644
geil	7	73	287	77	444
cool	16	32	172	97	317
toll	20	12	194	15	241
total	63	175	1290	1371	2879

Conclusions:

1. different preferential patterns (collocations), see frequencies in red;
2. strong collocation *sehr* and *gut*.

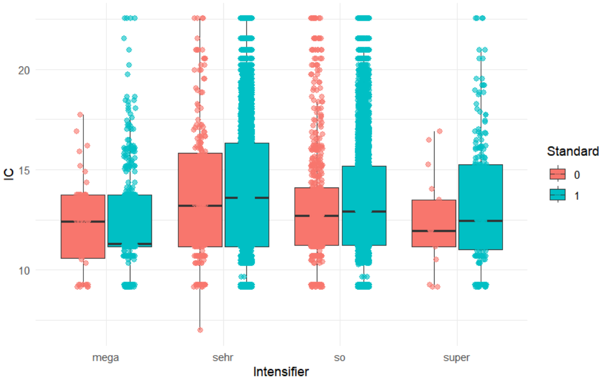
Boxplots of ln frequencies adjectives: four intensifiers by two variants



Observations:

1. higher frequencies standard intensifiers (of course); lower frequencies relative more frequent in non-standard;
2. distribution *sehr* standard seems different (more usual adjectives).

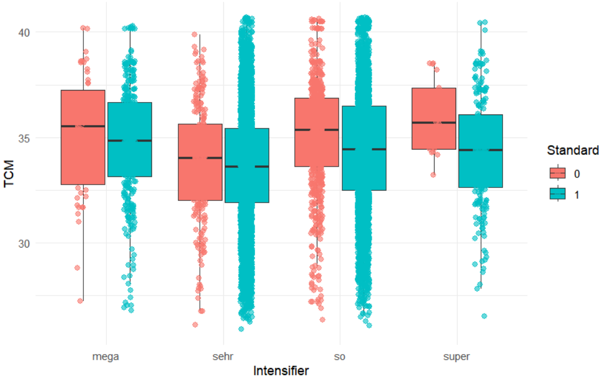
Boxplot IC following adjective



Observations:

1. no systematic difference standard versus non-standard;
2. *sehr* highest values.

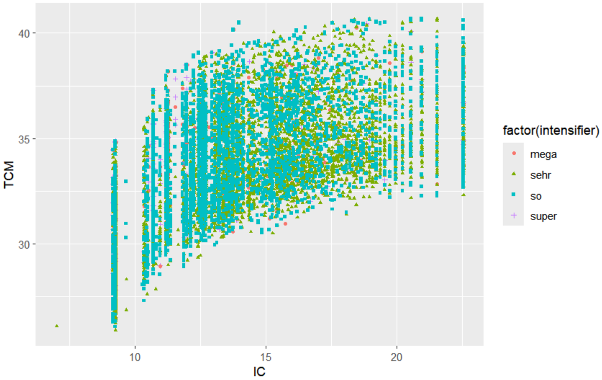
Boxplot TCM following adjective



Observations:

1. systematic difference standard versus non-standard;
2. adjectives as predictors of standard versus non-standard;
3. *sehr* lowest values.

Correlation IC and TCM following adjective



Observations:

1. correlation between IC and TCM;
2. *sehr* stands out: it has the lowest TCM-values;
3. the same IC has a range of TCM values.

ANOVA IC: statistical test

Tests of Between-Subjects Effects

Dependent Variable: IC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1055.555[a]	7	150.794	14.495	<,001	.009
Intercept	119176.812	1	119176.812	11456.158	<,001	.508
standard	37.485	1	37.485	3.603	.058	.000
intensifier	339.169	3	113.056	10.868	<,001	.003
standard * intensifier	14.793	3	4.931	.474	.700	.000
Error	115326.107	11086	10.403			
Total	2191607.522	11094				
Corrected Total	116381.663	11093				

a. R Squared = .009 (Adjusted R Squared = .008)

Observations:

1. Significant effect of the factor ‘intensifier’;
2. no significant effect of the factor ‘standard’;
3. no significant effect of the interaction between ‘standard’ and ‘intensifier’.

Anova TCM: statistical test

Tests of Between-Subjects Effects

Dependent Variable: TCM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2698.578[a]	7	385.511	50.869	<,001	.031
Intercept	832905.102	1	832905.102	109904.314	<,001	.908
standard	91.403	1	91.403	12.061	<,001	.001
intensifier	756.758	3	252.253	33.286	<,001	.009
standard * intensifier	48.616	3	16.205	2.138	.093	.001
Error	84014.773	11086	7.578			
Total	12967819.856	11094				
Corrected Total	86713.350	11093				

a. R Squared = .031 (Adjusted R Squared = .031)

Observations:

1. *Sehr* different; lower values than the other three intensifiers.

Anova TCM: posthoc analsysis

TCM

Tukey HSD[a,b,c]

intensifier	N	Subset 1	Subset 2
sehr	5629	33.618988586	
super	235		34.442993872
so	4837		34.529076397
mega	393		34.790312321
Sig.		1.000	.151

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 7.578.

a. Uses Harmonic Mean Sample Size = 556.774.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = ,05.

Concluding remarks: the hypotheses

- **Hypothesis 1:** Marked variants of intensifiers are less frequent and have a higher information value. A higher information value increases expressivity, also to announce a higher expressivity of the intensifier-adjective combination.
  YES. Overall effect for TCM. TCM is more sensitive / better than IC.
- **Hypothesis 2:** Marked/non-standard variants intensify in the same way as less frequent intensifiers and resemble intensifiers with a high information value.
  UNCLEAR. *So* has the same information values of the adjective as *super* and *mega*. On the other hand, the intensifier *so* has lower information values making the information gap larger. We need to investigate the information distribution in more detail.

References I

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003).
Latent dirichlet allocation.
Journal of machine Learning research, 3(Jan):993–1022.

Hale, J. (2001).
A probabilistic Earley parser as a psycholinguistic model.
In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics.

Levy, R. (2013).
Memory and surprisal in human sentence comprehension.
In *Sentence processing*, pages 78–114. Psychology Press.

Scheffler, T. and Seemann, H. J. (2024).
Das twiblocop. ein multimediales korpus aus blogposts und tweets .
Korpora Deutsch als Fremdsprache, 4:264–268.

Tribus, M. (1961).
Information theory as the basis for thermostatics and thermodynamics.
Appl. Mech. Mar 196, 28(1):1–8.