



Quantifying the Beauty of Words: A Neurocognitive Poetics Perspective

Arthur M. Jacobs^{1,2,3*}

¹ Department of Experimental and Neurocognitive Psychology, Freie Universität Berlin, Germany, ² Dahlem Institute for Neuroimaging of Emotion, Berlin, Germany, ³ Center for Cognitive Neuroscience Berlin, Berlin, Germany

In this paper I would like to pave the ground for future studies in Computational Stylistics and (Neuro-)Cognitive Poetics by describing procedures for predicting the subjective beauty of words. A set of eight tentative word features is computed via Quantitative Narrative Analysis (QNA) and a novel metric for quantifying word beauty, the *aesthetic potential* is proposed. Application of machine learning algorithms fed with this QNA data shows that a classifier of the decision tree family excellently learns to split words into beautiful vs. ugly ones. The results shed light on surface and semantic features theoretically relevant for affective-aesthetic processes in literary reading and generate quantitative predictions for neuroaesthetic studies of verbal materials.

Keywords: neurocognitive poetics, quantitative narrative analysis, machine learning, digital humanities, neuroaesthetics, computational stylistics, literary reading, decision trees

THE NEUROCOGNITIVE POETICS PERSPECTIVE

When a reader's brain processes information about single words like "LOVELY" or "SHRIEK," many neural circuits work together to enable meaning making. So far, practically all theoretical models have highlighted the neurocognitive processes underlying word recognition while neglecting the *affective-aesthetic* ones (for review: Hofmann and Jacobs, 2014; Jacobs et al., 2015). However, there is now abundant evidence that word recognition involves affective components from the first 100 ms of processing on (Kissler et al., 2007; Hofmann et al., 2009; for review see Citron, 2012). But there is practically no experimental research on aesthetic processes at the single word level (for exceptions, see Ponz et al., 2013; Jacobs et al., 2015). This is quite astonishing, given the success of neuroaesthetic research in other fields (e.g., Jacobsen et al., 2004; Jacobsen, 2006; Brattico et al., 2013; Leder, 2013; Nadal, 2013; Zeki et al., 2014; Marin, 2015) and work on the beauty of larger verbal materials, such as metaphors (McQuire et al., 2017), proverbs (Bohrn et al., 2013), idioms (Citron et al., 2016), or poems (Lüdtke et al., 2014; Hanauer, 2015).

The emerging field of Neurocognitive Poetics (Jacobs, 2015b,c; Willems and Jacobs, 2016) emphasizes such aesthetic processes during the reading of verbal materials in more natural and ecologically valid tasks and contexts and provides methods, e.g., QNA tools like the Berlin Affective Wordlist/BAWL (Võ et al., 2006, 2009), the DENN-BAWL (Briesemeister et al., 2011), EMOPHON (Aryani et al., 2013) or the Affective Norms for German Sentiment Terms/ANGST (Schmidtke et al., 2014a), as well as models for this field (e.g., the Neurocognitive Poetics Model/NCPPM; Jacobs, 2015a; Jacobs and Willems, 2017; Nicklas and Jacobs, 2017).

The methodological challenge for this perspective is immense given the complexity of the verbal materials and the focus on processes that recruit more than the usual language circuits in the brain (e.g., Keidel et al., 2013; Jacobs and Willems, 2017). However, recent developments in QNA methods and machine learning, as well as in fMRI data analyses promise rapid progress in this regard. Thus, applications of QNA-based machine learning tools have allowed successful prediction

OPEN ACCESS

Edited by:

Xiaolin Zhou,
Peking University, China

Reviewed by:

Thomas Jacobsen,
Helmut Schmidt University, Germany
Mireille Besson,
Institut de Neurosciences Cognitives
de la Méditerranée, France

***Correspondence:**

Arthur M. Jacobs
ajacobs@zedat.fu-berlin.de

Received: 23 October 2017

Accepted: 07 December 2017

Published: 19 December 2017

Citation:

Jacobs AM (2017) Quantifying the
Beauty of Words: A Neurocognitive
Poetics Perspective.
Front. Hum. Neurosci. 11:622.
doi: 10.3389/fnhum.2017.00622

of the liking of single words (Jacobs et al., 2016), classification of Shakespeare's 154 sonnets into motif categories (Jacobs et al., 2017), as well as predicting authorship, literariness and aptness of poetic metaphors (Jacobs and Kinder, 2017, in press), or subjective immersion into narratives (Jacobs and Lüdtke, 2017).

In this paper I show an application of such tools to predict the beauty/ugliness of single words from the Neurocognitive Poetics perspective in an attempt to motivate and generate more neuroscientific research on this issue.

MICROPOETRY: THE BEAUTY OF WORDS AND THE ORIGINS OF LUDIC READING

Beauty is an important human category, listed among the top features for almost all domains of aesthetic appreciation (Jacobsen and Beudt, 2017), the most prototypical aesthetic judgment (Jacobsen et al., 2004), and the most frequently used term for literature and poetry (Knoop et al., 2016). Readers often report the self-rewarding experience of beauty and harmony not only for entire poems (Jacobs, 2015b), but even for single words. This is documented in reports from the annual election of the most beautiful German word (Limbach, 2004). These examples show that words can be positive or negative, beautiful or ugly, and support the notion of *one-word poetry*, i.e., that single utterances or words—even outside lyrical contexts—can fulfill Jakobson's *poetic function* (Jakobson, 1960; Jacobs, 2015c; Jacobs and Kinder, 2015). However, there seems to be a single study so far that provides rating data on the beauty of single words, in German (Jacobs et al., 2015), while neuroimaging studies on that issue still are missing.

Understanding the neurocognitive bases of subjective feelings of the beauty of words and of micropoetic episodes is important for the investigation of more general and complex questions such as how language and emotion co-develop (Sylvester et al., 2016), how human beings come to like fiction (Jacobs and Willems, 2017), or how they acquire a taste for ludic reading and something like a lyrical sense (Jacobs and Kinder, 2015). Cognitive neuroscience so far has not even begun to shed light on the neural bases of the development of literary experiences (Jacobs, 2015c), although studies investigating the neural underpinnings of written language processing in children and adolescents are informative for the present purposes (e.g., Liebig et al., 2017).

PREDICTING THE BEAUTY OF WORDS

In the behavioral study reported by Jacobs et al. (2015) standard linear (stepwise) regression analyses suggested that word beauty was best predicted by valence and familiarity ratings ($R^2_{\text{lin}} = 0.77$; AICc = 608), while the other two considered features, arousal and imageability, did not account for a significant part of variance in the beauty ratings for that sample. Note that these predictors were themselves based on ratings and thus on “subjective” measures. The most beautiful word was LIBELLE (dragonfly) with a mean rating of 6.1/7, followed by MORGENRÖTE (aurora, 5.9), and MITTSOMMERNACHT

(midsummernight, 5.8). An additional hierarchical cluster analysis suggested that the most beautiful words described nine phenomena from nature (animals, flowers, rainbow etc.) and four states/objects of wellness (e.g., coziness), all rated high on beauty, valence, and imageability, and low on arousal. In contrast, the overall 24 ugliest words were almost all swear words associated with genitalia (see Jacobs et al., 2015, Supplementary Materials).

The multilevel hypothesis derived from the NCPM predicts that the liking of words, idioms, proverbs, sentences or entire poems is affected by nonlinear dynamic interactions of multiple features (or predictors) at multiple text levels, for example sublexical phonological features like phoneme salience with supralexical features like the global affective meaning (cf. Aryani et al., 2016). Powerful *decision tree* classifiers, e.g., extremely random trees/ERT which are the most accurate and efficient ones (Geurts et al., 2006) provide information about the importance of predictors from a large set (e.g., about 100; Jacobs et al., 2017), whether they are factorial or continuous, and even when there are more predictors than observations. They also work for unbalanced designs with high multicollinearity for which linear models are less appropriate (cf. Strobl et al., 2009; Tagliamonte and Baayen, 2012).

In contrast to Jacobs et al. (2015), here I exclusively used a set of QNA features that can directly be extracted from text corpora and the target words themselves by help of computer programs¹, i.e., no subjective rating data for quantifying lexical features were used as predictors. The machine learning programs (classifiers) were based on *scikit-learn* scripts (Pedregosa et al., 2011). The general procedure was similar to previous research in which we successfully classified verbal materials into motif or author categories or predicted response variables such as word liking and metaphor goodness ratings (Jacobs et al., 2016, 2017; Jacobs and Kinder, 2017, in press).

DATABASES AND FEATURES

The *sdeWaC* corpus (>40 million sentences, ~1 billion word tokens and 6 million types; Baroni et al., 2009) was used for computing reliable lexical indices (e.g., word frequency or orthographic neighborhood density/N), as well as other variables known to influence word recognition (e.g., Jacobs and Grainger, 1994) because its hit rate (overlap between words in database and 300 target words) was high: 74% (211/300). A complication was added by the relatively low hit rate of the German wordnet database (*GermaNet/GN*; Henrich et al., 2012)—crucial for computing word similarity based on semantic relatedness: 43%. Thus, overall 130 target words remained for final analysis [75 beautiful and 55 ugly ones; see S1 in Appendix (Supplementary Material)].

Anecdotal evidence (Limbach, 2004) and results from previous research (Jacobs et al., 2016) suggest that the liking and subjective beauty of even such simple verbal materials as single words can depend on quantifiable features in about all of the 16 cells of the 4 × 4 QNA matrix proposed in Jacobs (2015b). Thus, in the above mentioned book on the most beautiful German

¹Python scripts developed by the author.

words, a 9-year old boy explains why the German word LIBELLE (dragonfly) is the most beautiful for him by referring to features at the sublexical *phonological* level (e.g., the three “Ls” which make the word glide so well on the tongue), or the lexical, *affective-semantic* level (e.g., he loves seeing dragonflies wobble and finds that the word expresses this feeling, that it ensures that one is not afraid of these insects).

Given that more than 50 word features could already be quantified a decade ago for monosyllabic 4–6 letter words (e.g., Graf et al., 2005), a central issue in this field is to investigate which of the myriad of features are distinctive or potentially relevant in the aesthetic appreciation of poetry (cf. Knoop et al., 2016). Exploratory predictive modeling can help identify such features from a large candidate set (Jacobs et al., 2016, 2017; Jacobs and Kinder, 2017, in press). The present approach is basically an exploratory one using a limited set of features that can easily be computed from *sdeWaC* and *GN* or similar corpora for any given target at hand without recurring to standard rating-based word lists like the BAWL (whose hit rate for the present targets was far too low to be useful). Given these constraints and based on the results of pilot studies looking at potentially relevant predictors of the beauty ratings from Jacobs et al. (2015), I selected the following eight tentative features, also in an attempt to keep things as simple as possible and to facilitate follow-up studies, especially of the experimental kind (complementing the present computational one)². The two sublexical (i.e., syllable-based) features were *number of syllables* and *sonority score* (cf. Jacobs and Kinder, in press; see Appendix A in Supplementary Material). The six lexical features were *word length* (number of letters), *surprisal* ($-\log_2$ of *sdeWaC*-based word frequency), *orthographic neighborhood density* (N), *word similarity* (i.e., GN-based semantic relatedness between all 130 target words), *valence* (parametric positivity/negativity value), and *aesthetic potential* (AP; see Appendix B in Supplementary Material).

CLASSIFIER STUDY AND RESULTS

Each target was transformed into a vector based on the eight features, then used as input for machine learning tools classifying each word into one of two categories. Based on successful previous applications (cf. Jacobs and Kinder, 2017, in press), I used the ERT classifier to predict binary categorical ratings (i.e., beautiful vs. ugly; see Appendix C in Supplementary Material). As shown in **Figure 1A**, the performance of the classifier when training and test set were identical, as assessed by a confusion matrix, is flawless. When using the stratified k-fold cross validation method for evaluating the classifier’s predictive performance (prediction of test data on basis of training data), the classifier’s performance was excellent with parameter set 1 and perfect with parameter set 2 (Area Under Curve/AUC

= 0.94 and 1.0, respectively; see **Figures 1B,C** and Appendix C in Supplementary Material). As an additional check against overfitting, I applied a second model evaluation technique. Using a permutation test I checked that the classifier’s performance was around chance level (AUC = 0.5) when the labels “beautiful vs. ugly” were randomly attributed to the 130 target words (see **Figure 1D**).

The ERT classifier allows an estimation of the feature importances (which can be interpreted as a descriptive ranking of the predictor variables, Strobl et al., 2009). This ranking suggests that one out of the eight features was of minor importance for the classifier’s performance (importance <0.1: N), while *word length* (in letters and number of syllables) and *AP* (all >0.15) appear to be vital predictors, followed by *sonority score* and *surprisal* (0.12), as well as *word similarity* and *valence* (0.11)³.

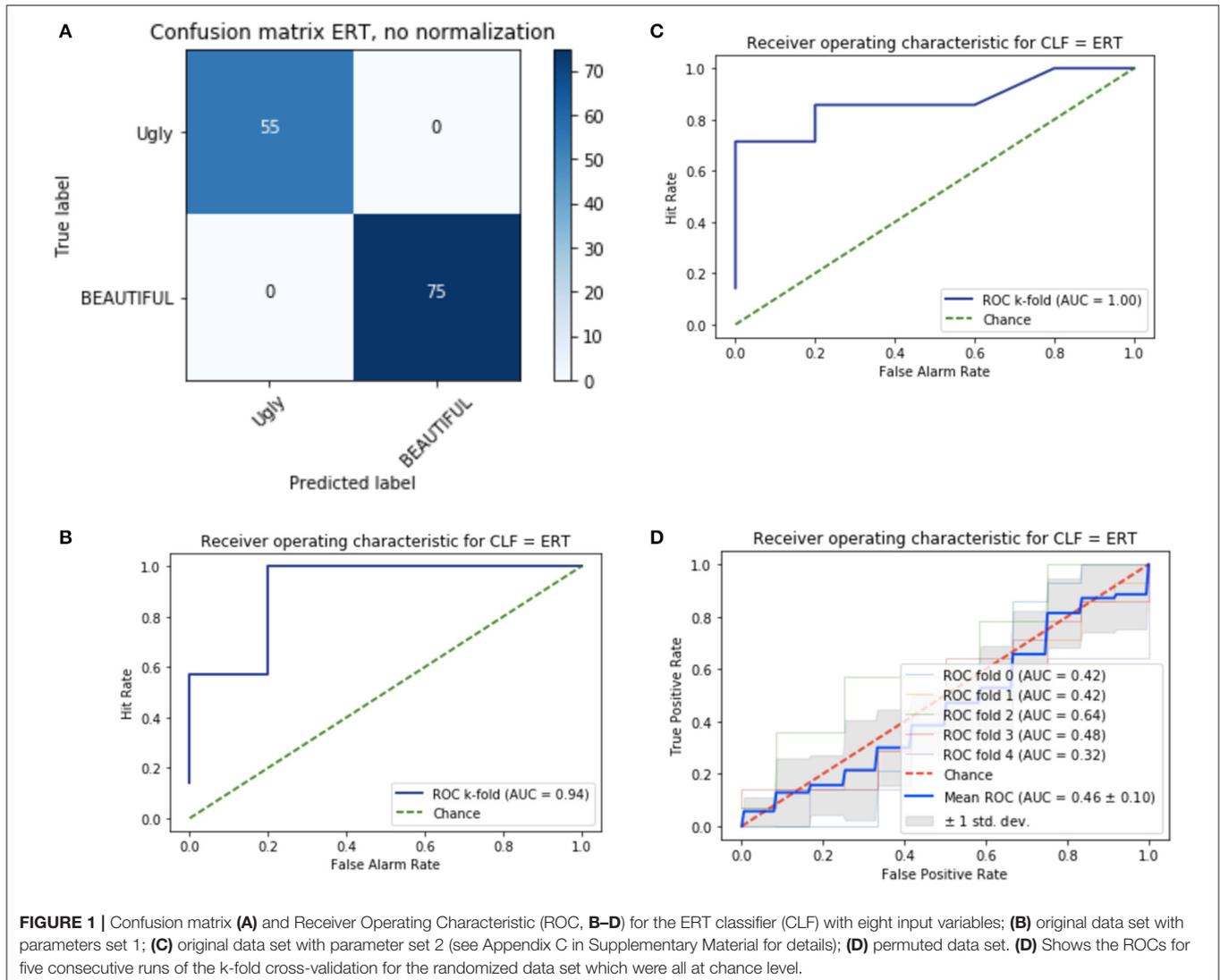
DISCUSSION

The results show that a potent classifier fed with eight input features can excellently predict whether a German word from the present database is judged as beautiful or ugly, generalizing perfectly from a training to a test data set. Two predictors seem crucial for the classification at hand: a surface feature (word length) and a semantic one (AP). The AP is a novel feature introduced in this paper specifically for assessing the aesthetic potential of words. In a one-way ANOVA, AP was significantly higher for beautiful than for ugly words [*z*-values: 0.25 vs. -0.33 ; $F_{(1, 128)} = 12.06$, $p < 0.0007$, $R^2_{adj.} = 0.08$], although the effect of this feature alone is very small. Still, its success as an important predictor of word beauty –in concert with seven others– is first validating evidence for the proposed list of 124 labels and should motivate future use in studies on reading literature. The number of syllables as crucial predictor is notable, since –much like number of letters– its mean value did not differ significantly between the two word groups and it was not strongly correlated with number of letters ($R^2 < 0.57$). Still, nonlinear, nonparametric supervised learning methods like *decision trees* can produce results largely differing from linear analyses due to their power of detecting hidden structure in complex data sets, e.g., by recursively scanning and (re-)combining variables (LeCun et al., 2015), and of dealing with complex interactions that are difficult to model in a mixed-effects logistic framework (Tagliamonte and Baayen, 2012).

The sonority score is a sublexical feature estimating the phonological aesthetic potential of words and phrases (Jacobs and Kinder, in press). Poetic language expertly plays with the sound-meaning nexus (Schrott and Jacobs, 2011; Aryani et al., 2013, 2016; Schmidtke et al., 2014b; Jacobs, 2015b,c; Jacobs et al., 2015; Ullrich et al., 2017) and thus it would not be surprising that words judged to be more beautiful show higher sonority

²I am aware of potential shortcomings of such a (non-exhaustive) QNA-based feature set and have discussed them extensively elsewhere (see limitations and outlook sections in Jacobs and Kinder, 2017; Jacobs et al., 2017). Still, I know of no other study that proposes an alternative set and would like to motivate just this: further research attempting to identify distinctive features driving the beauty of words and figurative language in general (cf. Jacobs, 2015b,c).

³Note that this ranking can vary with repeated runs and alternative parameter sets of the classifier due to its probabilistic nature. The values reported here are mean values averaged across 10 repetitive runs with parameter set 2 which yielded optimal performance (see Appendix C in Supplementary Material). Of course, much as for multiple linear regressions, adding or deleting features to the list may alter this ranking.



scores –just as the anecdotal evidence reported above suggests. This was indeed the case [beautiful: 3.12 vs. ugly: 2.9; $F_{(1, 128)} = 4.6$, $p < 0.033$, $R^2_{\text{adj.}} = 0.03$]. Through a process of phonological recoding in silent reading (Ziegler and Jacobs, 1995; Braun et al., 2009) which may play a key role especially in reading poetic texts (Kraxenberger, 2017), the implicit sonority of a written word could more or less unconsciously influence its beauty ratings, a speculation to be tested in future studies.

Surprisal has successfully predicted eye movement or brain wave parameters and correlates positively with reading time (Frank, 2013). Here it also predicted beauty ratings. Regarding word similarity the issue behind the GN-based measure was whether beautiful and ugly words differ in their within-group semantic relatedness. Although the difference was not significant in a linear regression ($p = 0.83$), the classifier makes use of this feature in concert with the other seven as it does with valence. The fact that descriptively valence was not as important as AP may in part be due its computation being based on altogether

36 labels (instead of 124 for AP). Moreover, based on fMRI and EEG results by Briesemeister et al. (2014, 2015) and Kuhlmann et al. (2016), respectively, we proposed that valence itself is a super-feature likely to be derived from core affects like joy or disgust, which is indirectly supported by the present results for the AP feature (Jacobs et al., 2016). Since valence and AP are not correlated ($R^2 = 0.005$), it could be used in its own right in future studies interested in affective lexical semantics (e.g., Sylvester et al., 2016) rather than aesthetics.

In sum, while none of the eight features on its own accounts for much variance in the data, when processed by the ERT classifier, they seem to fit almost perfectly together in predicting word beauty and perhaps reflect what Kintsch (2012) called *harmony* –how well parts fit the whole. Thus, if a German word features an optimal length (in this corpus: about 12 letters), a specific combination of sonorous syllables, semantic associations with words like ANMUT (grace) or FREUDE (joy) and is rather surprising, it has an increased likelihood of being classified as

beautiful. If these features fit together well and, additionally, also with the object that the word denotes, e.g., an aurora, then the word is likely beautiful. While other aspects not considered in the present analyses may also play a role (e.g., arousal, imageability), the present computational eight-feature model of word beauty can serve as a “null-model” against which to test more sophisticated future process models.

SOME PREDICTIONS FOR NEUROCOGNITIVE POETICS

Each of the eight features can, in principle, be used as a parametric regressor in fMRI studies on literary materials, e.g., to investigate whether similar neural networks in the ventral striatum and medial prefrontal cortex that were associated with beauty ratings of German proverbs (Bohrn et al., 2013) also are responsive to at least some of the present eight features, in particular the AP. It would also be interesting to run an fMRI decoding study (e.g., Haynes, 2015) in which the present ERT classifier is used to predict whether a word was beautiful or ugly on the basis of the participants’ brain activity patterns and where the present feature importances could be compared with estimates of neuronal variable importance (e.g., Oh et al.,

2003). As concerns the more general issues⁴ (i) to what extent beauty ratings reflect the beauty of the words and/or that of their referents, and (ii) whether similar results can be obtained in other languages (e.g., French or Chinese), future cross-cultural neuroimaging studies could address a question raised previously (Jacobs et al., 2016): to what extent an AP value is computed in the brain from (1) neural activation patterns distributed over the sensory-motor representations of a word’s referents (*experiential* aspect) and (2) the size and density of their context (*distributional* aspect), as computationally modeled using co-occurrence statistics, for example (Hofmann and Jacobs, 2014).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2017.00622/full#supplementary-material>

⁴I thank reviewer 2 for this suggestion.

REFERENCES

- Aryani, A., Jacobs, A. M., and Conrad, M. (2013). Extracting salient sublexical units from written texts: “Emophon,” a corpus-based approach to phonological iconicity. *Front. Psychol.* 4:654. doi: 10.3389/fpsyg.2013.00654
- Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., and Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychol. Aesthetics Creativity Arts* 10, 191–204. doi: 10.1037/aca0000033
- Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky Wide Web: a collection of very large linguistically processed web-crawled corpora. *Lang. Resour. Eval.* 43, 209–226. doi: 10.1007/s10579-009-9081-4
- Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2013). When we like what we know—a parametric fMRI analysis of beauty and familiarity. *Brain Lang.* 124, 1–8. doi: 10.1016/j.bandl.2012.10.003
- Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* 4:206. doi: 10.3389/fpsyg.2013.00206
- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., and Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Hum. Brain Mapp.* 30, 1977–1989. doi: 10.1002/hbm.20643
- Briesemeister, B. B., Kuchinke, L., and Jacobs, A. M. (2011). Discrete emotion norms for nouns - Berlin Affective Word List (DENN-BAWL). *Behav. Res. Methods* 43, 441–448. doi: 10.3758/s13428-011-0059-y
- Briesemeister, B. B., Kuchinke, L., and Jacobs, A. M. (2014). Emotion word recognition: discrete information effects first, continuous later? *Brain Res.* 1564, 62–71. doi: 10.1016/j.brainres.2014.03.045
- Briesemeister, B. B., Kuchinke, L., Jacobs, A. M., and Braun, M. (2015). Emotions in reading: dissociation of happiness and positivity. *Cogn. Affect. Behav. Neurosci.* 15, 287–298. doi: 10.3758/s13415-014-0327-2
- Budanitsky, A., and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Comput. Linguist.* 32, 13–47. doi: 10.1162/coli.2006.32.1.13
- Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* 122, 211–226. doi: 10.1016/j.bandl.2011.12.007
- Citron, F. M. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., and Jacobs, A. M. (2016). When emotions are expressed figuratively: psycholinguistic and affective norms of 619 idioms for German (PANIG). *Behav. Res.* 48, 91–111. doi: 10.3758/s13428-015-0581-4
- Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. *Top. Cogn. Sci.* 5, 475–494. doi: 10.1111/tops.12025
- Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1
- Graf, R., Nagler, M., and Jacobs, A. M. (2005). Factor analysis of 57 variables in visual word recognition. *Z. Psychol.* 213, 205–218. doi: 10.1026/0044-3409.213.4.205
- Hanauer, D. I. (2015). Beauty judgments of non-professional poetry: regression analyses of authorial attribution, emotional response and perceived writing quality. *Sci. Study Lit.* 5, 183–199. doi: 10.1075/ssol.5.2.04han
- Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* 87, 257–270. doi: 10.1016/j.neuron.2015.05.025
- Henrich, V., Hinrichs, E., and Suttner, K. (2012). Automatically linking germanet to wikipedia for harvesting corpus examples for germanet senses. *J. Lang. Technol. Comput. Linguist.* 27, 1–19.
- Hofmann, M. J., and Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: from behavioral to brain data. *Neurosci. Biobehav. Rev.* 46, 85–104. doi: 10.1016/j.neubiorev.2014.06.011
- Hofmann, M. J., Kuchinke, L., Tamm, S., Vö, M. L. H., and Jacobs, A. M. (2009). Affective processing within 1/10th of a second: high arousal is necessary for early facilitative processing of negative but not positive words. *Cogn. Affect. Behav. Neurosci.* 9, 389–397. doi: 10.3758/9.4.389
- Jacobs, A. (2015c). The scientific study of literary experience: sampling the state of the art. *Sci. Study Lit.* 5, 139–170. doi: 10.1075/ssol.5.2.01jac

- Jacobs, A. M. (2015a). "Towards a neurocognitive poetics model of literary reading," in *Towards a Cognitive Neuroscience of Natural Language Use*, ed R. Willems (Cambridge, MA: Cambridge University Press), 135–159.
- Jacobs, A. M. (2015b). Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Front. Hum. Neurosci.* 9:186. doi: 10.3389/fnhum.2015.00186
- Jacobs, A. M., and Grainger, J. (1994). Models of visual word recognition: sampling the state of the art. *J. Exp. Psychol. Hum.* 20, 1311–1334. doi: 10.1037/0096-1523.20.6.1311
- Jacobs, A. M., Hofmann, M. J., and Kinder, A. (2016). On elementary affective decisions: to like or not to like, that is the question. *Front. Psychol.* 7:1836. doi: 10.3389/fpsyg.2016.01836
- Jacobs, A. M., and Kinder, A. (2015). "Worte als Worte erfahren: wie erarbeitet das Gehirn Gedichte (Experience words as words: how the brain constructs poems)," in *Kind und Gedicht (Child and Poem)*, ed A. Pompe (Berlin: Rombach), 57–76.
- Jacobs, A. M., and Kinder, A. (2017). The brain is the prisoner of thought: a machine-learning assisted quantitative narrative analysis of literary metaphors for use in Neurocognitive Poetics. *Metaphor Symb.* 32, 139–160. doi: 10.1080/10926488.2017.1338015
- Jacobs, A. M., and Kinder, A. (in press). What makes a metaphor literary? Answers from two computational studies. *Metaphor Symbol.*
- Jacobs, A. M., and Lüdtke, J. (2017). "Immersion into narrative and poetic worlds: a neurocognitive poetics perspective," in *Narrative Absorption*, eds F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, and M. M. Doicaru (Amsterdam: John Benjamins), 69–96.
- Jacobs, A. M., Schuster, S., Xue, S., and Lüdtke, J. (2017). What's in the brain that ink may character: a quantitative narrative analysis of Shakespeare's 154 sonnets for use in neurocognitive poetics. *Sci. Study Literat.* 7, 4–51. doi: 10.1075/ssol.7.1.02jac
- Jacobs, A. M., Vö, M. L.-H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., et al. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes? *Front. Psychol.* 6:714. doi: 10.3389/fpsyg.2015.00714
- Jacobs, A. M., and Willems, R. M. (2017). The fictive brain: neurocognitive correlates of engagement in literature. *Rev. Gen. Psychol.* doi: 10.1037/gpr000106. [Epub ahead of print].
- Jacobsen, T. (2006). Bridging the arts and sciences: a framework for the psychology of aesthetics. *Leonardo* 39, 155–162. doi: 10.1162/leon.2006.39.2.155
- Jacobsen, T., and Beudt, S. (2017). Domain generality and domain specificity in aesthetic appreciation. *New Ideas Psychol.* 47, 97–102. doi: 10.1016/j.newideapsych.2017.03.008
- Jacobsen, T., Buchta, K., Köhler, M., and Schröger, E. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychol. Rep.* 94, 1253–1260. doi: 10.2466/pr0.94.3c.1253-1260
- Jakobson, R. (1960). "Closing statement: linguistics and poetics," in *Style in Language*, ed T. A. Sebeok (Cambridge, MA: MIT Press), 350–377.
- Keidel, J. L., Davis, P. M., Gonzalez-Diaz, V., Martin, C. D., and Thierry, G. (2013). How Shakespeare tempests the brain: neuroimaging insights. *Cortex* 49:913e919. doi: 10.1016/j.cortex.2012.03.011
- Kintsch, W. (2012). Musing about beauty. *Cogn. Sci.* 36, 635–654. doi: 10.1111/j.1551-6709.2011.01229.x
- Kissler, J., Herbert, C., Peyk, P., and Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychol. Sci.* 18, 475–480. doi: 10.1111/j.1467-9280.2007.01924.x
- Knoop, C. A., Wagner, V., Jacobsen, T., and Menninghaus, W. (2016). Mapping the aesthetic space of literature "from below." *Poetics* 56, 35–49. doi: 10.1016/j.poetic.2016.02.001
- Kraxenberger, M. (2017). *On Sound-Emotion Associations in Poetry*. Ph.D. Thesis, Freie Univ. Berlin.
- Kuhlmann, M., Hofmann, M. J., Briesemeister, B. B., and Jacobs, A. M. (2016). Mixing positive and negative valence: affective-semantic integration of bivalent words. *Sci. Rep.* 6:30718. doi: 10.1038/srep30718
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Leder, H. (2013). Next steps in neuroaesthetics: which processes and processing stages to study? *Psychol. Aesthetics Creativity Arts* 7, 27–37. doi: 10.1037/a0031585
- Liebig, J., Froehlich, E., Morawetz, C., Braun, M., Jacobs, A. M., Heekeren, H. R., et al. (2017). Neurofunctionally dissecting the developing reading system. *Dev. Cog. Neurosci.* 27, 45–57. doi: 10.1016/j.dcn.2017.07.002
- Limbach, J. (2004). *Das Schönste Deutsche Wort (The Most Beautiful German Word)*. Freiburg: Verlag Herder.
- Lin, D. (1998). "An information-theoretic definition of similarity," in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML'98)* (Madison, WI), 296–304.
- Lüdtke, J., Meyer-Sickendiek, B., and Jacobs, A. M. (2014). Immersing in the stillness of an early morning: testing the mood empathy hypothesis in poems. *Psychol. Aesthetics Creativity Arts* 8, 363–377. doi: 10.1037/a0036826
- Marin, M. M. (2015). Crossing boundaries: toward a general model of neuroaesthetics. *Front. Hum. Neurosci.* 9:443. doi: 10.3389/fnhum.2015.00443
- McQuire, M., McCollum, L., and Chatterjee, A. (2017). Aptness and beauty in metaphor. *Lang. Cogn.* 9, 316–331. doi: 10.1017/langcog.2016.13
- Nadal, M. (2013). "The experience of art: insights from neuroimaging," in *Progress in Brain Research*, vol. 204, eds S. Finger, D. W. Zaidel, F. Boller, and J. Bogousslavsky (Amsterdam: Elsevier), 135–158.
- Nicklas, P., and Jacobs, A. M. (2017). Rhetorics, neurocognitive poetics and the aesthetics of adaptation. *Poetics Today* 38, 393–412. doi: 10.1215/03335372-3869311
- Oh, J., Laubach, M., and Luczak, A. (2003). "Estimating neuronal variable importance with random forest," in *Proceedings of the 29th Annual IEEE Bioengineering Conference, New Jersey Institute of Tech- Nology*, eds S. Reisman, R. Foulds, and B. Mantilla (Newark, NJ: IEEE Press), 33–34.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Ponz, A., Montant, M., Liegeois-Chauvel, C., Silva, C., Braun, M., Jacobs, A. M., et al. (2013). Emotion processing in words: a test of the neural re-use hypothesis using surface and intracranial EEG. *Soc. Cogn. Affect. Neurosci.* 9, 619–627. doi: 10.1093/scan/nst034
- Schmidtke, D. S., Conrad, M., and Jacobs, A. M. (2014b). Phonological iconicity. *Front. Psychol.* 5:80. doi: 10.3389/fpsyg.2014.00080
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., and Conrad, M. (2014a). ANGST: affective norms for german sentiment terms, derived from the affective norms for english words. *Behav. Res. Methods* 46, 1108–1118. doi: 10.3758/s13428-013-0426-y
- Schrott, R., and Jacobs, A. M. (2011). *Gehirn und Gedicht: Wie wir Unsere Wirklichkeiten Konstruieren (Brain and Poetry: How We Construct Our Realities)*. München: Hanser.
- Stenneken, P., Bastiaanse, R., Huber, W., and Jacobs, A. M. (2005). Syllable structure and sonority in language inventory and aphasic neologisms. *Brain Lang.* 95, 280–292. doi: 10.1016/j.bandl.2005.01.013
- Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348. doi: 10.1037/a0016973
- Sylvester, T., Braun, M., Schmidtke, D., and Jacobs, A. M. (2016). The Berlin Affective Word List for Children (kidBAWL): exploring processing of affective lexical semantics in the visual and auditory modalities. *Front. Psychol.* 7:969. doi: 10.3389/fpsyg.2016.00969
- Tagliamonte, S. A., and Baayen, H. R. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Lang. Var. Change* 24, 135–178. doi: 10.1017/S095439451200129
- Turney, P. D., and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inform. Syst.* 21, 315–346. doi: 10.1145/944012.944013
- Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., and Conrad, M. (2017). On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic texts—a case study using 57 poems of H. M. Enzensberger. *Front. Psychol.* 7:2073. doi: 10.3389/fpsyg.2016.02073
- Vennemann, T. (1988). *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: de gruyter.
- Vö, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., and Jacobs, A. M. (2009). The berlin affective word list reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538. doi: 10.3758/B.R.M.41.2.534

- Võ, M. L. H., Jacobs, A. M., and Conrad, M. (2006). Cross-validating the berlin affective word list. *Behav. Res. Methods* 38, 606–609. doi: 10.3758/BF03193892
- Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., and Jacobs, A. M. (2014). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: using large text corpora to compute valence, arousal, and the basic emotions. *Q. J. Exp. Psychol.* 68, 1599–1622. doi: 10.1080/17470218.2014.970204
- Willems, R., and Jacobs, A. M. (2016). Caring about Dostoyevsky: the untapped potential of studying literature. *Trends Cogn. Sci.* 20, 243–245. doi: 10.1016/j.tics.2015.12.009
- Zeki, S., Romaya, J. P., Benincasa, D. M. T., and Atiyah, M. F. (2014). The experience of mathematical beauty and its neural correlates. *Front. Hum. Neurosci.* 8:68. doi: 10.3389/fnhum.2014.00068
- Ziegler, J. C., and Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *J. Mem. Lang.* 34, 567–593. doi: 10.1006/jmla.1995.1026

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jacobs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.