sciendo
DOI 10.2478/jazcas-2025-0002

CONSTRUCTIONAL CREATIVITY – THE ROLE OF CORPUS DATA

THOMAS HOFFMANN
Department of English Language and Linguistics, Faculty of Languages and Literatures, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany
(ORCID: 0000-0001-6096-3296)

HOFFMANN, Thomas: Constructional Creativity – The Role of Corpus Data. Journal of Linguistics, 2025, Vol. 76, No 1, pp. 20 – 30.

Abstract: One domain of human cognition that has recently received considerable attention in cognitive linguistics is linguistic creativity (e.g. Bergs 2019; Hartmann and Ungerer 2024, 2025; Hoffmann 2024, 2025; Turner 2018). The present paper gives an overview of creativity research from the fields of linguistics and psychology and introduces the 5C model of constructional creativity (constructors, co-constructors, constructs, constructional blending and the constructional network; Hoffmann 2024, 2025a,b). A particular focus of the paper will be on the role of corpus linguistics for the investigation of constructional creativity.

Keywords: constructional network, coverage, creativity, vector space models

1 INTRODUCTION

In their daily interactions, speakers reuse a great number of formulaic, fixed expressions such as *Hi*, *Good morning*, *No worries* or *Love you*. Conklin and Schmitt point out that "studies suggest that formulaic language makes up between one third and one half of discourse" (2012, p. 46). Sinclair (1991) calls this tendency to draw on conventional words and phrases 'the idiom principle'. At the same time, Sinclair notes a competing tendency that he labels the 'open choice principle': People have the ability to combine words into novel utterances that they have never uttered before. For Chomsky (1965, p. 6), this "creative aspect" is "an essential property of language".

In the last couple of years, linguistic creativity has received considerable attention in cognitive linguistics (cf., e.g. Bergs 2018, 2019; Bergs and Kompa 2020; Hartmann and Ungerer 2024, 2025; Herbst 2018; Hoffmann 2018, 2019, 2020, 2022a; Norde and Trousdale 2025; Schneck 2018; Trousdale 2018; Turner 2018, 2020; Uhrig 2018, 2020). The present article first presents a cognitive linguistic model (the 5C model of constructional creativity; Hoffmann 2024, 2025a,b) for the holistic investigation of linguistic creativity (Section 2). Then, it will illustrate how corpus linguistic methods can be used for studies exploring diachronic innovation as well as synchronic creative utterances (Section 3).

2 THE 5C MODEL OF CONSTRUCTIONAL CREATIVITY

While in linguistics creativity is often reduced to productivity (see Sampson 2016 for a detailed discussion), the standard definition in psychology emphasizes that creative products (linguistic or otherwise) have to be **original/novel** as well as **appropriate/useful** (cf., e.g. Simonton 2012; Kaufman 2016). To illustrate this, imagine if I randomly typed away at my keyboard: I might end up with a completely novel sequence of letters that has never been produced before. Yet, while *sdsgd hsd sdaor,rf jejgfb jbwe* might be original it clearly is utterly inappropriate as it is unintelligible. From a linguistic point of view, utterances are therefore creative if they are both novel/original as well as appropriate. While corpus data can help in assessing the degree of novelty of an utterance (see Section 3), appropriateness/usefulness is a subjective criterion that depends on context as well as on the appreciation by listeners or readers (Giora 2003; Veale 2012). In a literary context, such as (1), e.g. a semi-random sequence might become acceptable (or even be considered 'high art'):

(1) The fall (bababadalgharaghtakamminarronnkonnbronntqnnerronntuonnt-hunntrovarrhounawnskawntoohoohoordenenthur-nuk!) of a once wallstrait old-parr […] (James Joyce, *Finegan's Wake*, 1939, 1; cit. in: Bergs 2018, p. 286)

As this brief discussion already illustrates, various perspectives have to be taken into account when assessing the creativity of any phenomenon. Drawing on Glaveanu's 5A model (Glaveanu 2013; Lubart et al. 2021), we can identify the following five major variables:

1. the individual that is creative (the 'actor'),
2. the people who interact with the actor and evaluate the creative product (the 'audience'),
3. the creative product or act (the 'artifact'),
4. the creative process that produces the artifact (the 'action'),
5. the environmental, material and contextual factors that influence the action (the 'affordances').

Hoffmann (2024, 2025a,b) has incorporated the 5As into a cognitive linguistic model of linguistic creativity known as the "5C model of constructional creativity". The theoretical foundation of the model is Construction Grammar (Goldberg 2006, 2019; Hilpert 2019; Herbst and Hoffmann 2024; Hoffmann 2022). The main tenet of Construction Grammar is the claim that constructions, i.e. symbolic pairings of FORM and MEANING are the central units of human language:

Constructions cover the entire lexis-syntax cline and range from words (FORM: /ˈbʊk/ ↔ MEANING: 'concept of a book') over partly schematic patterns

(*Un*-V construction: FORM: /ʌn$_1$-ˈV$_2$/ ↔ MEANING: 'reversing$_1$ an EVENT$_2$)' as in *unbind*, *unfreeze*, *undo*) to fully schematic templates (e.g. the RESULTATIVE construction: FORM: [SBJ$_1$ [V$_2$ OBJ$_3$ OBL$_4$]VP] ↔ MEANING: 'Agent$_1$ causes Patient$_3$ to become RESULT-GOAL$_4$ by V$_2$-ing'; adapted from Hoffmann (2022, p. 179), that licenses utterances such as *He wiped the table clean*. or *They cut the man free*.). All of a speakers' constructions are stored in the long-term memory of their constructional network (Diessel 2019). The production of (creative as well as routine) utterances then involves the activation and combination of various constructions into so-called 'constructs' in the working memory. This happens via the domain-general process of Conceptual Blending (Fauconnier and Turner 2002), which is postulated to be the only mental operation required to combine constructions (also known as constructional blending; Herbst and Hoffmann 2024; Hoffmann and Turner fc.).

Fig. 1 provides a visual representation of the full 5C model:

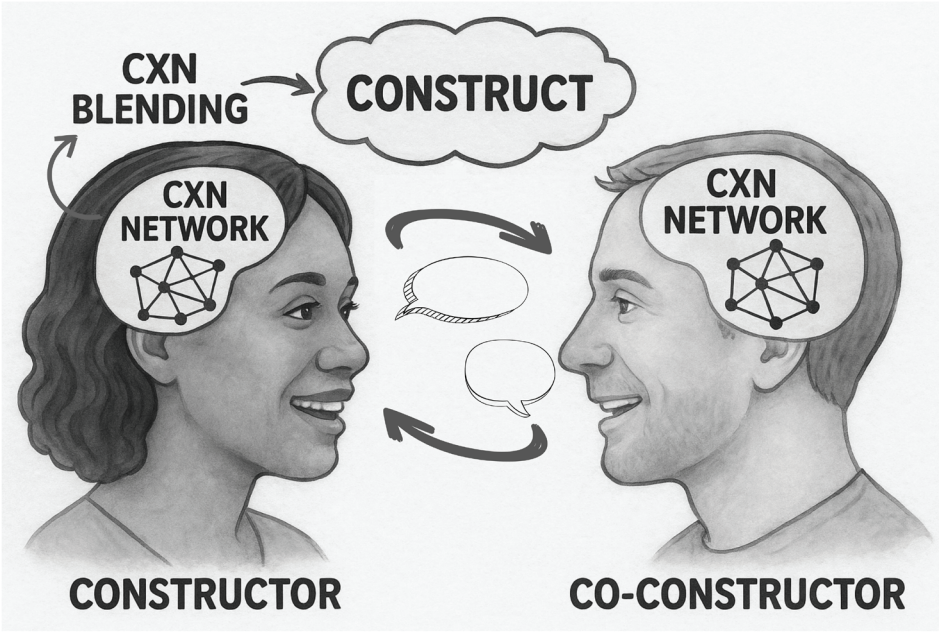

Fig. 1. The 5C Model of Constructional Creativity (available via open access at: https://osf.io/d9hbg)

Glaveanu's (2013) actor and audience are relabeled as 'constructor' and 'co-constructor' in the 5C model, and as Hoffmann (2024; 2025a,b) shows it is the dynamic interaction of the two that often leads to creative utterances (the parole elements that realize the mental constructs). The mental process that underlies the

production of creative utterances is 'constructional (cxn) blending' (see above). The particular focus of the present paper is on the final part of the 5C model – the 'constructional (cxn) network'.

Usage-based Construction Grammar approaches (such as Goldberg 2006, 2019; Hilpert 2019; Herbst and Hoffmann 2024; Hoffmann 2022) argue that constructions are acquired through language use and that the strength of mental storage (the 'entrenchment' of a construction) depends on frequency effects: If an utterance such as *Good Morning!* is repeatedly encountered without any variation it will become entrenched as a prefab/chunks. When a pattern such as the RESULTATIVE construction appears with many different lexicalizations (e.g. *They elected him president*., *He drank himself stupid*., *She danced herself happy.*[1]), it can give rise to the schema above. The specific utterances that underlie the schema are not forgotten immediately: Instead, "partially abstracted (lossy) structured exemplars dynamically cluster within our hyper-dimensional conceptual space" (Goldberg 2019, p. 51). These exemplar clouds, the "coverage" of a construction, play a crucial role when it comes to the acceptability of novel instances:

> Specifically, a potential productive use of an existing construction (a coinage) is acceptable to the degree that the category which would be required to include the previously attested examples and the coinage is well attested within the hyper-dimensional conceptual space in which exemplars cluster (Goldberg 2019, pp. 62–63).

Finally, constructions and their coverage are stored in the long-term memory network of constructions, where they have vertical as well as horizontal links (see, e.g. Diessel 2019; Sommerer and Van de Velde 2025): *Unfair*, *unholy* or *untrue* will be stored as specific taxonomic instances of a more schematic *Un*-Adjective construction. At the same time, a horizontal link to *fair*, *holy* and *true* as well as their schematic positive Adjective construction will encode their antonymic relationship in the network (see Hoffmann 2022, pp. 54–55).

3 CORPUS LINGUISTIC IMPLICATIONS

Most corpora are aggregate data collected from many different individuals. As Schmid (2020) emphasizes, they, therefore, only provide direct evidence for the conventionalization of constructions but not the level of individual mental entrenchment. At the same time, following Schmid's own 'from-corpus-to-cognition principle' (2000, p. 39), Stefanowitsch and Flach (2017, p. 122) argue that corpora at least provide an indirect window onto entrenchment. First of all, corpus data are

[1] *She painted a mess of some pale yellows and dirty greys, got the grown-ups placed, and she danced herself happy.* Source: https://admp.org.uk/wp-content/uploads/E-Motion-Spring-18_Vol-xxviii-No1.pdf [last accessed 08/08/2025].

necessarily the output of individual grammars and as such authentic corpus data can be used to draw "inferences about the mental representations underlying this behavior" (Stefanowitsch and Flach 2017, pp. 102–103; 'corpus-as-output' hypothesis). Secondly, according to the 'corpus-as-input' view, corpus data can also be seen as a proxy for the input that speakers of a speech community are exposed to and which consequently shape their mental construction networks (Stefanowitsch and Flach 2017, p. 103).

Corpora might, therefore, offer at least indirect evidence for mental constructional networks – but how can they be used in creativity research if creative constructs by definition are supposed to be novel, i.e. should not frequently be found in corpora? One potential use is the evolution of novel constructions in diachronic studies. Hoffmann and Trousdale (2022), e.g. investigated the rise of the construction in (2), which licenses constructs such as (3a,b):

(2) The *Hell*-construction
FORM: [NP_i V_j [the N_{TABOOj}]$_k$ [*out of* [NP_l]]
↔
MEANING: ['SEM_i $excessively_k$ $PRED_j$ SEM_l']
i = Subject/Agent
l = Oblique/Theme

function: [speaker's heightened emotion]
(adapted from Hoffmann and Trousdale 2022, p. 378)

(3) a. I got to beat the Devil out of you, child (2015, COCA)
b. I respect the hell out of those guys (2019, COCA)

The *Hell*-construction in (2) is an extravagant construction (Hartmann and Ungerer 2024, 2025) since it uses a taboo noun (*Devil* (3a), *hell* (3b)) and expresses a speaker's heightened emotion. Semantically, it expresses the same state of affairs as the TRANSITIVE construction (cf. *I got to beat you*. *I respect those guys*), albeit with the additional meaning that the event is carried out excessively ((3a) means the beating was excessive, (3b) that the level of respect was exceptionally high).

Earlier studies had identified single constructions as the source of the *Hell*-construction. Hoeksema and Napoli (2008), e.g. claimed that constructions that expressed a literal exorcism such as *I will preach the devil out of thee* (1835 COHA[2]) functioned as the source of (2). While the details of the corpus study by Hoffmann and Trousdale (2022) is beyond the scope of the present contribution, Fig. 2 at least

[2] Corpus of Historical American English (COHA). Accessible at: https://www.english-corpora.org/coha/ [last accessed 08/08/2025].

illustrates the various input constructions they identified as potential source constructions in their COHA data:

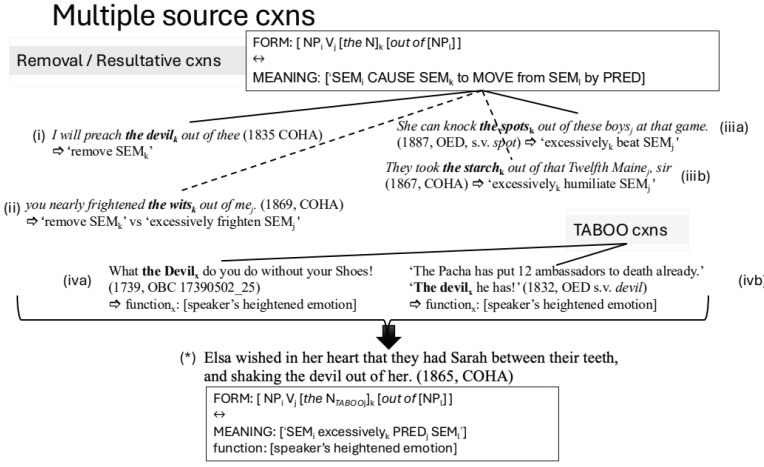

Fig. 2. The multiple sources of the modern Hell construction (based on the data of Hoffmann and Trousdale 2022)

Importantly, Hoffmann and Trousdale (2022) note that different speakers might have used different source constructions (e.g. (i) and (iiia) or (ii) and (iva,b) or other routes in Fig. 2) to innovate the novel construction. Similarly, hearers, upon encountering the construction for a first time, could rely on multiple routes in the constructional network to parse and then entrench the new constructions. Careful corpus research, however, is essential to uncover such multiple paths in language change.

Synchronically, corpora are also of great use for creativity studies. Particularly helpful are so-called 'vector space models' that use large corpora to produce so-called 'word embeddings' (see, e.g. Perek 2016; Surdeanu and Valenzuela-Escárcega 2024). Word embeddings are numerical vectors that encode the context in which a word appears and are an important component of large language models (Surdeanu and Valenzuela-Escárcega 2024, pp. 117–131). Since "words that occur in similar contexts tend to have similar meanings" (Surdeanu and Valenzuela-Escárcega 2024, p. 131), word embeddings can be seen as a proxy for the "semantic representation of words" (Surdeanu and Valenzuela-Escárcega 2024, p. 132). In line with the 'corpus-as-output' principle, the word embeddings of a constructional slot, such as the V slot in (2), allows researchers to establish the existing coverage of a construction. This can then be used to assess the creativity of experimental responses.

Hoffmann and Steinhauser (2025), e.g. expanded the Divergent Association Test (DAT; Olson et al. 2021) to constructional contexts. In an online experiment,

they asked subjects to "enter 10 verbs that are as different from each other as possible, in all meanings and uses of the words" in constructional contexts such as (4):

(4) Yesterday, the woman _____ the hell out of the man.

Hoffmann and Steinhauser (2025) randomized the subject and object position of *woman* and *man* (to preclude gender stereotype effects). All in all 55 subjects were recruited via the platform Prolific (age range 18–60, 37 females, 18 males, all L1 speakers of UK English). Again, the specific details of the study are beyond the current paper. Yet, once more, corpus data proved a useful tool to assess the creativity with which subjects responded. Fig. 3 gives the word cloud of all the verbs offered in the study, with the most frequent verbs being larger in size proportional to their frequency:

Fig. 3. Word cloud of experimental data of *Hell*-construction (from Hoffmann and Steinhauser 2025)

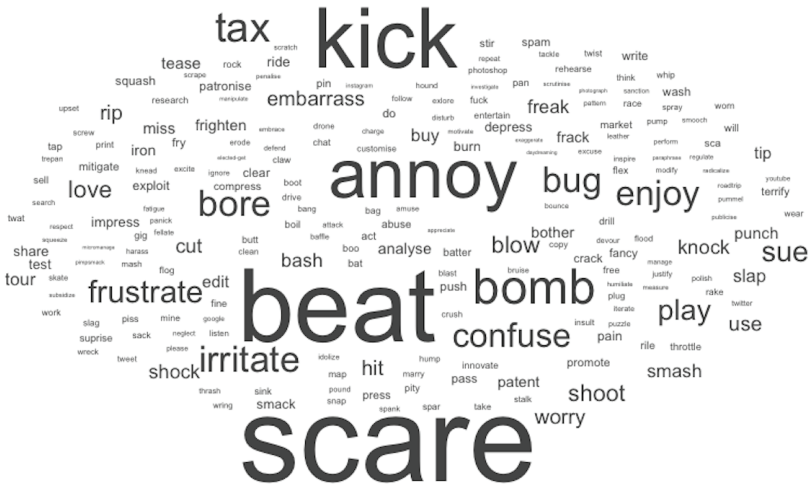

Fig. 4. Word cloud of corpus data of *Hell* construction (based on GB data from GloWbE; Hoffmann 2021)

Compare the results from Fig. 3 to the ones in Fig. 4, which are based on the coverage of the construction in the GB subcorpus of GloWbE (Hoffmann 2021): Since the verbs *beat* and *scare* (as well as their associated semantic frames) dominate the constructional coverage in the corpus, it is not surprising that these were also produced most frequently (and early) by subjects in the experiment.

Even more interesting perhaps is the possibility of using vector space models to plot the range of verbs that were produced in the experiment but did not appear in the corpus (Fig. 5):

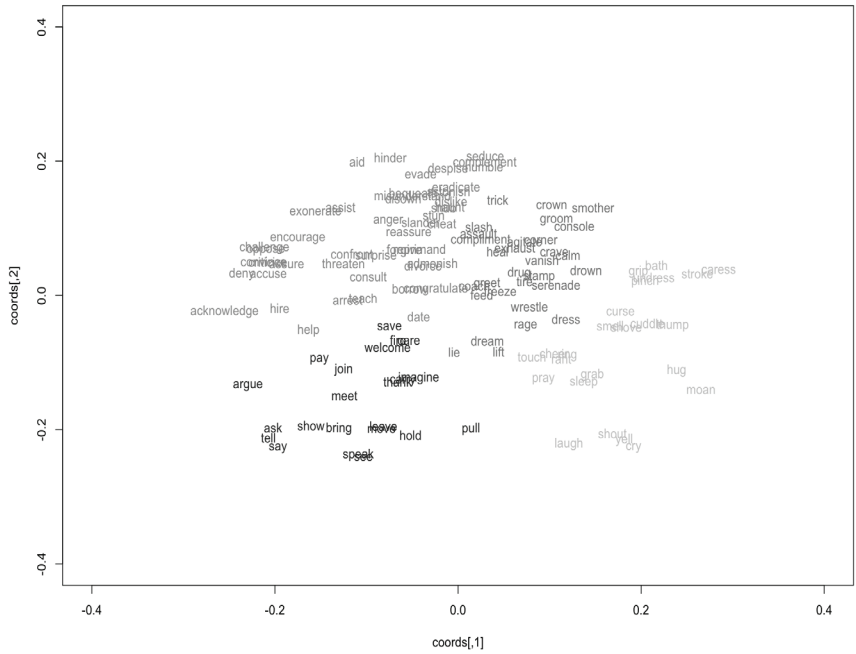

Fig. 5. Vector space model of verbs in *Hell* construction that were produced in the experiment but did not surface in the reference corpus[3] (data from Hoffmann and Steinhauser 2025 and Hoffmann 2021).

Several of the verbs seem to have been activated by the male-female subject-objects (e.g. *caress*, *date*, *seduce*). At the same time, there are many others (e.g. *acknowledge serande*, *vanish*) that warrant closer future attention. It is only through corpus-based vector space models that such relevant instances can be identified.

[3] The vector space that was used for reference was "verbs.coca.w5.skip.1000d.txt" by F. Perek. Accessible at: https://osf.io/3gynu/files/osfstorage/630e4092171132006f5a8dba [last accessed 08/08/2025].

4 CONCLUSION

The present paper has outlined that cognitive linguistic research on creativity must take into account the 5Cs (constructor, co-constructor, constructional network, construction blending, construct). A particular focus was placed on the constructional network and how corpus linguistic studies can fruitfully be used to investigate constructional creativity.

ACKNOWLEDGEMENTS

The research by Hoffmann and Steinhauser (2025) is part of the project Predictive Construction Grammar – Developing a neuro-cognitive model of linguistic creativity funded by the VolkswagenStiftung.

References

Bergs, A. (2018). 'Learn the rules like a pro, so you can break them like an artist' (Picasso): Linguistic aberrancy from a constructional perspective. Zeitschrift für Anglistik und Amerikanistik, 6(3), pp. 277–293.

Bergs, A. (2019). What, if anything, is linguistic creativity? Gestalt Theory, 41(2), pp. 173–183.

Bergs, A. and Kompa, N. (2020). Creativity within and outside the linguistic system. Cognitive Semiotics, 13(1).

Conklin, K., and Schmitt, N. (2012). The processing of formulaic language. Annual Review of Applied Linguistics 32, pp. 45–61.

Giora, R. (2003) On our mind: Salience, context, and figurative language. Oxford: Oxford University Press, 259 p.

Glaveanu, V. (2013). Rewriting the language of creativity: The five A's framework. Review of General Psychology, 17(1), pp. 69–81.

Hartmann, S., and Ungerer, T. (2024). Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. Journal of Linguistics, 60(3), pp. 599–634.

Hartmann, S., and Ungerer, T. (2025). Chaos theory, shmaos theory: Creativity and routine in English *shm*-reduplication. In: S. Arndt-Lappe – N. Filatkina (eds.): Dynamics at the lexicon-syntax Interface: Creativity and routine in word-formation and multi-word expressions. De Gruyter, pp. 295–322.

Fauconnier, G., and Turner, M. (2002). The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. New York: Basic Books, 464 p.

Goldberg, A. E. (2006). Constructions at Work: The Nature of Generalization in Language. Oxford: Oxford University Press, 286 p.

Goldberg, A. E. (2019). Explain Me This: Creativity, Competition and the Partial Productivity of Constructions. Princeton: Princeton University Press, 195 p.

Herbst, Th. (2018). Collo-Creativity and blending: Recognizing creativity requires lexical storage in constructional slots. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 309–328.

Herbst, Th., and Hoffmann, Th. (2024). A Construction Grammar of English: A Constructionist Approach to Syntactic Analysis (CASA). Amsterdam: John Benjamins, 315 p.

Hilpert, M. (2019). Construction Grammar and its Application to English. 2nd ed. Edinburgh: Edinburgh University Press, 296 p.

Hoeksema, J., and Napoli, D. J. (2008). Just for the hell of it: A comparison of two taboo-term constructions. Journal of Linguistics, 44(2), pp. 347–378.

Hoffmann, Th. (2018). Creativity and construction grammar: Cognitive and psychological issues. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 259–276.

Hoffmann, Th. (2019). Language and creativity: A construction grammar approach to linguistic creativity. Linguistics Vanguard, 5(1).

Hoffmann, Th. (2020). Construction grammar and creativity: Evolution, psychology and cognitive science. Cognitive Semiotics, 13(1).

Hoffmann, Th. (2021). The Cognitive Foundation Of Post-colonial Englishes: Construction Grammar as the Cognitive Theory for the Dynamic Model. (Cambridge Elements in World Englishes). Cambridge: Cambridge University Press.

Hoffmann, Th. (2022). Construction Grammar: The Structure of English. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press, 315 p.

Hoffmann, Th. (2024). The 5C model of linguistic creativity: Construction Grammar as a cognitive theory of verbal creativity. Journal of Foreign Languages and Cultures, 8(1), pp. 139–154.

Hoffmann, Th. (2025a). Creativity. In Reference Module in Social Sciences, Elsevier. Accessible at: https://doi.org/10.1016/B978-0-323-95504-1.00588-3.

Hoffmann, Th. (2025b). Cognitive approaches to linguistic creativity. In: X. Wen – Ch. Sinha (eds.): The Cambridge Encyclopedia of Cognitive Linguistics. Cambridge: Cambridge University Press.

Hoffmann, Th., and Steinhauser, M. (2025). Constructional Divergent Association Task (CXN-DAT): Using constructional contexts to measure linguistic creativity. Poster presented at the 2025 Society for the Neuroscience of Creativity (SfNC) conference, Paris Brain Institute (May 22 – 23, 2025).

Hoffmann, Th., and Trousdale, G. (2022). On multiple paths and change in the language network. Zeitschrift für Anglistik und Amerikanistik, 70(3), pp. 359–382.

Hoffmann, Th., and Turner, M. (fc.). Creative Construction Grammar. (Cambridge Elements in Cognitive Linguistics). Cambridge: Cambridge University Press.

Kaufman, J. C. (2016). Creativity 101. 2nd ed. New York: Springer Publishing Company, 368 p.

Lubart, T., Glăveanu, V., de Vries, He., Camargo, A., and Storme, M. (2021). Cultural perspectives on creativity. In: J. C. Kaufman – Robert J. Sternberg, (eds.): Creativity: An introduction. Cambridge: Cambridge University Press, pp. 128–151.

Norde, M., and Trousdale, G. (2025). Creativity, paradigms and morphological constructions: evidence from Dutch pseudoparticiples. Linguistics, 63(4), pp. 1029–1063.

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., and Webb, M. E. (2021). Naming unrelated words predicts creativity. Proceedings of the National Academy of Sciences of the United States of America, 118(25), e2022340118. Accessible at: https://doi.org/10.1073/pnas.2022340118.

Perek, F. (2016).Using distributional Semantics to study syntactic productivity in diachrony: A case study. Linguistics, 54(1), pp. 149–188.

Sampson, G. (2016). Two ideas of creativity. In: M. Hinton (ed.): Evidence, Experiment and Argument in Linguistics and Philosophy of Language. Bern: Peter Lang, pp. 15–26.

Schmid, H.-J. (2000). English Abstract Nouns as Conceptual Shells: From Corpus to Cognition. Berlin: Mouton de Gruyter, 468 p.

Schneck, P. (2018). Creative grammarians: Cognition, language and literature: An exploratory response. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 381–391.

Simonton, D. K. (2012). Creative productivity and aging. In: S. Krauss Whitbourne – M. J. Sliwinski (eds.): The Wiley-Blackwell Handbook of Adulthood and Aging. Malden, MA: Wiley-Blackwell, pp. 477–496.

Sommerer, L., and Van de Velde, F. (2025). Constructional networks. In: M. Fried – K. Nikiforidou (eds.): The Cambridge Handbook of Construction Grammar. Cambridge: Cambridge University Press, pp. 220–246.

Stefanowitsch, A., and Flach, S. (2017). The corpus-based perspective on entrenchment. In: H.-J. Schmid (ed.): Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge. Berlin: de Gruyter, pp. 101–127.

Surdeanu, M., and Valenzuela-Escárcega, M. A. (2024). Deep Learning for Natural Language Processing. A Gentle Introduction. Cambridge: Cambridge University Press, 344 p.

Trousdale, G. (2018). Creativity parallels between language and music. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 371–380.

Trousdale, G. (2020). Creativity, reuse, and regularity in music and language. Cognitive Semiotics, 13(1).

Turner, M. (2018). The role of creativity in multimodal construction grammar. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 357–370.

Turner, M. (2020). Constructions and creativity. Cognitive Semiotics, 13(1).

Uhrig, P. (2018). I don't want to go all Yoko Ono on you. Zeitschrift für Anglistik und Amerikanistik, 66(3), pp. 295–308.

Uhrig, P. (2020). Creative intentions – The fine line between 'creative' and 'wrong'. Cognitive Semiotics, 13(1).

Veale, T. (2012). Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity. London et al.: Bloomsbury, 184 p.