

# Theoretical stances shape empirical generalizations on inflection vs. derivation: quantitative evidence from Czech

Lukáš Kyjánek & Olivier Bonami  
Université Paris Cité, Paris, France

[l.kyjaneck@gmail.com](mailto:l.kyjaneck@gmail.com) & [olivier.bonami@linguist.univ-paris-diderot.fr](mailto:olivier.bonami@linguist.univ-paris-diderot.fr)

The distinction between inflection and derivation remains unresolved in morphology (Anderson 1982; Dressler 1989; Booij 1996; Haspelmath 1996; Corbett 2010; Spencer 2013; Štekauer 2015). Recent computational approaches have addressed this by contrasting the properties of sets of word pairs standing in the same morphological relation.

We point out that the studies in question differ in how they sample morphological relations, reflecting different theoretical stances on morphology. Bonami & Paperno (2018) adopt a fully paradigmatic view in the spirit of Bonami & Strnadova (2019), defining contrasts by content rather than form (Štekauer 2015), with no priority given to a ‘reference form’. Copot et al. (2022) rely on morphosemantic contrasts but limit attention to contrasts between a ‘reference form’ (the citation form or the base form) and another form (another inflected form or a derivative). Rosa & Žabokrtský (2019) and Haley et al. (2024) use a mixed approach: content-based contrasts to a reference form for inflection (e.g., word pairs like *hrad~hrad-u* ‘castle.NOM.SG~GEN.SG’) and *stroj~stroj-e* ‘machine.NOM.SG~GEN.SG’ are lumped together), individual processes for derivation (e.g. word pairs *přítel~přítel-kyně* ‘male~female friend’ and *učitel~učitel-ka* ‘male~female teacher’ are kept apart despite parallelism of content.).

In this study, we examine how these decisions influence conclusions about the inflection-derivation divide by applying Haley et al.'s methodology to Czech data with different ways of sampling morphological relations. Using extensive morphological resources of Czech (MorFlexCZ, Hajič et al. 2020; DeriNet, Vidra et al. 2021; Word embeddings, Kyjánek & Bonami 2022), we analyze over 20 types of morphological categories, exemplifying canonical inflection (e.g., grammatical case contrasts), derivation (e.g., verbs vs. agent nouns), and intermediate cases (e.g., diminutive formation, aspectual pairs, masculine-feminine pairs of human nouns).

**Keywords:** inflection, derivation, boundary, morphological relations, Czech

## References

- Anderson, S. R. (1982) Where's morphology? *Linguistic Inquiry*, 13.
- Dressler, U. W. (1989) Prototypical differences between inflection and derivation. *STUF – Language Typology and Universals*, 42(1):3–10.
- Booij, G. (1996) *Inherent versus contextual inflection and the split morphology hypothesis*. In: Yearbook of Morphology.
- Haspelmath, M. (1996) *Word-class Changing Inflection and Morphological Theory*. Yearbook of Morphology.
- Corbett, G. G. (2010) Canonical derivational morphology. *Word Structure*, 3.
- Spencer, A. (2013) *Lexical relatedness*.
- Štekauer, P. (2015) The delimitation of derivation and inflection. In: Word-Formation. *An International Handbook of the Languages of Europe*.
- Bonami, O., Paperno, D. (2018) Inflection vs. derivation in a distributional vector space. *Lingue e linguaggio*, 17(2).
- Bonami, O., Strnadova, J. (2019) Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2).

- Coppot, M., Mickus, T., Bonami, O. (2022) Idiosyncratic frequency as a measure of derivation vs. inflection. *Journal of Language Modelling*, 10(2).
- Rosa, R., Žabokrtský, Z. (2019) Attempting to separate inflection and derivation using vector space representations. In: *Proceedings of DeriMo2*.
- Haley, C., Ponti, E. M., Goldwater, S. (2024). Corpus-based measures discriminate inflection and derivation cross-linguistically. *Journal of Language Modelling*, 12(2).
- Hajič, J. et al. (2020) MorfFlex CZ 2.0. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Vidra, J. et al. (2021) DeriNet 2.1, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Kyjánek, L., Bonami, O. (2022) Package of word embeddings of Czech from a large corpus, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.