# Miloš Jakubíček, Michael Rundell, Cormac Breathnach, Pádraig Ó Mianáin

# DANTE Resurrected

## Large Open Source Lexical Data Database for English

### About DANTE

DANTE (Database of Analysed Texts of English) was initially developed in the years 2008–2010 (Atkins, Kilgarriff & Rundell, 2010) by a lexicographic team led by Sue Atkins, Adam Kilgarriff, Valerie Grundy and Michael Rundell. It was commissioned by Foras na Gaeilge, a governmental agency promoting the use of Irish language, for the purposes of the development of the New English Irish Dictionary (Mianáin & Convery, 2014). It was produced on the basis of an English corpus having about 1.7 billion words using the Sketch Engine toolchain (Kilgarriff et al., 2014). DANTE provides a very detailed lexicographic analysis of about 50,000 single word English entries as well as 45,000 compounds, with lexical units subject to the following structure:

- – wordclass
- – secondary grammar (inherent properties of headword)
- – informal definitions
- – syntactic constructions and arguments of the headword
- – lexical collocates based on corpus frequencies
- – support verb constructions
- – support prepositions
- – domain /subject field
- – regional variety
- – speaker/writer attitude
- – time
- – register
- – style
- – full example sentences (from corpus)

- variant forms
- derived forms
- cross-reference

**DANTE Resurrected**

DANTE was originally released on the IDM DPS platform[7] and until 2023 it was a closed proprietary product of Foras na Gaeilge. In this paper we present new development following the decision made by Foras na Gaeilge to release the content of DANTE under the terms of the CC-BY 4.0 open source license.[8]

We show a new dedicated web interface for DANTE based on the Lexonomy dictionary platform (Měchura et al., 2017; Rambousek et al., 2021) hosted at anonymized which interlinks DANTE with additional corpus based resources and discuss future uses of DANTE for the purposes of research in lexicography. We particularly focus on using DANTE for evaluation of automatic dictionary drafting, including by using large language models, and the full paper will provide an experimental evaluation of these methods based on DANTE data.

## magically

**1 adv** using magic
⇨ *Rather they enchant them* **magically** *and the horses serve willingly .*
⇨ *Maybe the Ring is* **magically** *conscious of Gandalf 's power : maybe , though , Frodo is already afraid that he will lose it .*
⇨ *Well-known in medieval Europe , the caladrius was said to herald the death and birth of kings . It would sit at the foot of the sick-bed and , if it gazed at the patient , its eyes would* **magically** *absorb the malaise so the man had a renewed chance of life .*

**2 adv** suddenly; as if by magic
⇨ *He punched some buttons and a ticket* **magically** *emerged .*
⇨ *The Baron , his eyes streaming with the hay fever , sneezed once more and the French guns* **magically** *ceased fire .*
⇨ *I had naively imagined that marriage would* **magically** *obliterate the origins of the wealth we shared , melting Dennis's laboriously acquired treasures down into a common heap of anonymous gold .*
⇨ *I asked to see a breakdown of the price quoted to me and found that I was being charged four times more than the normal retail price for curtain tape and hooks and was quoted for more material than I needed . When I questioned this , the sum was* **magically** *reduced .*
⇨ *They needed no professional help with the evening 's drinking but arose* **magically**

**Fig. 1:** Original interface for a sample DANTE adverb *magically*.

---

**Conclusions**

The 2010 paper on DANTE ends with the following description:

*"DANTE is a lexicographic project where the end-product is not a dictionary but an in-depth analysis to be used for creating one or more dictionaries. The users of DANTE are not the dictionary-using public but the lexicographic teams who will take this on to dictionary status."*

We believe that by open sourcing DANTE, Foras na Gaeilge made an important step towards the goals as initially envisaged and that DANTE is an important and welcome contribution to the international lexicographic community.

**References**

Atkins, B. S., Kilgarriff, A., & Rundell, M. (2010). Database of Analysed Texts of English (DANTE): the NEID database project. In *Proceedings of the XIV Euralex International Congress* (pp. 549–556).

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*(1), 7–36.

Měchura, M. B. et al. (2017). Introducing Lexonomy: an open-source dictionary writing and publishing system. In *Proceedings of the eLex 2017 conference* (pp. 19–21).

Mianáin, P. Ó., & Convery, C. (2014). From DANTE to Dictionary: The New English-Irish Dictionary. In *Proceedings of the 16th EURALEX International Congress* (pp. 807–817).

Rambousek, A., Jakubíček, M., Kosem, I., Cukr, M., Kallas, J., Krek, S., & Tiberius, C. (2021). New developments in Lexonomy. In *Electronic lexicography in the 21st century (eLex 2021)* (pp. 455–462).