CAFECONMIEL: Corpus Automático y Fenómenos de Contacto en Mallorca: Inteligencia, Entrenamiento y Lengua

David Sánchez

IFISC Winter Solstice Event (2022)

- **Call:** Research projects funded by the Balearic Islands Autonomous Community government (published in Dec 2020)

- Subproject 1 (IFISC at UIB): D. Sánchez IP1 and coordinator

- Subproject 2 (Spanish Linguistics at UIB): R. Miguel IP2

- **Budget:** 111742 Euro (IFISC team: 47521 Euro)
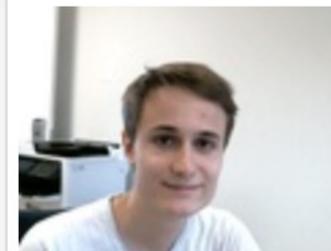
- Period: Dec 1 (2021) — Oct 15 (2024)

TOBIAS GALLA	JOSE JAVIER RAMASCO	DAVID SÁNCHEZ	THOMAS LOUF

Main: Create a general Spanish corpus in Mallorca using artificial intelligence methods to study language variation and change

- To apply classification techniques (both supervised and unsupervised) for cluster identification and **discovery of new features**

- To find characteristic lexical patterns that vary with geographical location and historical period

- To quantitatively assess the linguistic change with complex systems approaches

Example: Analysis of 18th-century texts from Mallorca (BAL) and mainland Spain (PEN)

BAL	â, á, à, testigo, ha, dixo, sabe, si, verdad, parte, declarante, decir, ante, alias, contra, dixó, haver, firmó, responde, porque
PEN	acordo, deesta, rexidores, vna, vn, maior, mill, vno, escriuano, desta, entendida, dela, enel, rreales, hauer, deeste, vso, setezientos, estte, treinta

THANK YOU

for your attention