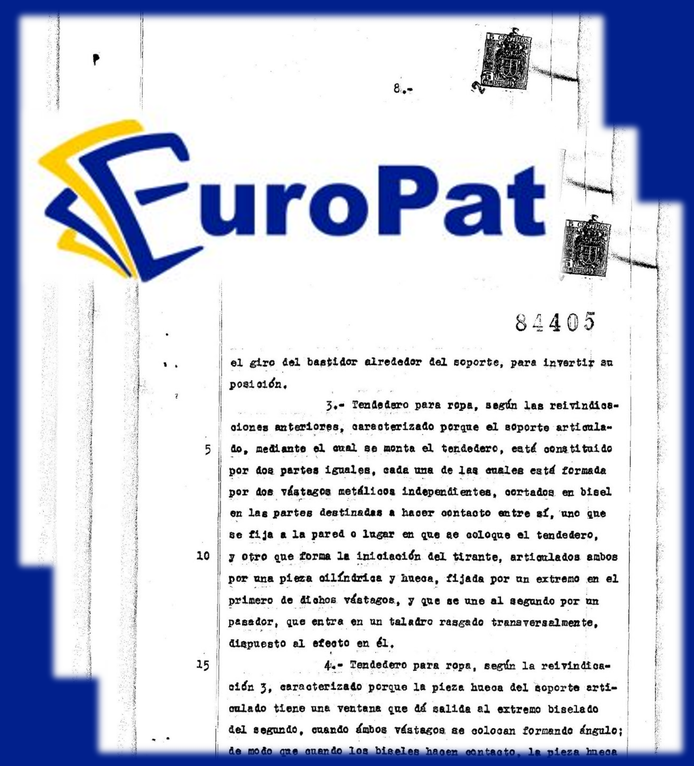
Parallel corpus from patents in German, Spanish, French, Croatian, Norwegian, and Polish, paired with English



Kenneth Heafield, **Elaine Farrow**, Jelmer van der Linde, Gema Ramírez-Sánchez, Dion Wiggins

INTRO

- Patents are a rich source of technical vocabulary and product names, complementing other data sources
- We mined parallel corpora by aggregating and aligning patents in six European languages paired with English

METHODS

1. Data acquisition: machine-readable text, scanned patent images
2. OCR and text extraction
3. Document alignment
4. Sentence alignment
5. Data cleaning
6. Domain grouping: International Patent Classification (IPC)
7. Corpus quality assessment: human and automatic
8. Data releases: RAW, TXT, and the translation memory standard TMX format

CORPUS SIZE

Language pair	Sentences (millions)
en-de	19.734
en-es	51.352
en-fr	11.098
en-hr	0.154
en-no	4.341
en-pl	0.332

NO20090845L

IPC: A23F5/08; A47J31/40 **Family:** 37517175

Fremgangsmate for raskere levering av en kort kaffe-ekstrakt fra kapsel

Abstract:

Fremgangsmåte for levering av en kort kaffeekstrakt raskere fra en lukket kapsel som inneholder malt kaffe ved injeksjon av vann under trykk i kapselen.

Kapselen blir fylt med malt kaffe, har en leveringsmembran og blir ekstrahert i en kaffeekstraksjonsanordning og trykksatt vann blir injisert i kapselen under trykk.

Kaffedrikken blir frigjort gjennom drikkeleveringsmembranen til kapselen med gripemidler som griper i og/eller mot membranen.

Trykktapet blir redusert i kaffebedet ved å tilveiebringe malt kaffe med en redusert prosentandel av finpartikler (F), avhengig av gjennomsnittspartikkelstørrelsen (D 4,3).

Ekstraksjonsutbyttet blir bibeholdt mellom 15 og 30%, og kaffeekstrakt på 25 eller 40 gram blir levert i løpet av en strømningstid på 20 sekunder eller mindre.

Description:

EP1882432B1

IPC: A23F5/08; A47J31/40 **Family:** 37517175

Method for delivering faster a short coffee extract from capsule

Abstract:

Method for delivering a short coffee extract faster from a closed capsule containing ground coffee by injecting water under pressure within the capsule.

The capsule is filled with ground coffee, has a delivery membrane and is extracted in a coffee extraction device and pressurized water is injected in the capsule under pressure.

The coffee beverage is released through the beverage delivery membrane of the capsule with engaging means engaging in and/or against the membrane.

The pressure loss is reduced in the coffee bed by providing ground coffee having a reduced percentage of fines (F) depending on the average particle size (D4,3).

The extraction yield is maintained between 15 and 30% and the coffee extract of 25 or 40 grams is delivered in a flow time of 20 seconds or less.

Description:

CONCLUSION

- **Creative Commons Zero, available at europat.net, OPUS, and ELRC-SHARE**
- The EuroPat corpus is expected to be widely useful for training high-quality machine translation systems, particularly for those targeting technical documents such as patents and contracts

THE UNIVERSITY of EDINBURGH

https://europat.net

prompsit

Omniscien TECHNOLOGIES

Co-financed by the European Union
Connecting Europe Facility