

## På objekt-jakt i korpusland

Portugisiske nullobjekt i tekstkorpus

(Objeto nulo no português-  
observações sobre a sua problemática)

Heidi Jansen og Diana Santos

## Hva er nullobjekt?

- «A Maria encontrou \_ no cinema.»  
(Maria møtte \_ på kino)
- Transitiv verbkonstruksjon uten uttrykt objekt  
(underforstått, fra tidl. setn. eller omgivelser)
- «Jeg skjønner» / «Jeg vet» - nullobjekt?
- Kinesisk, japansk, thai, koreansk, quechua,  
baskisk – og latin

## Tidligere forskning

Mye (chomskiansk) lingvistikk og beskrivelse av BP, mange nevner ikke PT eller påstår at det ikke finnes.

- Første (?) Omena i BR i 1978
- Huang 1984 for kinesisk
- Raposo (1986): Beskrivelse av PP. «Regra de predicação» er åpen for pragmatisk påvirkning.
- Cyrino: dr.grad (1994/97) NO i PB er annerledes enn i PP, og en slags ellipse. Mest brukt om ikke-animerte vesen og når antecedenten er en setning.
- Nunes (1996): «Cliticização» skiftet retning XIX årh.
- Luraghi (2004): Språk med nullobjekt har bare fulle (trykksterke) pronomen, ikke klitiske.

## Objektspronomen i portugisisk

- Trykksterke og trykksvake (klitiske)  
«Encontrei a ele» vs. «Encontrei-o»
- Etterstilt (ênclise) eller forstilt (próclise)  
«Encontrei-os» vs. «Não os encontrei»
- I Brasil: kun próclise (i uformelt språk)  
(forenkling): «Eu os encontrei / Encontrei (a) eles / Encontrei» ← Nullobjekt

## Hva er ikke nullobjekt?

- Ellipse: «Eu perdi [a chave] e o João encontrou a chave». (jeg mista [nøkkelen] og João fant nøkkelen)
- Ekkosvar: «-Compraste o livro? –Comprei!» (Kjøpte du boka? Jeg kjøpte!)
- Intransitiv bruk av transitive verb
- Faste uttrykk: «Logo veremos», «quem sabe» (vi får se, hvem vet)
- Topikalisering/topicalização: «Dele, sabe-se pouca coisa» (om ham vet man lite)
- Fokalisering/focalização: «Dele sabe-se pouca coisa».

## Også inkludert:

- Kopulaverb/nullpredikat: «Pois sou\_, meus senhores, sou eu o Bandarra.» (Så jeg er\_, mine herrer, det er jeg som er Bandarra.) A. Garrett
- Prep.objekt: «gostar de» «Estava gostando.» ((he/she) was liking).

Selv om de ikke strengt tatt er nullobjekt grammatisk sett, har jeg resonnet slik at for språkbrukerne vil det oppleves som samme fenomen som nullobjekt.

## Historisk portugisisk

- slutten av 1100-tallet – ca. 1400: português antigo
- ca. 1400 – ca. 1540: português médio
- ca. 1540 – ca. 1700: português clássico
- ca. 1700 – i dag: português moderno
- Oppdagelse av Brasil: år 1500, stor immigrasjon fra Portugal på 15- og 1600-tallet.

## Historisk/diakront sett

- Fra lat. «ille» (han/den der) og bøyninger
- De fleste rom. språk har klitiske pron. på I-
- Siden port. har ênclise som standard, og -I- forsvinner mellom vokaler, har pron. blitt til o/a/os/as. Forsvinner lettere i tale?
- Eller: Obj.pronomen kan ikke stå i próclise når de kommer først i setningen, derfor forsvinner de i Brasil som har det som standard?

## Hva var/er problemet med ON og korpus?

- Uenighet om definisjon
- Hvordan søke på noe som ikke eksisterer? De fleste korpus merker ikke null-partikler. (Burde de?)
- Vanskelig å få analyseverktøy til å se forskjell på nullobjekt og f.eks. ellipse; og enda verre å få språkvitere til å bli enige...

## Løsninger

- Først søke etter verb med objekt, så søke etter disse verbene brukt uten objekt
- Et «typisk transitivt» verb (trazer/levar) med eller uten klitika (mange resultater, mye siling)
- Typiske miljøer for nullobjekt: vanlige transitive verb, brukt uten klitiske pron. (i port. vanskelig å skille direkte og indirekte obj.-pron. i 1. og 2. person)
- I PANTERA: «det» som objekt i norske tekster.

## Materiale

- Korpus i Linguatca: CHAVE (journalistisk), Museu da Pessoa (muntlig)
- PortFalDA (muntlig)
- PANTERA (tospråklig)
- Tycho Brahe (historisk, 1502-1836)
- ISWOC (historisk, ca. 1350-1600)
- Grammatikkbøker (bl.a. Raposo et. al, 2013)

## Tycho Brahe

- <http://www.tycho.iel.unicamp.br/>
- Stort, historisk port. korpus, mange sjangre og perioder (ikke alle er syntaktisk annotert)
- Nullobjekt er annotert!
- Mange NO i tre skuespill (komedier), ca. 1850 og møtereferater fra BR på 1830-tallet
- (NP-ACC \*pro\*)

## ISWOC

- foni.uio.no:3000
- UiO-korpus med eldre tekster (Information Structure and Word Order Change in Germanic and Romance Languages)
- Ikke så stort, men annotert manuelt
- Gammelport.: ca. 1350-1600
- Hist. krøniker, hagiografier, rel. visjon, Décadas
- 1. Verb med obj., 2. Disse verbene uten obj.

## CHAVE

- Fra Linguateca: [www.linguateca.pt/ACDC/](http://www.linguateca.pt/ACDC/)
- To aviser, en fra PT og en fra BR, 1994-95
- Annotert automatisk med PALAVRAS
- «trazer» og «levar»
- “[pos=”PERS” & lema=”eu”] [lema=”trazer”] within s”
- [lema=”trazer(\+.+)\*”], 200 tilfeldige

## Museu da Pessoa

- I Linguateca
- Prosjekt som startet i Brasil, og fortsatt pågår
- Muntlig, transkribert av prosjektet
- 213 intervjuer med vanlige mennesker om deres liv (i Linguateca)
- PT og BR, ca. 1994-2010
- [lema="trazer(\+.+)\*"], 200 tilfeldige res.

## PortFalDA

- Oppr. produsert av Univ. de Lisboa (CLUL), med port.talende fra hele verden.
- Muntlig, transkribert, 70-, 80- og 90-tall
- Blitt tatt opp i Linguateca som PortFalDA (foreløpig bare PT og BR)
- 12 av 50 tekster analysert "manuelt", valgt med mest mulig variasjon
- "para" [temcagr=".\*INF.\*" & func="ICL.\*"]



## PANTERA

- Oversettelseskorpus POR-NOR-POR, alle bøker som er oversatt (ca. 10 s. fra hver)
- Skjønn- og faglitteratur
- 78 tekster, ca. 747.000 enheter (ord/tegn) pr. 1.9.16
- [word="det" & func="obj"] og [lema="det" & func=obj]

## Noen spørsmål:

- I hvilken grad er syntaktiske funksjoner merket, og hvor god er merkingen?
- Hvor mange resultater får man, hvor mye må filtreres, hvor lang tid tar det?
- Hva vet man ikke at man ikke har funnet? («recall», Ball 1994)
- Kan det i visse tilfeller være bedre å lese teksten på 'gamlemåten'?

## Funn i ulike hist. perioder:

- Crónica Geral de Espanha: (1344) 1
- Crónica de D. João I (1443): 7
- Teater (komedier, ca. 1850): 47
- Moderne muntl.korpus: 119+51
- Moderne journ. korpus: 2

## 3 slags resultater:

- Kvantitative: nullobjekt er vanligst i uformelt og muntlig språk i både PT og BR, men det finnes også i eldre språk. Veldig sjangerrelatert.
- Lingvistiske beskrivelser
  - «Se» (man) tar ikke klitisk DO-pronomen: «Não se \_ faz» - Man gjør (det) ikke. (evt. «Não se faz isso/Isso não se faz»)
  - «Para» (for) + pers.inf.: «para tu \_ fazeres» – for at du skal gjøre det. Ikke obl., men vanlig med ON.
- Ca. 250+ eksempler for (evt.) fremtidig forskning

## Å finne ting som ikke finnes

- Er fravær synliggjort ved korpusnotering?
- Er det tydelig i kontekster hvor det er lett å se?
- Kan det lett finnes kontrastivt?
- Må man undersøke alle (en undergruppe av) tilfeller?

Egentlig alle : det er ønskelig å bruke alle disse metodene for å løse mysteriet...

## Metodologi

- Hvor mange er nok?
- Hvorfor nullsubjekt når man ofte bruker subjekt? Fordi det er grammatisk mulig.
- Når er det at man sier *mesóclise* er ute?  
– *Dar-lho-ei, fá-lo-ias, encontrar-nos-emos*
- Hvor viktig er grammatikalitet når frekvensen er veldig lav?

## Omstridte meninger?

- Referansegrammatikk(er)?
- Noen problemer mer interessante enn andre?
- Alt henger sammen
- Forskjellige mål når det gjelder PB

## Eksempler på alle disse tilnærmingene

- Pro- og PRO-markering
- Relative pronouns in English  
Ball, Catherine N. 1994. Automated text analysis: Cautionary tales. *Literary and Linguistic Computing* 9, 295-302.
- Possessive datives in Portuguese using parallel corpora  
Santos, Diana. 2015 Os possessivos estão-me a complicar o ensino ;-)  
Um estudo do dativo possessivo baseado em corpos. *Linguística : Revista de Estudos Linguísticos da Universidade do Porto* 10, 107-130.
- Alle seriøse korpuslingvister 😊