

Variation in the pronominal ditransitive in British English Twitter messages

UKLVC 12

Queen Mary, University of London

Jon Stevenson

University of York

3rd September 2019

Outline

Overview of presentation

- ▶ The ditransitive in British English
 - ▶ Introduction to the structure
 - ▶ Geographical/Historical distribution
 - ▶ As feature of speech, data problem
- ▶ Twitter as solution
- ▶ Methodology
- ▶ Results
- ▶ Brief discussion
- ▶ Research directions

The ditransitive in British English

Verbs that encode three arguments, and can alternate: 'dative alternation'

Alternating verbs

e.g. *send, give, show*

GOAL-THEME ditransitive (GTD)

"John gave Mary the book."

Prepositional dative - (PDAT)

"John gave the book to Mary"

THEME-GOAL (TGD)

*?? "John gave the book Mary"

The pronominal ditransitive in British English

Focus on ditransitive with pronominal objects (pDit)

GOAL-THEME ditransitive (GTD)

“what was in that envelope when Malik **gave him it.**”

Prepositional dative - (PDAT)

“as if they **sent it to them** for free”

THEME-GOAL ditransitive (TGD)

“i **gave it him** and it had already melted”

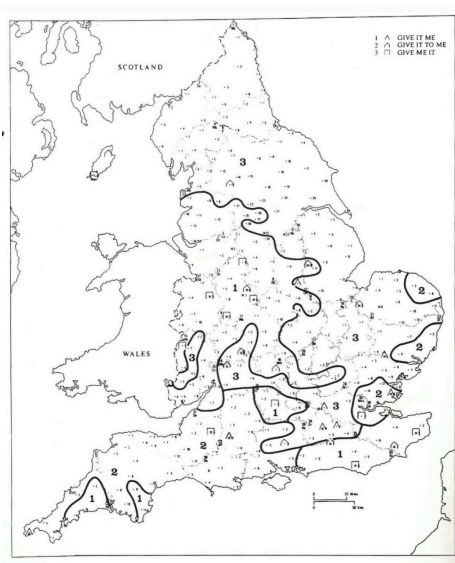
The ditransitive in British English

Regional variation

*“No better example exists of a syntactic puzzle than the quite definite regional preferences for the standard **give me it** in northern and eastern England, a non-standard **give it me** in the West Midlands, and an expanded **give it to me** in the south-west, as recorded by SED” (Upton, 2006, p.329)*

The ditransitive in British English

The Survey of English Dialects (SED)



from: Kirk (1985)

Changing use over time

Ditransitive with both objects as pronouns - cold water on GTD as Norse origin?

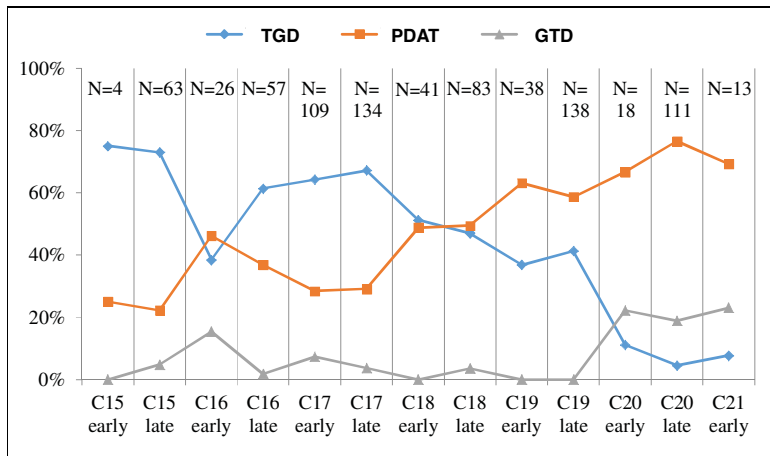


Figure: Figure from Yáñez-Bouza and Denison (2015)

pDit as a feature of speech - data problem

Explanation and implication

*“Conversational participants **share time and place, and they normally also share extensive personal background knowledge.** As a result, colloquial features like **pronouns and vague expressions are common.**” (Biber, Gray, & Staples, 2016, p.1).*

Implication for corpus study of pDit

- ▶ Rarity of pDits in written English means finding sufficient examples in spoken corpora may require prohibitively large datasets (Siewierska & Hollmann, 2007)
- ▶ Compounded by fact that syntactic features are already infrequent (compared to phonological features)
- ▶ More so when focusing on smaller areas
- ▶ Status of pronouns deemed essentially out of reach

Twitter as solution to data-problem

As 'Everyday' language, speech-like

Natural language corpus

- ▶ Spontaneous and unmonitored data
- ▶ A lot of it...!
- ▶ Location metadata allows messages to be mapped
- ▶ Includes person-to-person, vernacular written interaction conducted on mobile devices

Self-transcription

- ▶ Sense that users are 'self-transcribing', acutely aware of how they represent themselves
- ▶ Similar cross-platform with messages sent via mobile. Nothing special about Twitter other than ease of access to data

Twitter for dialectology

Kinds of Twitter message

Public directed tweets

- ▶ Tweets that are public facing: one-to-many
- ▶ Perhaps more deliberate / conscious

Conversation threads

- ▶ Followers of a user often respond to a tweet
- ▶ Series of messages will ensue: one-to-one/few
- ▶ It is in the conversation threads that we see most spontaneous data

Conversation threads

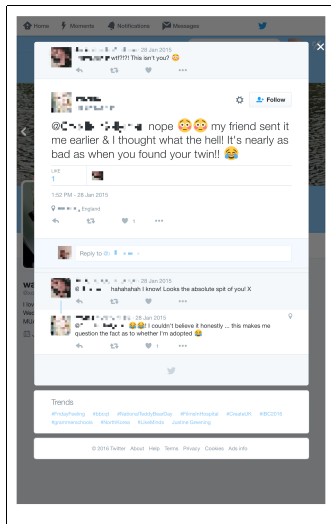


Figure: Example conversation thread from the dataset

Sample population

Twitter not as widespread as other social media

- ▶ Facebook now 2.3 billion users, Twitter has had stable 300 million since 2015¹
- ▶ 23% US adults use Twitter compared to 72% Facebook (Duggan, 2015)
- ▶ Twitter has young user-base: over 60% aged 18-34²

API

- ▶ 'Firehose': 100% data (paid-for)
- ▶ 'Stream' API: 1% of data (requires always-on connection)
- ▶ 'Search' API: also $\approx 1\%$ (bias to users with high follower-count)

¹<http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

²<http://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>

TAGS, API and a Note of caution

Be aware of sudden changes in API functionality!

Using 'TAGS': `tags.hawksey.info`

- ▶ TAGS runs in Google sheets
- ▶ Update every hour, runs indefinitely

Change in November 2014 - first dataset

- ▶ Fuzzy match no longer implemented. Only GPS encoded tweets.
- ▶ Took 15 months to gather 1500 tweets

Change again in 2016 - second dataset

- ▶ Now returns many more results, user inputted location much more readily available. Able to use well over 60% of data

TAGS interface

TAGS v6.0

NS - New Sheets

Created by mhawksey. Read more at <http://tags.hawksey.info>

readsheet you can:

ly pull results from a Twitter Search into a Google Spreadsheet

:

no TAGS menu click this button -->

ever run TAGS > Setup Twitter Access do so now (this should only need

"sent it me" OR "sent it
you" OR "sent it him"
OR "sent it her" OR
"sent it them" OR "sent
it us" OR "gave it me"
OR "gave it you" OR
"gave it him" OR "gave
it her" OR "gave it them"
OR "gave it us"

<- you can use serach operators
from:BarackObama' (without qu

ee off collection with TAGS > Run now! or set a trigger to collect every ho
en Tools -> Script Editor then Triggers -> Current script's triggers... and

Figure: <https://tags.hawksey.info/get-tags/>

Map generated from TAGS data

First dataset (GPS only)

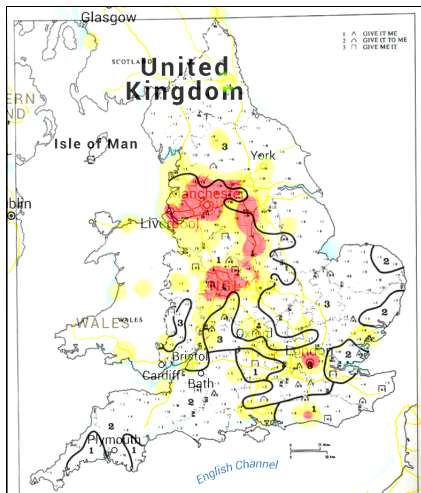


Figure: Heat map displaying TGD tweets generated using Google Fusion tables overlaid onto SED map (Kirk, 1985)

Super-regions

First dataset (GPS only): Comparing region to region by Chi-Square



Figure: 'Super-regions'

Super-regions

Cover wide, dialectally diverse regions

The larger regions contain, many different varieties
(Manchester/Sheffield/Birmingham all pattern together)

Regionally variant pragmatics

Where there is, in a given region, an apparent stable variation between variants, each variant likely carry pragmatic difference

Suggestion of boundaries/'faultlines'/'transition zones'

Linguistic rather than political boundaries that are indicative of historical and ongoing contact processes

GTD in East from Old Norse Gast (2007)

Proposal would fit the pattern found here

Second dataset

Second dataset (user-entered location)

Same methodology

But Twitter changed how it provides geographical data, allowing user-entered data resulting in more data

Size of Dataset three

Results in $\approx 60,000$ 'hits', with $\approx 35,000$ with usable, user-disclosed location data (after cleaning for duplicates/false positives etc.)

Interactive map generated using BatchGeo

Second dataset (user-entered location): confirm early findings, add detail by place

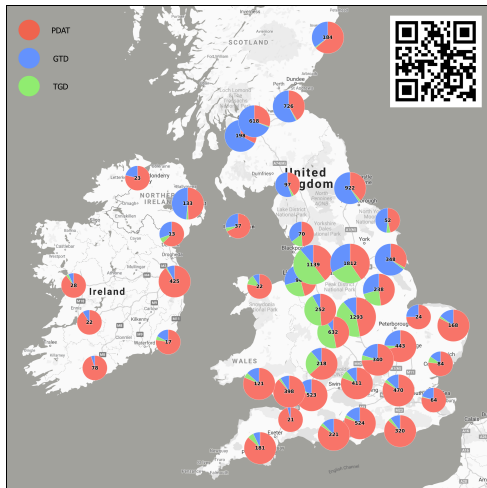


Figure: Generated in BatchGeo (paid), but possible to create similar maps using 'leaflet.minicharts' in R

Use by location

Second dataset (user-entered location) Focus on North-West and border areas

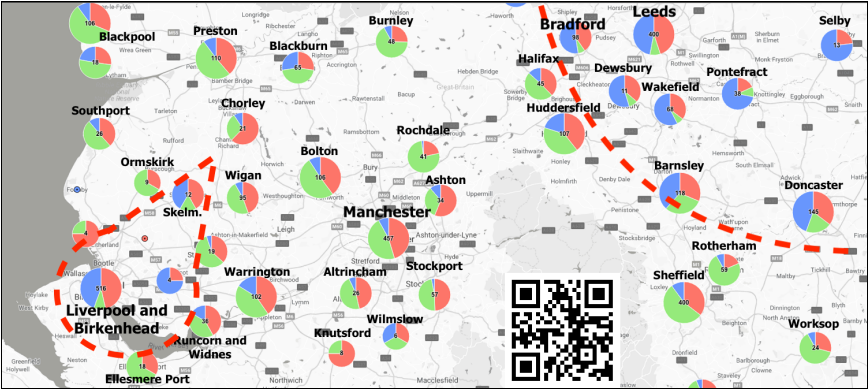


Figure: Detail of Northern England. Interactive version at: <https://batchgeo.com/map/7c71333a4373358465bf9fe7e71687c6>

Use by location

Second dataset (user-entered location): Towns organised from north to south by latitude



Figure: Variation by town ordered North to South

Variation by pronoun

Second dataset (user-entered location): Variation by goal pronoun, places with > 500 hits



Conclusions

- ▶ Language use on Twitter matches data gathered 50 years ago using traditional methods, showing persistence over time
- ▶ Very robust patterns indicative of longstanding natural language use (not transient 'netspeak' etc.)
- ▶ Demonstrates the validity of using Twitter for this kind of data gathering
- ▶ Reveals nuanced town-by-town variation with high level of detail, enough to reveal role of pronouns
- ▶ Again, patterns do lend support to Norse influence GTD, and arguably warrants further investigation
- ▶ However, recent mass-migrations to cities during industrial era and mixing of dialects must play a (large) role

Directions

Questions

- ▶ To what extent do acceptability judgments and predictive capacities of speakers align with frequency distributions?
- ▶ What geo-historical inferences can be drawn?
- ▶ Where there is apparent stable variation between variants, what pragmatic differences may be inferred?
- ▶ What evidence can be provided to support a 'single abstract source' for each alternation?
- ▶ Where different areas have a different 'abstract source' (e.g. Manchester/Liverpool), if this is perceptible to speakers/hearers (c.f. interface principle). Also, MacKenzie (2019) on 'covert representational variability'.

References I

- Biber, D., Gray, B., & Staples, S. (2016, January). Contrasting the Grammatical Complexities of Conversation and Academic Writing: Implications for EAP Writing Development and Teaching. *Language in Focus*, 2, 1-18. doi: 10.1515/lifjsal-2016-0001
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). MIT Press.
- Duggan, M. (2015). Mobile messaging and social media 2015. *Pew Research Center*.
- Gast, V. (2007). I gave it him – on the motivation of the 'alternative double object construction' in varieties of British English. *Functions of Language (special issue:Ditransitivity)*, 14(1), 31-56.
- Gerwin, J. (2013). Give it me!: Pronominal ditransitives in English dialects. *English Language and Linguistics*, 17, 445-463.
- Kirk, J. M. (1985). Linguistic Atlases and Grammar: The investigation and description of regional variation in English syntax. In J. M. Kirk & S. Sanderson (Eds.), *Studies in linguistic geography : The dialects of English in Britain and Ireland*. London ; Dover, N.H.: Croom Helm.
- MacKenzie, L. (2019, February). Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: a journal of general linguistics*, 4(1), 28. doi: 10.5334/gjgl.622
- Siewierska, A., & Hollmann, W. (2007). Ditransitive clauses in English with special reference to Lancashire dialect. *Structural-functional studies in English grammar. Amsterdam and Philadelphia: John Benjamins*, 83-102.
- Upton, C. (2006). Modern Regional English in the British Isles. In L. Mugglestone (Ed.), *The Oxford History of English*. Oxford: Oxford University Press.
- Yáñez-Bouza, N., & Denison, D. (2015). Which comes first in the double object construction? *English Language and Linguistics*, 19, 247-268. doi: 10.1017/s136067431500012x