

ANNOTATION AND COMPARATIVE ANALYSIS OF USAGE GUIDES: A BOTTOM-UP APPROACH

VIKTORIJA KOSTADINOVA, MARTEN VAN DER MEULEN, FOLGERT KARSDORP

LEIDEN UNIVERSITY, MEERTENS INSTITUTE AND RADBOD UNIVERSITY NIJMEGEN

21 JANUARY 2016, LEIDEN

WHAT

- explore data
- to be able to in the future:
 - study the characteristics of twentieth-century prescriptivism in language advice literature
 - analyse prescriptive (meta)language in the genre
 - a) how attitudes to usage are presented in these books
 - b) how these books incorporate facts of use and sociolinguistic considerations
 - c) provide a framework of analysis for comparison and study of prescriptivism over time and across contexts

WHY

- study newer forms of prescriptivism in language advice literature
- study the development of prescriptivism in the course of the twentieth century
- study aspects of prescriptivism in different contexts (American English and Dutch)
- a better understanding of prescriptivism and its relationship to temporal and historical context may shed further light on whether and how it affects language usage (and language users)

HOW

- data collected from language advice literature
 - comments on four usage features for each language
- developed annotation framework building on previous studies
 - texts annotated using BRAT rapid annotation tool
- used statistical models to explore associations between the various levels of annotations
- qualitative analysis of the same feature, double negation, in English and in Dutch data

DATA

	features	number of entries	total number of tokens	average number of tokens	
American English 1847 - 2014	ain't	35	122	27955	249.6
	double negation	22			
	literally	23			
	split infinitive	31			
Dutch 1897 - 2015	kost duur	35	99	8318	84.02
	omdat	28			
	werkloos	18			
	ontkenning	18			

ANNOTATION SCHEMA

- Treatment approving | disapproving | hedged
 - “It’s OK to split infinitives... notions change over time, and today almost everyone agrees that it is OK to split infinitives, especially when you would have to change the meaning of the sentence or go through writing gymnastics to avoid the split.” (Fogarty 2008)
 - “*Literal* means “true to the exact meaning of the words,” but many writers use it misguidedly ... *literal* is overused both as a disclaimer and an enhancer; it should be reserved to mean “without exaggeration”.” (Sutcliffe 1994)
 - “... the repetition of the negative by alternative means in close proximity helps to underscore the force and/or defiance of the utterance. Though negative concord is socially stigmatized in both American and British English, it has a long history of use and survives in casual conversation.” (Peters 2004)
- inter-annotator agreement on treatment (**Cohen’s Kappa**)

(Weiner 1988, Algeo 1991, Peters and Young 1997, Albakry 2007, Tieken-Boon van Ostade 2015, van der Meulen & Otto, ms,)

ANNOTATION SCHEMA

- **Treatment** approving | disapproving | hedged
- **Attitude** positive | negative
 - **Opinion** direct | reported
- **Argument** analogy | logic | history | authority | foreign influence | clarity | purity | usage | aesthetics
- **Sociolinguistic consideration** frequency | people | variety | register | mode | value
- **Examples** attributed | unattributed
- **Intertextual secondary sources** | general | media

(Weiner 1988, Algeo 1991, Peters and Young 1997, Albakry 2007, Tiekens-Boon van Ostade 2015, van der Meulen & Otto, ms,)

ANNOTATION EXAMPLE



1 **Hedged** **DOUBLE NEGATIVE** Most kinds of double negative are **Negative** inappropriate in **Mode** spoken and written **Variety** Standard English, except in jocular use: **Adhoc** Don't never say that again. **Adhoc** I can't do nothing about it.

People Eighteenth-century grammarians decided that **Logic** since two negatives made a positive in mathematics and logic, they must do so in spoken and written English too. **History** This was not always so, however, and the double negative remains one of the best illustrations of what **History** was once a perfectly acceptable locution being driven by the decisions of grammarians, not out of the language, but out of Standard use. **Variety** **Authority** Chaucer used double and even triple negatives, and so did **Authority** Shakespeare: these were **Positive** simply powerful, heavily stressed, multiple negatives.

People **Frequency** And many speakers still use these constructions today, even though they are now **Negative** shibboleths that **Negative** **Value** mark speakers of Vulgar English. Can't hardly is usually also classed as a double negative.

Adhoc You can't hardly expect her to be grateful, when analyzed, is doubly negative, in that You can't expect her to be grateful is renegated by the overlay of You can hardly expect her to be grateful. Other adverbs of a negative quality, such as scarcely, are also considered double negatives when used with a negative auxiliary such as can't or cannot. Another kind of double negative

Adhoc occurs mostly in relatively Formal writing and at the higher levels of speech: **Adhoc** Let me give you my address now, in the not unlikely event that the train is late.

Adhoc I was not uncomfortable on the terrace. The not un- construction can be **Value** the mark of the careful qualifier, **People** the thoughtful speaker or writer trying to achieve an accuracy that may be, in fact, not inappropriate, but is, not improbably, hard to achieve; after all, if an act is described as not unwitting, it must be deliberate. Avoid pussyfooting unless you really must be that precise.

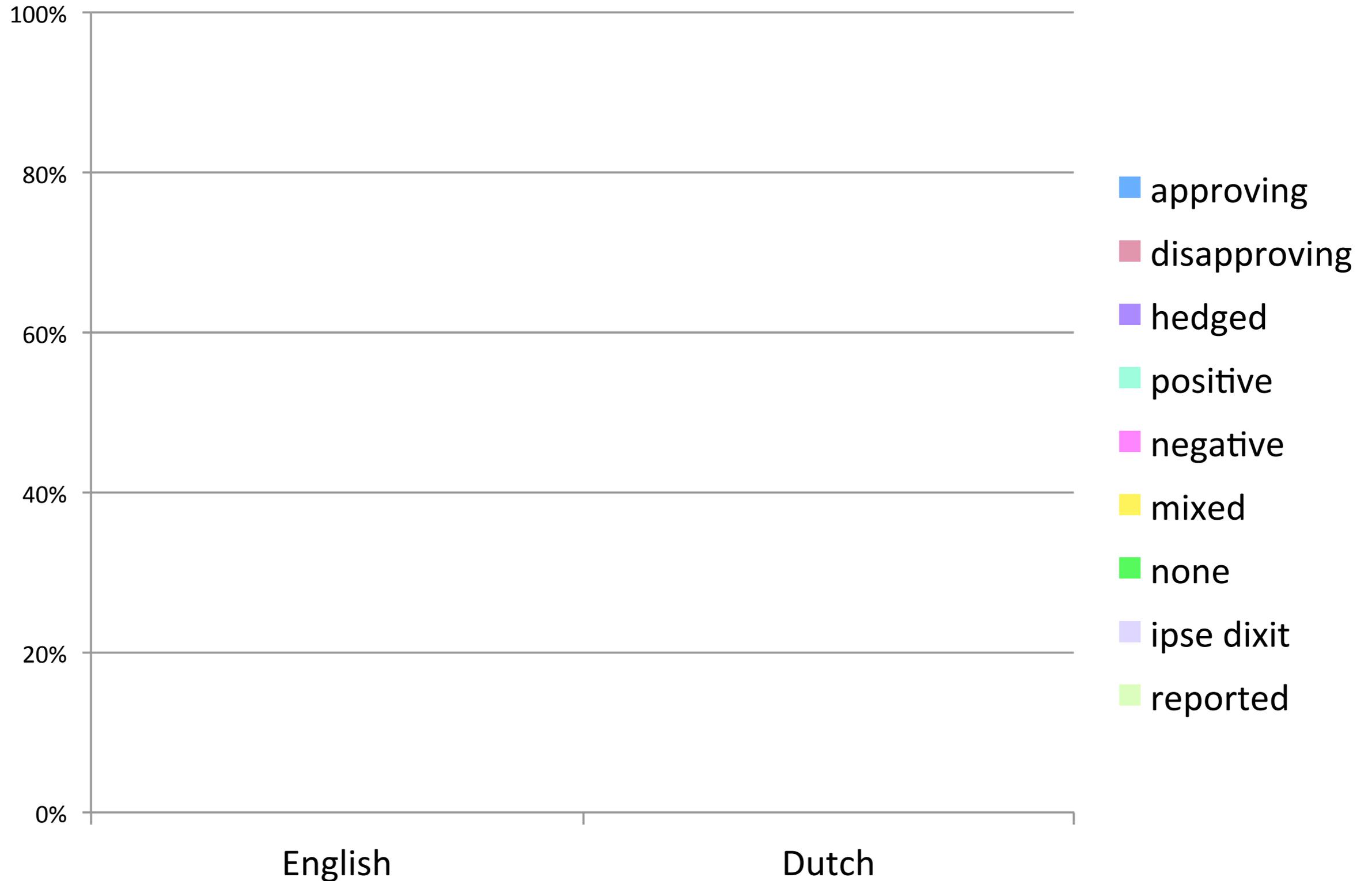
Reported Someone, somewhere, will dislike this sort of double negative because it is **Negative** needlessly opaque and **Attitude** sounds stuffy. **Negative** If it obtrudes, even when you think you are being amusing, it will not amuse. See LITOTES.

➤ number of annotations: 1389 (en) + 665 (nl) = 2054

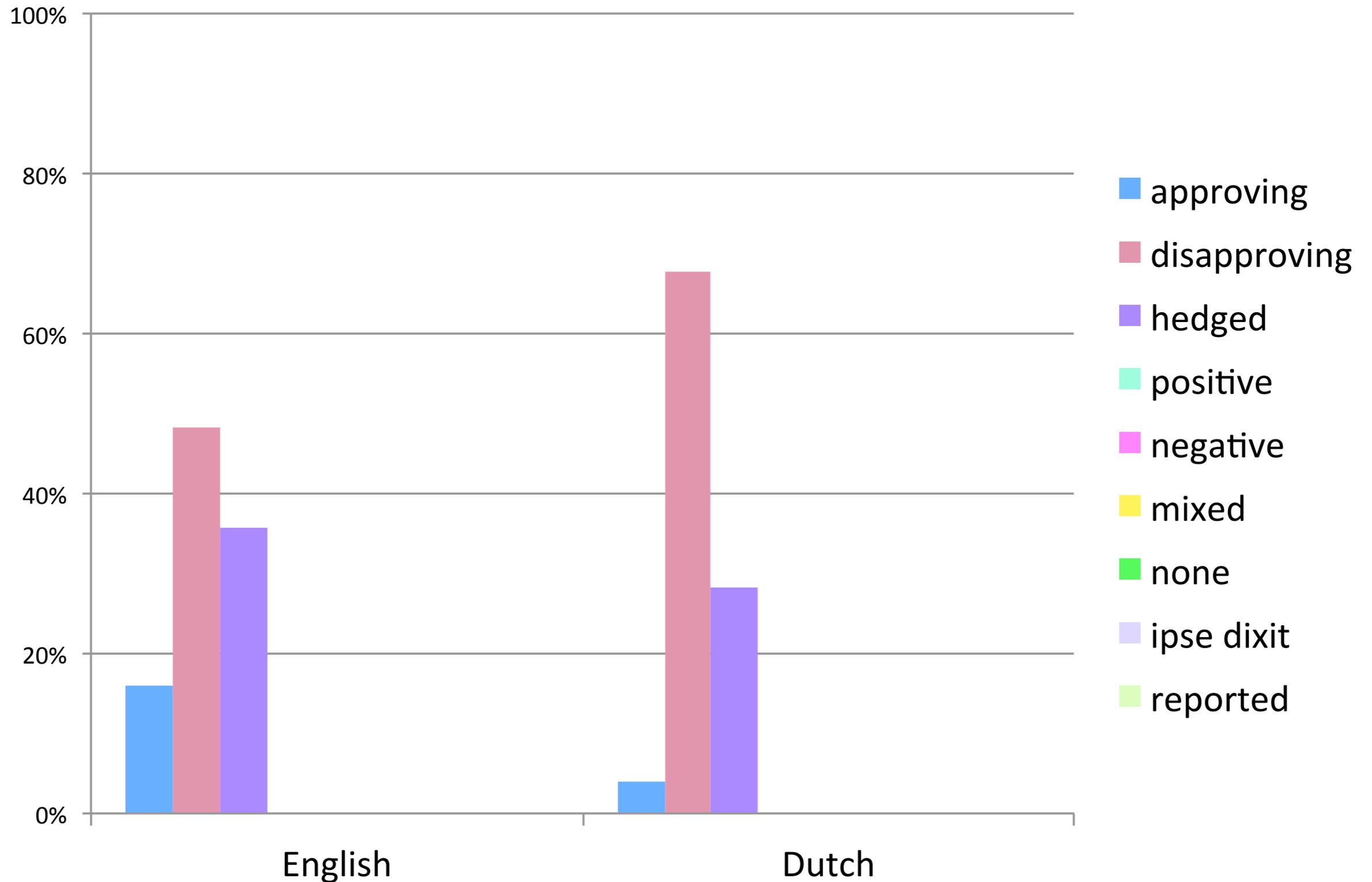
NOTES ON ANALYSIS

- entries were analyzed using BRAT rapid annotation tool
- inter-annotator agreement on **treatment (Cohen's Kappa)**
- *data was analyzed with the vcd library in R, for categorical data
- annotations can be used for qualitative (comparative) analysis

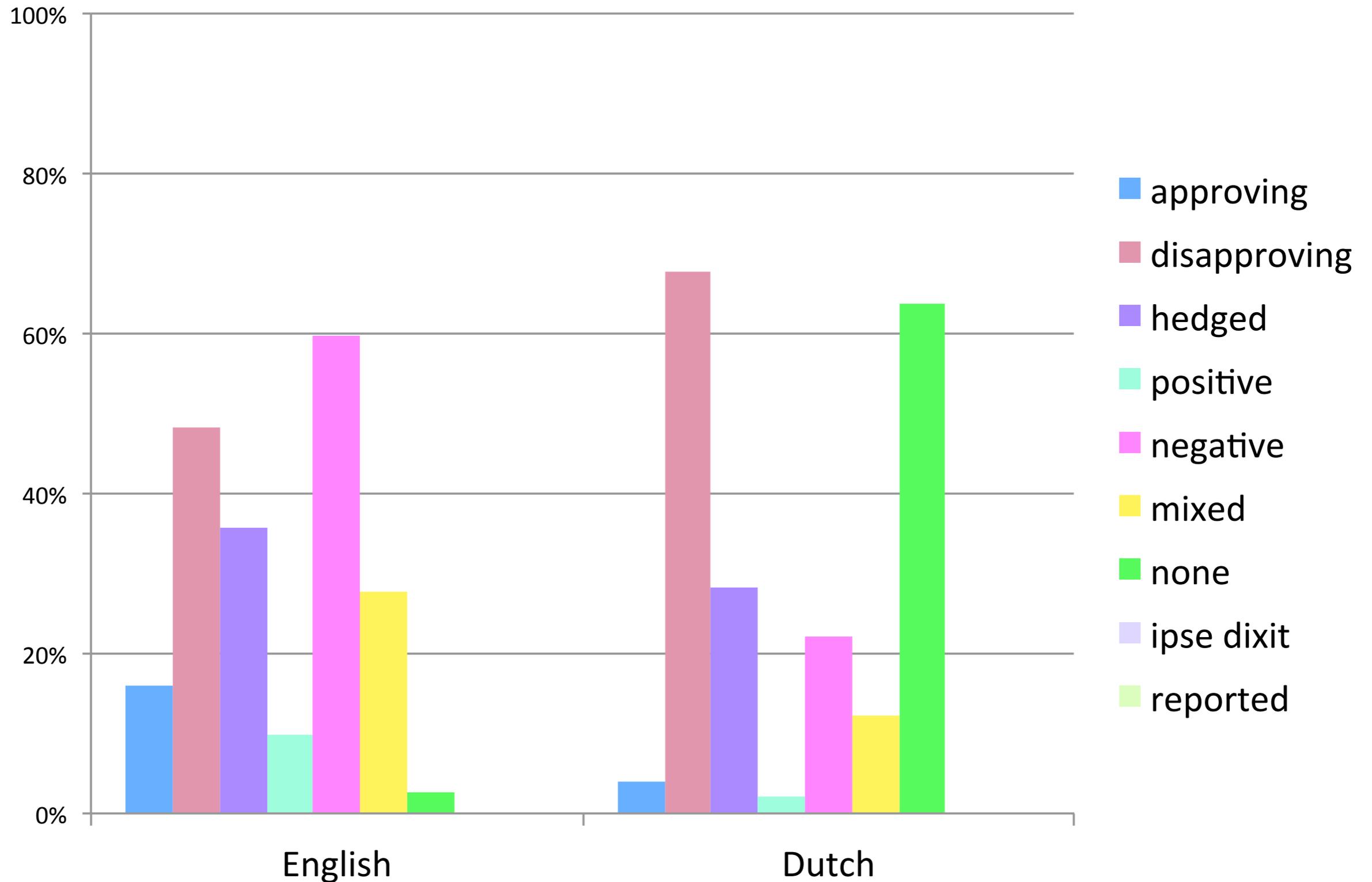
TREATMENT, ATTITUDES AND OPINIONS



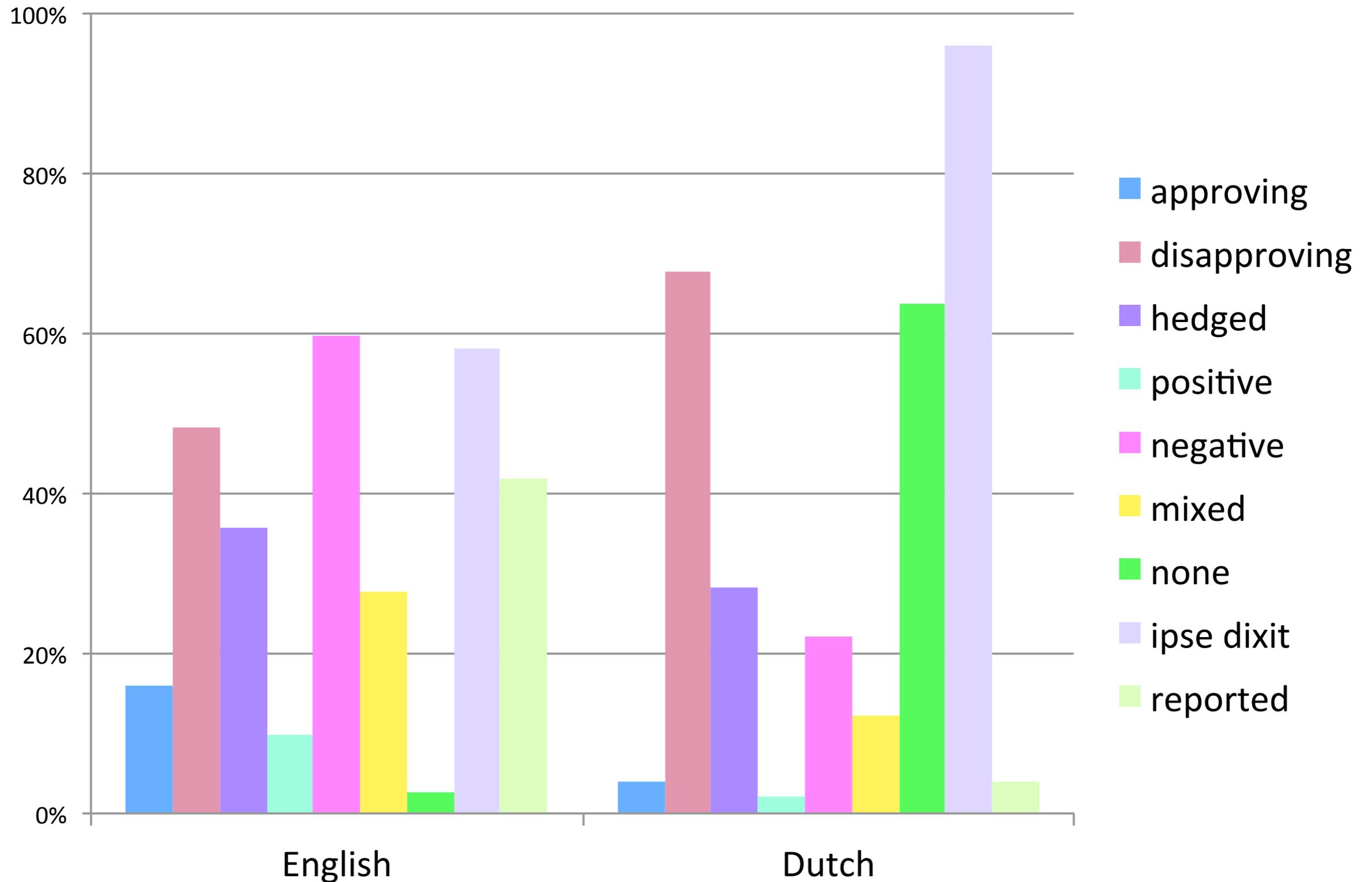
TREATMENT, ATTITUDES AND OPINIONS



TREATMENT, ATTITUDES AND OPINIONS



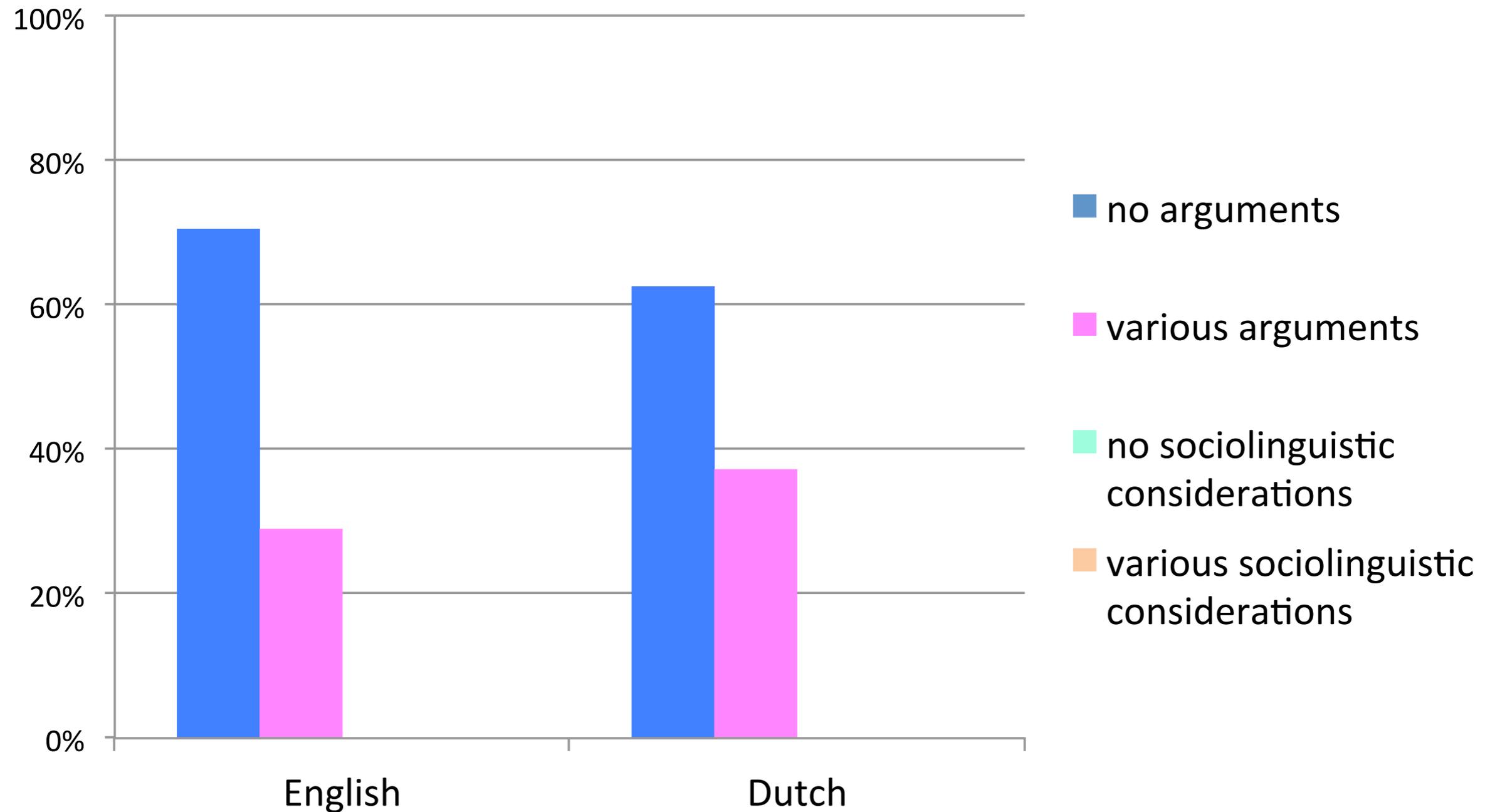
TREATMENT, ATTITUDES AND OPINIONS



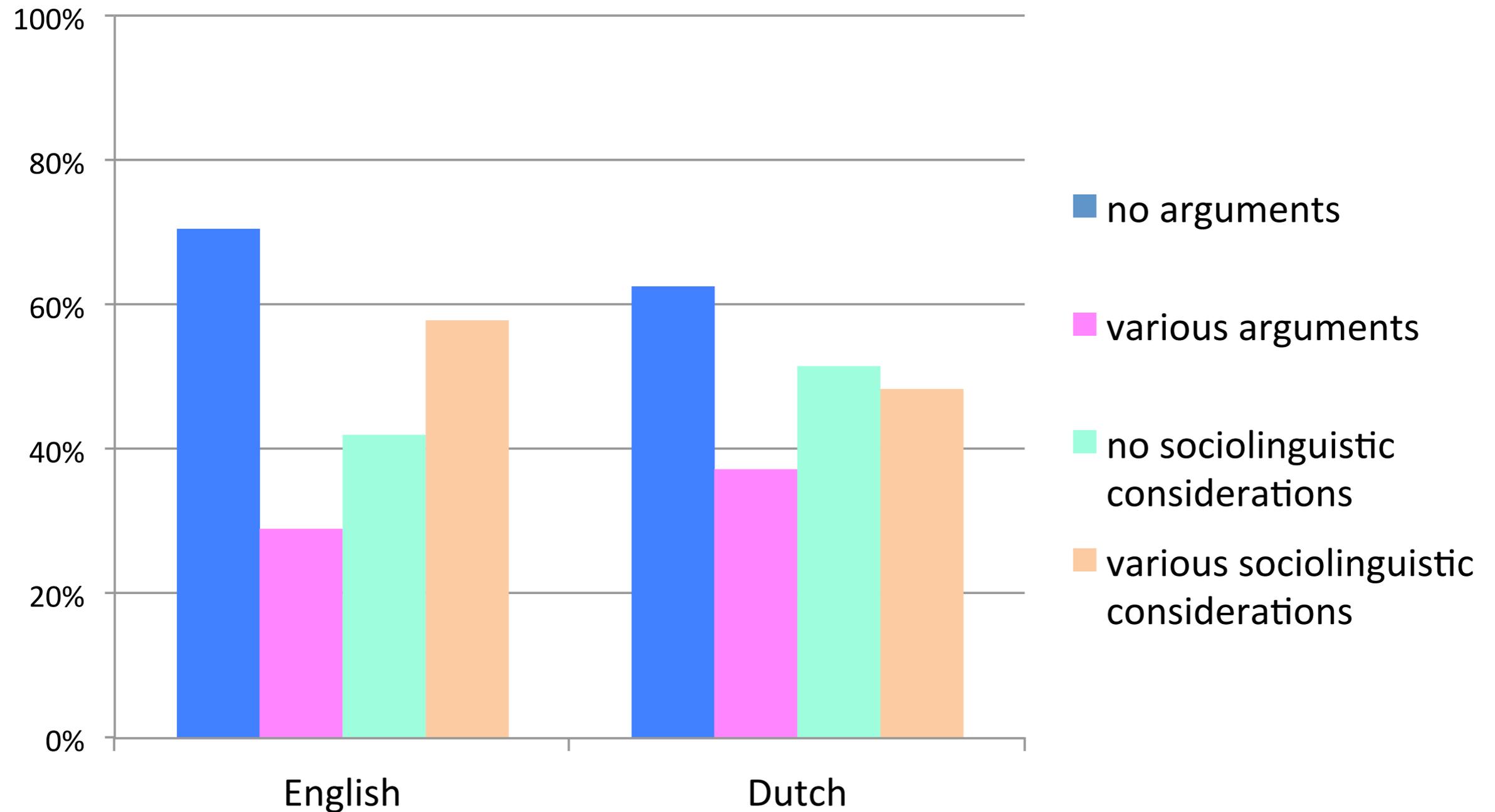
ARGUMENTS AND SOCIOLINGUISTICS CONSIDERATIONS



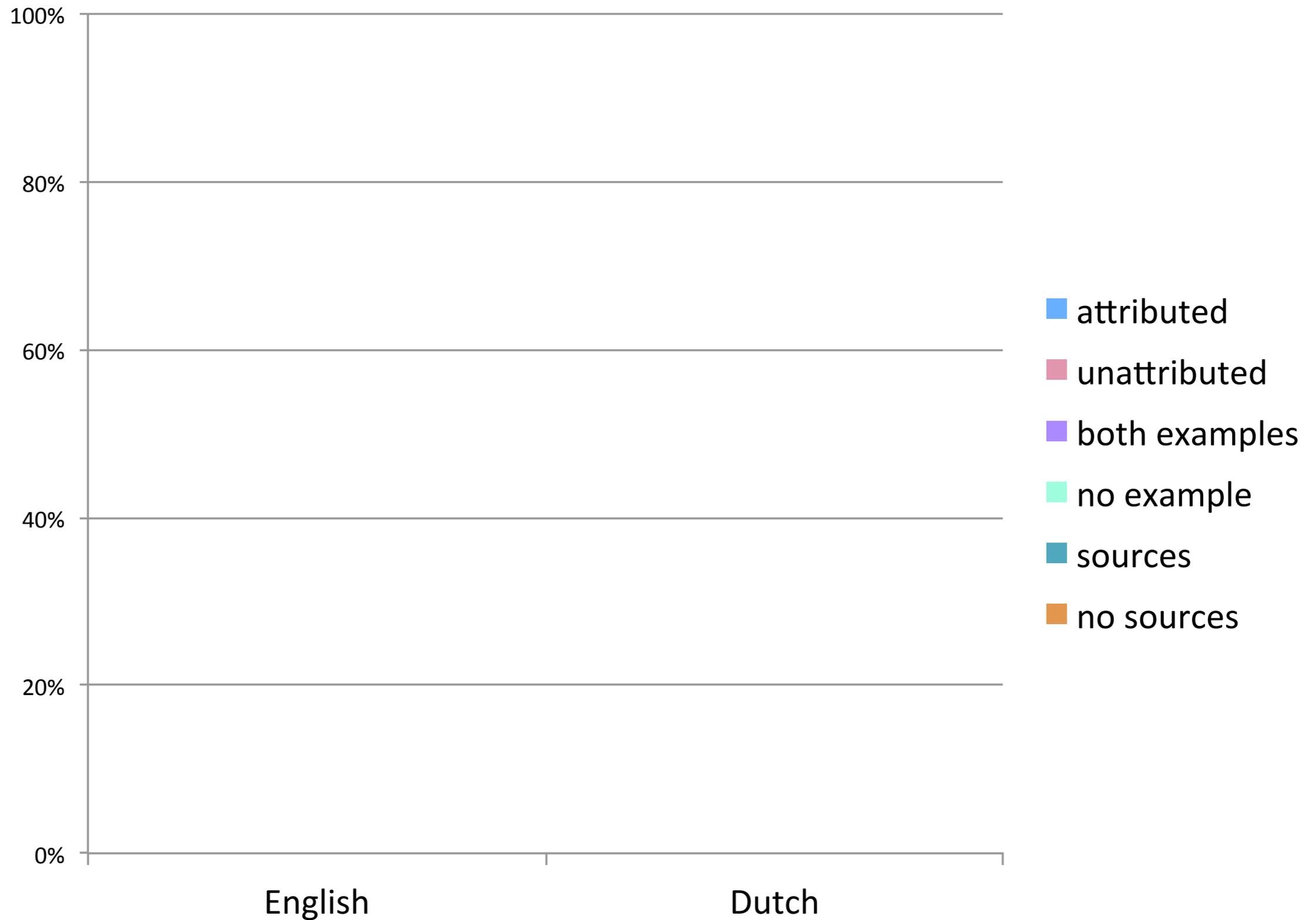
ARGUMENTS AND SOCIOLINGUISTICS CONSIDERATIONS



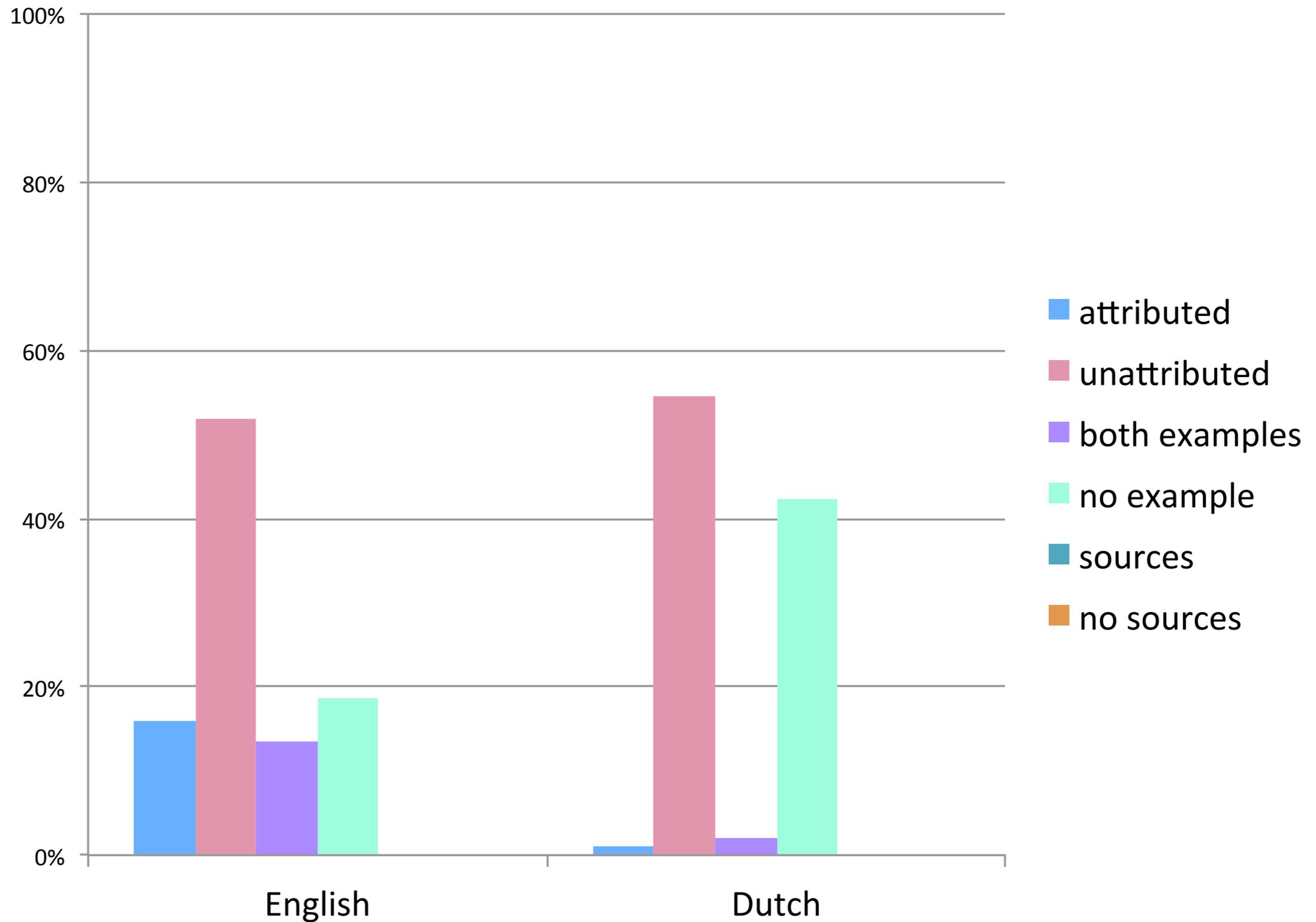
ARGUMENTS AND SOCIOLINGUISTICS CONSIDERATIONS



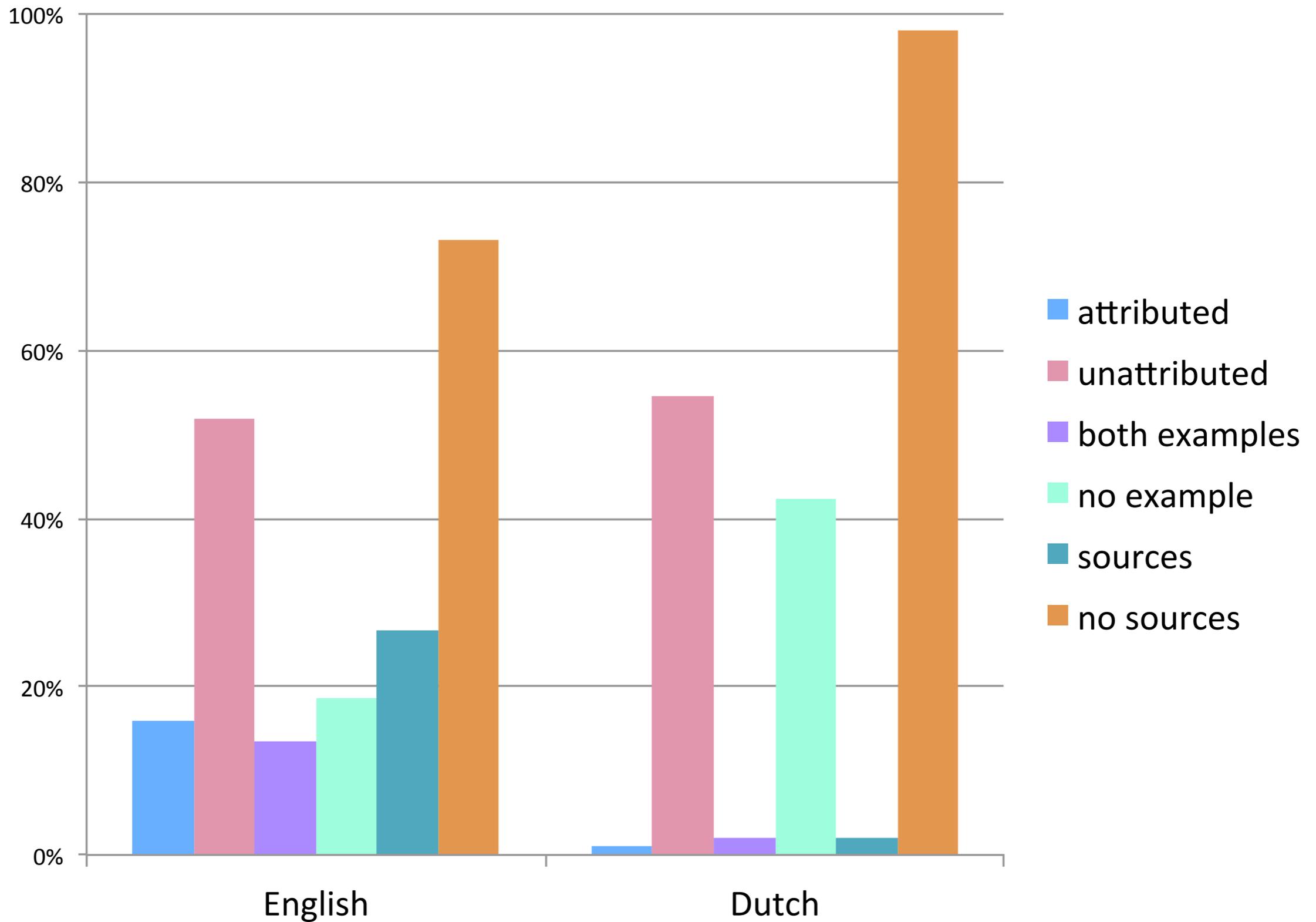
SOURCES AND EXAMPLES



SOURCES AND EXAMPLES



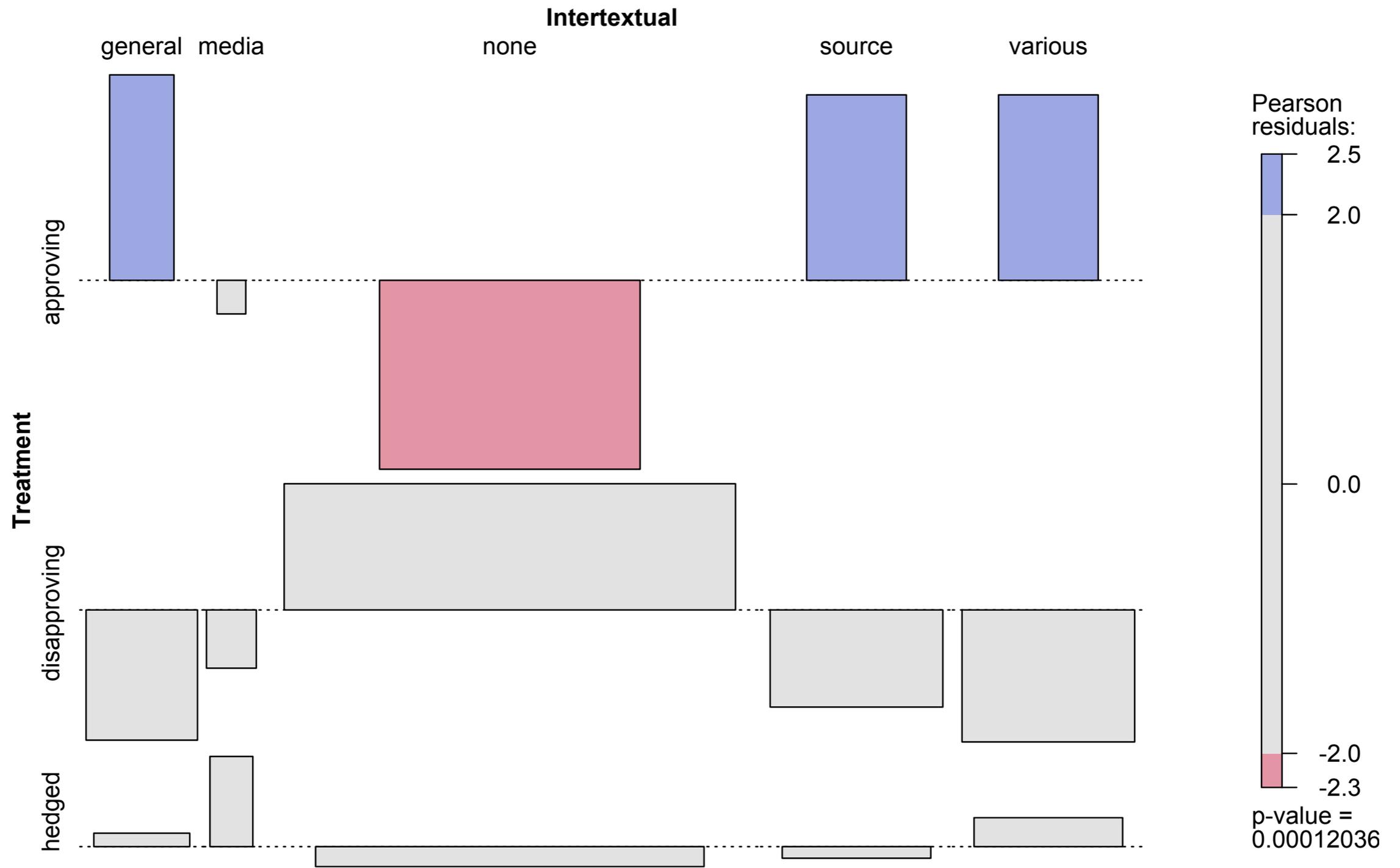
SOURCES AND EXAMPLES



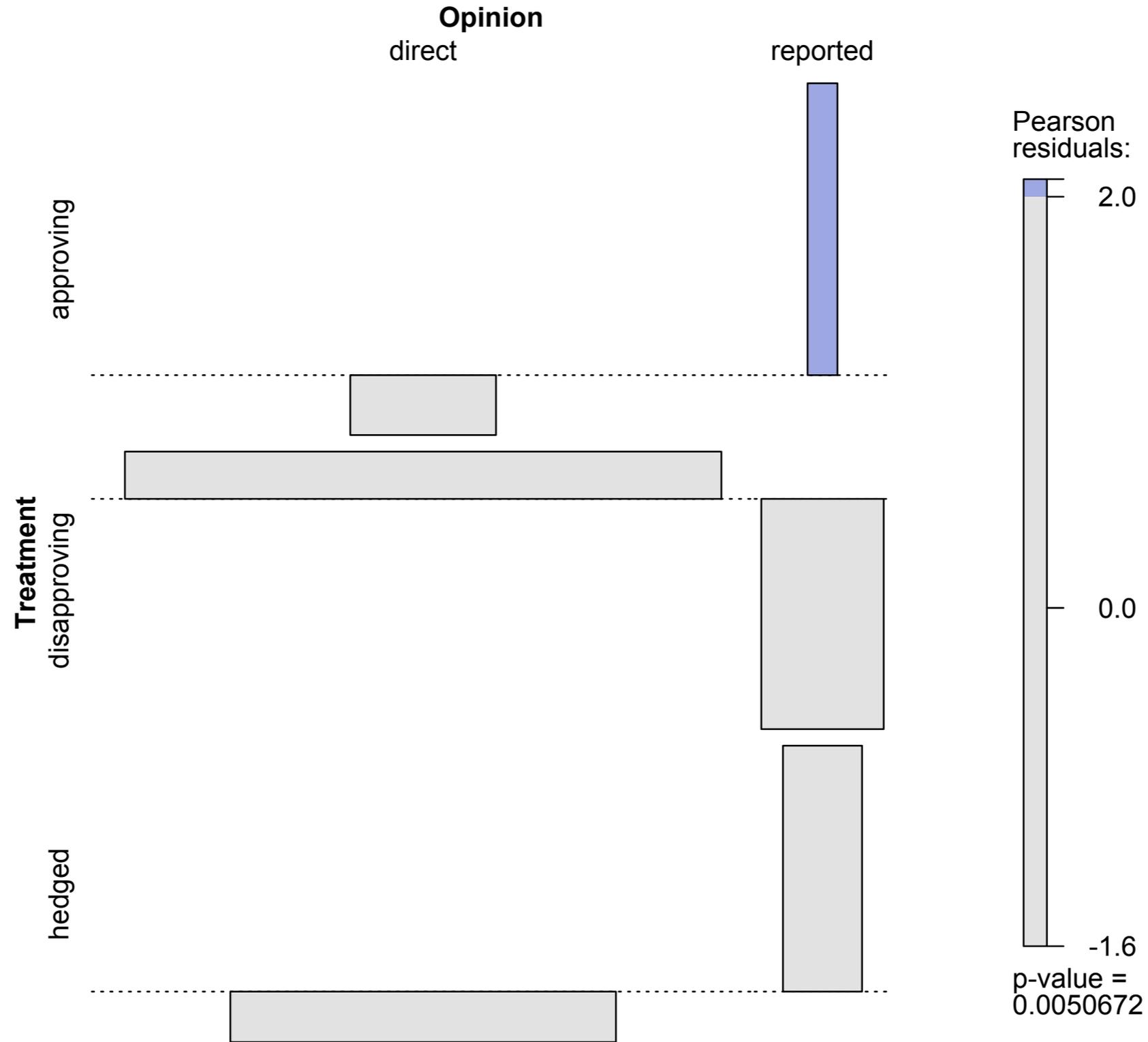
INTERIM SUMMARY

- disapproving treatment is more common in both sets of data; English data more evenly distributed
- English data contain more negative attitudes, while in the Dutch set there are no overtly expressed attitudes
- Ipse dixit statements are more characteristic of Dutch entries, although they still predominate in the English entries
- Around half of the entries both in English and in Dutch contain arguments and sociolinguistic considerations
- Unattributed examples seem to be prevalent, especially in Dutch
- Very often the treatment is based on no sources at all, which comes to about 96% of the Dutch entries

TREATMENT AND INTERTEXTUAL REFERENCES - ENGLISH DATA



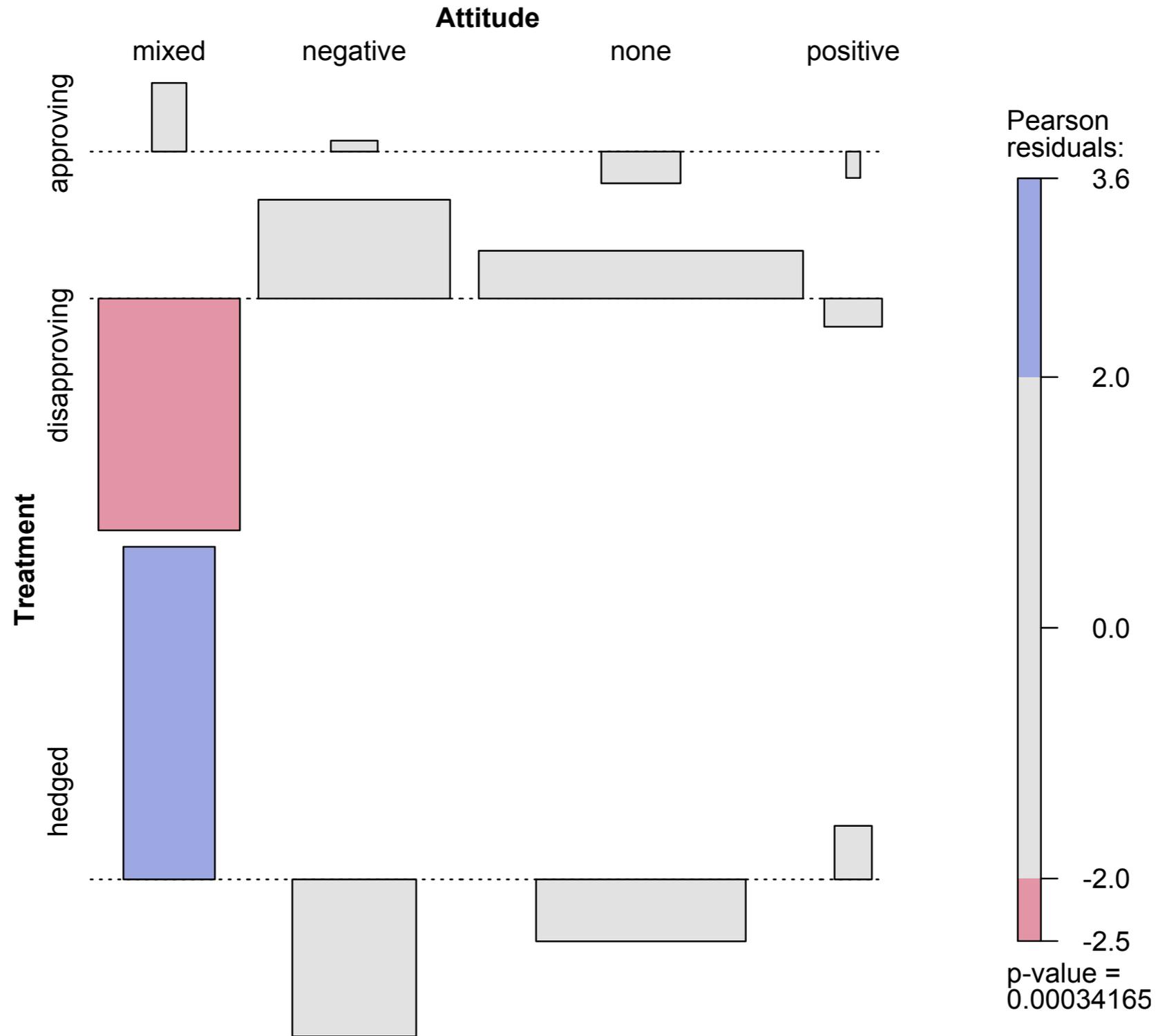
TREATMENT AND OPINION – ENGLISH DATA



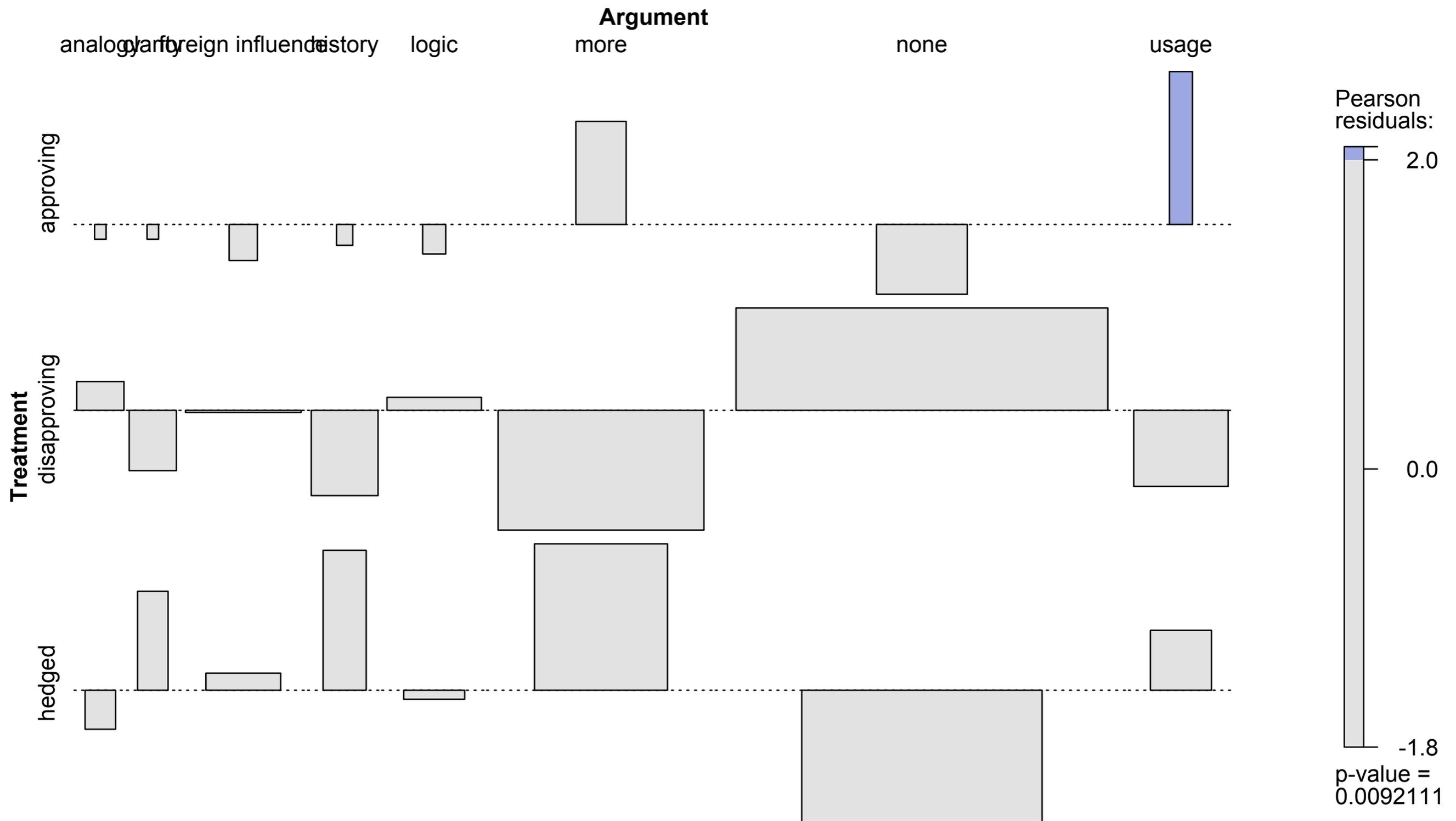
ENGLISH DATA SUMMARY

- approving entries make use of various secondary sources ($p=0.00012$), and tend to report opinions ($p=0.005$) instead of using ipse dixit pronouncements
- entries that use various arguments also use various sociolinguistic considerations ($p=0.00018$)
- entries that use various secondary sources also tend to use more attributed examples ($p=0.00095$)
- entries with reported opinions also contain various sociolinguistic references ($p=0.00085$)
- the use of various arguments correlates with the use of attributed examples ($p=0.0017$)

TREATMENT AND ATTITUDE - DUTCH DATA



TREATMENT AND ARGUMENT – DUTCH DATA



DUTCH DATA SUMMARY

- hedged entries present mixed attitudes, while disapproving entries tend to NOT present mixed attitudes ($p=0.0003$)
- approving entries report attitudes rather than expressing them directly - similar to English data ($p=0.005$)
- approving entries make use of the usage argument more than other entries ($p=0.009$)
- entries with mixed attitudes also tend to report their value judgements, rather than express them directly ($p=0.0009$)
- entries with mixed attitudes also refer more to the frequency of use of the feature and register ($p=0.007$)

DOUBLE NEGATION: ENGLISH AND DUTCH

- Negative attitudes more overtly expressed in English
 - “There are no exceptions about this taboo.” (Booher 1988)
- English treatment evokes historical and authority related arguments, while Dutch favors logic, clarity and purity, as well as foreign influence:
 - “For centuries, it was fine to pile one negative on top of another in the same sentence. Chaucer and Shakespeare did this all the time to accentuate the negative.” (O’Connor 1996)
 - “In het Nederlandsche taaleigen past de dubbele ontkenning niet of althans slecht en wie er prijs stelt, onze taal zuiver te spreken en te schrijven zal dien “dubbelzinnigen” en onnederlandschen vorm dan ook vermijden.” (Van Wageningen 1946)

DOUBLE NEGATION: ENGLISH AND DUTCH

- Speakers, register, mode and variety and social value of the double negative are more often mentioned, while in Dutch frequency and mode predominate:
 - “Many people still use these constructions today, even though they are now shibboleths that mark speakers of Vulgar English.” (Wilson 1993)
 - “Maar het feit blijft bestaan dat vele Nederlanders in de dubbele ontkenning een bevestiging voelen en daarmee dienen wij rekening te houden.” (Van Wageningen 1946)
 - “In de spreektaal wordt de dubbele ontkenning echter vaak gebruikt om een ontkenning kracht bij te zetten.” (Damsteegt 1954)

SUMMARY

- The results suggest differences in the two traditions of language advice literature
- The tradition is more developed in English than in Dutch and possibly more ideologically underpinned
- The genre of advice literature changes over time from becoming more conservative to presenting both sides of the usage coin
- Quantitative exploration of the various dimensions of metalinguistic works proves useful for revealing more general patterns
- Qualitative analysis is crucial to provide more substantial answers

WHAT NEXT

- Categories need to be further specified and tested for agreement reliability
- Expand to additional usage features and explore differences across features and sub-genres

WORKS CITED

- Albakry, M. 2007. Usage Prescriptive Rules in Newspaper Language. *Southern Journal of Linguistics* 31. 2. 26–51. Algeo, J. 1991b. Sweet are the usages of diversity. *Word*, 42. 1. 1–17.
- Booher, Dianna. [1988] 1992. *Good Grief, Good Grammar*. New York: Ballantine.
- Busse, U. and A. Schröder. 2009. Fowler's *Modern English Usage* at the interface of lexis and grammar. In Römer, U. and R. Schulze (eds.) *Exploring the Lexis-Grammar Interface*. Amsterdam: Benjamins. 69–87.
- Damsteegt, B.C. 1954. *In de doolhof van het Nederlands*. Zwolle: Uitgeversmaatschappij Tjeenk Willink.
- van der Meulen, M. and I. Otto. 201
- O'Conner, Patricia. [1996] 1998. *Woe is I*. New York: Riverhead Books.
- Peters, P. and W. Young. 1997. English grammar and the lexicography of usage. *Journal of English linguistics*. 25.4. 315–331.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou and Jun'ichi Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations Session at EACL 2012.
- Straaijer, R. et al. 2015. Hyper Usage Guide of English Database (HUGE). Available online at <http://huge.ullet.net>
- Tieken-Boon van Ostade, I. 2010. The Usage Guide: its birth and popularity. *English Today* 26. 02. 14–23.
- Tieken-Boon van Ostade, I. 2015. Five Hundred Mistakes Corrected: An early American English usage guide. In Dossena, M. (ed.) *Transatlantic Perspectives on Late Modern English*. Amsterdam: John Benjamins. 55–71.
- Van Wageningen, J.C. 1946. *Even tijd voor.... onze taal!* Leiden: A.W. Sijthoff's Uitgeverijmaatschappij.
- Weiner, Edmund. 1988. On editing a usage guide. In Stanley, E. G. and T. F. Hoad (eds.), *Words for Robert Burchfield's Sixty-Fifth Birthday*. Cambridge: D. S. Brewer. 171–183.
- Wilson, Kenneth G. 1993. *The Columbia Guide to Standard American English*. New York: Columbia University Press