

Pragmatics: Data trends

Jonathan Culpeper and Mathew Gillings (Lancaster University)

Abstract

This paper identifies the trends in the data used in pragmatics studies over the last 20 years, and thereby makes predictions about how pragmatics might develop in the next few years, other things being equal. To establish those trends, 200 papers from the *Journal of Pragmatics*, covering the period 1999 to 2018, were categorised. The categorisation scheme was designed to capture some of the key ways in which data in pragmatics varies, and included: the general focus of the paper (e.g. whether it is data-driven), the focal point in the data of the analysis, the quantity of the data, the medium of the data, the number of modes or channels represented in the data, the degree of interactivity of the data, the fictionality of the data, and the language of the data (specifically whether it only contains English). Trends discovered include: a continual strong focus on data, a shift of analytical focus towards more macro units, increasing use of greater quantities of data, decreasing dominance of purely spoken data, increasing use of multimodal data, increasing use of more interactive data, decreasing use of constructed examples, no increasing use of fictional data, and the continual diversification of the languages treated.

Keywords

data, *Journal of Pragmatics*, pragmatics

1. Introduction

No self-respecting historical linguist or philologist would try to predict the future. Witness the case of the word *hoolivan*, one of the last words to be added to the 1989 second edition of the *Oxford English Dictionary*. The editors had witnessed the recent development of this word in the spring of 1985 (<http://www.sirc.org/publik/fvtackle.html>). Blending the words *hooligan* and *van*, it was used to denote the van, stuffed with communications equipment and monitors linked to CCTV cameras, that the police took to football matches in order to carry out surveillance of potential hooligans. They predicted that this new word would continue and thus added it to the dictionary. But they got it wrong. Nevertheless, it seems a reasonable enterprise to establish clear trends and then predict that those will continue, *other things being equal*. This is what this paper will do.

Some years ago, the first author of this paper had the benefit of discussions with Jacob Mey, the founding editor-in-chief of the *Journal of Pragmatics*. According to Mey, amongst a number of criteria for acceptance into the journal, a paper must make contact with data. It does not matter what type of data (even constructed examples) or how slight the contact is (a few illustrative examples), but the paper must engage with data in some way at some point. This view is reflected in the very first sentence of a recent – and warmly recommended – overview of data in pragmatics research: “There is no research in pragmatics without data” (Jucker 2018: 3). Like Mey, Jucker’s (2018: 4) overview, and indeed the whole volume in which it appears (i.e. Jucker et al. 2018), is agnostic about the kind of data, starting from “the premise that there is no single best type of data”. Whilst we support this agnostic position, it is not one that all in the field of pragmatics share. For example, fictional data has often been seen as not genuine data or not naturally-occurring, and thus in some way substandard (see the discussion in Locher and Jucker 2017: 4-11). Besides, there are many reasons why pragmatics scholars may privilege some types of data over others, including issues of accessibility and personal interest. In this paper, we will reveal preferences for particular kinds of data, and, moreover, show whether they have changed over time.

The following section introduces our data method. In a nutshell, we categorise 200 papers from the *Journal of Pragmatics*, stretching from 1999 through to 2018. Our

categorisation scheme is designed to capture some of the key ways in which data in pragmatics varies. We cross-checked this with Jucker (2018) in order to make sure that the scheme we devised was reasonably comprehensive, and also tweaked it in the light of issues we suspected will be of interest to the pragmatics community (e.g. whether the language data was exclusively in English). The remaining sections report the results of our analysis. They cover the focus of the paper (e.g. whether it is data-driven), the focal point of the analysis (e.g. whether it focuses on a particular kind of linguistic unit), data quantity (e.g. whether a few short individual examples are used or a large corpus), data medium (e.g. whether it is speech or writing), the number of modes or channels represented in the data (e.g. whether it is single mode or multimode), the degree of interactivity of the data (e.g. whether it is monologic or dialogic), the fictionality of the data (e.g. whether it is real or fictional), and the language the data might be in, specifically, whether it is English or not.

2. Our data and method

This paper, and especially its method, is partly inspired by Jucker and Staley (2017). That paper was concerned with the methods used in politeness studies and how they have changed from 1985 through to 2014. They scrutinized papers from the *Journal of Politeness Research* and the *Journal of Pragmatics*, but with the latter restricted to papers containing the word ‘politeness’ in the title, abstract or keywords. Our data consists of 200 papers published in the *Journal of Pragmatics* in the years 1999 through to 2018 inclusive. The *Journal of Pragmatics* is the biggest (in terms of the number of papers it publishes) and longest running (since 1977) journal that caters for a mixed range of pragmatics research.¹ It has the added advantage of being fully searchable via the powerful engine SCOPUS. Nevertheless, we should remember that it does not represent all of pragmatics. For example, there are other, more specialist, pragmatics channels for publication, such as the journals *Intercultural Pragmatics* and *Pragmatics and Cognition*. This is to say nothing of book chapters and monographs. Nevertheless, given the importance of the *Journal of Pragmatics* to the field, not least if measured purely in terms of the number of papers it publishes, we hope that our results will be at least indicative of some general trends.

From the *Journal of Pragmatics*, we selected the top 50 most-cited papers from each of four periods: 1999-2003, 2004-2008, 2009-2013, and 2014-2018. It is worth pausing to note that we are, through the mechanism of citations, examining the papers that are deemed to be important by the readers of the *Journal of Pragmatics* and have had most exposure. One might wonder whether our conclusions would be the same if we had analysed a random sample of what the *Journal of Pragmatics* actually published. For example, if fictional data is not held in high regard in the pragmatics community, it is unlikely to be one of the most-cited papers. Thus, a lack of papers focussing on fictional data could be seen as a consequence of the citation criterion, rather than what scholars actually produce. To get a sense of any possible discrepancies, we analysed a random sample of 25 papers from the first period and a further sample of 25 papers from the third period. For both random samples of actual publications, the patterns that emerged did not contradict any of the conclusions we make in this paper. We had expected to find a discrepancy regarding the role of English as the focal point of papers. However, even here, the differences are minor, as we elaborate in section 3.8.

Papers were retrieved via SCOPUS. Editorials and introductions to special issues were excluded. Purely methodological or theoretical papers only feature in section 3.1, as they could not be judged against the same criteria as the data of other papers. Where papers were mixed or fuzzy edged, they were classified on the basis of their main aspect. For example, if a paper used the British National Corpus for its analysis, yet it also contained a couple of constructed illustrative examples, it would still be classified as “real” data. All tricky cases were discussed by the two authors at length. Occasionally, where categorisation was not possible, the paper was allotted to the Unclassified category.²

The upcoming sections not only report our results, but each section will start by motivating and then introducing the particular data categories they present.

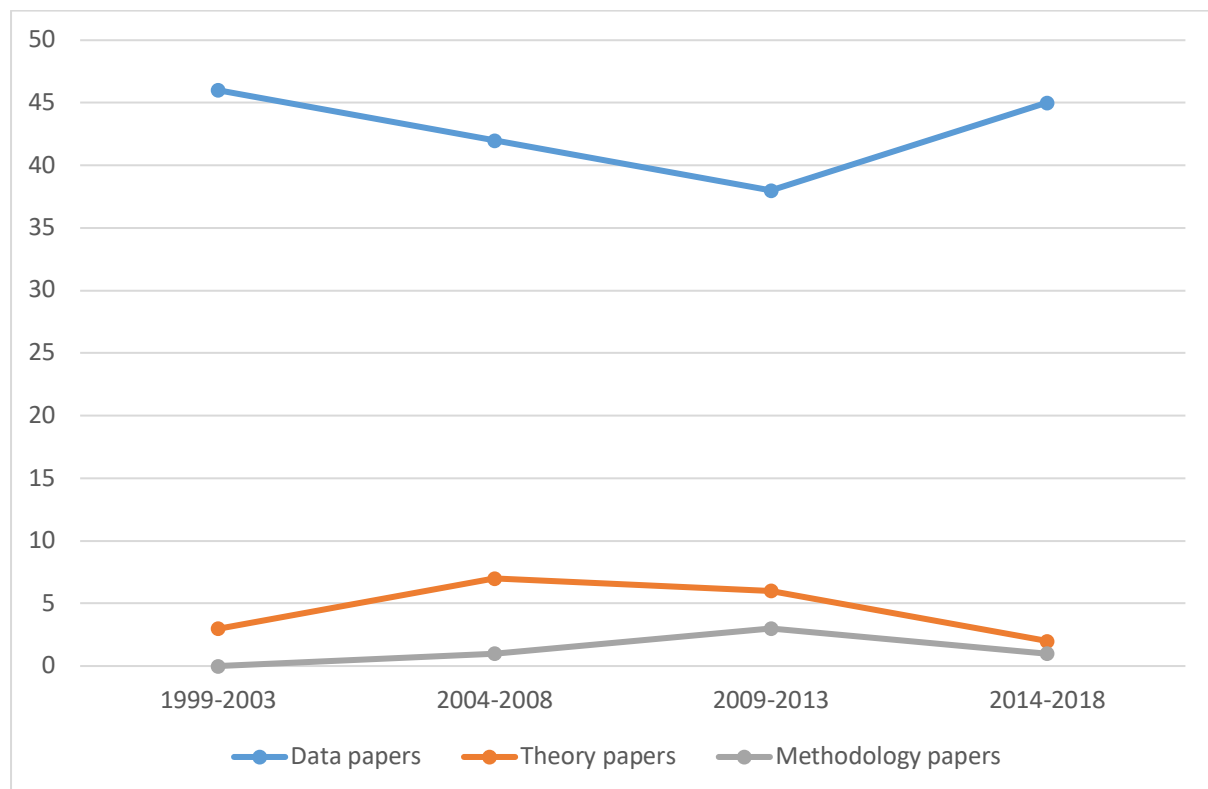
3. Results

3.1 Focus of the paper: Data, theory or methodology

This first results section has the broadest perspective. It addresses the role of data in the paper: to what extent are papers thoroughly oriented to the data? An obvious case of a data-focused paper is a paper in which concepts and/or descriptions are derived from the data. Other potential categories include theory-focused papers, where the concern is to describe, build and/or challenge theory, and also methodology-focused papers, where the concern is to develop and/or evaluate research methods and frameworks. However, operationalizing all these categories is not straightforward. Whilst we were able to pinpoint mainly theory-focused or methodology-focused papers, it was less easy to do the same for data-focused papers. In many papers, a combination of categories are involved – part of the paper would be data-focused, but that might be combined with theoretical developments, not to mention methodological ones. We thus made the decision to include in the data-focused category any paper that is mainly data-driven (in practice, this meant any paper that did not fit the theory-focused or methodology-focused categories would be placed here).

In the light of our comments in section 1 about the close connection between pragmatics and data, one might reasonably expect data-driven papers to be by far the most common category, and purely theoretical papers to be non-existent. Our results are displayed in Figure 3.1.

Figure 3.1. Data, theory or methodology papers in papers in the *Journal of Pragmatics*, 1999-2018



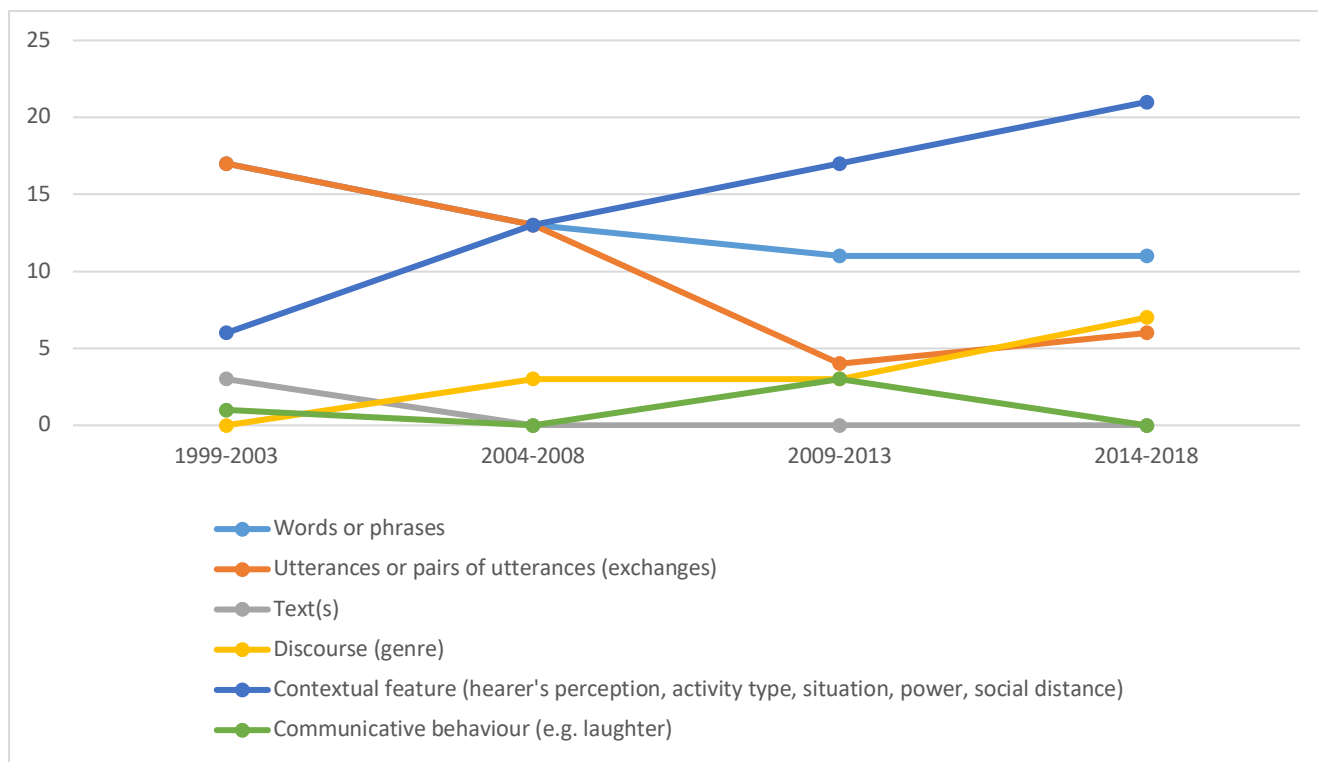
As expected, it is clear that data-focused papers have always been most popular. Theoretical papers are a minor affair. Considering the claim we cited in the Introduction of this paper about the close relation between pragmatics and data, one might be surprised that there are any such papers, as these papers contain an absence of data. One example is Bargiela-Chiappini (2003), which discusses politeness without supplying a single linguistic example, instead focusing on the theories and critiquing what others have said about them. Yet it would be difficult to deny that this is a pragmatics paper. One might argue that this paper, and perhaps all theoretical papers, deal with data at one or more removes; in other words, they engage or cite papers that do connect with data. Whatever the case, we are reminded that the claim about pragmatics and data is not an absolute but a robust generalisation. The fact that methodological papers look like an endangered species is also worthy of note: the field has paid remarkably little attention to this area, especially given its predilection for data. The publication of Jucker et al. (2018) is perhaps a move to rectify this situation. Let us hope that it stimulates endeavor. As yet, no clear diachronic shifts for any type of focus, data, theory or method, can be detected.

3.2 Focal point of the analysis: Utterance, discourse, context or something else?

Early work in pragmatics, such as Grice (1975), was very much concerned with utterance meaning and interpretation. Even in cases where examples consisted of an exchange, or occasionally a series of exchanges, the focal point was a particular utterance (or occasionally utterances). This is still very much the case with certain strands of pragmatics, notably, relevance theory (Sperber and Wilson 1995). What we wish to consider here is the extent to which the focal point of analysis is the utterance, or whether another kind of linguistic unit or pragmatic feature has become more important. Our general prediction is that the focus has become considerably broader than the utterance.

The utterance is not in fact the smallest linguistic unit that has been the focus of attention in pragmatics papers. Discourse markers are a particular case in point. They often consist of just one word. Similarly, there are papers on expressions of stance, interjections, pronouns, expressions of definiteness, and so on – all of which often consist of one word. Above the utterance in the size hierarchy, we have texts, that is, a whole text (e.g. a complete literary work, a political speech, an interview). “Whole text” here does not mean that the whole text is quoted; it means that the author is discussing or drawing conclusions about the whole text. With this category, we made the decision to include papers that included more than one text but treated them all individually. Above the individual text, we have genre. Here, we included papers that were constructed to draw conclusions about a whole genre or type of discourse (e.g. newspaper articles, courtroom proceedings). Aside from these linguistic units, we took two pragmatic features into consideration. One category was for papers that focused on communicative behaviours, such as silence or laughter. Another was for papers that focused on a contextual aspect (or aspects), such as the hearer’s perception of appropriateness or power relations. Our results are displayed in Figure 3.2.

3.2 The dominance of particular kinds of linguistic unit or pragmatic feature as the analytical focus in data in papers in the *Journal of Pragmatics*, 1999-2018³



Utterances, as the focal point of analysis, have declined quite considerably, and the decline, though somewhat less marked, can be seen in words and phrases. Counter to this, there has been a steady and marked increase in papers dealing with particular contextual features. The other categories remain quite minor features, though perhaps genre is undergoing a gentle increase. What all this suggests is a general broadening of analytical focus. Perhaps this is evidence of a shift from micro pragmatics to macro pragmatics.

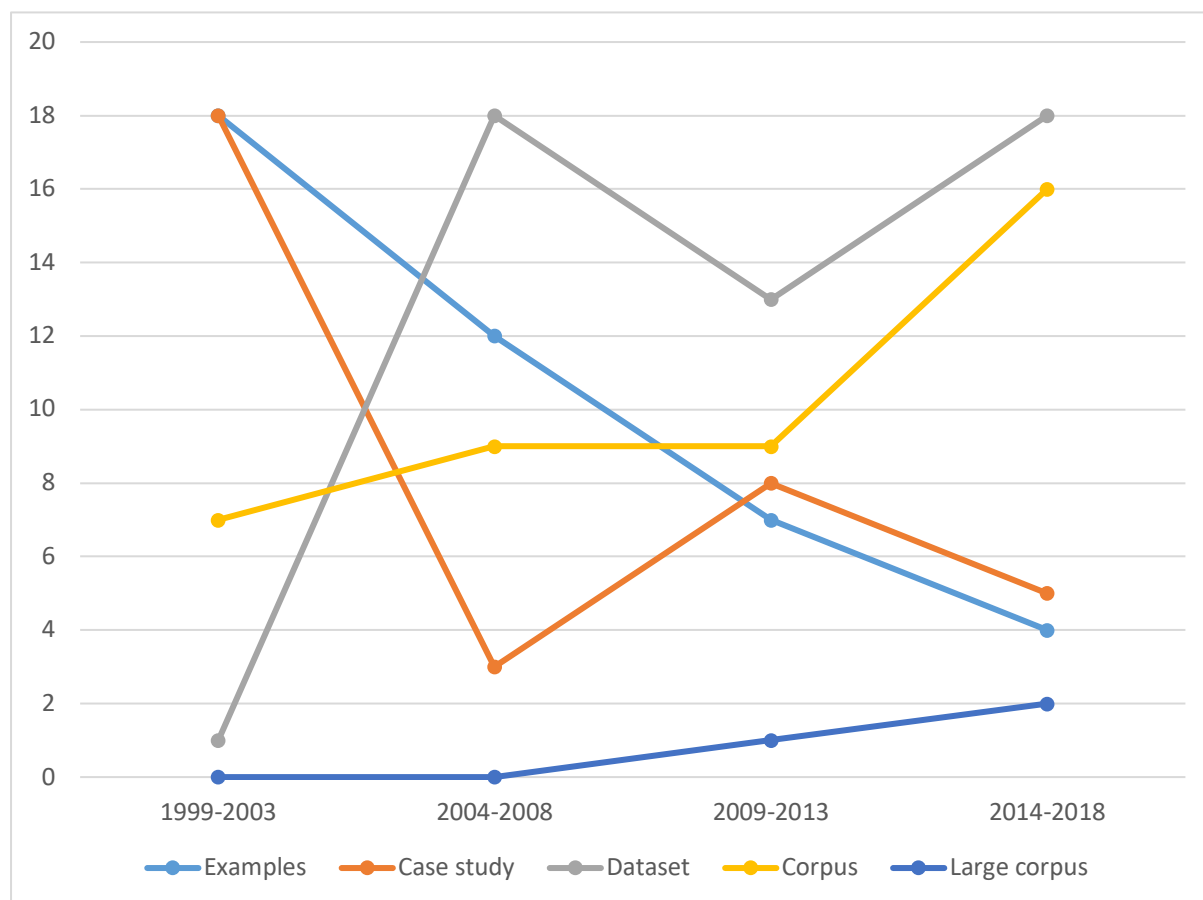
3.3 Data quantity: Individual examples, a large corpus or something else?

The advent of the computer has made it easier to manage larger datasets. This is not simply a matter of storage on one's computer. A computer can facilitate the analysis of data with software programs designed for qualitative or quantitative analyses. Moreover, the development of email and Skype has made it easier for teams to pursue large collaborative projects, often with large datasets. In addition, the advent of the field of corpus linguistics, and indeed corpus pragmatics, has resulted in the ready availability of huge corpora and the tools with which to analyse them. With all of this, it seems likely that pragmatics papers are generally encompassing larger datasets than they used to.

Devising categories which form points on a scale relating to quantity of data is a tricky business. It was not practical, and in some cases impossible, to compute the quantity of data used in each paper. Instead, we formed points on the scale that could be identified and would give at least an approximation of the overall quantity of data used in a particular paper. The category with the smallest quantity of data concerns papers that only use individual and short examples (usually not longer than a sentence or two), typically to illustrate a particular point been made. Next, we formed a case study category: papers that focused on one individual text in-depth, or several individual texts in depth but with no attempt to analyse those texts as a group. Following on from that category, we formed a dataset category, which included papers that involved the study of and conclusions about a group of texts, but which did not fit our criteria for a corpus. Distinguishing a dataset from a corpus is difficult, not least because the label "corpus" seems to have become fashionable and is sometimes applied to what formerly might have been described as a dataset. Admission to the corpus category

involved the paper fulfilling at least two of the following criteria: the paper used corpus linguistic methods/processes to carry out its analysis (e.g. corpus annotation, corpus software); it analysed a corpus recognised by the community as a corpus; or it was constituted by at least 100,000 words but not more than 1 million words.⁴ The reason for the upper threshold of 1 million words was to accommodate our final category, the large corpus, such as the British National Corpus, which consists of 100 million words.⁵ Our results are displayed in Figure 3.3.

Figure 3.3 The dominance of differing quantities of data in papers in the *Journal of Pragmatics*, 1999-2018



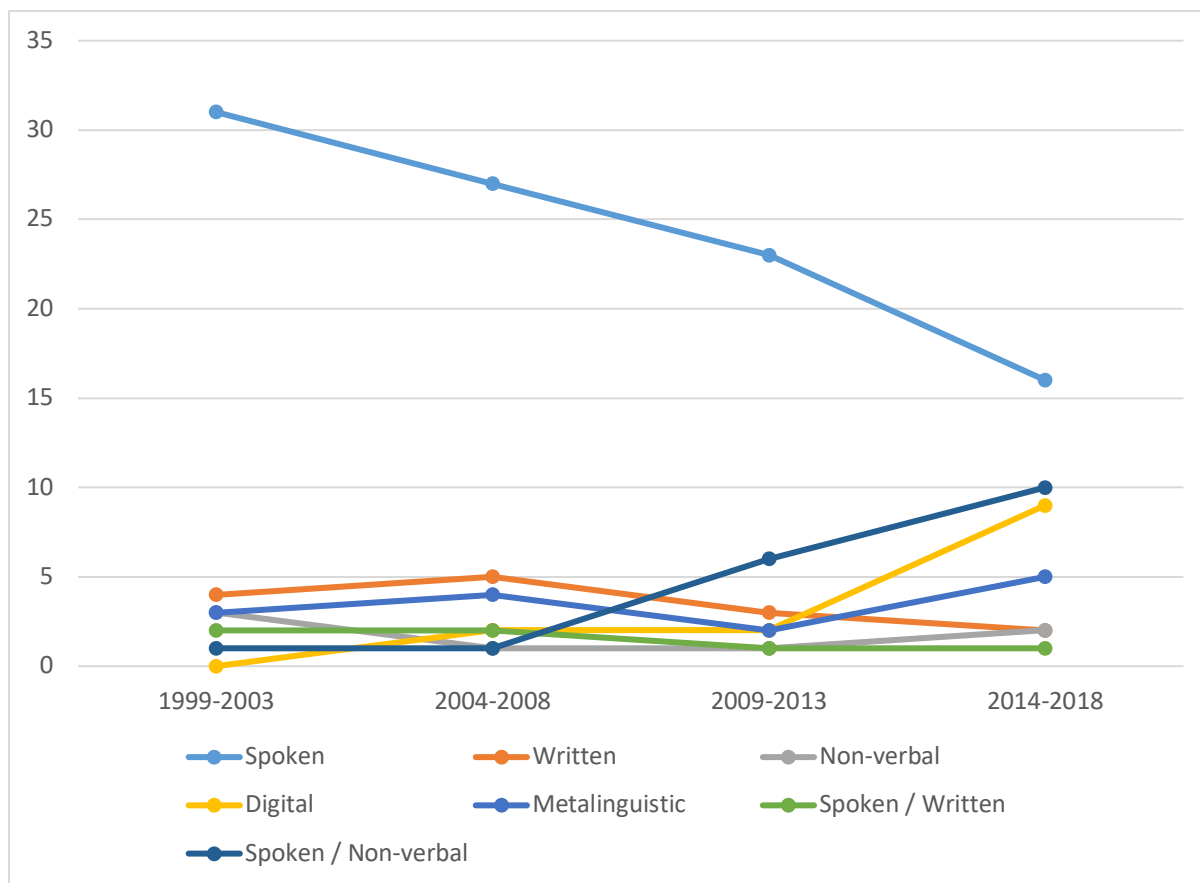
The clearest and steadiest shift is the decline in papers which use a sprinkling of short individual examples. Somewhat similar, though less steady, is a decline in papers that are based on an individual case study (or case studies). These declines seem to be matched by an increase in the use of datasets and, to a lesser extent, an increase in the use of corpora, especially since 2009-2013. This is in tune with Jucker and Staley’s (2017: 422) comment regarding politeness research in the *Journal of Pragmatics*: “Corpus-based methods are increasing but still marginal in politeness research”. Their study only considered papers up to 2014, so they would not have observed a possible acceleration in more recent years. Large corpora are hardly being used, perhaps, one might speculate, because their very size seems to obscure any detail regarding the contexts of the texts that comprise them, thus making many types of pragmatic study difficult. Overall, however, it does indeed seem to be the case that pragmatics papers are encompassing larger datasets.

3.4 Data medium: Speech, writing or something else?

The focus of early pragmatist scholars on the utterance, as mentioned in section 3.2, entails a spoken linguistic unit. Similarly, Searle's (1969) work on speech act theory denotes work on speech. Traditionally, then, pragmatics has been strongly biased towards spoken language. Nowadays pragmatics is clearly broader in scope, encompassing communicative behaviours regardless of medium. Jucker (2018: 12) suggests that corpus linguistics has had "the effect of shifting the primacy of description [of the spoken language] to the written language because language that already exists in written form is much more readily available for corpus compilation". It may also be the result of a number of works that have demonstrated the richness of pragmatic phenomena in even the most "written" of written texts (see, for example, Myers 1989 on the pragmatics of politeness in scientific writing). We should also remember that digital media (email, Facebook, Twitter, etc.) are often principally in written form, and that they have understandably attracted much interest. Overall, then, we would predict a relative decline in papers focusing solely on speech, and a diversification of modes, with a particular rise in digital media.

Our spoken category was constituted by papers whose main data was transmitted in speech, including, for example, via the telephone.⁶ Our written category involved papers where the main focus was on data transmitted in writing. We had a mixed spoken/written category, for papers which featured both spoken and written datasets. We distinguished digitally transmitted datasets by placing them in their own category. Communicative behaviours, of course, are not confined to speech or writing. We created a non-verbal category to accommodate papers whose main dataset involved non-verbal features (e.g. gestures). Non-verbal features are often combined with spoken features in pragmatic papers, so we created a category for that. Finally, it became clear that we needed a category for papers focusing on metalinguistic data (e.g. the language used as prompts in experiments). Our results are displayed in Figure 3.4.

Figure 3.4 The dominance of speech, writing and other media data in papers in the *Journal of Pragmatics*, 1999-2018

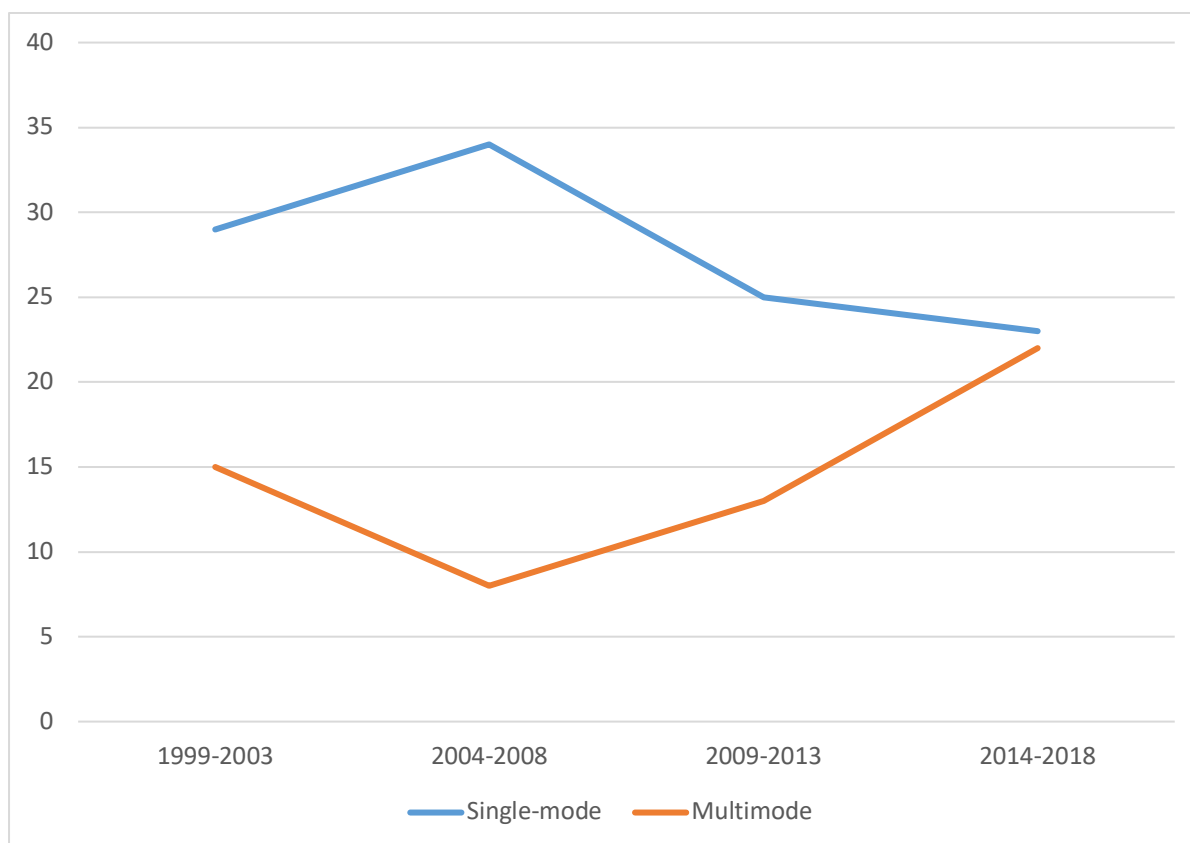


The dominance of spoken language is stark and remains so across the periods. However, it is also apparent that spoken language oriented papers are experiencing a clear and steady decline over the years, as was predicted. What is evident from the period 2004-2008 onwards is that part of that decline is being countered by papers combining data on both spoken features and non-verbal features. Other media remains a fairly low level. Papers examining solely non-verbal data are strikingly few. Contrary to predictions, written data even slightly declines after the period 2004-2008. However, in tune with what one would expect, we should note the marked rise in digital data after the period 2009-2013.

3.5 Number of modes: Single mode or multimode?

Multimodal data was apparent in the previous section: we created the category spoken/non-verbal to accommodate that multimodal mix.⁷ But it is not, of course, the only possible mix. Digital communication often involves images and symbols, along with words (in fact, this did not occur in our sample). The objective of this section is to assess the quantities of papers that focus on a single mode, usually linguistic in nature but not always, versus the quantities of papers that focus on multimodes, usually a linguistic feature plus non-verbal aspects. We would predict that multimode papers have been gaining ground over the period, partly because the tools for managing multimodal data have made significant progress, but also because a number of studies have drawn attention to the multimodal nature of spoken interaction (e.g. Norris 2004; Stivers and Sidnell 2005). Our results are displayed in Figure 3.5.

Figure 3.5 The dominance of single mode and multimode data in papers in the *Journal of Pragmatics*, 1999-2018



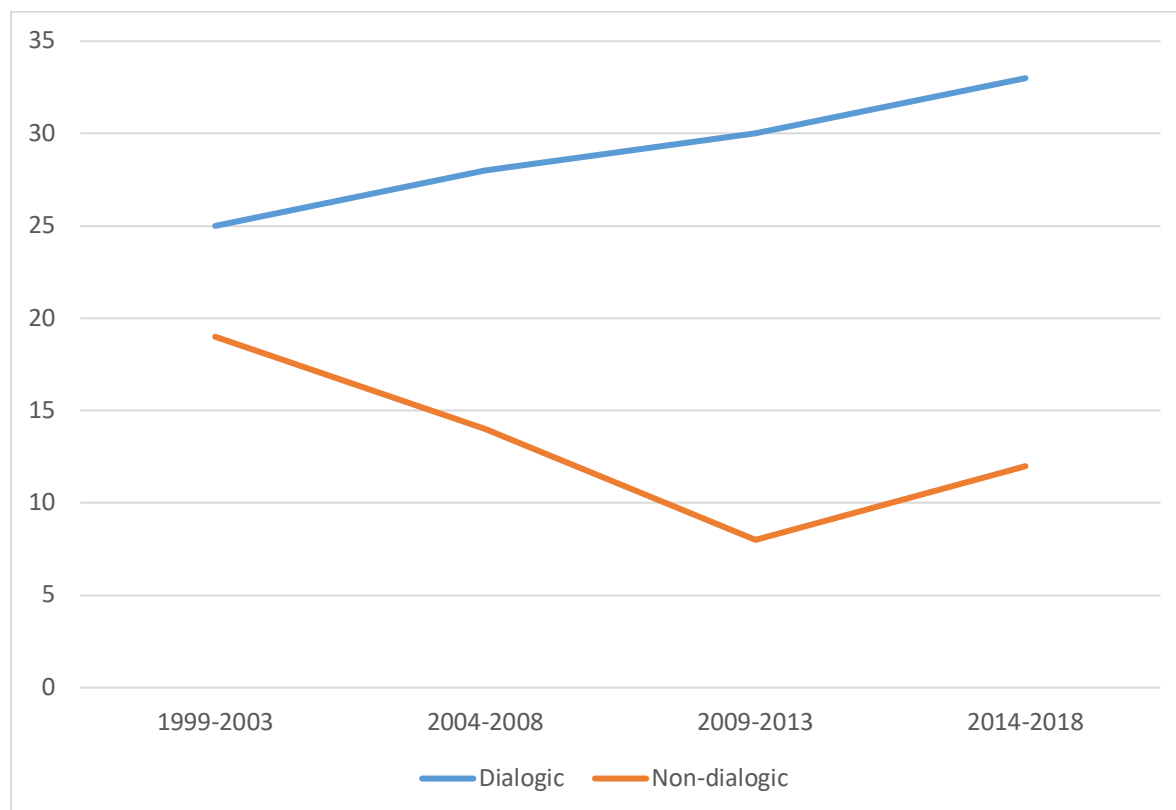
Papers written in the early 2000s seemed much more focused on single mode communication. However, the situation has changed. Now, there is almost a 50/50 split between papers examining single mode data and those examining multimode. Clearly our prediction was correct. What our analysis shows is that the key period of change has been the last 10 or so years.

3.6 Interactivity: Monologues or dialogues?

The pragmatics of interaction or interactional pragmatics seems to have been gaining ground. Volume 4 of the *Handbook of Pragmatics Highlights*, published in 2009 by John Benjamins, was entitled *The Pragmatics of Interaction* (edited by Sigurd D'hondt, Jan-Ola Östman and Jef Verschueren). In the opening survey chapter, D'hondt (2009: 1) states that the first requirement for inclusion in that volume is that the chapters are "concerned with the empirical investigation of how human beings organise their exchanges in natural settings". Note that we are dealing with exchanges as a key unit, not utterances. Furthermore, a requirement is that the chapters take their analysis "beyond the level of 'what is said' to that of the interactional organisation of speech and action" (D'hondt 2009: 1). As one might imagine, Conversation Analysis is a major input into this area, though not the only one, others being, for example, Goffman's work on the interaction order and interactional sociolinguistics. Traditional pragmatics – conversational implicature, speech act theory, etc. – is "in the background", but "continues to exert its influence" (D'hondt 2009: 1). This is not, of course, the only way of defining interactional pragmatics. Consider that the 1995 pragmatics book written by Jenny Thomas was entitled *Meaning in Interaction* and could be construed as a precursor to a type of interactional pragmatics, but has little to do with Conversation Analysis. Whatever its provenance or definition, our prediction is that the rise

of interactional pragmatics has caused an increase in the use of dialogic data. By dialogic data we refer to data that represents exchanges between two or more participants. This contrasts with non-dialogic data where utterances or words are analysed in isolation. Our results are displayed in Figure 3.6.

Figure 3.6 The dominance of dialogic and monologic data in papers in the *Journal of Pragmatics*, 1999-2018



As predicted, there has been a steady increase in the number of papers using dialogic data, and a fairly steady decrease in the number of papers focusing on non-dialogic data, with a particular dip in the 2009-2013 period.

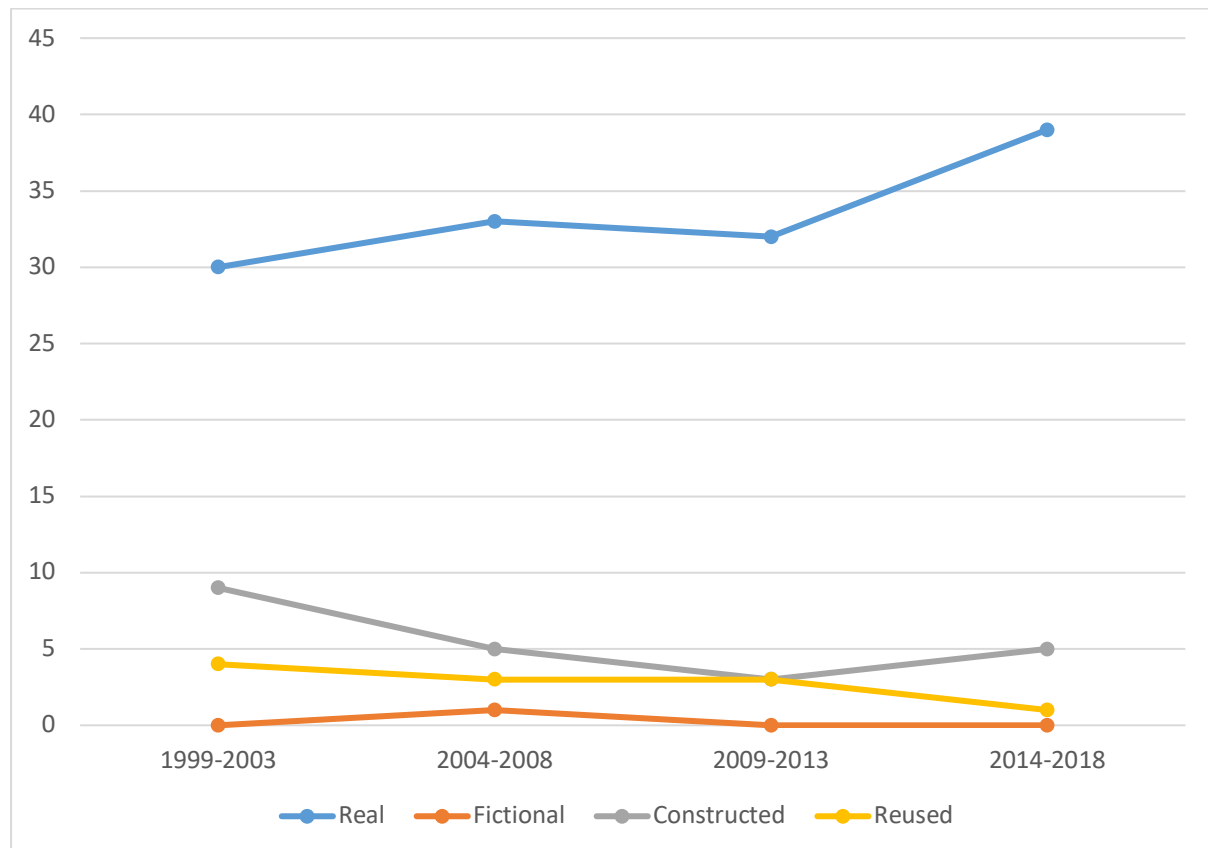
3.7 Fictionality: Real, fictional or something else?

As briefly mentioned in our introduction, fictional data have not been held in high regard by the pragmatics community. However, fiction intrudes into multifarious aspects of our daily lives and cannot be easily separated from what is supposedly "real". Consider a personal advertisement, a "story" about what happened at work, or a film about a real person. Even canonical literary forms – poems, plays and prose – have a claim to be naturally-occurring, and are full of pragmatic interest. The recent volume by Locher and Jucker (2017) fully demonstrates the pragmatic riches of fictional data. But perhaps it is too early for such works to have changed attitudes much. We predict a slight increase in the use of fictional texts.

Our category "real" data is for papers that contain data from real sources, that is, real people experiencing real events and things. The fictional category, in contrast, relates to papers that contain data from fictional sources, that is, non-real people experiencing non-real events and things. In addition, we created two further categories. One concerns papers that contain constructed data, typically short examples to illustrate a particular point. These have no basis in actuality. Such data, in our view, constitute the paradigm case of what is not

naturally-occurring. The other category, driven by what we found in the data, concerns papers that reused examples, that is, examples reused from other publications. It is important to recollect our general policy here, namely, that categorizations are based on what the majority of the data are. Thus, for reused examples, one might suppose it to be fairly common for papers to occasionally reuse examples from other publications for illustrative purposes, but uncommon for the majority of examples to be reused. Our results are displayed in Figure 3.7.

Figure 3.7 The dominance of real, fictional and other kinds of data in papers in the *Journal of Pragmatics*, 1999-2018



As indicated, the use of real data has always been the most popular choice. Our analysis shows that it has become even more popular in recent years. In contrast, the use of fictional data has always been low. In fact, just one paper in the whole dataset focused on fictional data, fictional prose, in the 2004-2008 period. Contrary to our tentative prediction, there is no sign of a reprieve for fictional data in more recent years. The use of constructed examples has seen a fairly steady decrease, perhaps, one might speculate, because of a halo effect from areas of linguistics such as corpus linguistics or usage-based models of language which emphasize actual language use. Papers dominated by reused language data are also minority, and one that is slowly shrinking.

3.8 Language: English or something else?

Pragmatic theory – like theories of phonology, morphology, syntax and so on – is not bound by data pertaining to a particular language. On this basis, one might reasonably expect a kaleidoscope of languages to populate the data found in the *Journal of Pragmatics*. Pragmatic descriptions, of course, can focus on a particular language (or particular aspect of a particular

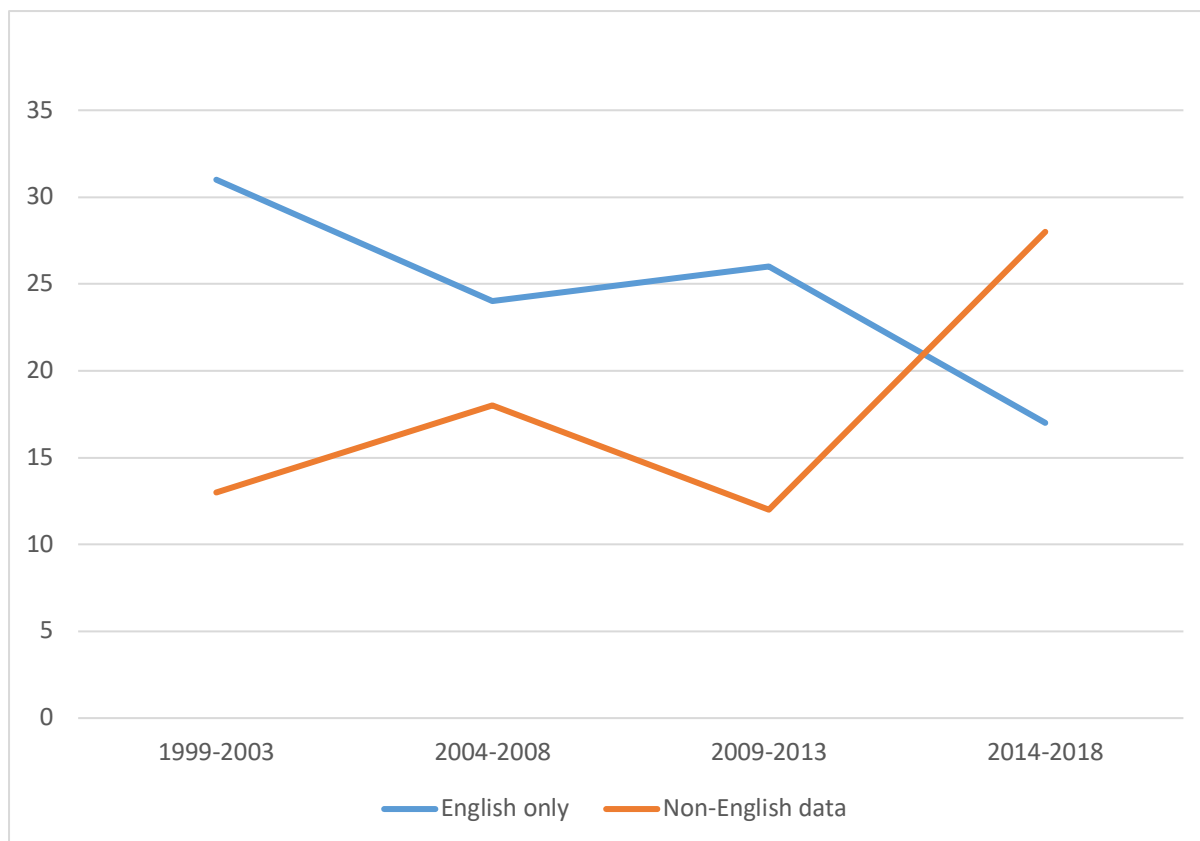
language), but it is difficult to think of a purely linguistic reason for why any one language will deserve more pragmatic description than another. In spite of all this, it hardly takes a scholarly study to ascertain that there is a bias in the world of pragmatics in favour of English. The dangers of this situation are manifold. In the introduction to a special issue in the *Journal of Pragmatics* on “emancipatory pragmatics”, the editors state (Hanks et al. 2009: 1-2):

It is our shared conviction that pragmatics as an analytic enterprise has been dominated by views of language derived from Euro-American languages and ways of speaking. Speech acts defined in terms of standard illocutionary forces and felicity conditions, implicatures explained on the basis of the Gricean cooperative principle and maxims, politeness defined in terms of a universal notion of ‘face’, and the very idea that speech is driven by the exchange of information are all examples of the problem. While these research traditions have enriched the field of pragmatics, they also have tended to rely uncritically on the common sense of speakers of modern Western languages, with the attendant premises of individualism, rationality, and market economy. That is, while they are presented as general models of rational language use, they in fact rely heavily on the native common sense of their authors and practitioners.

One only needs to think of the seminal works by scholars such as Austin or Searle to realise that we are mainly talking about English. How much of a bias is there towards English? Egbert et al. (2016) examined 100 papers in the *Journal of Pragmatics* from 2011 through to 2013, focusing especially on the issue of the accessibility of non-English data. They report that “half of the hundred articles investigate interactional data in English”, and that “38 articles study only English data in comparison with 62 studying non-English data” (Egbert et al. 2016: 104). However, their study is different from ours in that they only included articles that analyse transcripts of human interaction. As is clear from section 2.1, we have a much more inclusive policy.

We sorted our 200 papers into two categories (we excluded the very few papers that did not contain any data; see section 2.2): English data only and non-English data (possibly in combination with English). The results are displayed in Figure 3.8.

Figure 3.8. The language of the data: English only or non-English data



Although there is a slight reversal of the trends in the 2009-2013 data, overall there has been a decrease in the number of papers focusing purely on datasets in English, and a commensurate increase in the number of papers that focus on a language other than English, or another language along with English. Strikingly, in the 2014-2018 dataset, there are now more papers discussing languages other than English than English alone. One might speculate that this is partly due to the efforts of emancipatory pragmatics scholars, along with papers like Egbert et al. (2016), in raising awareness.

As we mentioned in section 2, the papers we have been analysing in these sections are the 50 top-cited ones (in a random sample for each period), and so the papers that seem to be having most impact. We also analysed a random sample of 25 papers for the 1999-2003 period and another 25 for the period 2009-2013. Our results matched our findings remarkably well. With regard to papers with English-only data, we had expected to find a greater number in the randomly sampled data, hypothesising that papers focussing on English might attract more citations and hence skew our sample towards English-only papers. If we combine the numbers of English-only papers in our random samples, we get 26 English-only papers out of the 50, which, expressed as a percentage, is 52%. If we do the same for the cited (and then randomly sampled) datasets, we get 58 out of 100, or, in other words, 58%. The difference is clearly far from big. If we look at the individual periods of the random samples, we find a decline in English-only papers, as indeed noted in the previous paragraph for our cited datasets. However, that decline is somewhat more marked. In percentage terms, the English-only papers in the cited (and then randomly sampled) datasets shrink from 62% of the total of papers in the first period to 54% in the second, whereas in the randomly sampled datasets, they shrink from 60% to 44%. Still, this is hardly evidence that the cited datasets are skewed. If it were, we would also expect to see a difference in the first period, instead of the similar figures of 62% and 60%. Instead, what all this mainly points to is that the papers in the *Journal of Pragmatics* for the period 2009-2013 are a little bit different.

3. Conclusion

In the light of the results of our analyses, we will make the following very tentative predictions, which, of course, depend upon all other things being equal.

- *Focus of the paper*: Pragmatics will remain overwhelmingly focussed on data. No increase in the few theory-focused and methodology-focused papers is likely in the near future.
- *Focal point of the analysis*: Pragmatics will continue to broaden its analytical focus beyond the utterance, shifting from the more micro to the more macro, incorporating in particular contextual features.
- *Data quantity*: Pragmatics will continue to increase the quantities of data it investigates, moving away from a sprinkling of individual examples. However, no increase in extremely large datasets is likely in the near future.
- *Data medium*: Pragmatics will continue to be dominated by spoken data, but that dominance will decline. Instead, there will be a rise in research on data formed of multiple media, and a particular rise in research on digital media data.
- *Number of modes or channels represented in the data*: Pragmatics will continue to move away from single mode data towards multimodal data.
- *Degree of interactivity of the data*: Pragmatics will continue to increase its use of more interactive, more dialogic data.
- *Fictionality of the data*: Pragmatics will continue to be dominated by real data and become even more so. The use of constructed examples will continue to decline. The use of fictional data will not increase in the near future.
- *Language*: Pragmatics will continue to diversify the range of languages it treats.

Needless to say, we must remember that these predictions are based on the trends of 200 papers in one journal, the *Journal of Pragmatics*, 50 for each period. A larger sample would, of course, help clarify the picture we have developed, especially the minor trends. Also, we should remember that our dataset consisted of the 50 most-cited papers from a random sample of each time period. The citation criterion may have had an effect on some of our categories, although scrutiny of a completely random sample of 50 papers, 25 for each of two periods, failed to reveal any but minor differences.

References

- Bargiela-Chiappini, Francesca (2003). Face and politeness: New (insights) for old (concepts). *Journal of Pragmatics* 35 (10-11): 1453-1469. [https://doi.org/10.1016/S0378-2166\(02\)00173-X](https://doi.org/10.1016/S0378-2166(02)00173-X)
- Bednarek, Monika (2011). Approaching the data of pragmatics. In: Wolfram Bublitz and Neal R. Norrick (eds.), *Foundations of Pragmatics*. Vol. 1 of *Handbooks of Pragmatics*. Berlin: De Gruyter Mouton, 537-559.
- D'hondt, Sigurd (2009). The pragmatics of interaction: A survey. In: Sigurd D'hondt, Jan-Ola Östman and Jef Verschueren (eds.), *The Pragmatics of Interaction*. Vol. 4 of the *Handbook of Pragmatics Highlights*. Amsterdam/Philadelphia: John Benjamins, 1-19.
- D'hondt, Sigurd, Östman, Jan-Ola and Jef Verschueren (eds.) (2009). *The Pragmatics of Interaction*. Vol. 4 of the *Handbook of Pragmatics Highlights*. Amsterdam/Philadelphia: John Benjamins.
- Egbert, Maria, Yufu, Mamiko and Fumiya Hirataka (2016). An investigation of how 100 articles in the *Journal of Pragmatics* treat transcripts of English and non-English

- languages. *Journal of Pragmatics* 94: 98-111.
<http://dx.doi.org/10.1016/j.pragma.2016.01.010>
- Grice, H. Paul (1975 [1989]). Logic and conversation. In: Peter Cole and Jerry Morgan (eds.), *Syntax and Semantics, Volume 3. Speech Acts*. New York: Academic Press, 41-58.
- Jucker, Andreas H. (2018). Data in pragmatic research. In: Andreas H. Jucker, Klaus P. Schneider and Wolfram Bublitz (eds.), *Methods in Pragmatics*. Vol. 10 of Handbooks of Pragmatics. Berlin: De Gruyter Mouton, 3-36.
- Jucker, Andreas H., and Larssyn Staley (2017) (Im)politeness and developments in methodology. In: Jonathan Culpeper, Michael Haugh and Dániel Kádár (eds.). *The Palgrave Handbook of Linguistic (Im)Politeness*. London: Palgrave, 403-429.
- Jucker, Andreas H., Schneider, Klaus P. and Wolfram Bublitz (eds.) (2018) *Methods in Pragmatics*. Vol. 10 of Handbooks of Pragmatics. Berlin: De Gruyter Mouton.
- Locher, Miriam A. and Andreas H. Jucker (eds.) (2017). *Pragmatics of Fiction*. Vol. 12 of Handbooks of Pragmatics. Berlin: De Gruyter Mouton.
- Myers, Greg (1989). The pragmatics of politeness in scientific text. *Applied Linguistics* 10: 1-35. <https://doi.org/10.1093/applin/10.1.1>
- Norris, Sigrid (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. New York and London: Routledge.
- Searle, John R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sperber, Dan and Deirdre Wilson ([1986]1995). *Relevance: Communication and Cognition*. (2nd edn) Oxford, U.K. and Cambridge, U.S.A.: Blackwell.
- Stivers, Tanya and J. Sidnell (2005). Introduction: Multimodal interaction. *Semiotica* 156: 1-20. <https://doi.org/10.1515/semi.2005.2005.156.1>
- William, William F., Sachiko Ide and Yasuhiro Katagiri (2009). Towards an emancipatory pragmatics. *Journal of Pragmatics* 41(1): 1-9.
<https://doi.org/10.1016/j.pragma.2008.02.014>

¹ An alternative choice might have been *Pragmatics*, the journal of the International Pragmatics Association (IPrA), which has been running since 1990. This, however, is not so easily searchable according to the criteria of our method (see Section 2). For example, it is only fully searchable via SCOPUS from 2009. Furthermore, although it accommodates all kinds of pragmatics, in practice it tends to attract papers written by European scholars and papers which fit a rather more “Continental European” view of pragmatics.

² The Unclassified category was never populated by more than three instances for any data feature in any one of the four periods. For this reason, this category is not represented in the Figures, in order to reduce clutter.

³ The scores for Words and phrases and Utterances are identical for the first two periods, and hence their relevant lines overlap. This is why there seems to be five lines on the graph for the first two periods and six for the second two.

⁴ Not every criterion was equally easy to apply to every paper, and especially so for the final criterion. Some papers which appeared to have large datasets only specified the overall quantity in terms of the number of hours of recording. As a very rough rule of thumb, we took 10 hours of recording to be equivalent to our 100,000 written word cut-off point.

⁵ Some of our distinctions here are inspired by Bednarek (2011), who distinguishes amongst case studies of one or just a few texts, small-scale corpora and large-scale corpora. Aside from our additional category of short individual examples, we carve out space between case studies and small-scale corpora with our datasets category.

⁶ This also included one paper which analysed a political speech, which would have been written to be spoken.

⁷ The category of “spoken/non-verbal” features in the counts for section 3.4 and in the multimodal counts for this section, 3.5. To have simply counted it here as multimodal data, would have been misleading with regard to our claims in the previous section about spoken data. One might wonder whether there was a “written/non-verbal” category (some combination of writing and pictures, for example). This did not occur in our sample.