From pre-owned printers to pristine Porsches

A corpus linguistic analysis of eBay item descriptions

Andrew Kehoe,[i] Matt Gee,[i] and Ursula Lutzky[ii]

[i] Birmingham City University | [ii] Vienna University of Economics and Business

Recent years have seen a considerable increase in e-commerce, with sales forecast to continue rising over the coming years. This study provides a corpus linguistic analysis of item descriptions on eBay's UK website, on which both members of the public and businesses can offer goods for sale. It is based on two corpora of items sold on the site in 2015 and 2020 which together contain 412,601 item descriptions and over 57 million words of text. The analysis applies corpus linguistic methods to gain further insight into the diachronic development of language use on eBay, to explore linguistic features in item descriptions and across product categories, and to relate word choice to product selling price. Its findings offer new understanding of the changing language of online selling and indicate how a corpus linguistic methodology may be used to explore the impact of linguistic features on sales figures.

Keywords: e-commerce; eBay; item descriptions; word choice; selling price

1. Introduction

Founded in 1995, eBay is an international online marketplace for the sale of a wide range of goods. The site has 135 million active buyers worldwide, with 1.7 billion items listed for sale at any given time (eBay Inc., 2022). In its early years, eBay was often thought of as an auction site where members of the public could sell unwanted gifts and household clutter to the highest bidder, a view that persisted until relatively recently. Examples 1–3 from our corpus of articles from the *Guardian* newspaper (described in Kehoe & Gee, 2009) capture this perception:

(1) At *eBay*, the auction site, people can make money selling their *old junk*. (31 March 2002)

(2) Sell *unwanted junk* at a car boot sale or on auction site *eBay* you'll have more space and cash. (18 January 2003)

Published in *International Journal of Corpus Linguistics*: https://doi.org/10.1075/ijcl.24063.keh

> (3) Fast forward to today, and we can now buy the services of "*declutter*" life coaches, drop off a box of unwanted items on the doorstep of our local charity shop, or flog all our *junk* on *eBay* (so that someone else can own more stuff). (29 July 2009)

However, eBay has changed significantly in recent years. The company's 2015 Annual Report (eBay Inc., 2015) stated "[w]hile eBay was once an auction site selling vintage items, today 80% of the items sold on eBay are new". By 2020, the company was reporting that this figure had increased to 81%, and that 91% of items were listed at a fixed price rather than as an auction (eBay Inc., 2020a).[1] This reflects the fact that many businesses are now using eBay to sell their products to the public, including small businesses that use the site as their main "shop window" and large retailers that use it as an additional online sales channel. Figure 1 shows two examples of this. The first, Vinyl Tap, is an independent record shop that has traded in Yorkshire since 1985, with over 210,000 items listed on eBay at the time of writing, all as fixed price "Buy it Now" listings (Vinyl Tap eBay Store, 2024). The second example is Tesco, the largest retailer in the UK, which has 200 items listed on eBay, mainly electrical items at a fixed price (Tesco eBay Store, 2024).

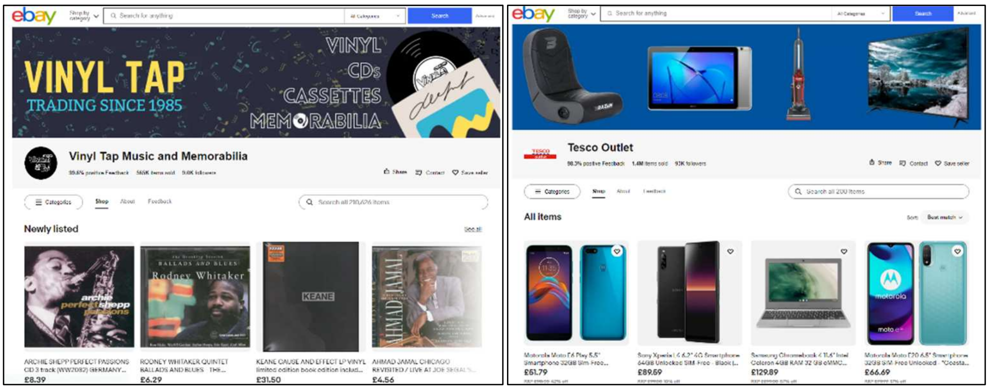

ebay
Shop by category
Search for anything
All Categories
Search
VINYL TAP
TRADING SINCE 1985
VINYL
CDs
CASSETTES
MEMORABILIA
Vinyl Tap Music and Memorabilia
Share
Contact
Save seller
Categories
Shop
About
Feedback
Search all 210,626 items
Newly listed
See all
ARCHIE SHEPP PERFECT PASSIONS CD 3 track (WW2082) GERMANY...
£8.39
RODNEY WHITAKER QUINTET BALLADS AND BLUES THE...
£6.29
KEANE CAUSE AND EFFECT LP VINYL limited edition book edition includ...
£31.50
AHMAD JAMAL CHICAGO REVISITED / LIVE AT JOE SEGAL'S...
£4.56
ebay
Shop by category
Search for anything
All Categories
Search
Advanced
Tesco Outlet
98.3% positive Feedback
1.4M items sold
93K followers
Share
Contact
Save seller
Categories
Shop
About
Feedback
Search all 200 items
All items
Sort
Best match
Motorola Moto E6 Play 5.5" Smartphone 32GB Sim-Free...
£51.79
Sony Xperia L4 6.2" 4G Smartphone 64GB Unlocked SIM-Free - Black (...
£89.59
Samsung Chromebook 4 11.6" Intel Celeron 4GB RAM 32 GB eMMC...
£129.89
Motorola Moto E20 6.5" Smartphone 32GB SIM-Free Unlocked - "Coasta...
£66.69

Figure 1. Examples of companies with eBay stores: Vinyl Tap and Tesco

Crucially, all eBay sellers, whether they be individuals or large companies, are encouraged to describe in their own words the items they list for sale. In this article, we analyse a large sample of these item descriptions and present the first large-scale corpus linguistic study of eBay. Our analysis is based on two corpora of item descriptions drawn from all categories in which items are offered for sale ("Computers", "Sporting Goods", "Baby", etc.) on eBay's UK website: a

[1] The company no longer publishes such detailed figures on its website or in other publicly-accessible reports.

14-million-word corpus from 2015 and a 43-million-word corpus from 2020 (as detailed in Section 3).

The aim of our analysis is to assess the extent to which eBay has shifted from an amateur auction site to a professional e-commerce platform and the impact this has had on the language used in item descriptions over time. We explore linguistic variation between the descriptions of different types of item in the various eBay categories, between sellers, and between selling prices, to gain further understanding of the language of online selling, which is vital as e-commerce continues to grow worldwide. We do so by addressing the following research questions (RQs):

1. To what extent has language use on eBay changed between 2015 and 2020?
2. How do item descriptions frame the nature of used products listed for sale?
3. How do they address the stereotype of eBay being associated with fake products?
4. To what extent do item descriptions across eBay product categories exhibit linguistic variation?
5. Can a relationship be identified between word choice and selling price?

2. Business communication and the study of e-commerce

Recent years have seen a considerable increase in e-commerce, which includes all forms of business transactions conducted online. Sparked in particular by the Covid-19 pandemic, e-commerce has grown by 50% between 2019 and 2021 (Goldberg, 2022). While "global retail e-commerce sales reached an estimated 5.8 trillion U.S. dollars" in 2023, they are forecast to "surpass eight trillion dollars by 2027" (Chevalier, 2024). E-commerce has therefore not only become an established component of business practice but is expected to continue to grow in importance. Consequently, it is crucial to know more about the mechanics of e-commerce and the role language plays in online sales. Before we explore the nature of item descriptions on the eBay platform in detail, we will situate our study within the field of business communication and discuss previous research on the communicative aspects of e-commerce.

Business communication is an established field of study that explores "all formal and informal communication within a business context, using all possible media, involving all stakeholder groups, operating both at the level of the individual employee and at that of the corporation" (Louhiala-Salminen, 2009: 312). As a field, business communication is characterised by its interdisciplinary nature and has been studied from a range of perspectives including management, marketing, communication studies, and linguistics. Therefore, research has used diverse methodological approaches when exploring examples of business communication including ethnography, surveys, and experiments. Compared to other approaches, corpus linguistics has not been widely used in the study of business communication to date (Jaworska, 2017). This is related to the nature of business communication data, which are often confidential or difficult to access. As a result, several of the main corpora in the field, such as the Wolverhampton Business English Corpus (2001), which focuses on written business communication, and the Cambridge and Nottingham Corpus of Business English (McCarthy & Handford, 2004; Handford, 2010), which is based on spoken language data, are not publicly available.

Despite these challenges in data availability, the benefits of corpus linguistics to the study of business communication have increasingly been recognised, not least due to the usefulness of its tools in exploring real world contexts. This is shown, for example, by handbooks on corpus linguistics now including chapters on business communication in general and digital contexts (see e.g. Mautner, 2020; Lutzky, 2020; Fuoli & Lutzky, forthcoming). In addition to creating small, specialised corpora (Koester, 2022), digital business data that is available in the public sphere has opened up new opportunities for corpus linguistic studies in the field. Lutzky and Kehoe (2022), for example, illustrate how a corpus linguistic methodology may be used when studying a corpus of customer service exchanges on the social networking platform Twitter (see also Lutzky, 2021).

In a digital context, online consumer reviews have grown in importance for both online and offline sales, as consumers increasingly turn to reviews when seeking information about products to support their purchasing decisions. As a consequence, online consumer reviews have been the focus of research investigating the linguistic and communicative features of this emerging genre (e.g. Vásquez, 2014). Studies have tried, for example, to find ways of identifying fake reviews by exploring the linguistic differences between shill and normal reviews, finding that the latter rank higher on readability and subjectivity (Ong et al., 2014). The problem of fake reviews is one that concerns the e-commerce platform eBay. Recent investigations by the consumer rights group *Which?* found that eBay's product review system allows "unscrupulous sellers to mislead shoppers" (Walsh, 2020). This is because reviews for the same product can be shared by different sellers, offering the possibility of displaying reviews for a new product in listings for used or even counterfeit items. The platform thus bears the potential of sellers abusing the review sharing system and of consumers being misled by reviews bearing little relation to the item listed. Given the non-transparent nature in which product information is shared on eBay, consumers also reported that they found themselves buying “a fake product by mistake” (King, 2011).

In fact, Meinl (2014) studied complaints in eBay’s British and German feedback forums and found that one of the main reasons consumers complain on eBay is that the item they received was different from what they expected. An important feature on eBay which allows users to find the products they are looking for in the first place, concerns naming conventions. Bodoff et al. (2017) study the emergence of naming conventions on eBay, basing their study on data from the eBay Big Data Lab which allows researchers access to samples of eBay’s data archive. They investigate the keywords used by sellers in item titles on eBay’s US website and by customers in product searches in 2012 and 2013, finding that the association between word and object became more precise over time, the number of synonyms used declined, and the word mappings of sellers and consumers became more similar.

While this development should have a positive effect on product searches on the platform, studies have also explored the role of reputation on eBay and found that the seller’s reputation has a small but statistically significant impact on selling price (Melnik & Alm, 2003). More specifically, Standifird (2001) finds that if a seller’s reputational rating is positive, it has little influence on the final selling price. However, a negative reputational rating turned out to be “highly significant and detrimental” (p. 293), underlining the importance of a positive reputation in e-commerce. In addition to reputation, research has investigated the influence of usernames on the perceived trustworthiness of eBay sellers, finding that, in a German speaking context, sellers are rated as more trustworthy if they have usernames that are short and easy to pronounce according to German phonotactic rules (Silva et al., 2017).

All of these considerations are of particular importance in the context of eBay, which began as an online auction site operating without an auctioneer. Traditional auctions are among “the most communication-intensive methods of selling” (Boyd, 2001: 287) and auctioneers play a prominent role in the process by introducing auction items, facilitating the bidding process, and maintaining order. Nevertheless, Boyd finds that eBay has managed to compensate for the absence of this key figure by imitating an oral style and providing detailed item descriptions that inform potential buyers about the qualities of a listed product. A key difference to traditional auction houses, which often issue an auction catalogue describing the items for sale, is that eBay item descriptions are written by individual sellers themselves. The importance of these descriptions is acknowledged by the platform. eBay has its own item description policy (eBay Inc., 2024), which introduces item descriptions to sellers as “one of the most important parts of your listing because it helps buyers decide whether to buy your item and what to expect if they do". Analysing completed listings for comic books from a discourse analytic perspective, Rawlins and Johnson (2007) find that readability is key when it comes to item descriptions. Achieving readability involves providing details about products and their quality in a clear manner, avoiding abbreviations and jargon, and writing in full sentences.

While previous research has explored various features of language use on eBay, there has only been one corpus linguistic study to date of product descriptions on the site. Knight et al. (2017) study a corpus of 650,000 words and 8,000 product descriptions from the "Shoes" category on eBay. They focus on linguistic differences in the descriptions of items listed as new and used, and in those listed by sellers regarded as experienced and inexperienced. Their findings indicate that politeness markers tend to be used more frequently in descriptions for used items – especially by novice sellers – whereas modal verbs tend to characterise descriptions of new products listed by experienced sellers. With regard to pronoun usage, new items listed by experienced sellers show a higher density of third-person pronouns, whereas used items listed by inexperienced sellers use first-person pronouns more frequently, indicating a more personalised approach to online selling. In the present study we aim to complement these initial findings about language use on eBay with a large-scale corpus linguistic analysis of item descriptions across all product categories in two time periods. We describe the corpora used in our analysis in Section 3.

3. Data

The corpora used in this study contain the main textual content (i.e. the item description) from listings on the eBay UK website. The initial data collection was performed in 2015, with a subsequent collection period in 2020 to facilitate comparison over time. The listing data was collected via the eBay Application Programming Interface (API), which enables machine-readable access to eBay data. Via the API, we searched for completed listings (either fixed price or auction), saving the results in an XML format (described below).

To avoid bias towards any one type of product and to facilitate comparison, we identified the 35 top-level categories on the eBay UK website (Table 1) and downloaded listings separately for each category. We limited our searches to the UK eBay site, but note that it is possible for listings from other worldwide eBay sites to be made available to UK users (with the US being the most frequent of the non-UK sites in our data). There are subtle differences between the UK and US product category names, some of which are unsurprising (e.g. "Mobile Phones" versus "Cell Phones" and "Property" versus "Real Estate") while others are more unexpected (e.g. “Sound & Vision” versus “Consumer Electronics”, and “Clothes” versus “Clothing”). Throughout this paper we use the UK names shown in Table 1. These linguistic differences had no impact on data collection as the category ID numbers remain the same across all eBay sites. The top-level categories also remained the same between 2015 and 2020.

Table 1. The 35 top-level product categories on eBay UK

Antiques	Collectables	Holidays & Travel	Sound & Vision
Art	Computers/Tablets & Networking	Home, Furniture & DIY	Sporting Goods
Baby	Crafts	Jewellery & Watches	Sports Memorabilia
Books, Comics & Magazines	Dolls & Bears	Mobile Phones & Communication	Stamps
Business, Office & Industrial	DVDs, Films & TV	Music	Toys & Games
Cameras & Photography	Events Tickets	Musical Instruments	Vehicle Parts & Accessories
Cars, Motorcycles & Vehicles	Everything Else	Pet Supplies	Video Games & Consoles
Clothes, Shoes & Accessories	Garden & Patio	Pottery, Porcelain & Glass	Wholesale & Job Lots
Coins	Health & Beauty	Property	

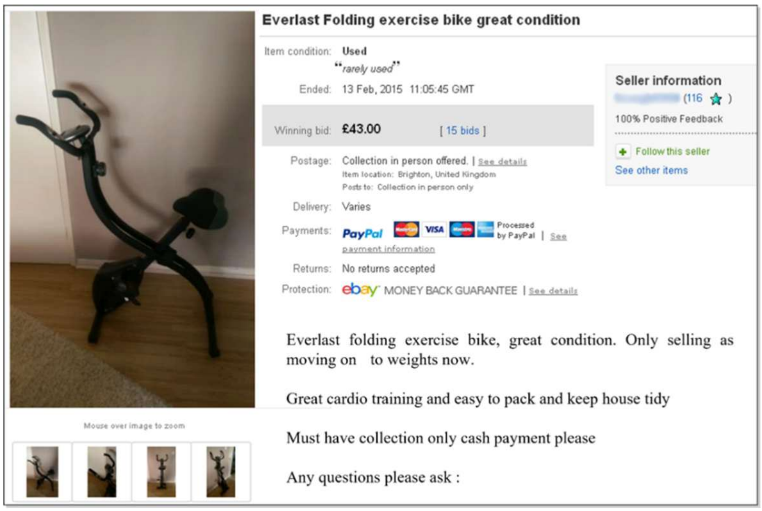
To construct the 2015 corpus within the limits set by the API for daily usage, we downloaded 100 items from each of the 35 categories each day between 13 January and 22 April 2015 (excluding occasional failures where no data was collected due to connectivity issues). This resulted in an initial corpus of 245,000 items. However, some duplicates were returned on consecutive days which, once removed (based on the unique ID given to each listing), left circa 235,000 items with 51 million tokens in the item descriptions.

Each listing was stored in XML, capturing the text of the item description and useful metadata, including whether the listing ended with a sale or not, sale price, category, location, and bid counts. Figure 2 shows a listing for what might once have been considered a typical eBay product: a second-hand exercise bike.[2] The XML formatted equivalent is shown in Figure 3. The listing ended with a sale, which can occur either through a successful bid during an

[2] Note that we anonymise seller names throughout this paper. Thus, they are blurred out or replaced with dashes in Figures 2 and 3.

auction or a “Buy it Now” fixed price purchase. In this case, the purchase was completed through a successful bid (after 15 bids) at a sale price of £43. As can be seen in the XML data, the item’s category contains four levels of detail (“Sporting Goods: Fitness, Running & Yoga: Cardio Machines: Exercise Bikes”) separated by colons within the <category> XML tag. The price and category information will be explored during the linguistic analysis in Section 4, alongside the item description text.

eBay allows for a great deal of flexibility in how item descriptions are written. Unlike e-commerce sites such as Amazon, where the product details are often provided by the manufacturer, the item descriptions on eBay are written by individual sellers (see Section 2) and constitute the main part of the listing through which the seller can present the required information to potential buyers. In Figure 2, the description is the main body of text at the bottom of the figure. The same description is shown in the <desc> XML tag in Figure 3. This sample description shows some of the information which may be present, including what the product is, its condition, why the seller is selling, the beneficial features of the product, and how to collect and pay.



Everlast Folding exercise bike great condition

Item condition: Used
“rarely used”
Ended: 13 Feb, 2015 11:05:45 GMT

Winning bid: £43.00 [15 bids]

Postage: Collection in person offered. | See details
Item location: Brighton, United Kingdom
Posts to: Collection in person only

Delivery: Varies

Payments: PayPal Processed by PayPal | See payment information

Returns: No returns accepted

Protection: ebay MONEY BACK GUARANTEE | See details

Seller information
[redacted] (116 ★)
100% Positive Feedback

Follow this seller
See other items

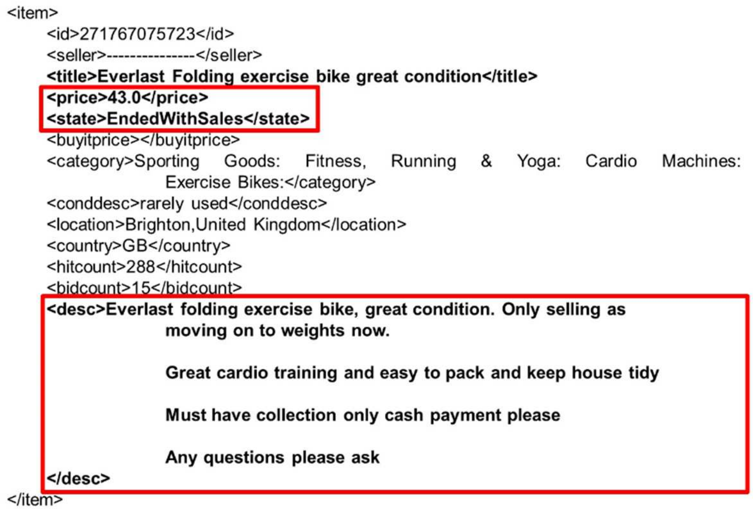
Everlast folding exercise bike, great condition. Only selling as moving on to weights now.

Great cardio training and easy to pack and keep house tidy

Must have collection only cash payment please

Any questions please ask :

Figure 2 Example eBay listing (at time of 2015 crawl, through the desktop web interface)



```
<item>
    <id>271767075723</id>
    <seller>---------------</seller>
    <title>Everlast Folding exercise bike great condition</title>
    <price>43.0</price>
    <state>EndedWithSales</state>
    <buyitprice></buyitprice>
    <category>Sporting Goods: Fitness, Running & Yoga: Cardio Machines:
              Exercise Bikes:</category>
    <conddesc>rarely used</conddesc>
    <location>Brighton,United Kingdom</location>
    <country>GB</country>
    <hitcount>288</hitcount>
    <bidcount>15</bidcount>
    <desc>Everlast folding exercise bike, great condition. Only selling as
              moving on to weights now.

              Great cardio training and easy to pack and keep house tidy

              Must have collection only cash payment please

              Any questions please ask
    </desc>
</item>
```

Figure 3 Details recorded during our crawl (main details highlighted)

One goal of this study (see RQ 5) is to explore the relationship between selling price and word choice in the descriptions. This requires the corpus to be further filtered to only the items which ended with a sale, which were 29% of the total items in the original 2015 corpus. Thus, we use the sub-corpus of sold items for all 2015 analyses, which consists of 67,817 items and 14,389,534 tokens (Table 2). All item descriptions were part-of-speech (POS) tagged using the *TreeTagger* software (Schmid, 1994). We use our *WebCorpLSE* software (Kehoe & Gee, 2007) to perform the analysis.

We are also interested in whether the patterns we observe for 2015 continue (or not) in following years. To achieve this, we extended the corpus considerably with data from 2020. Again, this was captured via the eBay API, following the same process as described above for the 2015 data. However, for the 2020 data capture we only downloaded items which were classified as sold, matching our analysis of the 2015 data. By filtering out unsold items at the point of data capture, we were able to build a larger useable corpus of 344,784 listings and 42,955,094 tokens (Table 2). The 2020 corpus was captured during the period 8 June to 14 October.

Table 2. Corpus size information

Year	Status	Items	Tokens	Types	Average description length (tokens)
2015	**Sold**	67,817	14,389,534	221,386	212
2020	**Sold**	344,784	42,955,094	599,065	125

In Table 2 we see a reduction in the average length of item descriptions from 212 tokens in 2015 to 125 tokens in 2020. This coincides with changes in user behaviour and the eBay interface. The 2020 Annual Report (eBay Inc., 2020b: 15) notes that "[a] significant and growing portion of our users access our platforms through mobile devices". At the same time, the app and mobile sites were modified such that the seller's item description is either not shown or only shown in brief unless the prospective buyer clicks a link to read it in full. Examples of some very brief descriptions from 2020, in which the brevity may have been motivated by the interface changes, are discussed in our analysis. Our approach to this analysis is outlined in Section 4, along with our findings.

4. Linguistic analysis of eBay item descriptions

In this section we carry out a large-scale analysis of linguistic variation on eBay, employing corpus linguistic methods and drawing upon examples from our 2015 and 2020 corpora. We begin in Section 4.1 by establishing the general lexicon of eBay by examining the most frequent words across all product categories. This is achieved using word frequency lists, with lexical changes explored by comparing word list ranks between 2015 and 2020 (see RQ 1 above). We then present two case studies concerning important aspects of eBay's business model, focusing on two terms mentioned above (see RQs 2 and 3): *used* (Section 4.2) and *fake* (Section 4.3). In these case studies, we use span-based collocation to establish the linguistic contexts in which the terms occur, measured using z-scores at span four left and right. In Section 4.4 we look at linguistic variation across product categories (see RQ 4), with examples from Antiques, Baby, Computers, and Cars. The keywords approach is used to compare word frequencies across product categories (e.g. Scott & Tribble, 2006; Culpeper, 2009). Thus, we compare each product category in turn against the remaining categories, using the log-likelihood measure (Dunning, 1993), and filtering the results by part-of-speech where appropriate. Finally, in Section 4.5, we explore the relationship between language use and selling price (see RQ 5), illustrating this with examples from the Wristwatches and Headphones categories. In this analysis, we present keywords for each category ordered by median selling price, enabling us to compare the price of items based on the occurrence of keywords in the item descriptions. In all our analyses, we highlight differences between the 2015 and 2020 corpora which may be indicative of wider trends in the use of eBay.

4.1 General eBay lexicon (RQ 1)

Table 3 shows the 30 most frequent words (types) across all 35 product categories in our 2015 and 2020 eBay corpora. All words are converted to lower case, meaning that case variants appear as a single entry in the table, and a stop word filter is applied to remove high frequency grammatical words.[3] Frequencies are shown per million words (fpm) and differences in ranks between 2015 and 2020 are indicated (diff).

Table 3. Most frequent words in 2015 and 2020 eBay corpora, excluding stop words

	2015 corpus				**2020 corpus**		
	word	**fpm**	**diff**		**Word**	**fpm**	**diff**
1	please	8,542.60	=		please	7,747.72	=
2	item	5,767.18	=		item	4,738.72	=
3	shipping	5,149.30	=		shipping	4,658.49	=
4	items	4,811.41	+1		condition	4,390.28	+7
5	days	3,826.95	+3		items	4,052.23	-1
6	payment	3,815.06	+1		new	2,916.37	+4
7	contact	3,203.30	+2		payment	2,793.34	-1
8	ebay	3,132.42	+2		days	2,719.96	-3
9	delivery	2,405.29	+5		contact	2,653.24	-2
10	new	2,380.41	-4		ebay	2,625.30	-2
11	condition	2,286.73	-7		mail	2,280.06	+18
12	paypal	2,155.59	+7		used	2,245.16	+16
13	free	1,976.99	=		free	2,159.33	=
14	return	1,959.90	+8		delivery	2,096.00	-5
15	postage	1,944.96	+6		good	1,885.29	+9
16	order	1,913.82	+7		royal	1,830.45	+81
17	feedback	1,883.59	+8		class	1,823.30	+98
18	time	1,841.48	+2		dispatched	1,750.95	+239
19	day	1,588.86	+7		paypal	1,737.42	-7
20	uk	1,565.86	+11		time	1,702.36	-2
21	service	1,540.43	+9		postage	1,686.44	-6
22	working	1,457.80	+10		return	1,654.61	-8
23	note	1,452.58	+6		order	1,462.13	-7
24	good	1,450.92	-9		see	1,452.19	+14
25	returns	1,399.84	+11		feedback	1,427.89	-8
26	refund	1,342.16	+13		day	1,386.82	-7
27	use	1,340.07	=		use	1,355.46	=
28	used	1,301.50	-16		size	1,339.66	+12

[3] The stop word list used can be found at https://www.webcorp.org.uk/wcx/lse/guide#also

29	mail	1,298.24	-18		note	1,334.44	-6
30	email	1,295.73	+52		service	1,283.88	-9

Despite the fact that our two corpora are separated by five years, the word frequency rankings are remarkably consistent, with the words in the top three positions remaining the same and 24 of the top 30 words in 2015 remaining in the top 30 in 2020. This suggests that the words in the table form the core eBay lexicon, used often by many users to discuss the logistics of trading on the site. A buyer will purchase an *item* or *items* (in *new* or *used condition*) from a seller (perhaps after checking their *feedback*), often making *payment* using the *PayPal* service that was once owned by eBay. A *shipping* fee may be charged (also referred to much less frequently by the British English term *postage*, perhaps reflecting the site's US origins) but shipping will often be *free*. *Delivery* will take place next *day* or within a certain number of *days*, and some sellers will accept *returns* if the buyer is unhappy with the purchase. The fact that the word *used* is less frequent than *new* in both corpora may reflect the shift in the focus of eBay that was discussed in our Introduction. However, we note that *used* actually increases in rank by 16 places between 2015 and 2020, with its frequency per million words rising from 1,301.50 to 2,245.16. We examine the word *used* in more detail in Section 4.2, where we consider the fact that there are other possible ways of referring to used products (*nearly new* being one which, of course, increases the frequency of *new* in Table 3).

Other differences between 2015 and 2020 in Table 3 require further explanation. The one word with a significantly higher rank in 2015 is *email*, ranked 30th in 2015 but 82nd in 2020. We believe this is due to a shift in eBay's policies and processes. Whereas buyers and sellers were previously free to contact each other by email, more recently they have been strongly encouraged to communicate only through eBay's own messaging service. Meanwhile, several of the words ranked significantly higher in 2020 than in 2015 are actually part of the same phrase: *dispatched with Royal Mail*, sometimes followed by either *first class* or *second class*. The phrase *dispatched with Royal Mail* appears only 0.76 times per million words in our 2015 corpus but 1302.29 times per million words in our 2020 corpus. We see some very brief descriptions in 2020 containing this phrase, as in Example (4), and even some listings where the phrase is included in the item title as well as the description.

(4) 4 Simpsons Comic books. Condition is Used. *Dispatched with Royal Mail* 2nd Class.[4]

[4] All corpus examples are reproduced verbatim, except for added italics.

We are not alone in observing this relatively recent phenomenon, with a 2019 post on the WordReference.com Language Forums[5] asking "In the eBay UK [sic], almost all sellers write 'Dispatched with Royal Mail' in the description. Shouldn't it be, 'Will be dispatched with Royal Mail'?". We have been unable to find a clear explanation, but it may the case that changes to the eBay user interface after 2015 made postage/shipping information less visible elsewhere, forcing sellers to include it in their own product descriptions. There may also exist a third-party listing template or software tool which encourages sellers to describe their items in this format.

Aside from the growth in frequency of this phrase, we find that the core eBay lexicon remains stable between 2015 and 2020. It may be surprising to see that *please* is the most frequent word in both years. Although high frequency grammatical words were removed from Table 3, *please* is more frequent in both corpora than common words such as *are*, *with*, and *be*. A routinised marker of politeness, the particle *please* is a strong indicator of the speech act of request. By examining concordance lines for *please* from the corpora, we find that it often appears with this function in constructions such as *please note* (2015: 952.36 occurrences per million words; 2020: 847.33 occurrences per million words), *please don't/do not hesitate to contact me* (2015: 24.53; 2020: 27.19), *please feel free to contact me* (2015: 23.91; 2020: 20.46), and *if you have any questions please ask* (2015: 15.29; 2020: 15.32). In their regular use of such constructions, sellers are setting out the terms and conditions of sale but also attempting to present themselves as approachable and trustworthy, thus instilling confidence in potential buyers. The concept of trust is an important one for any e-commerce platform and on eBay this is managed through the *feedback* system where buyers and sellers are encouraged to give each other positive or negative ratings, with a feedback score displayed next to usernames. We examine further examples related to trust in Section 4.3.

4.2 Case study: *used* (RQ 2)

Our first case study concerns the term *used*, a frequent word in both the 2015 and 2020 eBay corpora which we explore by conducting a collocational analysis across all listings in each corpus. We consider the words frequently appearing within a span of four words to the left and right of *used*, with the results shown in Table 4.[6]

[5] Retrieved January 5, 2023 from https://forum.wordreference.com/threads/dispatched-with-royal-mail.3552805/

[6] The z-scores are higher for 2020 because this corpus is larger. Our focus is on differences in rank between 2015 and 2020.

Table 4. Top 25 collocates of *used* at span 4, sorted by z-score

	2015			2020	
	Collocate	**z-score**		**Collocate**	**z-score**
1	condition	144.26		condition	1165.28
2	hardly	119.66		dispatched	757.48
3	good	102.93		royal	627.93
4	brand-new	73.49		mail	555.09
5	barely	55.56		good	337.02
6	considered	53.85		hardly	258.40
7	previously	41.74		excellent	161.50
8	widely	39.62		barely	157.32
9	new	39.43		books	150.36
10	commonly	39.33		widely	148.31
11	times	37.68		previously	141.27
12	excellent	37.37		lightly	140.17
13	lightly	36.46		handful	139.03
14	twice	33.84		twice	101.99
15	handful	33.83		times	94.74
16	fine	32.90		buying	92.92
17	opened	32.73		new	89.38
18	gently	30.22		gently	85.89
19	postally	26.17		priority	85.40
20	signs	25.82		great	84.22
21	couple	25.77		latest	79.93
22	acceptable	24.21		shipped	79.78
23	pre-owned	23.33		once	78.11
24	once	22.94		working	77.20
25	rarely	22.78		couple	72.77

The collocates of *used* on eBay remain fairly constant between 2015 and 2020, the biggest difference being the addition of *dispatched*, *royal*, and *mail* at high ranks in 2020 accounted for by the recent phenomenon discussed in the previous section, and specifically by examples like (4) where the shipping method is described immediately after the item condition (i.e. *dispatched*, *royal*, and *mail* all occur within four words to the right of *used*). The words *priority* and *shipped* (referring to shipping methods other than Royal Mail) are collocates for the same reason in 2020.

What is most striking about the collocates of *used* in both years is that many are adverbs signalling a lack of use: *hardly*, *barely*, *lightly*, *gently*, *rarely*. On a related note, we see references to items which have been used a *handful* of *times*, a *couple* of *times*, *once*, or *twice*. Similarly, items are described as being in *fine*, *acceptable*, *good*, *great*, or *excellent* used *condition*. The term *postally used* refers specifically to stamps, one of the item categories in our corpora. Overall, then, it seems that when sellers refer to their item as *used*, they are keen to highlight the fact that it has not been used very much.

In the 2015 collocates in Table 4 we also see an antonym of *used* in *brand-new*, and what appears to be a synonym in *pre-owned*. To explore the use of these and other related terms in more detail, we looked for sellers' own definitions of the terms. We did this by searching our 2015 corpus for examples of the word *used* appearing in single or double quotes and/or collocating with the word *mean(s).* Using this method we found numerous examples such as the following, all from separate sellers:

> (5) Vintage/*Used* means pre-owned even if near perfect there is always some evidence of wear
>
> (6) What does "*USED*" mean? A product is listed as '*Used'* or 'Unsealed' when the manufacturers seal on the product box has been broken
>
> (7) This guitar has never been played! It is lightly stamped "*used*" on the back of the headstock. This designation means the guitar has a minor cosmetic blemish
>
> (8) This is a great phone – it served my husband well (he recently upgraded). It is "*used*" so it has some light scratches but absolutely no cracks or anything else alarmingly worrying about it. (a weeny, barely visible chip in one corner which you might make out on the 1st picture above)
>
> (9) When an item is classified "*used*", "vintage" "antique" "old" or "pre-loved", please do not expect that it will be in perfect or new condition. It has had a previous life, so naturally it will have some signs of use/wear/dirt or other minimal damage. If you are looking for 'perfect vintage' items, that is not what we sell, unless stated as such

We see from these examples that different sellers define the term *used* in different ways, and that the use of synonyms and antonyms is common in such definitions. In Example (5) the seller classes *used* as a synonym of *vintage*, defining both terms as *pre-owned* which means there is “some evidence of wear”. Example (9) is similar except that this seller extends the list of synonyms for *used*, including *vintage* but adding *antique*, *old*, and *pre-loved*. This seller then contrasts all of these terms with the antonyms *perfect* and *new* (see also *perfect vintage*). Examples (6) and (7) include the term *used* to describe items that have not actually been used at all. In (6), from a listing posted by the large electrical retailer Currys, *used* simply means that the box has been opened (synonym *unsealed*), while in (7) the seller describes a guitar which actually has *used* stamped on it, but has never been played and is therefore not “used” in any meaningful sense. This can be contrasted with Example (8), where the seller lists a mobile phone which appears to be very heavily used. These examples illustrate that the term *used* can mean different things not only to different sellers, but also when describing different kinds of item. Linguistic variation by product category is a topic we explore in more depth in Section 4.4.

Another point that this case study illustrates is that the same passage of text may be included in multiple item listings by the same seller. In the above examples, only (8) is unique to a single listing, while the others all appear on every listing posted by that seller (Example (7) is a seller specialising in guitars rather than a one-off guitar seller). There is thus a risk that such boilerplate text will skew the word frequencies which form the basis of all our corpus linguistic analyses. This is something we bear in mind when drawing any conclusions from our eBay corpus.

4.3 Case study: *fake* (RQ 3)

In this second case study, we consider a second popular belief about eBay discussed above: that it is a common source of *fake* products (see Walsh, 2020). We thought it unlikely that sellers would describe their own items as *fake*, but we noted that this word occurred with a relatively high frequency in both our corpora (2015: 15.479 per million words; 2020: 16.552 per million words), so we wanted to analyse its use in more detail. To do this we adopted the same approach as in our first case study, this time considering the span four collocates of the word *fake* in both corpora (Table 5).

Table 5. Top 25 collocates of *fake* at span 4, sorted by z-score

	2015			**2020**	
	Collocate	**z-score**		**Collocate**	**z-score**
1	sell	16.55		sell	68.41
2	beware	12.94		real	64.79

3	accounts	10.89		plated	52.90
4	suspended	9.97		counterfeit	36.89
5	artificial	8.89		imitating	31.98
6	foliage	6.97		identification	30.67
7	copy	6.54		increasing	29.74
8	bidding	6.42		artificial	29.36
9	flower	5.75		products	28.21
10	real	5.51		items	27.83
11	counterfeit	4.99		sold	25.03
12	flowers	4.79		coins	22.80
13	buy	4.02		imitation	18.95
14	cheap	3.83		guaranteed	16.95
15	plants	3.74		buy	15.14
16	products	3.49		recognizing	14.99
17	genuine	3.26		shared	14.81
18	eyelashes	2.98		tan	14.78
19	diamonds	2.96		laws	14.74
20	judge	2.96		experienced	14.53
21	confuse	2.95		artefacts	13.91
22	reproduction	2.95		high	13.24
23	knowingly	2.92		purchase	12.29
24	spot	2.91		flower	11.30
25	authentic	2.83		genuine	11.24

There is slightly more variation between years for *fake* than there was for *used* in the previous section but, nonetheless, we can observe some common patterns. The collocational analysis reveals that in a limited number of cases it is perfectly acceptable for sellers to describe their own products as *fake*: *foliage*, *flower(s)*, *plants*, *eyelashes*, *tan*. In these product categories, *fake* is synonymous with *artificial* and *imitation* rather than with *counterfeit*. In some product categories, fake products are less acceptable, but we find that sellers in those categories often attempt to justify themselves as in Example (10) from a listing for *diamonds* (which also appears as a collocate in 2015 in Table 5):

> (10) It is important to remember that they are not "*fake*", they are simply man made (2015)

In some cases, sellers will use an alternative, more palatable, term such as *reproduction* in addition to *fake*, as in Example (11):

(11) Up for sale is a reproduction of a rare stamp. This stamp is *fake* but will make a great space filler until the real one comes along! (2015)

In this example the seller acknowledges that the stamp is *fake* but attempts to reassure potential buyers that it still has some value. The word *copy* in Table 5 is used in a similar way. Another technique used by sellers is to warn prospective buyers about rival sellers who do sell fake items and to contrast themselves with these sellers, as in Examples (12) to (15):

(12) Quit worrying about buying *fake* autographs on ebay and buy a REAL one here! (2015)

(13) There are many *fake* items littered throughout the net and on ebay unfortunately. (2015)

(14) We would like to point out once again to our customers that *fake* digital cameras, digital SLR cameras and digital camcorder batteries (counterfeit batteries) are available on the market. (2020)

(15) Beware of an increasing amount of *fake* British coins being sold on eBay. I have made an eBay guide on this subject. (2020)

This is potentially a risky strategy as it reminds buyers that eBay is known for fake products. However, by acknowledging this and presenting themselves as the “good guys”, sellers are hoping to instil confidence and encourage people to buy from them rather than the competition. Overall, we find that there is a fine line between acceptable fakes (*artificial*, *reproduction*) and unacceptable fakes (*counterfeit* products) and, once again, this varies between product categories. As we will go on to discuss in the next section, some categories have their own euphemisms for describing fake items.

4.4 Linguistic differences between eBay product categories (RQ 4)

While the above analyses studied language use on eBay as a whole, here we explore the patterns of linguistic variation detectable across the 35 eBay product categories in more detail. To achieve this, we adopt a keywords approach (Scott & Tribble, 2006), comparing each category in turn with all the other categories combined. For space reasons, we limit this section of the

Published in *International Journal of Corpus Linguistics*: https://doi.org/10.1075/ijcl.24063.keh

analysis to the 2015 corpus only. We illustrate the results of the category comparison for this corpus in Table 6, which shows the top keywords in the Antiques category.

Table 6. Keywords in the Antiques category in the 2015 eBay corpus

Keyword	Log-likelihood
antique	2642.98
I	1617.52
vintage	1404.35
silver	1101.76
age	812.38
chinese	757.93
measures	676.78
sterling	666.39
brass	626.12
old	561.52
art	517.83
piece	506.29
my	493.85
century	492.00
carved	472.71
porcelain	463.68
victorian	457.04
me	388.10
plate	384.76
beautiful	368.75
very	357.73
hallmarked	343.38
tall	327.82
wood	321.36
lovely	301.13

Unsurprisingly, several of these keywords are topic-related, that is, they are terms which are more likely to be used to describe antique items than items in other product categories. These terms often relate to age (*vintage*, *century*, *victorian*, *old*, *age*, and the word *antique* itself) or material (*silver*, *sterling*, *brass*, *porcelain*, *hallmarked*, *wood*). What is perhaps more surprising is that the first-person pronouns *I*, *me*, and *my* are keywords in the Antiques category, i.e. significantly more frequent in this category than across the other categories. This can be explained by the fact that sellers listing an item for sale in this category are much more likely to have had a long-term personal connection to that item, as in Example (16):

Published in *International Journal of Corpus Linguistics*: https://doi.org/10.1075/ijcl.24063.keh

(16) It has been in *my* family for over 60 years.

This is in line with the finding by Knight et al. (2017) that descriptions of used items in eBay's Shoes category are more likely to include first-person pronouns than descriptions of new items in that category. The difference is, of course, that all items in the Antiques category are, by definition, used, and the length of the connection between seller and item is likely to be much longer for an antique than for a pair of shoes.

Table 6 also includes two evaluative adjectives which are significantly more frequent in the Antiques category: *beautiful* and *lovely.* We extended this analysis across all categories in the 2015 corpus, using *TreeTagger* to restrict our output to adjectival keywords. We present results for three categories in Table 7: Baby; Computers/Tablets & Networking; and Cars, Motorcycles & Vehicles.

Table 7. Adjectival keywords in three product categories in the 2015 eBay corpus

	Baby		**Computers**		**Cars**	
	Keyword	**Log-likelihood**	**Keyword**	**Log-likelihood**	**Keyword**	**Log-likelihood**
1	lovely	516.40	compatible	1321.05	rear	1977.71
2	cute	461.90	universal	1064.99	electric	1074.75
3	newborn	422.11	hard	1002.21	front	852.59
4	pink	413.40	functional	663.12	last	710.11
5	soft	404.01	optical	612.20	new	676.57
6	little	402.47	lcd	305.94	good	601.69
7	washable	233.34	rigorous	288.79	heated	449.65
8	white	225.07	floppy	285.16	previous	424.85
9	gorgeous	200.94	octagonal	242.46	tidy	331.31
10	excellent	195.98	external	207.83	clean	290.11
11	new	188.66	video	187.54	central	227.01
12	blue	166.03	non-working	169.44	cheap	214.76
13	good	161.00	audio	168.03	full	208.91
14	adjustable	153.29	dual	167.40	old	176.55
15	striped	129.39	usable	166.58	rust	164.66
16	easy	119.19	mini	164.63	economical	152.32
17	removable	117.96	unaltered	163.44	low	137.72
18	beautiful	115.19	pre-owned	148.38	reliable	125.07
19	bnwt	110.27	sellable	143.82	metallic	122.57
20	unisex	94.29	genuine	137.17	diesel	110.21

21	reversible	88.55	serial	130.74	automatic	102.41
22	great	87.92	generic	130.35	manual	95.72
23	clean	85.60	integrated	128.32	honest	88.08
24	matching	83.33	cosmetic	127.31	reluctant	86.48
25	adorable	81.41	non-original	127.04	private	86.30

In Table 7 we see that the Baby category shares two adjectival keywords with Antiques (*beautiful*, *lovely*)[7] but has many unique keywords of its own. These include words used to describe specific baby products (*newborn*, *little*, *unisex*), some of which reflect stereotypical gender norms (*pink*, *blue*) and some of which are evaluative forms typical for this category such as *cute*, *gorgeous*, and *adorable*. The term *bnwt* is found across all clothing categories on eBay and stands for "brand new with tags" (in contrast with *used*).

The adjectival keywords for Computers/Tablets & Networking in Table 7 contain several IT-related terms, as would be expected in this product category. We find synonyms for *used* in the form of *pre-owned*, and potentially also in *usable*, *functional*, and *cosmetic* (damage). In addition, there are several synonyms, or euphemisms, for *fake*: *compatible*, *universal*, *generic*, and *non-original*. In contrast to *genuine*, all of these terms refer to computing parts or peripherals which are not made by the original manufacturer but should work just as well.

The adjectival keywords from the Cars, Motorcycles & Vehicles category in Table 7 again contain some topic-related words (*electric*, *metallic*, *diesel*, *automatic*, *manual*) but also words specific to the lexicon of selling used vehicles, which was an established market many years before eBay existed. This includes phrases such as one *previous* owner, *low* mileage, and the words *honest*, *private*, and *reluctant*, which collocate with *sale* to indicate that the vehicle is being sold by an individual for genuine reasons rather than by a dealer for profit. One word surprisingly absent is *second-hand*, an established term in the offline sale of used cars. In fact, *second-hand* occurs in all forms (with a hyphen, with a space, or as a solid compound) only nine times in the entire Cars, Motorcycles & Vehicles category. This equates to a frequency of 47 per million words, which is significantly lower than the frequency of the same term in other categories, including Books (170 per million words) and Video Games (321 per million words). We would suggest that while it is considered acceptable to use the term *second-hand* for the

[7] This is possible because our method calculates keywords in each category in turn, compared against all other categories combined. These words are significantly more frequent in *both* the Antiques and Baby categories than in the rest of the corpus.

sale of smaller items on eBay, there is a stigma associated with the term when it comes to larger items, particularly cars where it has had a long association with "dodgy dealing".

4.5 Linguistic variation by selling price (RQ 5)

In this final section of our analysis, we explore variation in language use across price bands in both our eBay corpora. The average selling price of items in our corpora increases from £97.15 in 2015 to £147.38 in 2020, but we find considerable variation across price bands in both corpora (Figure 4).

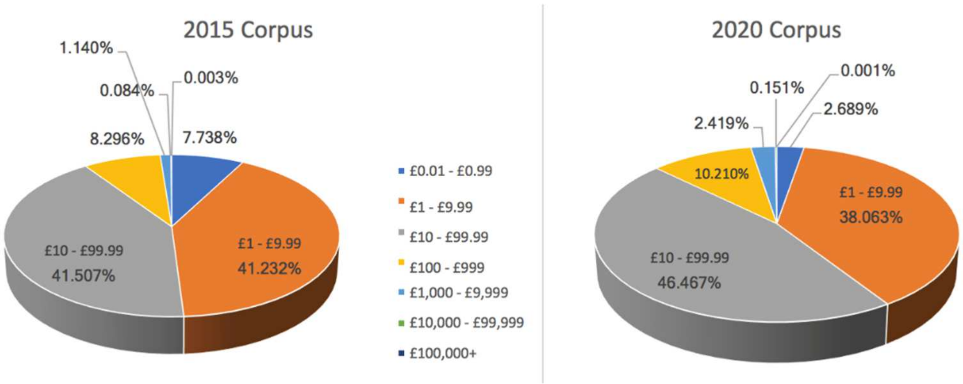


Figure 4. Selling prices of items in 2015 and 2020 eBay corpora (across all product categories)

As Figure 4 shows, almost half the items in our 2015 corpus sold for under £10, with 7.7% at under £1 and 41.2% at between £1 and £9.99. This proportion had reduced by 2020, with only 2.7% of items in that corpus selling for under £1, and 38.1% selling for between £1 and £9.99. This is likely to be the result of inflation over the five-year period, with a larger proportion of items selling for all price bands between £10 and £99,999 in 2020. It may also be the case that increases in eBay fees and other overheads have made it less financially viable for people to sell low price items. Examples of higher priced items in our corpora include a villa in Spain (£105,000), a 2007 Porsche 911 Carrera (£25,995), a 1 kilo Pamp Suisse 24k gold bullion bar (£23,700), and 24 pieces of commercial gym equipment (£13,574).

We conducted a further analysis to explore the relationship between language use and selling price. To do so we added information about the median selling price of items in each product category to our keywords analysis as follows:

i. Extract keywords for the category as above, filtering out grammatical words, product names, and words or numbers related to quantities or measurements (e.g. *ml*, *hz*, *mm*).

ii. For each keyword, find all items in the category with descriptions containing that keyword and calculate the median selling price (in GBP) of those items, filtering out keywords that occur in fewer than 10 item descriptions.

iii. Select the top remaining keywords[8] based on keyness score and plot these in order of median selling price.

To prevent price differences being unduly influenced by product type, we chose more specific categories than those explored above. For example, the Computers category includes products ranging from cheap peripherals for under £1 to expensive top of the range computers, and the Cars category includes car parts as well as complete vehicles. We focus our analysis here on two sub-categories that cover a less diverse range of products than the top-level categories: Wristwatches (Figure 5) and Headphones (Figure 6).

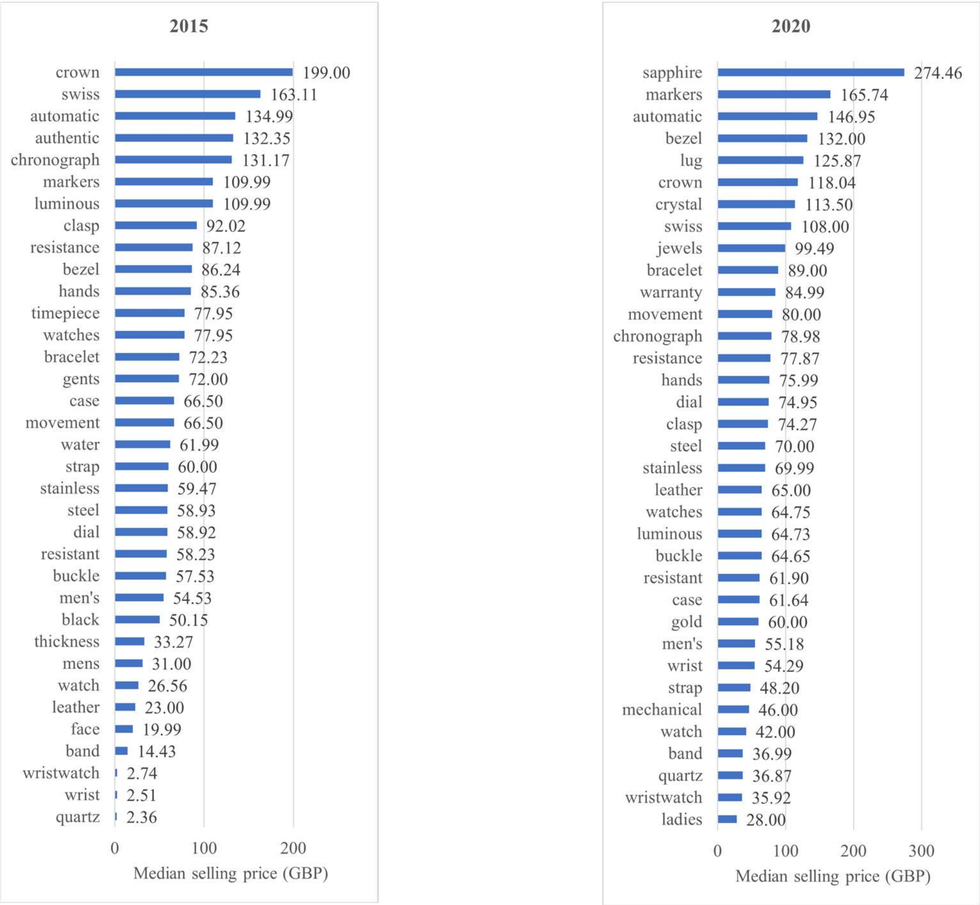


Figure 5. Keywords in Wristwatches category, ordered by median selling price

[8] We choose 35 keywords for a practical reason, namely, to fit within a readable chart.

Wristwatches is a sub-category of Jewellery & Watches, with an overall median selling price of £26.54 in our 2015 corpus and £42.00 in our 2020 corpus. Figure 5 shows the median selling price of items in that sub-category which contain specific keywords within their description. Although there is some variation between corpora, we find that listings across both years that sold for a higher price tend to mention particular watch parts: *crown* (the knob on the side of the watch used to change the time and date), *bezel* (the ring surrounding the *dial/face*), *lug* (where the *strap*, *band,* or *bracelet* attaches to the *case*), clasp (the mechanism for fastening a *bracelet*), and *crystal* (the type of glass covering the *dial/face*). Although *face* and *dial* are synonyms in the context of watches, the latter appears in listings which attract a higher median selling price (£58.92 in 2015, £74.95 in 2020). In both years we find that the specific term *chronograph* tends to be used to describe items which attract a higher price than the generic *wristwatch* and *watch*, with *timepiece* taking the middle ground in 2015. Watches described as having a particular country of origin – *swiss* – also attract a significantly higher price, and men's watches sell for more than women's watches. In the 2015 corpus the word *men's* appears in listings with a median selling price of £54.53, but the misspelling *mens* attracts only £31.00. The synonym *gents* is associated with an even higher median price of £72.00, perhaps due to the suggestion of class and sophistication which the neutral *men's* does not carry. The terms *women's* and *ladies* are not at high enough rank to be included for 2015, but the latter does appear in 2020 with a lower median price of £28.00.

When it comes to Headphones (Figure 6), a sub-category of Sound & Vision, the median selling price of items increases from £20.00 in 2015 to £28.00 in 2020. We again see desirable features – *wireless*, *bluetooth*, *noise* (cancelling) – attracting a higher median selling price than generic features: *wired*, *mic(rophone)*, *stereo*. There is one noticeable shift between our 2015 and 2020 corpora. In 2015, *on-ear* headphones (median price £73.99), such as those produced by the Beats by Dre brand, were fetching a higher price than *in-ear* headphones (£24.99). Since then the trend has reversed, with (ear) *buds* (£39.00) – written variously as *earbud* (£27.50) and *earbuds* (£21.75) – growing in popularity and, although not shown in Figure 6, *on-ear* dropping to a median price of £34.99 in 2020.

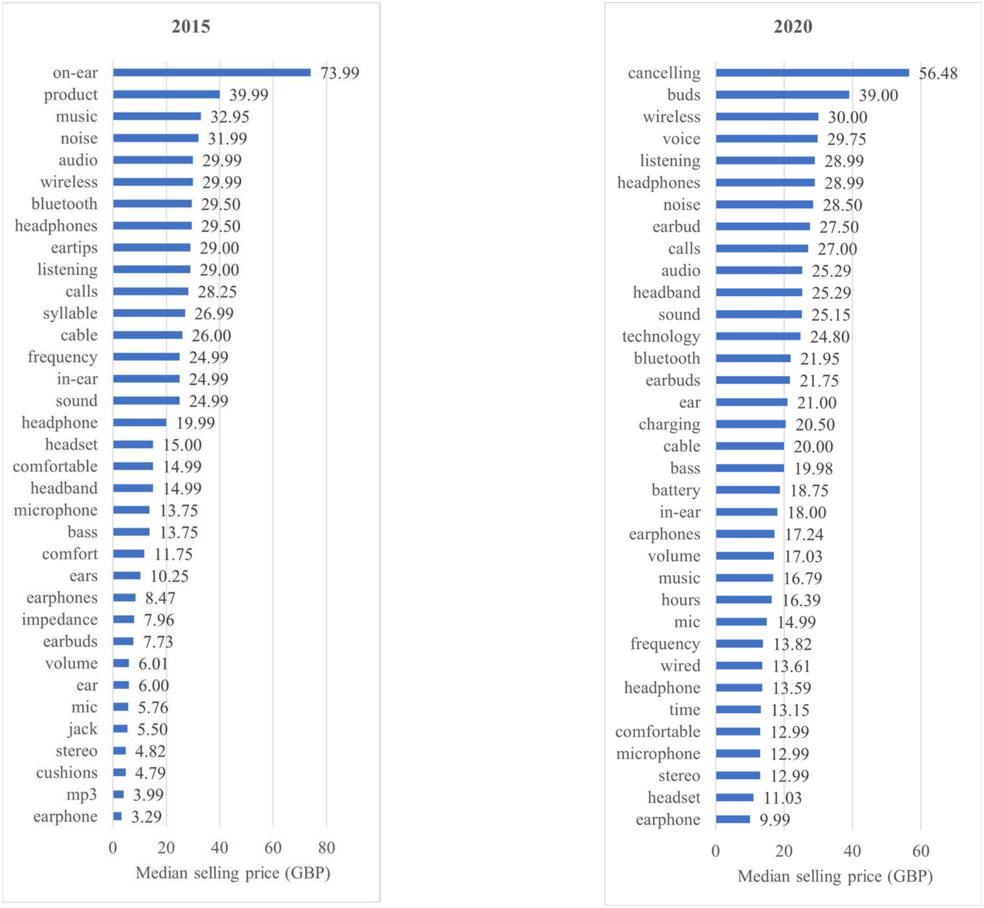


Figure 6. Keywords in Headphones category by median selling price

5. Conclusions

In this paper, we have conducted the first large-scale corpus linguistic analysis of eBay item descriptions in order to provide a deeper understanding of the language of online selling. Addressing RQ 1, we began by demonstrating that the core eBay lexicon, across product categories and sellers, remains relatively stable over a five-year period, where the frequently used non-grammatical words are those relating to the logistics of buying and selling on the platform. Interestingly, we found the most frequent word in both 2015 and 2020 to be *please*: a marker of politeness and strong indicator of the speech act of request. Our analysis revealed that sellers use this word both to set out the terms and conditions of sale (*please note*) and to present themselves as approachable and trustworthy (*please don't hesitate to contact me*). In answer to RQ 3, we also highlight the importance of trust in e-commerce through our collocational analysis of the word *fake*, where we find numerous examples of sellers warning potential buyers of fake items being sold by other people both on eBay and elsewhere online. In addition, our contextual analysis of *fake* reveals subtle differences in the meaning of the word, ranging from artificial (e.g. flowers) to compatible (computing hardware) to counterfeit (coins).

We addressed RQ 2 through a collocational analysis for the word *used*. This analysis revealed that while the word itself increased in frequency on eBay between 2015 and 2020, it continued to collocate with terms such as *barely*, *hardly*, and *lightly*, indicating a lack of use. We also demonstrated that *used* is defined differently by different sellers and has particular synonyms in specific product categories and contexts. For instance, *pre-owned* is a keyword in Computers, while *pre-loved* tends to describe clothes and *second-hand* appears to be stigmatised when selling used cars.

In addition, our analysis explored linguistic variation between item descriptions in different product categories (RQ 4) and, subsequently, at different price points (RQ 5). By adopting a keywords approach to compare categories we have demonstrated that while there are obvious differences in frequent topic-related words between categories (primarily nouns), there are also significant differences in adjective use, for example, the use of *lovely*, *cute*, and *adorable* to describe Baby products. By introducing an innovative approach to keyword analysis across categories and prices, we have been able to show the interplay between word choice in item descriptions and the price at which products were sold, demonstrating the impact stylistic choices may have on income. While it is a commonly accepted fact in linguistics that word choice matters and a word's connotation can affect the perception and therefore success of a message (see e.g. Kopf, 2017), this study is one of the first to demonstrate the immediate impact it can have on selling price. This illustrates the opportunities corpus linguistics has to offer to individuals and businesses selling their products on e-commerce platforms such as eBay.

There is scope for further work examining the boilerplate text which appears in multiple item descriptions by the same and different sellers, comparing the language of sold and unsold eBay listings, and comparing the seller-authored item descriptions for new items on eBay with professionally written descriptions for similar products on other e-commerce sites such as Amazon. Nonetheless, the findings of our study provide a deeper understanding of the language of online selling, which is vital as e-commerce continues to grow worldwide.

References

Bodoff, D., Bekkerman, R., & Dai, J. (2017). Evolution of language: An empirical study at eBay Big Data Lab. *PLoS ONE*, *12*(12), e0189107.

Boyd, J. (2001). Virtual orality: How eBay controls auctions without an auctioneer's voice.

American Speech, *76*(3), 286–300.

Chevalier, S. (2024, May 22). *Retail e-commerce sales worldwide from 2014 to 2027*. Statista. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, *14*(1), 29–59.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

eBay Inc. (2015). *2015 annual report*. https://www.annualreports.com/HostedData/AnnualReportArchive/e/NASDAQ_EBAY_2015.pdf

eBay Inc. (2020a). *eBay fast facts*. https://web.archive.org/web/20200805205205/https://investors.ebayinc.com/fast-facts/default.aspx

eBay Inc. (2020b). *2020 annual report/form 10-K*. https://ebay.q4cdn.com/610426115/files/doc_financials/2020/ar/2020-Annual-Report.pdf

eBay Inc. (2022). *eBay Marketplace fast facts at-a-glance (Q3 2022) – shareholders' report* https://web.archive.org/web/20230101002825/https://investors.ebayinc.com/fast-facts/default.aspx

eBay Inc. (2024). Item Description Policy. https://www.ebay.co.uk/help/policies/listing-policies/item-description-policy?id=4372

Fuoli, M., & Lutzky, U. (forthcoming). Business communication. In G. Brookes & M. Mahlberg (Eds.), *The Bloomsbury handbook of corpus linguistics*. Bloomsbury.

Goldberg, J. (2022, February 18). *E-Commerce sales grew 50% to $870 billion during the pandemic*. Forbes. https://www.forbes.com/sites/jasongoldberg/2022/02/18/e-commerce-sales-grew-50-to-870-billion-during-the-pandemic/?sh=1095ff6b4e83

Handford, M. (2010). *The language of business meetings*. Cambridge University Press.

Jaworska, S. (2017). Corpora and corpus linguistic approaches to studying business language. In G. Mautner & R. Franz (Eds.), *Handbook of business communication: Linguistic approaches* (pp. 583–608). Walter de Gruyter.

Kehoe, A., & Gee, M. (2007). New corpora from the web: making web text more 'text-like'. In P. Pahta, I. Taavitsainen, T. Nevalainen, & J. Tyrkkö (Eds.), *Towards multimedia in* *corpus studies*. University of Helsinki. https://varieng.helsinki.fi/series/volumes/02/kehoe_gee/

Kehoe, A., & Gee, M. (2009). Weaving web data into a diachronic corpus patchwork. In A. Renouf & A. Kehoe (Eds.), *Corpus linguistics: Refinements and reassessments* (pp. 255–279). Brill Rodopi.

King, M. (2011, December 7). *Christmas shoppers warned over flood of counterfeit toys*. The Guardian. https://www.theguardian.com/money/2011/dec/07/christmas-shopping-counterfeit-toys

Knight, D., Walsh, S., & Papagiannidis, S. (2017). I'm having a spring clear out: A corpus-based analysis of e-transactional discourse. *Applied Linguistics*, *38*(2), 234–257.

Koester, A. (2022). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed.). (pp. 48–61). Routledge.

Kopf, S. (2017). The Transatlantic Trade and Investment Partnership (TTIP) in the *Kronen Zeitung*. *Discourse, Context & Media*, *20*, 45–51.

Louhiala-Salminen, L. (2009). Business communication. In F. Bargiela-Chiappini (Ed.), *The handbook of business discourse* (pp. 305–316). Edinburgh University Press.

Lutzky, U. (2020). Digital media and business communication. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 394–407). Routledge.

Lutzky, U. (2021). *The discourse of customer service tweets: Planes, trains and automated text analysis*. Bloomsbury.

Lutzky, U., & Kehoe, A. (2022). Using corpus linguistics to study online data. In C. Vásquez (Ed.), *Research methods for digital discourse analysis* (pp. 219–236). Bloomsbury.

Mautner, G. (2020). Business discourse. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 319–333). Routledge.

McCarthy, M., & Handford, M. (2004). "Invisible to us": A preliminary corpus-based study of spoken business English. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 167–201). John Benjamins.

Meinl, M. E. (2014). *Electronic complaints: An empirical study on British English and German complaints on EBay.* Frank & Timme.

Melnik, M., & Alm, J. (2003). Does a Seller's eCommerce reputation matter? Evidence from eBay auctions. *The Journal of Industrial Economics*, *50*(3), 337–349.

Ong, T., Mannino, M., & Gregg, D. (2014). Linguistic characteristics of shill reviews. *Electronic Commerce Research and Applications*, *13*(2), 69–78.

Rawlins, C., & Johnson, P. (2007). Selling on eBay: Persuasive communication advice based on analysis of auction item descriptions. *Journal of Strategic E-Commerce*, *5*(1/2), 75–81.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing, Vol. 12* (pp. 44–49).

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins.

Silva, R. R., Chrobot, N., Newman, E., Schwarz, N., & Topolinski, S. (2017). Make it short and easy: Username complexity determines trustworthiness above and beyond objective reputation. *Frontiers in Psychology*, *8,* Article 2200. https://doi.org/10.3389/fpsyg.2017.02200

Standifird, S. S. (2001). Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management*, 27(3), 279–295.

Tesco eBay Store. (2024). https://www.ebay.co.uk/str/tescooutlet

Vásquez, C. (2014). *The discourse of online consumer reviews*. Bloomsbury.

Vinyl Tap eBay Store. (2024). https://www.ebay.co.uk/str/vinyltapmusicandmemorabilia

Walsh, H. (2020, March 13). *How eBay's review system is promoting fake, counterfeit and even dangerous products.* Which? https://www.which.co.uk/news/article/ebay-customer-reviews-aDd0j0g9ewIk

Wolverhampton Business English Corpus. (2001). https://catalogue.elra.info/en-us/repository/browse/ELRA-W0028/

Address for correspondence

Andrew Kehoe
College of English and Media
Birmingham City University
Millennium Point
Birmingham
B4 7XG
United Kingdom

andrew.kehoe@bcu.ac.uk
@ayjaykay

Co-author information

Matt Gee
College of English and Media
Birmingham City University
Millennium Point
Birmingham
B4 7XG
United Kingdom

matt.gee@bcu.ac.uk

Co-author information

Ursula Lutzky
Department of Business Communication
Vienna University of Economics and Business
Welthandelsplatz 1
1020 Vienna
Austria

ursula.lutzky@wu.ac.at
@UrsulaLutzky